A Large Language Model Guides the Affinity Maturation of Variant Antibodies Generated by Combinatorial Optimization

Faisal Bin Ashraf[1], Zihao Zhang[2], Karen Paco[2], Mariana P. Mendivil[2], Jordan A. Lay[2], Animesh Ray[2,3], and Stefano Lonardi[1]

[1]Department of Computer Science and Engineering, University of California, Riverside, 92521, CA, United States
[2]Riggs School of Applied Life Sciences, Keck Graduate Institute - 535 Watson Dr. Claremont CA
[3]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA

Abstract

Machine learning-based antibody design and optimization by computational affinity maturation is emerging as a promising approach to combating infectious diseases. This has been possible because of significant advances in artificial intelligence methods and a surge in experimental datasets on antigen-antibody interaction. The ability of an antibody to bind an antigen with sufficient strength (measured by *binding affinity*, the inverse of the equilibrium dissociation constant) and specificity are critical properties in the design of neutralizing antibodies. Here we introduce Ab-Affinity, a new large language model in conjunction with a genetic algorithm and simulated annealing for diversity generation and fitness optimization, which can accurately predict the binding affinity of specific antibodies against a target peptide within the SARS-CoV-2 spike protein. When trained on large datasets of existing antibodies that bind to certain antigens, we show that Ab-Affinity can generate novel antibodies with more than a 160-fold enhancement in binding affinities over those obtained experimentally. The predicted biophysical properties of the synthetic antibodies demonstrate their robustness. Molecular docking and molecular dynamics simulation of the binding interactions of the best candidate synthetic antibodies showed enhanced interactions and stability on the target peptide epitope. In general, antibodies generated by Ab-Affinity appear to be superior to those obtained with other existing computational methods.

1 Introduction

The antibody sequence motifs that specifically bind to an antigen target sequence or *epitope* are embedded within the complementarity determining regions (CDR) of the light and heavy chains, respectively [70]. The binding of an antigen by the antibody can trigger various immune responses, including the neutralization of a live virus, a toxin or an antigen present on a diseased cell surface (*, for example*, a cancer antigen), leading to the eventual elimination of the antigen or cell from the organism [70]. Direct measurement of the binding affinity of an antibody to its target antigen, which is critical to therapy development, incurs significant time and cost. Computational approaches that use structural databases of known antigen-antibody complexes and machine learning algorithms are increasingly replacing several time-intensive steps to narrow the list of candidate antibodies for final laboratory testing [46].

Predicting the specificity and strength of an antigen-antibody interaction is a subproblem of predicting the physical parameters of specific protein-protein interactions [46]. However, the first problem is hard because the epitope and the 'complementary' sequence motif on the antigen that is bound (the *paratope*) are often structurally flexible and contain relatively mobile intrinsically disordered protein regions (IDPRs), at least when not bound to each other [67, 41]. Structures of flexible and mobile areas of proteins are relatively poorly represented in protein structure databases because they are technically harder to determine; training data for learning structural rules are consequently sparse. Predicting these interactions from amino acid sequences can be advantageous if the structural and functional roles of amino acids in the sequence are effectively identified and modeled.

The problem of computationally predicting the binding affinity of protein-protein interaction was a grand challenge in biology until five years ago. Classical machine learning approaches for predicting the free energy of binding extract the biophysical characteristics of structured protein-protein interaction surfaces, side chain mobility, and interatomic distances at the binding interface, and assume incomplete models of force-field potential energy approximations and water structure for training purposes [71, 48, 34, 34]. However, it was recognized that the amino acid sequences of polypeptide chains at the binding interface between the epitopes and their cognate paratopes might contain enough information to accurately predict interaction if sufficient training data on the interacting protein sequences existed. The amino acid sequences on all observed binding interfaces constitute a minute fraction of the global protein sequence space; therefore, the interaction surfaces must embody regions of low Shannon entropy. Thus, the rules that govern the sequence requirements should be learnable. This idea led to the development of machine learning approaches that take properties of amino acid sequences at the binding interface as inputs for feature extraction. For example, ISLAND [1] used comprehensive extraction of amino acid sequence characteristics combined with regression, but achieved modest predictive accuracy. DeepDTA [53] improved upon ISLAND by relying on Convolutional Neural Networks (CNN) for protein-ligand interaction.

Predicting accurate protein-protein interaction between novel antigen and antibody pairs depends on predicting accurate structures of the interacting proteins–a computationally hard problem. A major advance occurred when AlphaFold [63] was enabled on the DeepMind neural network with features extracted from the amino acid sequences of proteins combined with known structural information, and from multiple alignments of the protein sequences with others exhibiting evolutionary relatedness. This approach, using the deep neural network architecture of Deepmind, led to the prediction of distances between pairs of amino acids and the distribution of the peptide bond torsion angles of the polypeptide scaffold. A gradient descent algorithm on the protein-specific potential energy field then optimized the packing of the side chain atoms. A more recent version, AlphaFold2 [32], achieved even more remarkable accuracy comparable to experimentally determined structures, and garnered widespread attention for its transformative impact on structural biology. Capitalizing on the successful prediction of single protein structures, Alphafold-Multimer [18] could predict the structures of interacting proteins with modest success; shortly thereafter, Alphafold3 [2] vastly improved prediction by incorporating a diffusion module for approximating atomic positions, in addition to re-framing the representational learning of physico-chemical properties of proteins and their evolutionary relationships from multiple sequence alignment. While Alphafold3 significantly improved the prediction accuracy of multi-chain protein complexes, the accuracy of predicting antigen-antibody complexes has been limited, producing about 60% correct prediction and 30-40% high accuracy prediction for antibodies with low sequence homology to known structures. These successes soon led to learning by natural language processing the "grammatical rules" by which polypeptide chains spontaneously fold on natural protein sequence strings, which evidently have been selected by evolution for optimized function. In this framing of the problem, the amino acids are the alphabets, short amino acid stretches are words and the entire sequences are the sentences in which the probability distribution of the immediate neighbors of an amino acid within the polypeptide string is guided by grammatical rule(s) that might have been used and shaped by evolution. Moreover, long-distance dependencies in amino acid positions must also exist and have been captured by a stochastic context-free grammar [14, 59, 20, 42]. The first successful Large Language Model (LLM) for predicting protein structure, ESMfold, used an unsupervised transformer model trained on nearly 15 billion parameters, which learned the rules of inter-atomic contacts among amino acids of natural proteins without having access to any structural information or by using explicit physics-inspired models [38]. More recent models such as DG-Affinity [77] and CSM-AB [50] integrate LLMs and graph-based techniques, thus enhancing the accuracy of prediction of antigen-antibody interaction. Tag-LLM [64] and FAbCon [6] further refined these capabilities using task-specific tags and fine-tuned LLMs. The use of pre-trained protein language models in DG-Affinity [77] has shown superior performance in predicting antibody binding affinities compared to several other methods. LLMs were recently used to predict the binding affinity of natural antibodies against the SARS-CoV-2 spike protein [37]. However, due to the high specificity of antibodies needed for combating viral diseases such as SARS-CoV-2, existing multi-target affinity prediction models are unsuitable. Therefore, an efficient binding affinity prediction model is essential for targeting this or any other virus.

Recent advances in generative AI/ML have brought significant progress to *de novo* protein and antibody design–proteins that do not exist in nature. Generative AI has allowed the designing of novel proteins that perform new functions (see, e.g. [19, 44]). Protein structure-based diffusion models were used to optimize the sequence and structure

simultaneously to attain the target function [40]. Several LLMs have demonstrated superior performance in capturing protein properties directly from the sequence [52]. For instance, ProtGPT2 [20] and ProGEN2 [42] are autoregressive language models trained on protein sequences that can generate new proteins with a desired function. ESMDesign [27] employs a masked large language model combined with simulated annealing optimization and ESMFold-based scoring to refine protein designs. ESMFold includes the large language model ESM-2, which was trained on the UniRef50 database [38]. Diffusion models such as RFDiffusion [74] and EvoDiff [4] offer more refined controls: RFDiffusion allows conditioning on attributes such as chain length and binding characteristics, while EvoDiff focuses on multiple sequence alignment families for evolutionary consistency.

Generative AI/ML is also revolutionizing the process of designing therapeutic antibodies (see, e.g. [8, 73, 35, 51, 33]). Hybrid approaches employing both laboratory methods and ML algorithms have been tested (see, e.g. [43, 13, 39, 54, 3, 46, 49, 22, 61, 7]). RefineGNN [30] uses graph neural networks trained on antibody sequences and structures, while IgLM [65] and DiffAb [40] provide enhanced control through conditioning on chain type, species, and specific target antigens. In a recent study, LLMs were trained and the binding affinity landscape of a library of mutant antibody sequences was generated from a starting antibody gene sequence, thereby improving upon and optimizing the binding characteristics of the generated antibodies [37]. Notably, general Protein Language Models can guide affinity maturation of antibodies using evolutionary information alone and can achieve competitive performance compared to in vitro methods [28].

Here we demonstrate for the first time, to our knowledge, the generation of novel antibody molecules with strong binding affinities against the HR2 epitope of the SARS-CoV-2 spike protein, by generating CDR diversity using a genetic algorithm and simulated annealing, both guided by an LLM specifically trained on antigen-antibody binding data.

2 Method

We introduce a large language model called Ab-Affinity to estimate the binding affinities for complete (both heavy and light) concatenated single-chain variable sequences, and validated by protein-docking and molecular dynamics simulation. We leveraged the prediction abilities of Ab-Affinity to generate *de novo* antibody sequences from a seed antibody. We developed three approaches for the generation, namely (i) a Position Specific Score Matrix (PSSM) for directed evolution, (ii) a Genetic Algorithm (GA) incorporating mutation and crossover operations, and (ii) a Simulated Annealing (SA) approach that explores sequence space through mutation strategies such as PSSM-based substitution, random substitution, and an antibody-specific substitution matrix (AB Model [47]). For sanity check, we used uniform probabilities to generate new sequences and predicted their affinities using Ab-Affinity.

