Pretrainable Geometric Graph Neural Network for Antibody Affinity Maturation

Huiyu Cai[1,2,3†], Zuobai Zhang[2,3†], Mingkai Wang[5,6†], Bozitao Zhong[2,3†], Yanling Wu[5,6*], Tianlei Ying[5,6*] and Jian Tang[1,2,4*]

[1]BioGeometry, Office 1505, 8 Haidianbeierjie St., Beijing, 100083, China.
[2]Mila - Québec AI Institute, 6666 St-Urbain, Montréal, H2S 3H1, Québec, Canada.
[3]Department of Computer Science and Operations Research, Université de Montréal, 2920 chemin de la Tour, Montréal, H3T 1J4, Québec, Canada.
[4]Department of Decision Sciences, HEC Montréal, 3000 chemin de la Côte-Sainte-Catherine, Montréal, H3T 2A7, Québec, Canada.
[5]Shanghai Engineering Research Center for Synthetic Immunology, Fudan University, 131 Dong'an Street, Shanghai, 200032, China.
[6]School of Basic Medical Sciences, Fudan University, 138 Yixueyuan Street, Shanghai, 200031, China.

*Corresponding author(s). E-mail(s): yanlingwu@fudan.edu.cn; tlying@fudan.edu.cn; tangjian@biogeom.com;
Contributing authors: huiyu.cai@mila.quebec; zuobai.zhang@mila.quebec; 20111010061@fudan.edu.cn; bozitao.zhong@mila.quebec;
†These authors contributed equally to this work.

Abstract

In the realm of antibody therapeutics development, increasing the binding affinity of an antibody to its target antigen is a crucial task. This paper presents GearBind, a pretrainable deep neural network designed to be effective for *in silico* affinity maturation. Leveraging multi-level

geometric message passing alongside contrastive pretraining on protein structural data, GearBind capably models the complex interplay of atom-level interactions within protein complexes, surpassing previous state-of-the-art approaches on SKEMPI v2 in terms of Pearson correlation, mean absolute error (MAE) and root mean square error (RMSE). *In silico* experiments elucidate that pretraining helps GearBind become sensitive to mutation-induced binding affinity changes and reflective of amino acid substitution tendency. Using an ensemble model based on pretrained GearBind, we successfully optimize the affinity of CR3022 to the spike (S) protein of the SARS-CoV-2 Omicron strain. Our strategy yields a high success rate with up to 17-fold affinity increase. GearBind proves to be an effective tool in narrowing the search space for *in vitro* antibody affinity maturation, underscoring the utility of geometric deep learning and adept pre-training in macromolecule interaction modeling.

Keywords: GearBind, Geometric Deep Learning, Contrastive pretraining, Antibody Affinity Maturation, Computational Biology

1 Introduction

Antibody plays a crucial role in the human immune system and serves as a powerful diagnostic and therapeutic tool, due to its ability to bind selectively and specifically to target antigens with high affinity. *In vivo*, antibodies go through affinity maturation, where the target-binding affinity gradually increases as a result of somatic hypermutation and clonal selection [1]. When a new antigen surfaces, therapeutic antibody leads repurposed from known antibodies or screened from a natural or *de novo* designed library often require *in vitro* affinity maturation to enhance their binding affinity to a desired, usually sub-nanomolar, level.

Wet lab experimental methods for *in vitro* antibody affinity maturation usually involve several rounds of mutation and selection [2–5]. These methods, while significantly improved during the past few years, are still labor-intensive and costly in general, especially when the combinatorial search space of possible mutations is considered [6]. There are usually 50-60 residues on the complementarity-determining region (CDR) of an antibody, which are hypervariable *in vivo* and contribute to the majority of the binding free energy ΔG_{bind} [7]. Previous works show that multiple point mutations are often needed for successful affinity maturation [8, 9]. Performing experiments on all combinations of the ∼1000 possible point mutations in CDR region per antibody (60 residues × 19 residues per residue) is difficult if not prohibitive. Therefore, a fast and accurate computational method for narrowing down the search space is much desired.