2.1 Dataset

The training used a dataset of single-chain fragment variable (scFv) antibody sequences with the associated binding scores against a peptide in the SARS-CoV-2 HR2 region [16]. The SARS-CoV-2 spike protein has mutated, forming variants such as Wuhan, Alpha, Delta, Omicron, and others. The selected HR2 peptide, conserved across all variants of SARS and MERS spike proteins, is crucial for evaluating antibodies against the broader coronavirus group. The dataset was obtained by introducing one, two, and three amino acid changes into the sequences of three candidate antibodies obtained from a phage display library that bound to the HR2 antigen polypeptide: Ab-14-VH and Ab-14-VL, Ab-91-VH, Ab-95-VH and Ab-95-VL. The binding affinity (K_D) of each of the 104,972 resulting variants to the selected target peptide was estimated in three independent biological replicates [16]. Each K_D value was the equilibrium dissociation constant estimated by an indirect competitive binding assay. In this assay, a population of antibodies (see above) enabled the mating between two yeast strains one of which displayed the antibody and the other displayed the antigen on the cell surface. The extent to which the two yeast strains expressing the binding proteins could mate was proportional to the binding affinity. The proportions of the mated population were then estimated by counting the proportion of DNA sequences of the corresponding antibody genes and the relative sequence enrichment values were compared against a set of interacting protein pairs (internally controlled) with known K_D values measured directly by protein bi-layer interferometry [16]. This provided the estimated K_D values in nanomolar for each unknown interaction by regression and extrapolation.

The experimental assay was conducted in triplicate for each antibody, and the full dataset contained three K_D values for each of 314,925 antigen-antibody pairs. Of these 104,972 corresponded to unique antibodies. We preprocessed the interaction data by taking the arithmetic mean of the two closest K_D values for each antigen-antibody interaction and by eliminating the third value to reduce the outlier effect. Antibodies that had missing values across all three replicates (33,138) were disregarded from the analysis. After preprocessing, the data from the remaining 71,834 unique antibodies were utilized for model training. Ab-Affinity was trained on all antibody sequences (from each seed antibody) and their corresponding log-transformed binding affinity $\log_{10}(K_D)$. Distributions of training antibodies are shown in **Supplementary Figure 1**.

2.2 Model Architecture

Ab-Affinity's architecture is based on BERT [12], as implemented in ESM-2 [38] (see Figure 1D). We chose the BERT architecture because the dataset consists of amino acid sequences with a few point mutations. BERT effectively captures long-range dependencies, providing meaningful representations that align with binding affinity changes caused by these mutations. Ab-Affinity contains N sequential layers of encoder blocks. Each encoder block consists of multi-head attention layers [72] followed by feed-forward layers. The value of N determines the model size. We have tested $N = 6, 12$, and 33, which resulted in a model with $8M, 35M$, and $650M$ parameters, respectively. We chose these values for N based on the ESM-2 study [38], which demonstrated the impact of model size on the performance of this BERT-based architecture. We used the last encoder layer output as the sequence representation (i.e., the *embedding*). Depending on the choice of N the embedding of the sequence had $320, 480$ and 1280 dimensions, respectively. We added one fully connected layer to predict the binding affinity from the embedding of the sequence.

2.3 Training

We used Mean Squared Error (MSE) as the loss function, and Adam optimizer to optimize the parameters. We used 85% of the data to train the model (maintaining the distribution of affinity values). The remaining 15% was used to validate the model. We trained our model using four cores NVIDIA A100 (80GB) with a batch size of 128 for 100 epochs. For the amino-acid contact analysis, we used the method by Rao *et. al.* [55]. We fine-tuned the pretrained ESM-2 protein language model encoder layers, leveraging the knowledge acquired from the entire protein database to train our model. Additionally, we trained a model with randomly initiated weights to understand the impact of pretrained protein knowledge on binding affinity prediction. We saved the best-performing model based on the Pearson correlation coefficient on the validation set for each setup of training. Ab-Affinity, with 33 layers and fine-tuned from ESM-2, is the best-performing model among all the trained models. **Supplementary Figure 2** contains the performance of different trained models.

2.4 t-SNE visualization of embeddings

The T-Distributed Stochastic Neighbor Embedding (t-SNE) [68] was used to reduce high dimensional sequence representation to 2-D space to plot the antibodies. Dimensionality reduction was conducted using Python scikit-learn package, setting perplexity at 200.

2.5 Generation of Synthetic Antibodies

We generated *de novo* synthetic antibody sequences by altering the CDRs of the seed antibodies Ab-14, Ab-91, and Ab-95, and using Ab-Affinity to predict the binding affinity as their fitness score. Lower $log\ K_D$ values are considered to have higher fitness (Eq. 1). We explored four strategies to generate new antibodies, namely (i) random mutations based on a uniform distribution, (ii) random mutations according to a position-specific scoring matrix, (iii) mutations introduced by a genetic algorithm, and (iv) mutations driven by a simulated annealing method.

$$Fitness(s) = \frac{1}{f(s)} \tag{1}$$

where, s is the complete antibody sequence (scFv), and $f(s)$ is the predicted binding affinity of s using Ab-Affinity.

Uniform. To generate a random mutation to a protein s of length $|s|$, (1) we randomly selected a position $i \in [1, |s|]$ to be mutated with uniform probability, then (2) we mutated the amino acid $s[i]$ with one of the remaining 19 symbols with uniform probability. We chose this naive approach as a null model to assess the capability of generating *de novo* optimized sequences.

Position-specific scoring matrix (PSSM). To create a random mutation to a protein s, given a position-specific scoring matrix (PSSM) [23] P of dimension $|s| \times 20$ we (1) we randomly selected a position $i \in [1, |s|]$ with uniform probability, then (2) mutated amino acid $s[i]$ with one of the remaining 19 symbols using the probability distribution $P[i]$. The PSSM was obtained by computing the position-specific frequency of each amino acid in the training set. The pseudo-code for generating new sequences using PSSM can be found in the Supplementary material.

Genetic Algorithm (GA). We designed a genetic algorithm that starts from an initial population of randomly mutated seed antibody sequences. In each generation, sequences from the current set undergo crossover and mutation operations. The fitness of each sequence is evaluated using Ab-Affinity, which calculates its binding affinity ($\log\ K_D$). The genetic algorithm selects the highest fitness sequences (i.e., the lowest $\log\ K_D$ values) along with a subset of lower fitness sequences, as inputs for the following generation (see Figure 1B). Three substitution strategies were employed to mutate a position, namely, Random, PSSM, and AB model. The random and the PSSM approaches are described above. The AB model operates similarly to PSSM but employs a different substitution matrix specifically tailored for antibody sequences [47]. Our algorithm uses a temperature parameter that controls the diversity of the generated sequences. By tuning the temperature, we control the degree of randomness in the selection process, thereby fine-tuning the exploration of the fitness landscape. We set a population size of 50, number of generations to 1000 , one mutation per generation, and a temperature of 5.0. To determine the sequences for the next generation, we selected the top 20% best-fit sequences, while the remaining 80% were chosen randomly from the rest. This strategy allowed us to retain the best-fitting sequences while also exploring other sequences in the fitness landscape. The Supplementary material contains the pseudo-code for the Genetic Algorithm, Crossover and Mutation process.

Simulated Annealing (SA). We employed simulated annealing [69] to generate improved synthetic antibodies sampled from the binding affinity landscape. The method mutates a sequence by accepting moves that are less favorable with a small probability, thus preventing the steepest-descent algorithm to get stuck in local minima/maxima. Figure 1C provides schematic representation of this process. The algorithm starts from a seed antibody sequence. In each iteration, it generates neighboring sequences using the mutation techniques described above for the genetic algorithm. Among all the neighbors in the neighborhood, it selects the sequence with the lowest $\log\ K_D$ (greedy steepest-descent). The size of the neighborhood is defined by the allowed mutational distance. The algorithm always accepts the new neighbor if its Ab-Affinity predicted $\log\ K_D$ is lower. When the predicted value of the new neighbor is higher, it accepts it with probability $e^{\frac{\Delta E}{T}}$ where ΔE is the change in predicted binding affinity and T is the temperature parameter. This second case allows the algorithm to occasionally explore uphill moves, preventing the algorithm from getting stuck in a local optimum. The temperature parameter is progressively decreased by a cooling rate parameter, following the concept of annealing in physics. The algorithm continues until a maximum number of iterations is reached. In each iteration, the antibodies with the best-predicted affinity are recorded. A detailed pseudo-code of our implementation can be found in the Supplementary material.

2.6 Biophysical Properties

We used Biopython to compute the isoelectric point, hydrophobicity, instability index, and secondary structure contents for our synthetic antibodies [9]. The hydrophobicity score for each amino acid was computed using the Kyte & Dolittle index [36]. The hydrophobicity scores for whole antibodies were obtained by computing the average hydrophobicity over all the amino acids. The instability index [25] estimates whether the antibody will be stable in solution. We computed the fraction of alpha helix, beta sheets and the number of peptide backbone turns in the antibodies.

2.7 Structural Docking

HADDOCK v2.4 [10], a web-based tool, was used to dock the synthetic antibodies against the SARS-CoV-2 spike protein's HR2 domain (8CZI in the PDB database). First, we computed the predicted 3D structure for our synthetic antibodies using ESMFold [38]. Then, we used HADDOCK to carry out the docking between the predicted 3D structure and the target HR2 peptide structure [75]. We marked the regions corresponding to the CDRs of the antibody and the peptide of the epitope so that HADDOCK v2.4 could provide specific interactions between these two domains. We scored the quality of the docking using the HADDOCK score that combines Van der Waals intermolecular energy, electrostatic intermolecular energy, desolvation energy, and distance restraints energy. We also recorded the buried surface area, which is the fraction of solvent-exposed surface area of the unbound protein that becomes solvent-inaccessible (*i.e.*, buried) upon binding and contributes to the Gibbs free energy of binding through entropic and enthalpic components [11].

2.8 Molecular Dynamics Simulation

The top docking-scored models of the 8 best-generated candidates, the seed antibodies, and 4 randomly chosen weak-binding antibodies were used as the initial coordinates for 100 ns of molecular dynamics simulations. All molecular dynamics simulations were performed using the CUDA-accelerated version of OpenMM package on a computer cluster equipped with NVIDIA A100 GPUs [15]. Each scFv-epitope complex structure was inspected for missing atoms, and hydrogen atoms were assigned to the residues according to their protonation states at pH 7.0 using the OpenMM-Setup application [15]. The preprocessed structures were subsequently immersed in explicit solvent boxes of water, extending $10°$A from the protein boundary, using the Amber14 force field and its implicit TIP3P (Transferable Intermolecular Potential with Three Interaction Sites) water model [31]. The charge of the systems was neutralized by adding sodium ions.

A 5000 iteration energy minimization was performed to relieve steric clashes from the newly added solvent and optimize the initial geometry of the system. Then, an equilibrium phase of 3 ns was conducted using a 2 fs time step under the NPT ensemble to stabilize the system at 1 atm and 300 K [45]. The MD simulations were performed for 100 ns under constant pressure and the Long-range periodic electrostatic interactions were calculated using the smooth Particle-Mesh Ewald (PME) method [17], applying a real-space cutoff of $10°$A. All chemical bonds involving hydrogen atoms were constrained with tolerance of 1×10^{-6} using SHAKE algorithm [60]. The trajectory was recorded every 10,000 steps (20 ps) in a DCD file. The simulation trajectories were analyzed and visualized using the Visual Molecular Dynamics software (VMD) [29]. Hydrogen bonds were identified using the Hbond plugin of VMD. The cutoff distance and angle were set as 3.5°A and 30 degrees respectively. To analyze the conformational change and overall stability of the complexes, the root mean square deviation (RMSD) was calculated for CDR loops and epitope using MDAnalysis package in Python [21].

2.9 Sequence Analysis

We estimated the sequence similarity between the CDRs of the synthetic antibodies to the CDRs of all known antibodies (stored in the SAbDab database [62]) by computing the Levenshtein edit distances. We also compared the frequency distributions of amino acid residues in the synthetic CDRs with those in all known CDRs. Jensen-Shannon Divergence (D_{JS}) and Hellinger Distance (D_H) were used to compare the distribution of amino acids in the known CDRs (P) with the synthetically generated CDRs(Q).

$$D_{JS}(P,Q) = \frac{1}{2}D_{KL}(P \parallel M) + \frac{1}{2}D_{KL}(Q \parallel M) \tag{2}$$

where $M = \frac{P+Q}{2}$ and, $D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$

$$D_H(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_i \left(\sqrt{P(i)} - \sqrt{Q(i)}\right)^2} \tag{3}$$

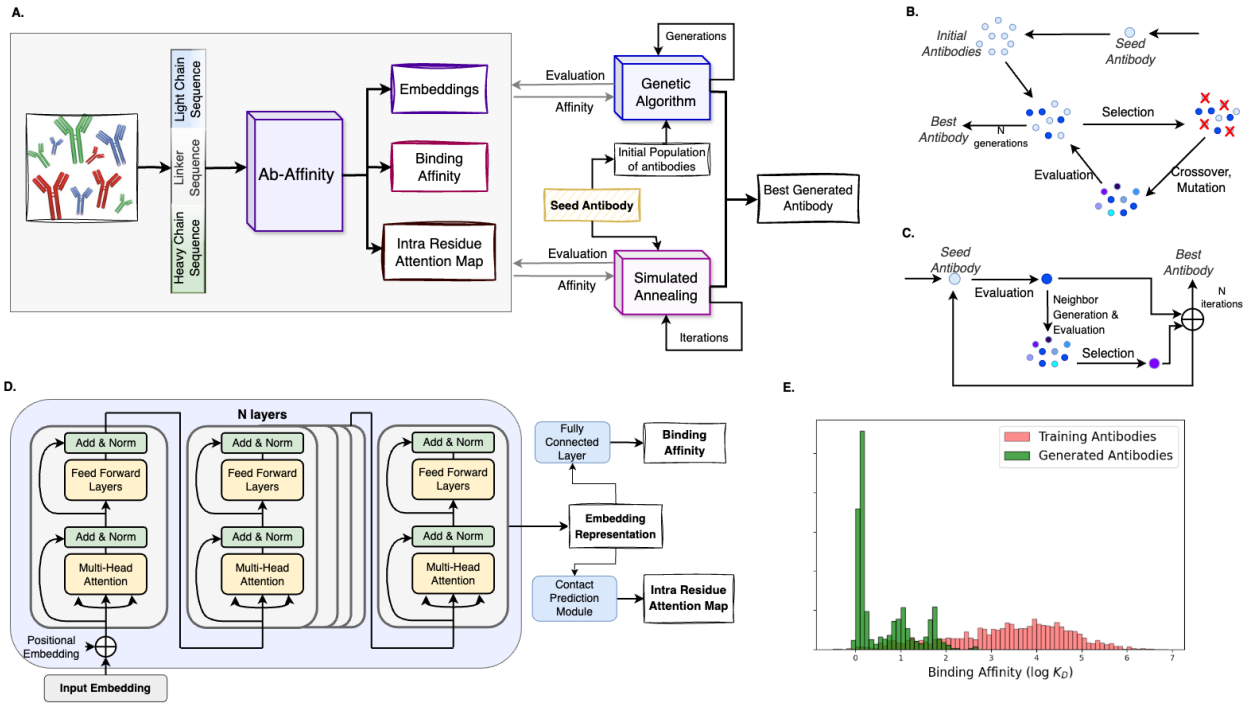


Figure 1: **A.** Ab-Affinity predicts the binding affinity against a specific target peptide; it also provides residue-residue contact maps and an embedding of the input sequence; a genetic algorithm and a simulated annealing method in combination with Ab-Affinity generates new synthetic antibodies; **B.** The genetic algorithm generates synthetic antibodies starting from a seed antibody by introducing mutations and crossover operations, guided by the predicted affinity scores from Ab-Affinity; **C.** The simulated annealing method starts from a seed antibody and explores its sequence neighbors, guided by the predicted affinity scores from Ab-Affinity; **D.** Model Architecture of Ab-Affinity. N sequential layers, each containing Multi-Head attention form the Encoder part of the model that provides embeddings representation. The Fully connected layer and the contact prediction module are on top of the encoder to predict binding affinity and residue-residue attention map; **E.** Binding Affinity distribution of the generated antibodies and training antibodies. Generated antibodies have strong binding affinity (lower $log\ K_D$). Note: all K_D values considered are in units of $10^{-9}M(nM)$.

3 Experimental Results

3.1 Overview of the framework

Our computational framework for generating and optimizing antibodies against the target antigen begins with a population of natural antibody sequences, including the heavy and light chains and linker sequences. These sequences are used to train the "Ab-Affinity" model to predict embeddings, binding affinity, and intra-residue attention maps. The outputs are used to evaluate antibody performance. To optimize antibody design, we employed two methods: genetic algorithm and simulated annealing. The genetic algorithm starts with an initial population of antibodies, then iteratively applies evaluation, selection, crossover, and mutation to identify the best-performing candidates. Simulated annealing begins with a seed antibody, then generates and evaluates its nearest neighbors iteratively, and selects candidates that improve affinity. These iterative processes ultimately converge on the optimal candidate antibodies against the HR2 peptide of the SARS-CoV-2 spike protein. Figure 1 illustrates the overall framework of our strategy.

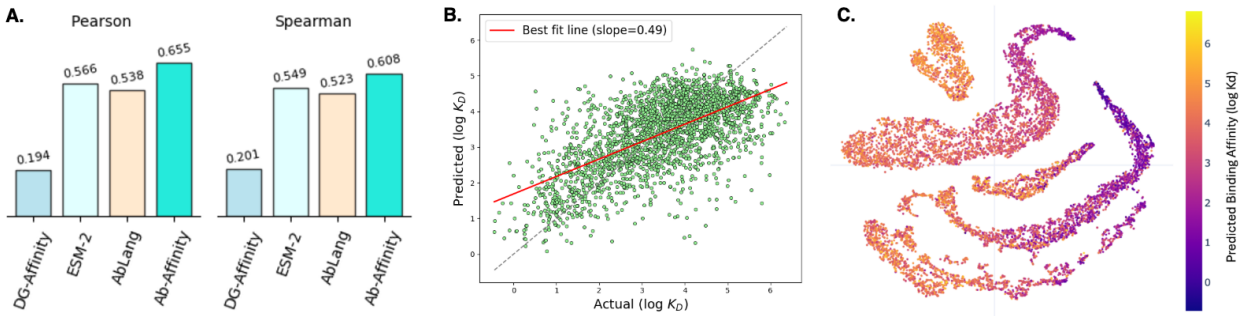


Figure 2: **A.** Pearson and Spearman correlation for DG-Affinity (p-values = 3.86×10^{-14}, and 3.38×10^{-15}), ESM-2 (p-values = 8.03×10^{-198}, and 8.02×10^{-198}), AbLang (p-values = 1.24×10^{-175}, and 1.02×10^{-163}) and Ab-Affinity (p-values = 4.03×10^{-261}, and 8.65×10^{-217}); **B.** scatter plot for actual vs. predicted binding affinity using Ab-Affinity; **C.** t-SNE representation of the embedding produced by Ab-Affinity, antibodies are colored according to their binding affinity ($log\ K_D$).

3.2 Ab-Affinity Predicts the Binding Affinity of Antibodies Against SARS-CoV-2

After training, we investigated whether Ab-Affinity creates meaningful latent space features (i.e., embeddings) for the antibodies. We first compared Ab-Affinity's ability to predict binding affinity with those of three other LLM-based methods, namely DG-Affinity, ESM-2, and AbLang. The same test dataset, which was held out from training and not utilized during model development, was used to compare the predicted binding affinities against experimentally observed values by all three methods. Note that DG-Affinity uses a pre-trained language model combined with a ConvNext-based architecture for prediction, which outperforms 26 other methods on an independent antibody dataset [77]; ESM-2 was pre-trained on the UniRef50 protein database and generated 640-dimensional antibody sequence embeddings, which we fed into a simple linear regression model; AbLang uses separate embeddings for the heavy and light chains of antibodies, which we concatenated into a 1534-dimensional vector and used in a linear regression model to predict the affinity score. Figure 2A illustrates that the Ab-Affinity model outperforms all other methods. The highest Pearson/Spearman correlation coefficients were observed with the predictions from Ab-Affinity . Scatter plots (Figure 2B) show the distribution of the predicted $\log\ K_D$ relative to the experimentally determined values. While DG-Affinity exhibited the lowest Pearson correlation coefficient (0.194), this was presumably because despite using an LLM pre-trained on antibodies, the architecture head itself was not trained with any SARS-CoV-2-specific data. Surprisingly, ESM-2, which was pre-trained on "generic" proteins, produced high Pearson/Spearman correlation coefficients. AbLang, which is trained on antibody sequences, also exhibited high Pearson/Spearman correlation. The scatter plots of actual vs. predicted binding affinity show that the best-fit lines for Ab-Affinity have the highest slope, indicating the best predictive accuracy (**Supplementary Figure 4**). Pearson correlation coefficient measures the linearity of the relationship between the observed and the predicted values, which is sensitive to outliers. On the other hand, Pearson rank correlation is insensitive to outlier values and is robust to nonlinear and monotonic trends. The scatter plots show a relative absence of extreme outliers in the test dataset, suggesting the linearity of prediction with observed values by all three methods.