Nevertheless, it is nontrivial for computational affinity maturation methods to balance speed and accuracy. Molecular dynamics methods based on empirical force fields [10, 11] rely on human knowledge and abstractions to

evaluate binding free energy changes after mutations. However, accurate models are often too slow to be used for ranking thousands of mutations (let alone their combinations), and simpler and faster ones fail to fully capture the delicate atomic interactions, resulting in lower reliability. Recent years have demonstrated the potential of machine learning, particularly deep learning, as a formidable tool capable of tackling this complex dilemma. Many machine learning methods [12–17] formulate the affinity maturation problem as a structure-based binding free energy change ($\Delta\Delta G_{\text{bind}} := \Delta G^{(\text{mt})}_{\text{bind}} - \Delta G^{(\text{wt})}_{\text{bind}}$, where *wt* is short for *wild type* and *mt* denotes *mutant*) prediction problem. However, despite the importance of protein side-chain conformation to protein-protein interaction, most existing methods either disregard atom-level information or model it indirectly. These approaches inadequately address the intricate interplay between side-chain atoms. Another critical problem is that machine learning models usually require a massive amount of paired data in order to become accurate and reliable. To the best of our knowledge, the largest publicly available protein-protein binding free energy change dataset, Structural Kinetic and Energetic database of Mutant Protein Interactions (SKEMPI) v2.0 [18], contains only 7085 $\Delta\Delta G_{\text{bind}}$ measurements, a tiny amount compared to the training set sizes of successful deep learning models such as AlphaFold2 [19] and ESM2 [20].

To tackle the aforementioned challenges, we introduce GearBind, a deep neural network that leverages multi-level geometric message passing to model the nuanced antibody-antigen interactions. Furthermore, utilizing contrastive learning-based pretraining on protein structural data allows the incorporation of vital structural insights for predicting $\Delta\Delta G_{\text{bind}}$-values (Fig. 1). Through a series of *in silico* experiments, we validate GearBind's superior performance and demonstrate the benefit of pretraining. GearBind outperforms previous state-of-the-art methods in multiple evaluation metrics. With the help of contrastive pretraining, it is sensitive to binding affinity changes in mutants and reflects amino acid substitution tendency. We then use a GearBind-based pipeline to perform antibody affinity maturation and successfully optimize the binding affinity of antibody CR3022 to the receptor binding domains (RBD) of Delta and the spike (S) protein of Omicron SARS-CoV-2 strains by up to 17 fold. These results underscore the importance of geometric deep learning and effective pretraining on antibody affinity maturation and, more generally, macromolecule interaction modeling.

2 Results

2.1 The GearBind Framework

The GearBind framework is designed to extract geometric representations from wild-type and mutant structures and predict the binding free energy change $\Delta\Delta G_{\text{bind}}$, enhanced with self-supervised pretraining on unlabeled protein structures. Adopting a perspective that extends beyond a singular

focus, GearBind builds a geometric encoder in a multi-relational and multi-granularity manner (Fig. 1b). Specifically, it leverages information within a protein complex at three different levels with complementary insights, namely, atom-level information which renders precise spatial and chemical characteristics, edge-level information which captures critical angular relationships signifying spatial arrangement of atoms, and residue-level information which highlights broader context within sequences of the protein complex. Merging these distinct yet interconnected tiers of information allows a more holistic view of protein complexes and has the potential to substantially enhance the predictive capabilities of binding affinity change.

More formally, to encode a protein, an interface graph is first constructed to model the interactions within the complex, with multiple edge types indicating sequential and spatial proximity between atoms. Atom-level representations can be obtained by applying a geometric relational graph neural network (GearNet [21]) on the graph. On top of that, a line graph is constructed by connecting edges with a common end node atom and encoding angular information between them. Then, edge-level interactions are captured by performing edge message passing, a sparse version of triangle attention [19]. Finally, after aggregating atom and edge representations for each residue, a geometric graph attention layer is applied to pass messages between residues with relative positional information being encoded. This *multi-level message passing* scheme fuses learned representations with detailed structural information, thus making them highly useful for the task of binding energy change prediction.