We visualized Ab-Affinity's embeddings in a 2D space using t-distributed Stochastic Neighbor Embedding (t-SNE) maps. To evaluate Ab-Affinity's performance, we also visualized the embeddings produced by ESM-2, the best existing method. In the t-SNE plots in Figure 2C each point is an antibody and the color illustrates its corresponding binding affinity. As shown in Figure 2C, the embeddings produced by Ab-Affinity display a a smooth gradient of binding affinity, i.e., the value of $\log K_D$ monotonically decreases. By contrast, the ESM-2 embeddings (**Supplementary Figure 3**) do not show a clear gradient descent along the t-SNE components, suggesting that compared to ESM-2, Ab-Affiity embedding is more suited for describing antigen-antibody interaction.

To explore whether the prediction methods are sensitive to the choice of the test dataset, we compared the performance of Ab-Affinity with those of the other three methods using a different dataset. Table 1 compares the performance of CNN-based and LLM-based methods for predicting binding affinity to the same peptide. We used the 14H dataset (heavy chain sequences) and 14L dataset (light chain sequences), which are the mutated versions of Ab-14 heavy and light chain sequences, respectively. To fit the heavy chain sequences from the 14H dataset, the seed light chain

Table 1: Comparing Ab-Affinity against other binding affinity predictors on 14H and 14L dataset

Model	Ref	14H		14L	
		Pearson	Spearman	Pearson	Spearman
Ens-Grad	[39]	0.601	0.476	0.637	0.645
ESM-F	[26]	0.634	0.516	0.674	0.681
AntiBERTa2	[5]	0.623	0.545	0.673	0.684
AbMAP	[66]	0.606	0.510	0.674	0.685
A2Binder	[26]	0.642	**0.553**	0.683	0.688
Ensembles-14H	[37]	N/A	0.512	N/A	N/A
Ensembles-14L	[37]	N/A	N/A	N/A	0.688
Ab-Affinity	[this]	**0.652**	0.526	**0.712**	**0.713**

sequence of Ab-14 was paired with different heavy chains from 14H to form scFv sequences. Similarly, the heavy chain seed was paired with different light chains from 14L to form additional scFv sequences. We observed that the Pearson correlation coefficient of both 14H and 14L antibodies for Ab-Affinity were the highest among all models. While the best Spearman correlation for 14H was obtained by A2Binder, that of Ab-Affinity closely followed. Scatter plots for actual vs. predicted for these methods on both 14H, and 14L dataset are given in **Supplementary Figures 5 and 6**. Note that these correlation coefficients are specific to predicting the binding affinity for antibodies derived from the Ab-14 seed antibody. However, predictions from Ab-Affinity not only show a superior correlation for Ab-14 but also for all three seed antibodies.

3.3 Ab-Affinity's Embeddings enable Downstream Classification Tasks

We used Ab-Affinity embeddings to conduct two downstream classification tasks related to antibody binding characteristics, namely (i) the problem of determining the antibody binding affinity classes (High, Medium, Low), and (ii) the problem of determining whether the binding affinity of an antibody was stronger than the binding affinity of the corresponding seed antibody.

In the first task, Ab-Affinity achieved an AUC of 0.92 for strong antibody classification, outperforming ESM2's 0.78, while for weak antibodies, the AUCs were 0.78 and 0.71 for Ab-Affinity and ESM2, respectively (**Suppelmentary Figure 7**). In the second task, Ab-Affinity also demonstrated superior performance with an AUC of 0.91 relative to ESM2's 0.74 (**Suppelmentary Figure 7**). These results suggest that Ab-Affinity's embeddings are more informative for antigen-antibody optimization.

3.4 Ab-Affinity's Attention Maps for Strong and Weak Binding Antibodies

We extracted attention maps from the Ab-Affinity model to determine which residue-residue interactions were more important for the prediction of binding affinity. We produced residue-residue contact maps for heavy and light chains of antibodies with strong binding affinity (i.e., $\log K_D < 0.5$) and antibodies with weak binding affinity (i.e., $\log K_D > 5.5$). **Suppelmentary Figure 8** illustrates the residue-residue differences between the attention maps for strong and weak binding antibodies for both heavy and light chains where the CDRs are highlighted in pink, green and purple. Observe that many of the strongest differences occur in CDR-H1, CDR-H2, CDR-L1 or the regions immediately adjacent to them.

3.5 The Thermal Stability Feature Co-Clusters with Ab-Affinity Representation

Robustness to changes in temperature is important for therapeutic antibodies as it determines their structural integrity and functionality. While thermal stability (thermostability) was not an explicit feature on which Ab-affinity was trained, we investigated whether representation from Ab-Affinity exhibit correlation with thermostability. We use as reference a dataset of experimentally-determined thermostability of 26 antibodies against SARS-CoV-2 [28, 57]. In the 2D visualization of the antibody embeddings (**Suppelmentary Figure 9**) each point is an antibody and the

	Ab-14		Ab-91		Ab-95	
	Predicted Binding ($\log K_D$)	Fold Improvement	Predicted Binding ($\log K_D$)	Fold Improvement	Predicted Binding ($\log K_D$)	Fold Improvement
Seed Antibody	1.39	1	2.45	1	2.22	1
Uniform	4.25	-	3.36	-	3.29	-
PSSM	0.28	∼13	1.91	∼3.5	0.85	∼24
GA - Random	0.14	∼18	0.86	∼39	0.04	∼150
GA - PSSM	0.13	∼19	1.75	∼5	0.72	∼31
GA - ABModel	0.12	∼19	0.98	∼29	0.05	∼147
SA - Random	0.13	∼**19**	0.85	∼40	-0.01	∼**168**
SA - PSSM	0.16	∼17	0.79	∼**45**	0.04	∼150
SA - ABModel	0.13	∼**19**	1.39	∼12	0.03	∼154

Table 2: Comparing the predicted binding affinity (and fold improvement over the seed antibody) of synthetic antibodies generated by various methods within Ab-Affinity (bold font indicates the best results)

color illustrates its corresponding experimentally computed thermostability value. In Ab-Affinity's embedding, the antibodies are separated into two distinct clusters, each with relatively similar thermostability values. This is not the case for the ESM-2 embeddings, where the clusters display mixed values of thermostability.

3.6 Ab-Affinity's Synthetic Antibodies have Stronger Binding Affinity

We explored eight approaches to generate synthetic single-chain variable fragment sequences. The Uniform method generated random amino acids selected with uniform probability at random positions, also selected with uniform probability. In the PSSM method, we used a position-specific scoring matrix to drive the selection of amino acids. The other six approaches fall into two groups. In the first group, we employed a genetic algorithm that can either use (i) random mutations with uniform probability, (i) random mutations based on the probability distribution in the PSSM, and (iii) random mutations based on the ABModel matrix [47]. In the second group, we used a simulated annealing strategy, again using all three mutation techniques (i-iii) above.

Table 2 reports the main statistics for the top-performing antibodies generated by the eight methods. Each method was started from one of the three seed antibodies: Ab-14, Ab-91, and Ab-95. The table shows the predicted binding ($\log K_D$) and the fold improvement relative to the seed antibody. The Uniform method serves as the baseline. As shown in the Table 2, this approach fails to produce meaningful improvements in binding affinities. The PSSM method showed good improvements in binding affinity. For instance, Ab-14's binding improved about 13-fold, Ab-91 about 3.5-fold, and Ab-95 about 24-fold. The methods based on the Genetic Algorithm and Simulated Annealing yielded substantial improvements. For instance, GA-ABModel method for Ab-95 reached a 147-fold improvement, corresponding to $\Delta\Delta G = -2.95$ kcal/mol relative to the seed antibody, and SA-Random achieved the highest (168-fold) enhancement, corresponding to $\Delta\Delta G = -3.3$ kcal/mol or about 1-3 additional hydrogen bonds.

The density plots presented in Figure 3A show the distribution of predicted binding affinity ($\log K_D$) for all synthetic antibodies generated by the GA and SA methods from seed antibody Ab-91. The vertical dotted line indicates the binding affinity of the seed antibody. We also included the distribution of binding affinity for the antibodies in the training set. As expected, Ab-Affinity improves upon the affinity of the seed antibodies. The density plots for all seed antibodies (Ab-14, Ab-91, Ab-95) show sharp peaks of improved antibodies by all generation methods relative to antibodies in the training libraries (broad and flat distributions of each) (see **Suppelmentary Figure 10**). Importantly, the GA and SA methods, especially when combined with PSSM and ABModel, have density distributions peaking at significantly higher affinities (lower $\log K_D$ values).

Next, we investigated how the different approaches explored the sequence landscape within the same number of trials. To this end, we computed the mutational distance of synthetic antibodies with a relatively stronger binding affinity than the seed sequence. We found that some diversity generation methods were more conservative than others

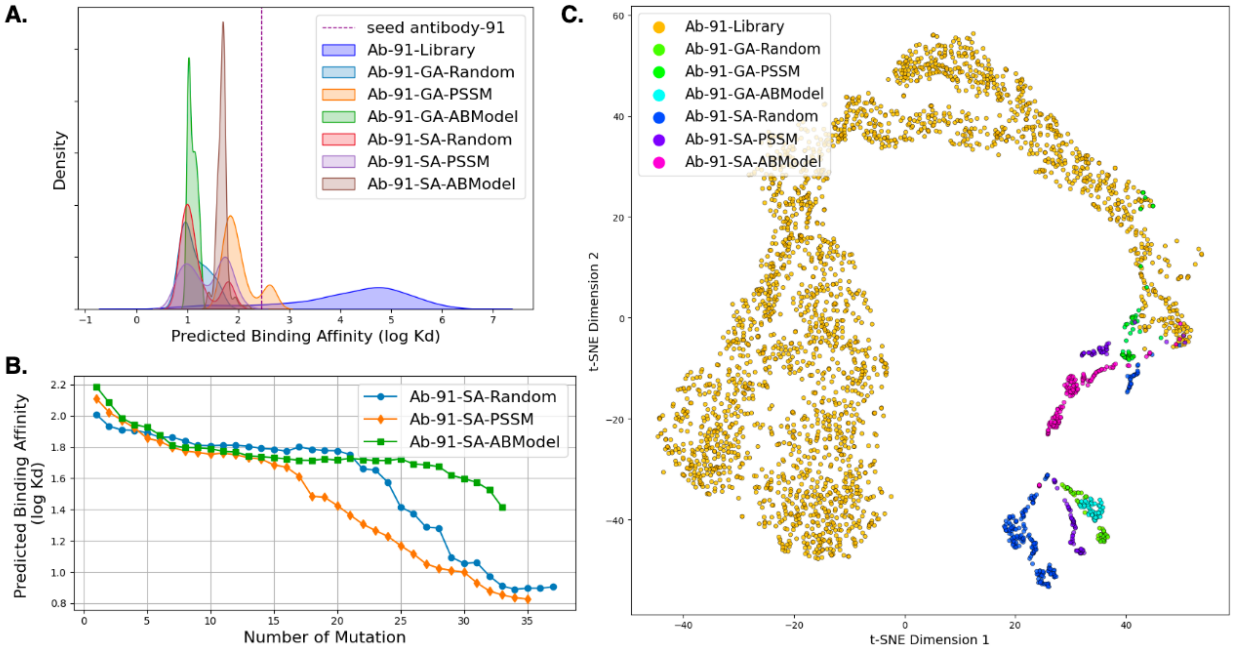


Figure 3: Analysis of generated antibodies from seed antibody Ab-91. **A.** Distribution of predicted binding affinity for the antibodies in the training library compared to the synthetic antibodies generated by the GA and SA methods; **B.** Predicted binding affinity of the synthetic antibodies as a function of the number of mutations from the corresponding seed antibody; **C.** t-SNE projection for the Ab-affinity embedding of the antibodies in the training library and the synthetic antibodies.

in exploring the fitness landscape. For example, the PSSM approach generated antibodies with up to 9 mutations, whereas the genetic algorithm had a broader exploratory reach, spanning mutational distances of 24, 34, and 37 for Ab-14, Ab-91, and Ab-95, respectively. Simulated annealing generated even more diversity, producing sequences up to 26, 37, and 44 mutational distances for the respective antibodies (see **Suppelmentary Figure 11**). We analyzed the relationship between the number of mutations and the predicted binding affinity using simulated annealing to estimate how the exploratory space affects selection. For Ab-14 and Ab-95, the binding affinity improved significantly as more mutations were explored. In the case of Ab-91 (see Figure 3B), only SA-Random and SA-PSSM showed consistent improvements in binding affinity as the number of mutations increased.

3.7 Ab-Affinity's Synthetic Antibodies are Novel

To compare synthetic antibodies to antibodies in the original library, we projected the embeddings in a 2D space using t-SNE (Figure 3C and **Suppelmentary Figure 12**). For example, the t-SNE plot for Ab-91, the synthetic antibodies generated by GA-Random, GA-PSSM, GA-ABModel, SA-Random, SA-PSSM, and by SA-ABModel, respectively, form distinct clusters that are clearly separated from the library antibodies (golden). Similar observations were made for Ab-14 and Ab-95 (**Suppelmentary Figure 12**). Our approaches generate new antibodies with up to 44 mutational distances from the seed sequence, whereas the training dataset contains antibodies with up to 3 mutational distances. Thus, our method can explore novel subspaces of the sequence diversity landscape, generating candidates whose sequence/embedding is far from the antibodies in the original library. Not surprisingly, the candidate antibodies produced by our methods have binding affinities significantly better than those in the training library (see **Suppelmentary Figure 12**).

3.8 Ab-Affinity's Synthetic Antibodies have Optimized Biophysical Properties

Next we analyzed the biophysical properties of synthetic antibodies, which are crucial to their overall performance against a target antigen. The isoelectric point (pI) impacts the antibody-antigen interaction because the net charge of the antibody relative to that of the surrounding solvent affects the binding kinetics. Human antibodies typically have pI 7.5-9 [24]. Synthetic antibodies generated using GA-Random, GA-ABModel, SA-ABModel, and SA-PSSM from the Ab-14 seed had pI 7.5-9.0 (**Supplementary Figure 13**). From the Ab-91 seed, all GA and SA had pI > 7.0. The hydrophobicity index determines solubility; synthetic antibodies generated by SA are less hydrophobic compared to library antibodies for all three target antigens, suggesting that former antibodies are likely to be more soluble. The pattern of *hydrophobicity* of a protein, especially on its surface, influences its thermal stability and solubility– less the surface hydrophobicity the more stable and soluble [73]. The *instability index* measures the overall stability of the protein structure: ≥ 40 indicates instability [25]. Our synthetic antibodies exhibit a lower instability index (< 40) compared to that of the library antibodies. Finally, the secondary structure provides insights into the antibody's binding abilities. For example, efficient binding is expected for antibody variable regions that contain a higher content of polypeptide backbone turns, rather than possessing mostly alpha helices and/or beta sheets, suggesting the presence of flexible and globular regions. The higher backbone turn contents of our synthetic antibodies compared to library antibodies, relative to the proportion of α-helices and β-sheets, suggest better structural properties of their interaction surfaces relative to those of the starting library antibodies **Supplementary Figure 14**.

3.9 Structural Docking and Molecular Dynamics Simulation Validates the Top Synthetic Antibodies

We carried out molecular docking to assess the interaction of our synthetic antibodies with the target antigen. We used ESMFold to obtain the predicted 3D structures of our synthetic antibodies, which were docked to the SARS-CoV-2 spike protein using HADDOCKv2.4. HADDOCK generated several bindings by considering different conformations, which were clustered based on energy and physical properties. Finally, we computed the HADDOCK-Score, which captures the combined energy of the protein complex (a lower HADDOCK-Score indicates higher complex stability). With one exception (SA for Ab-95), all synthetic antibodies were predicted to form more stable complexes than the seed antibody (Figure 4A). We computed the *buried surface area* inaccessible to the solvent due to complex formation (which is directly proportional to $\frac{1}{K_D}$, [11]) of each antigen-antibody complex. As anticipated, all except one synthetic antibody yielded higher buried surface area than that of the corresponding seed antibody when complexed with the antigen (Figure 4B). Figure 4C, 4D, 4E illustrates the docking between the synthetic antibodies and the target peptide, highlighting the improved binding features of the generated antibodies.

To confirm optimized complex formation, we conducted molecular dynamics (MD) simulations up to 300 ns for the complex containing the predicted Ab-95-GA-random-3, and analyzed the trajectories of CDR as well as the target epitope residues, respectively, during the final 100 ns of simulation (Figure 5A-D). We started with the highest-scoring docked complex; residue-averaged root mean square deviation (RMSD) values jumped within the first 40 ns, but stabilized by 200 ns. Whereas the aggregate RMSD value in the final 100 ns was significantly reduced for the epitope residues, we did not observe a similar trend with the CDR residues. Note that the target epitope was a contiguous stretch of 14 amino acids near the N-terminus of a 55-residue long HR2 region that is mostly an alpha-helix; on the other hand, the CDR region is substantially longer (63 amino acids), contains several flexible regions, and only a minority of the CDR residues is expected to be in contact with the epitope. Consistent with the expectation for a more efficient complex formation, we observed that the variance in RMSDs of Ab-95-GA-random-3 CDR residues is significantly reduced relative to the RMSD variance of its seed, Ab-95 (Figure 5E). Significantly, the most stable bond-forming conformation in the Ab-95-GA-random-3:Epitope complex predicted a 2.57Åhydrogen bond between Asp249 of the CDR and Lys103 of the epitope, and another 2.85Åhydrogen bond between of Asn257 of CDR and Val99 of the epitope (Figure 5F). Hydrogen bonds are typically in the range of 2.5–3.5Å. A bond length of 2.57Åsuggests a strong interaction, which likely has an electrostatic component. Asp (acidic) and Lys (basic) residues are commonly involved in ionic (salt bridge) interactions, especially in physiological pH (MD was run at pH 7.0). The short bond length and the nature of the residues make it likely that the interaction is not purely a hydrogen bond but also has an ionic component. In many cases, especially for charged amino acids like Asp and Lys, both interaction types can coexist, contributing to the overall stability of the interaction. These results strongly suggest the presence of an ionic

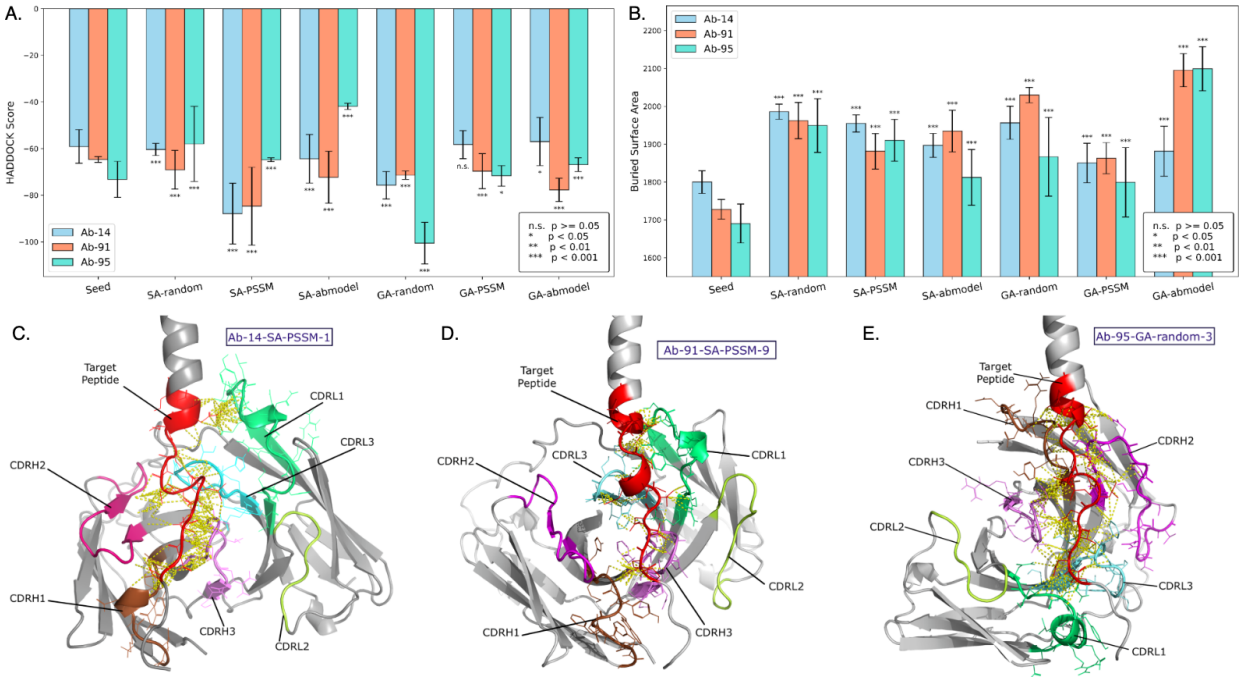
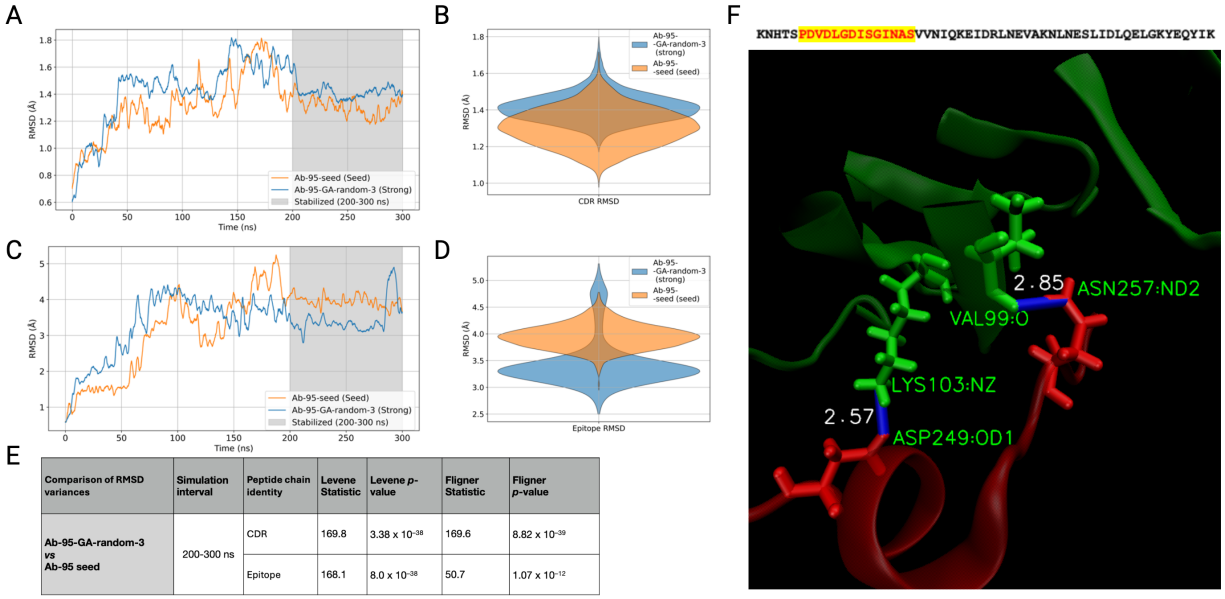