Although the geometric encoder is able to utilize labeled protein complex structures in $\Delta\Delta G_{\text{bind}}$ datasets, training on a limited set of mutation data could result in overfitting and poor generalization. To address this problem, we further propose a self-supervised pretraining task to exploit large amounts of unlabeled protein structures in PDB [21, 22]. In the pretraining stage, the encoder is trained to model the distribution of these native protein structures via noise contrastive estimation [23]. Specifically, we maximize the probability (*i.e.* push down the energy) of native proteins in PDB while minimizing the probability of randomly mutated structures (Fig. 1c). The mutant structures are generated by randomly mutating residues and randomly rotating side-chain torsional angles according to a rotamer library [24]. This carefully-designed contrastive learning scheme aims to distinguish native, low-energy protein structures from randomly generated mutant structures for identifying beneficial mutations when performing *in silico* binding affinity maturation.

2.2 *In Silico* Validation

We validated GearBind performance via a split-by-complex five-fold cross validation on SKEMPI v2.0. Note that such a splitting method is more realistic than the split-by-mutation method (where the wild-type protein complexes and even the mutation sites in the test set could appear during training) because, in real-world affinity maturation cases, we typically do not have experimental $\Delta\Delta G_{\text{bind}}$ data on the antibody we wish to optimize. In Table 1, we report

results on all single-point mutations in SKEMPI v2.0 following [9]. Leveraging its powerful interface graph construction and multi-level geometric message passing modules, GearBind performs best on 3 of the 4 evaluation metrics (MAE, RMSE and PearsonR), with a remarkable PearsonR of 0.620, surpassing previous state-of-the-art methods. Notably, our observations indicate that pretraining grants significant improvements, enhancing MAE, RMSE, and SpearmanR by 5.6%, 4.8%, and 9.6% respectively, highlighting the advantages our model gains from self-supervised geometric pretraining.

We also present the performance of an ensemble model, where we average the predictions of FoldX, Flex-ddG, Bind-ddG, GearBind and GearBind-P. The ensemble model enjoys the best performance on all four metrics, with its predictions highly correlated with ground-truth free energy changes (PearsonR = 0.683). Hence, we use this model for *in silico* affinity maturation in subsequent experiments.

Table 1: Cross validation performance of different methods on the single-point mutation subset of SKEMPIv2 ($n = 4060$). "+P" means "with geometric pretraining".

Model	MAE↓	RMSE↓	PearsonR↑	SpearmanR↑
FoldX [10]	1.118	1.764	0.430	0.434
Flex-ddG [11]	1.229	1.852	0.512	**0.490**
Bind-ddG [9]	1.070	1.539	0.580	0.356
GearBind	1.023	1.474	**0.620**	0.357
GearBind+P	**0.966**	**1.403**	0.615	0.395
Ensemble	**0.894**	**1.321**	**0.683**	**0.533**

* MAE: Mean average error. RMSE: root mean square error.

2.3 Unsupervised Pretraining Makes GearBind Sensitive to Mutation and Binding Affinity Changes

To showcase the effectiveness of our pretraining approach, we applied the pretrained GearBind to analyze a randomly selected subset of the PDB. The visualizations of protein representations and their corresponding mutants, generated through random mutations, are presented in Fig. 2a. In the figure, we observe a distinct separation between the wild-type and mutant proteins. For example, the two proteins, namely 5ds9 (Fig. 2b) and 5dsd (Fig. 2c), exhibit substantial alterations in their side chain conformation due to mutations after relaxation, resulting in a clear separation depicted on opposite sides of the figure (Fig. 2a). This demonstrates the ability of our pretraining method to distinguish structurally unstable mutations.