Figure 4: Molecular Docking of the antibodies with SARS-CoV-2 spike protein. **A.** Average HADDOCK-Score for the best synthetic and seed antibody [lower score indicates stronger binding]; **B.** Average buried surface area [higher the better] for the best synthetic and seed antibody; the error bar indicates the standard deviation; significance p-values compared to seed antibody are shown in asteric; **C, D, E.** HADDOCK docking for three synthetic antibodies against the target peptide from the SARS-CoV-2 spike protein; yellow lines connects residues within 4°A.

(electrostatic) interaction between Asp249 and Lys103, reinforced by the hydrogen bond. This dual interaction would significantly contribute to the stabilization of these residues in the structure of the complex. Thus, the total free-energy contribution of just these two interactions is approximately –5.42 kcal/mol, or a third of a typical free-energy of antigen-antibody interaction. Without considering any additional interactions, such as hydrophobic interaction in the looped region between the two hydrogen-bonded residues and additional entropic effects, this value is indicative of a relatively stable complex.

To examine whether MD simulation was generally consistent with the predicted robustness and specificity of the generated antibodies, we similarly used the highest-scoring docked models of antibody-antigen complexes of the 8 top-scoring antibodies, the three seed antibodies, and four randomly chosen weak-binding antibodies (Supplementary Figure 15, 16). For most complexes, the trajectories stabilized after 60-70 ns, indicating convergence to equilibrium conformations, with minor fluctuations suggesting loop flexibility. RMSDs of the 'weak' affinity antibody group displayed notable fluctuations after 70 ns, with RMSD values ranging between $1.4\AA$ and $2.7\AA$. In contrast, the candidates generated above maintained stable conformations beyond 70 ns. Although the RMSD of epitope residues was generally reduced in the complexes with the generated antibodies compared to the complexes with the corresponding seed antibodies, we did not observe a consistent reduction in the RMSD of the CDR residues in the generated antibody complexes (Supplementary Figure 15, 16). However, we did observe a decrease in the variance of RMSD values of the CDR residues during the terminal 30 ns of the simulated trajectories in complexes containing the generated antibodies relative to CDR residue variances in the corresponding seed antibodies (**Supplementary Table 3**). The distribution of the RMSD values of the residues is significantly different within complexes containing the generated antibodies compared to complexes with the corresponding seed antibodies (**Supplementary Table 3**).

Hydrogen bond analysis (**Suppelementary Figure 16**) demonstrated that the 'top' antibody group consistently maintained a higher number (mean 2.8, range 2-5) of significant hydrogen bonds, most of which have one or more hydrogen bonds with occupancy $\geq$ 50% during the last 30 ns of simulation. In contrast, the 'weak' group exhibited 1.25



Comparison of RMSD variances	Simulation interval	Peptide chain identity	Levene Statistic	Levene *p*-value	Fligner Statistic	Fligner *p*-value
Ab-95-GA-random-3 *vs* Ab-95 seed	200-300 ns	CDR	169.8	3.38×10^{-38}	169.6	8.82×10^{-39}
		Epitope	168.1	8.0×10^{-38}	50.7	1.07×10^{-12}

Figure 5: Molecular Dynamics simulation of the antibodies in complex with the HR2 region of SARS-CoV-2 spike protein. **A.** RMSD values of the amino acid residues of the epitope in complex with two antibodies (blue: Ab-95-GA-random-3; yellow: Ab-95 seed). **C.** CDR RMSD values of amino acid residues (colors are as above). **B, D.** Violin plots of RMSD distributions in the trajectories for the epitope, and the CDR for the final 100 ns of simulation (shaded regions on the left panels), respectively. **E.** Statistical analysis of difference between the variances in RMSD values shown in panels **B** and **D** above. **F.** The interaction surface with two identified hydrogen bonds between the synthetic antibody Ab-95-GA-random-3 and the epitope. ASN257 and ASP249 are residues of the CDR region of the antibody (in red); Val99 and Lys103 are two residues of the epitope (in green). The target epitope sequence is shown above the image (red, in yellow highlight) of the HR2 region. The close approach of Lys103 epsilon amine nitrogen (LYS103:NZ) and the ionizable oxygen atom of the carboxylate group of Asp249 (ASP249:OD1), and the fact that both groups are ionized at neutral pH at which MD simulation was conducted, together indicate that there is also electrostatic interaction between these two groups, further stabilizing the complex.

hydrogen bonds (range 0-3), indicating less stable interactions between antibodies and their target antigen. Detailed hydrogen bond analysis data are available in **Supplementary Table 2**.

3.10 Ab-Affinity's Synthetic Antibodies Recapitulates Properties of Natural Antibodies

We compared the sequence similarity of CDRs of our synthetic antibodies with the CDRs of all known antibodies present in the SAbDab database, available through the Protein Data Bank (PDB) [58]. Jensen-Shannon Divergence between these two distributions was 0.187, and Hellinger Distance was 0.191, suggesting that the distributions are comparable (**Supplementary Figure 17**). The distributions are indistinguishable except those for alanine (A) and lysine (K). **Supplementary Table 1** lists the top twenty hits from the SAbDab database. Our synthetic antibodies have $> 71\%$ sequence identity to their natural counterparts, especially to antibodies targeting antigens in diseases caused by Sarbecoviruses.

4 Discussion

4.1 Ab-Affinity Outperforms Existing Language Models

We have developed Ab-Affinity, a large language model that can predict the binding affinity of antibodies against the SARS-CoV-2 spike protein. Ab-Affinity demonstrated stronger predictive performance than those generated by previously published methods: DG-Affinity, ESM-2, and AbLang. The t-SNE visualization of Ab-Affinity's embeddings suggested that it captured the relationships between antibody sequences and binding affinities to the target antigen. The model architecture incorporates 33 layers of BERT modules, leveraging self-attention mechanisms to analyze and learn the impact of mutations across the entire antibody sequence. This ability to evaluate mutations with long global reaches may be critical, as changes in antibody sequences, particularly in the complementarity-determining regions (CDRs), can alter their structural conformation in these regions and binding efficacy with the target peptide. Unlike existing approaches that focus on generalized models for different antibody-antigens complexes (i.e., DG-Affinity [77]), Ab-Affinity is tailored to capture subtle effects of mutations in CDRs on the binding affinities against the HR2 segment of SARS-Cov-2 spike protein. This capability is important for addressing the specificity of the antibody, a challenge that conventional models often fail to address. This is especially important in predicting binding affinities for mutated antibodies from a seed antibody. The superior performance of Ab-Affinity over ESM-2 and AbLang is likely because the latter are trained on protein sequences in general (including antibody sequences). Unlike these earlier models, which are designed primarily to extract features represented by amino acid sequences, Ab-Affinity is purpose-built to predict binding affinities. The self-attention mechanism in Ab-Affinity focuses specifically on learning the effects of amino acid replacement on the antigen binding affinity. This targeted learning enables Ab-Affinity to outperform general protein and antibody language models.

Synthetic antibodies were generated using combinatorial optimization and Ab-Affinity, and we were able to generate new antibodies with > 160 fold binding improvement. Molecular docking of our synthetic antibodies with the target peptide demonstrated that the synthetic antibodies had enhanced interactions with the target peptide and formed more stable complexes compared to the seed antibody. We further validated the stability and specificity of the antibody-antigen complexes through MD simulations (100-300 ns). MD analysis revealed that the synthetic antibodies are stable and form, on average, more hydrogen bonds in their interactions with the SARS-CoV-2 spike protein HR2 peptide, the target epitope. These findings together suggest that Ab-Affinity can generate novel antibodies with enhanced binding affinities and produce favorable biophysical properties (see below). These results hold significant implications for the development of novel therapeutic and diagnostic interventions targeting SARS-CoV-2 and potentially other infectious agents.

Ab-Affinity's embeddings enabled a strong performance on several classification tasks (binding activity, and binding improvement prediction), emphasizing the utility of our model beyond just affinity prediction. Since Ab-Affinity is fine-tuned from ESM-2, a pre-trained protein language model that has the ability to represent protein sequences, Ab-Affinity inherently retains this capability. The fine-tuning process further equips Ab-Affinity to learn the effects of mutations.

The residue-residue attention maps extracted from Ab-Affinity reveal that our model focuses on the CDRs and/or the immediately neighboring regions of the antibodies. The disparities in the residue-residue attention maps highlight Ab-Affinity's proficiency in distinguishing between strong and weak binders. The model's self-attention mechanism inherently learns to identify regions within the antibody that influence binding affinity. While the CDRs are the most attended regions, the model also assigns attention to select positions outside the CDRs, reflecting its capacity to capture the effects of mutations across the entire sequence. This observation suggests that the model may be informing antibody design by learning and implementing critical molecular interaction rules governing binding affinity.

Ab-Affinity demonstrated a better "understanding" of antibody thermal stability than ESM-2 despite not being explicitly trained on this attribute. One explanation of this finding is that the fine-tuned language model has learned intrinsic rules that not only govern the physics of antigen-antibody interaction but also the physics of protein stability that contributes to the stable interactions over a range of temperatures. Since the thermal stability of proteins depends on long-range interactions and epistasis or non-additive influences [76, 56], the embedding provided by Ab-Affinity appears to have captured these complex nonlinear interactions better than generic protein language model embeddings.

While this manuscript was being prepared, an antibody-antigen binding prediction model, A2binder, was made public [26]. A2binder predicts the binding affinity of antibody sequences against various epitopes of SARS-CoV-2's

spike protein. However, there are significant differences between Ab-Affinity and A2binder. At the methodological level, there are two main differences: First, A2binder uses three pre-trained language models for antibody heavy chain, light chain and antigen epitope, respectively, followed by a complex CNN module. Instead, Ab-Affinity uses a single language model with no additional module. Second, A2binder requires antigen sequence for each variant of SARS-CoV-2, whereas Ab-Affinity targets a common peptide in all variants of SARS-CoV-2. There are also significant differences at the experimental level. First, Ab-Affinity provides a meaningful representation (i.e., embedding) of antibodies that align with both binding properties and structural robustness, and can be used to generate the binding landscape of antibodies. Second, Ab-Affinity provides intra-residue attention maps of antibodies which can be used to explain the differences of residue-residue interactions for strong and weak binding.

4.2 The Basis of Robustness of the Generated Antibodies

In our pursuit of designing "better" antibodies by introducing mutations to the seed antibodies, we explored eight generative approaches, employing Ab-Affinity as the metric for evaluation. To establish a baseline, we used a uniform random generator. We tested position-specific scoring matrices (PSSM), genetic algorithms (GA), and simulated annealing (SA). Our methods yielded antibodies with improved predicted binding affinities (up to 168-fold) compared to the seed antibodies, with the exception of the uniform random generation. In fact, the synthetic antibodies had better binding affinity than all the antibodies in the training library. Among all approaches, PSSM shows the least improvement because it relies on the amino acid distribution from the training dataset and introduces randomness (via a temperature parameter) to replace amino acids, resulting in synthetic antibodies whose sequences remain similar to the seed and the training antibodies. In contrast, GA and SA explore the antibody fitness landscape differently: GA's recombination mechanism generates new antibodies whose sequences are substantially different from the seed antibodies over generations, while SA's annealing mechanism iteratively explores the neighbors of the seed antibody in a less abrupt manner than GA. Both approaches control the amino acid distribution of synthetic antibodies through mutation. While PSSM restricts substitution to closely resemble the training antibodies, random substitution generates antibodies with a highly different distribution. By contrast, AbModel's substitution matrix confines synthetic antibodies to align more closely with the antibody sequence database's amino acid frequency distribution. These three substitution strategies enable varying degrees of distribution diversity in the synthetic antibodies, allowing for the generation of novel antibodies compared to the seed and training antibodies. At the same time, the amino acid distributions are not drastically different from existing antibodies, as shown by their concordance with established substitution probability matrices. The analysis across Ab-14, Ab-91, and Ab-95 demonstrates that GA and SA methods, especially when combined with PSSM and ABModel, significantly enhance binding affinities. This suggests that some antibodies that are more distant from the seed have better binding affinity. However, while exploring the landscape, it is critical to move toward the regions with better binding affinities. Both GA and SA incorporate mechanisms to balance exploration and exploitation by selecting weak-binding antibodies and further refining them, ensuring that the generation process consistently progresses toward regions with improved binding antibodies.

Despite remarkable advances in antibody design enabled by LLMs and diffusion models, the challenge of computationally generating antibodies capable of neutralizing pathogens requires further research. Given the propensity of SARS-CoV-2 and other viruses to constantly mutate, later strains might prove to be more virulent and more adept at evading vaccine- or past infection-induced antibodies. It is crucial to not only monitor and predict the trajectory of the virus' evolution but also to design and rapidly develop antibody therapeutics to effectively neutralize novel strains of the virus.

5 Data and Code Availability

The Ab-Affinity model is readily available for use and can be easily installed on any machine via PyPi. For convenience, users can access and install Ab-Affinity directly from the Python Package Index (PyPi) using the following link: https://pypi.org/project/AbAffinity/. This ensures seamless integration into various research environments and workflows. The training and test datasets are publicly available.

References

[1] Wajid Arshad Abbasi, Adiba Yaseen, Fahad Ul Hassan, Saiqa Andleeb, and Fayyaz Ul Amir Afsar Minhas. Island: in-silico proteins binding affinity prediction using sequence information. *BioData Mining*, 13:1–13, 2020.

[2] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pages 1–3, 2024.

[3] Mauricio Aguilar Rangel, Alice Bedwell, Elisa Costanzi, Ross J Taylor, Rosaria Russo, Gonçalo JL Bernardes, Stefano Ricagno, Judith Frydman, Michele Vendruscolo, and Pietro Sormanni. Fragment-based computational design of antibodies targeting structured epitopes. *Science Advances*, 8(45):eabp9540, 2022.

[4] Sarah Alamdari, Nitya Thakkar, Rianne van den Berg, Alex Xijie Lu, Nicolo Fusi, Ava Pardis Amini, and Kevin K Yang. Protein generation with evolutionary diffusion: sequence is all you need. *bioRxiv*, pages 2023–09, 2023.

[5] Justin Barton, Jacob D Galson, and Jinwoo Leem. Enhancing antibody language models with structural information. *bioRxiv*, pages 2023–12, 2024.

[6] Justin Barton, Aretas Gaspariunas, David A Yadin, Jorge Dias, Francesca L Nice, Danielle H Minns, Olivia Snudden, Chelsea Povall, Sara Valle Tomas, Harry Dobson, et al. A generative foundation model for antibody sequence understanding. *bioRxiv*, pages 2024–05, 2024.

[7] Nathaniel R Bennett, Brian Coventry, Inna Goreshnik, Buwei Huang, Aza Allen, Dionne Vafeados, Ying Po Peng, Justas Dauparas, Minkyung Baek, Lance Stewart, et al. Improving de novo protein binder design with deep learning. *Nature Communications*, 14(1):2625, 2023.

[8] Paul J Carter and Arvind Rajpal. Designing antibodies as therapeutics. *Cell*, 185(15):2789–2805, 2022.

[9] Brad Chapman and Jeffrey Chang. Biopython: Python tools for computational biology. *ACM Sigbio Newsletter*, 20(2):15–19, 2000.

[10] Vicky Charitou, Siri C Van Keulen, and Alexandre MJJ Bonvin. Cyclization and docking protocol for cyclic peptide–protein modeling using haddock2. 4. *Journal of Chemical Theory and Computation*, 18(6):4027–4040, 2022.

[11] Jieming Chen, Nicholas Sawyer, and Lynne Regan. Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area. *Protein Science*, 22(4):510–515, 2013.

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[13] Robby Divine, Ha V Dang, George Ueda, Jorge A Fallas, Ivan Vulovic, William Sheffler, Shally Saini, Yan Ting Zhao, Infencia Xavier Raj, Peter A Morawski, et al. Designed proteins assemble antibodies into modular nanocages. *Science*, 372(6537):eabd9994, 2021.

[14] Witold Dyrka and Jean-Christophe Nebel. A stochastic context free grammar based framework for analysis of protein sequences. *BMC bioinformatics*, 10:1–24, 2009.

[15] Peter Eastman, Raimondas Galvelis, Raúl P Peláez, Charlles RA Abreu, Stephen E Farr, Emilio Gallicchio, Anton Gorenko, Michael M Henry, Frank Hu, Jing Huang, et al. Openmm 8: molecular dynamics simulation with machine learning potentials. *The Journal of Physical Chemistry B*, 128(1):109–116, 2023.

[16] Emily Engelhart, Ryan Emerson, Leslie Shing, Chelsea Lennartz, Daniel Guion, Mary Kelley, Charles Lin, Randolph Lopez, David Younger, and Matthew E Walsh. A dataset comprised of binding interactions for 104,972 antibodies against a sars-cov-2 peptide. *Scientific Data*, 9(1):653, 2022.

[17] Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh ewald method. *The Journal of chemical physics*, 103(19):8577–8593, 1995.

[18] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, et al. Protein complex prediction with alphafold-multimer. *biorxiv*, pages 2021–10, 2021.

[19] Juan Ferrando and Lee A Solomon. Recent progress using de novo design to study protein structure, design and binding interactions. *Life*, 11(3):225, 2021.

[20] Noelia Ferruz, Steffen Schmidt, and Birte Höcker. Protgpt2 is a deep unsupervised language model for protein design. *Nature communications*, 13(1):4348, 2022.

[21] Richard J Gowers, Max Linke, Jonathan Barnoud, Tyler John Edward Reddy, Manuel N Melo, Sean L Seyler, Jan Domanski, David L Dotson, Sébastien Buchoux, Ian M Kenney, et al. Mdanalysis: a python package for the rapid analysis of molecular dynamics simulations. Technical report, Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2019.

[22] Jordan Graves, Jacob Byerly, Eduardo Priego, Naren Makkapati, S Vince Parish, Brenda Medellin, and Monica Berrondo. A review of deep learning methods for antibodies. *Antibodies*, 9(2):12, 2020.

[23] Michael Gribskov, Andrew D McLachlan, and David Eisenberg. Profile analysis: detection of distantly related proteins. *Proceedings of the National Academy of Sciences*, 84(13):4355–4358, 1987.

[24] Priyanka Gupta, Emily K Makowski, Sandeep Kumar, Yulei Zhang, Justin M Scheer, and Peter M Tessier. Antibodies with weakly basic isoelectric points minimize trade-offs between formulation and physiological colloidal properties. *Molecular pharmaceutics*, 19(3):775–787, 2022.