Moreover, we selected three protein complexes from SKEMPIv2 with more than 200 mutants, and presented the visualizations of the pretrained GearBind

embeddings for these mutants in Fig. 2d. Remarkably, even in the absence of any supervision from experimental $\Delta\Delta G_{\text{bind}}$ data, our pretraining methodology successfully clusters mutants with both high and low $\Delta\Delta G_{\text{bind}}$ values. This further substantiates the effectiveness of our contrastive pretraining approach, showing that our method has effectively elucidated crucial insights for distinguishing between beneficial and detrimental mutations.

2.4 Analysis of the Amino Acid Substitution Tendency through Pretrained GearBind Embeddings

Based on the pretrained GearBind, we analyzed the propensity of mutations among varying types of residues and scrutinized their congruence with the empirical BLOSUM62 mutation matrix. The propensity of a mutation is measured by the average similarity of the residue embeddings before and after mutations, with higher similarity indicating similar effects of the residue types. Fig. 2e and 2f furnish a side-by-side presentation of the BLOSUM62 and GearBind matrices, offering a clear comparative illustration. Our results identified that several mutations, considered compatible in terms of biochemical traits and in the BLOSUM62 matrix, were similarly accommodated by GearBind. Remarkably, mutations within the two hydroxyl-bearing amino acids, serine and threonine, displayed high acceptance rates in both BLOSUM62 and GearBind. An analogous trend was also observed for the two acidic amino acids (aspartic acid and glutamic acid), three basic amino acids (histidine, arginine, lysine), and three aromatic amino acids (phenylalanine, tyrosine, tryptophan). Furthermore, mutations amongst the four hydrophobic amino acids (methionine, isoleucine, leucine, valine), which were classified as acceptable in the BLOSUM62 matrix, were similarly deemed tolerable by GearBind. These findings suggest that pretraining efficiently unveils the notable mutation propensities among different types of residues, enriching our comprehension of protein structure and function in light of mutations.

2.5 GearBind Successfully Optimizes Binding Affinity of CR3022

To validate the efficacy of our methodology, we chose CR3022, isolated from a convalescent SARS patient [25] but later discovered to also bind to SARS-CoV-2 [26, 27], as the affinity maturation target. We sought to optimize the binding affinity of CR3022 to SARS-CoV-2 through our GearBind-based pipeline.

In the first round of experiment, a total of 12 CR3022 mutants were picked out for experimental validation according to the prediction from the ensemble model. Among these mutants, 10 have single point-mutation in their heavy chains or light chains, and the remaining 2 have double mutations in both heavy and light chains. Firstly, we tested the binding affinity of these mutants to the RBD of SARS-CoV-2 Delta strain with the concentration starting at 100 nM. We found all of the candidates targeting RBD with sub-nanomolar to low nanomolar affinities. (Fig. A3a). 9 of the 12 candidates exhibited improved

binding affinities than the wild-type CR3022 (Fig. A2a). These 9 candidates were picked out for further validation. We reduced the RBD concentration to 10 nM and found all selected mutants with the EC50 values lower than the wild-type CR3022 (Fig. A3b). Furthermore, we observed that multiple point mutations seem to have a synergistic effect on increasing the binding affinity, as most double mutants typically had lower EC50 values than single mutants (Fig. 3a, c).

Next, we further designed and synthesized 8 candidates with double or triple mutations (Fig. A3c). All of the eight CR3022 mutants exhibited outstanding binding ability, with 7 of the 8 having lower EC50 values than the wild-type, especially a triple mutant candidate named SH100D+SH103Y+SL33R, with an EC50 of only 0.06 nM (Fig. 3b, d). These results demonstrated the extraordinary performance of GearBind in antibody affinity optimization.

We next employed the spike (S) protein of the Omicron strain as the target antigen to examine if the antibody candidates possess the desired cross-reactivity. The results (Fig. A3d) show that all of the 8 CR3022 mutants displayed superior binding ability to the Omicron S protein with sub-nanomolar affinities, of which the SH100D+SH103Y+SL33R triple mutant also exhibiting the best performance, with approximately 17-fold increase in affinity compared to the wild type (Fig. 3b, d).