[25] Kunchur Guruprasad, BV Bhasker Reddy, and Madhusudan W Pandit. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering, Design and Selection*, 4(2):155–161, 1990.

[26] Haohuai He, Bing He, Lei Guan, Yu Zhao, Feng Jiang, Guanxing Chen, Qingge Zhu, Calvin Yu-Chian Chen, Ting Li, and Jianhua Yao. De novo generation of sars-cov-2 antibody cdrh3 with a pre-trained generative large language model. *Nature Communications*, 15(1):6867, 2024.

[27] Brian Hie, Salvatore Candido, Zeming Lin, Ori Kabeli, Roshan Rao, Nikita Smetanin, Tom Sercu, and Alexander Rives. A high-level programming language for generative protein design. *bioRxiv*, pages 2022–12, 2022.

[28] Brian L Hie, Varun R Shanker, Duo Xu, Theodora UJ Bruun, Payton A Weidenbacher, Shaogeng Tang, Wesley Wu, John E Pak, and Peter S Kim. Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology*, 42(2):275–283, 2024.

[29] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[30] Wengong Jin, Jeremy Wohlwend, Regina Barzilay, and Tommi Jaakkola. Iterative refinement graph neural network for antibody sequence-structure co-design. *arXiv preprint arXiv:2110.04624*, 2021.

[31] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[32] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

[33] Parismita Kalita, Timir Tripathi, and Aditya K Padhi. Computational protein design for covid-19 research and emerging therapeutics. *ACS Central Science*, 9(4):602–613, 2023.

[34] Panagiotis L Kastritis, Iain H Moal, Howook Hwang, Zhiping Weng, Paul A Bates, Alexandre MJJ Bonvin, and Joël Janin. A structure-based benchmark for protein–protein binding affinity. *Protein Science*, 20(3):482–491, 2011.

[35] Jisun Kim, Matthew McFee, Qiao Fang, Osama Abdin, and Philip M Kim. Computational and artificial intelligence-based methods for antibody development. *Trends in Pharmacological Sciences*, 44(3):175–189, 2023.

[36] Jack Kyte and Russell F Doolittle. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157(1):105–132, 1982.

[37] Lin Li, Esther Gupta, John Spaeth, Leslie Shing, Rafael Jaimes, Emily Engelhart, Randolph Lopez, Rajmonda S Caceres, Tristan Bepler, and Matthew E Walsh. Machine learning optimization of candidate antibody yields highly diverse sub-nanomolar affinity antibody libraries. *Nature Communications*, 14(1):3454, 2023.

[38] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[39] Ge Liu, Haoyang Zeng, Jonas Mueller, Brandon Carter, Ziheng Wang, Jonas Schilz, Geraldine Horny, Michael E Birnbaum, Stefan Ewert, and David K Gifford. Antibody complementarity determining region design using high-capacity machine learning. *Bioinformatics*, 36(7):2126–2133, 2020.

[40] Shitong Luo, Yufeng Su, Xingang Peng, Sheng Wang, Jian Peng, and Jianzhu Ma. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. *Advances in Neural Information Processing Systems*, 35:9754–9767, 2022.

[41] Christopher A MacRaild, Jack S Richards, Robin F Anders, and Raymond S Norton. Antibody recognition of disordered antigens. *Structure*, 24(1):148–157, 2016.

[42] Ali Madani, Ben Krause, Eric R Greene, Subu Subramanian, Benjamin P Mohr, James M Holton, Jose Luis Olmos, Caiming Xiong, Zachary Z Sun, Richard Socher, et al. Large language models generate functional protein sequences across diverse families. *Nature Biotechnology*, 41(8):1099–1106, 2023.

[43] Emily K Makowski, Patrick C Kinnunen, Jie Huang, Lina Wu, Matthew D Smith, Tiexin Wang, Alec A Desai, Craig N Streu, Yulei Zhang, Jennifer M Zupancic, et al. Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. *Nature communications*, 13(1):3788, 2022.

[44] Enrique Marcos and Daniel-Adriano Silva. Essentials of de novo protein design: Methods and applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(6):e1374, 2018.

[45] Glenn J Martyna, Douglas J Tobias, and Michael L Klein. Constant pressure molecular dynamics algorithms. *The Journal of chemical physics*, 101(5):4177–4189, 1994.

[46] Derek M Mason, Simon Friedensohn, Cédric R Weber, Christian Jordi, Bastian Wagner, Simon M Meng, Roy A Ehling, Lucia Bonati, Jan Dahinden, Pablo Gainza, et al. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nature Biomedical Engineering*, 5(6):600–612, 2021.

[47] Alexander Mirsky, Linda Kazandjian, and Maria Anisimova. Antibody-specific model of amino acid substitution for immunological inferences from alignments of antibody sequences. *Molecular biology and evolution*, 32(3):806–819, 2015.

[48] Iain H Moal, Rudi Agius, and Paul A Bates. Protein–protein binding affinity prediction on a diverse set of structures. *Bioinformatics*, 27(21):3002–3009, 2011.

[49] Yoochan Myung, Douglas EV Pires, and David B Ascher. mmcsm-ab: guiding rational antibody engineering through multiple point mutations. *Nucleic Acids Research*, 48(W1):W125–W131, 2020.

[50] Yoochan Myung, Douglas EV Pires, and David B Ascher. Csm-ab: graph-based antibody–antigen binding affinity prediction and docking scoring function. *Bioinformatics*, 38(4):1141–1143, 2022.

[51] Richard A Norman, Francesco Ambrosetti, Alexandre MJJ Bonvin, Lucy J Colwell, Sebastian Kelm, Sandeep Kumar, and Konrad Krawczyk. Computational approaches to therapeutic antibody design: established methods and emerging trends. *Briefings in bioinformatics*, 21(5):1549–1567, 2020.

[52] Tobias H Olsen, Iain H Moal, and Charlotte M Deane. Ablang: an antibody language model for completing antibody sequences. *Bioinformatics Advances*, 2(1):vbac046, 2022.

[53] Hakime Öztürk, Arzucan Özgür, and Elif Ozkirimli. Deepdta: deep drug–target binding affinity prediction. *Bioinformatics*, 34(17):i821–i829, 2018.

[54] David Prihoda, Jad Maamary, Andrew Waight, Veronica Juan, Laurence Fayadat-Dilman, Daniel Svozil, and Danny A Bitton. Biophi: A platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning. In *MAbs*, volume 14, page 2020203. Taylor & Francis, 2022.

[55] Roshan M Rao, Joshua Meier, Tom Sercu, Sergey Ovchinnikov, and Alexander Rives. Transformer protein language models are unsupervised structure learners. *bioRxiv*, 2020.

[56] Manfred T Reetz, Pankaj Soni, Juan Pablo Acevedo, and Joaquin Sanchis. Creation of an amino acid network of structurally coupled residues in the directed evolution of a thermostable enzyme. *Angewandte Chemie International Edition*, 48(44):8268–8272, 2009.

[57] Angelo Rosace, Anja Bennett, Marc Oeller, Mie M Mortensen, Laila Sakhnini, Nikolai Lorenzen, Christian Poulsen, and Pietro Sormanni. Automated optimisation of solubility and conformational stability of antibodies and proteins. *Nature communications*, 14(1):1937, 2023.

[58] Peter W Rose, Chunxiao Bi, Wolfgang F Bluhm, Cole H Christie, Dimitris Dimitropoulos, Shuchismita Dutta, Rachel K Green, David S Goodsell, Andreas Prlić, Martha Quesada, et al. The rcsb protein data bank: new resources for research and education. *Nucleic acids research*, 41(D1):D475–D482, 2012.

[59] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215, 2019.

[60] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman JC Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics*, 23(3):327–341, 1977.

[61] Constantin Schneider, Andrew Buchanan, Bruck Taddese, and Charlotte M Deane. Dlab: deep learning methods for structure-based virtual screening of antibodies. *Bioinformatics*, 38(2):377–383, 2022.

[62] Constantin Schneider, Matthew IJ Raybould, and Charlotte M Deane. Sabdab in the age of biotherapeutics: updates including sabdab-nano, the nanobody structure tracker. *Nucleic acids research*, 50(D1):D1368–D1372, 2022.

[63] Andrew W Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander WR Nelson, Alex Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020.

[64] Junhong Shen, Neil Tenenholtz, James Brian Hall, David Alvarez-Melis, and Nicolo Fusi. Tag-llm: Repurposing general-purpose llms for specialized domains. *arXiv preprint arXiv:2402.05140*, 2024.

[65] Richard W Shuai, Jeffrey A Ruffolo, and Jeffrey J Gray. Iglm: Infilling language modeling for antibody sequence design. *Cell Systems*, 14(11):979–989, 2023.

[66] Rohit Singh, Chiho Im, Yu Qiu, Brian Mackness, Abhinav Gupta, Taylor Sorenson, Samuel Sledzieski, Lena Erlach, Maria Wendt, Yves Fomekong Nanfack, et al. Learning the language of antibody hypervariability. *bioRxiv*, pages 2023–04, 2023.

[67] Vladimir N Uversky and Marc HV Van Regenmortel. Mobility and disorder in antibody and antigen binding sites do not prevent immunochemical recognition. *Critical Reviews in Biochemistry and Molecular Biology*, 56(2):149–156, 2021.

[68] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[69] Peter JM Van Laarhoven, Emile HL Aarts, Peter JM van Laarhoven, and Emile HL Aarts. *Simulated annealing*. Springer, 1987.

[70] MHV Van Regenmortel. The concept and operational definition of protein epitopes. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 323(1217):451–466, 1989.

[71] Anna Vangone and Alexandre MJJ Bonvin. Contacts-based prediction of binding affinity in protein–protein complexes. *elife*, 4:e07454, 2015.

[72] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[73] Bo Wang, Sachith Gallolu Kankanamalage, Jianbo Dong, and Yue Liu. Optimization of therapeutic antibodies. *Antibody therapeutics*, 4(1):45–54, 2021.

[74] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

[75] K Yang and AT Brunger. Cryo-em structure of the sars-cov-2 hr1hr2 fusion core complex with extended hr2. electron microscopy data bank, 2022.

[76] Haoran Yu and Paul A Dalby. Coupled molecular dynamics mediate long-and short-range epistasis between mutations that affect stability and aggregation kinetics. *Proceedings of the National Academy of Sciences*, 115(47):E11043–E11052, 2018.

[77] Ye Yuan, Qushuo Chen, Jun Mao, Guipeng Li, and Xiaoyong Pan. Dg-affinity: predicting antigen–antibody affinity with language models from sequences. *BMC bioinformatics*, 24(1):430, 2023.