A potential problem for affinity optimization of a cross-reactive antibody is that it may alleviate the binding affinity for other cross-reactive antigens while improving the binding to the target antigen. To evaluate the degree of this side effect, we chose the eight candidates with double or triple mutations and measured their binding to SARS-CoV RBD (Fig. A3e). We found that almost all of the CR3022 mutants demonstrated no significant changes in SARS-CoV RBD binding compared to the wild-type, except for SH103Y+SL33R+TL59H triple mutant. In summary, these above results demonstrate the outstanding prediction specificity of GearBind in antibody affinity optimization.

2.6 Structural Characteristics of Optimized Antibodies

Comprehending the sequence-structure-function relationship of deep learning-designed mutations not only enhances our modeling capabilities but also facilitates the interpretation of their biological significance, providing insight into the fundamental principles that govern binding between antibodies and antigens. To achieve this, we conducted structural analysis on the predicted mutant and wild-type structures.

In accordance with the experimental results presented in Sec. 2.5, we were able to identify five key mutations that showcased a substantial positive influence on the antigen-antibody binding process. These notable mutations comprise S100D and S103Y in the heavy chain, in addition to S33R, I34W, and T59H located within the light chain. Of particular interest, the S103Y mutation prompted the formation of a potential new π-π interaction within the binding interface with the wild-type (WT) receptor-binding domain (RBD)

protein as depicted in Fig. 4c. This unique interaction significantly facilitated the stabilization of the rotamer of mutated residues, thereby resulting in an increase in binding affinity.

Furthermore, our analysis revealed that new polar contacts between the RBD and the antibody were formed in the S100D (Fig. 4b) of the heavy chain and S33R (Fig. 4d), I34W (Fig. 4e) mutations of the light chain. Additionally, the positive charge introduced by the T59H (Fig. 4f) mutation in the light chain and the negative charge in the S100D (Fig. 4b) mutation also contributed to enhancing the binding process. The newly induced charges can aid in binding the RBD and antibody by enhancing electrostatic interactions.

Interestingly, the GearBind contribution as displayed in the corresponding figure further provided insight into the formation of potential contacts for these designed mutations. The contribution obtained using the GearBind tool was found to be consistent with our deductions based on the protein structure, thereby corroborating our findings. The π-π interaction in S103Y was also evident in the contributions, providing further validation to our findings and aligning well with our deductions based on protein structure.

Our findings underscore the importance of understanding the structural and functional relationships of mutations in the antibody-antigen binding process. This knowledge could prove valuable in the development of more effective therapeutic strategies.

3 Discussion

We introduced GearBind, a novel approach for predicting changes in protein-protein binding free energy. Our approach goes byond the common residue-level message passing; it employs atom-level message passing to explicitly model interactions between protein side chains, which play a crucial role in protein binding. It also incorporates edge-level message passing to introduce angular information and model interactions between residue pairs. These novel message-passing schemes are pivotal for accurately predicting $\Delta\Delta G_{\text{bind}}$.

In addition, our method incorporates a unique pretraining algorithm based on contrastive learning. This algorithm harnesses the abundant collection of unlabeled single- and multi-chain protein structures found in the Protein Data Bank (PDB), enabling the model to learn the intrinsic features of protein structures and detect destabilizing structural mutations. Our method surpasses previous approaches on SKEMPIv2 in terms of Pearson correlation, MAE and RMSE, demonstrating its superior performance. Furthermore, when combined with other models in an ensemble, it achieves state-of-the-art performance.

Experimental validation demonstrated GearBind's applicability in antibody binding affinity optimization. As exemplified in the CR3022 case study, GearBind-proposed mutations led to successful enhancements in target protein binding affinity. Notably, 7 out of 10 single site mutations and 9 out of 10 multiple mutations resulted in a significant increase, with up to a 3.82-fold binding increase for the SARS-CoV-2 Delta RBD and a remarkable 17.00-fold

for the Omicron Spike protein. Thus, GearBind proves to be an efficient and powerful tool for the design of antibodies with enhanced binding affinities.

Looking ahead, the potential applications of GearBind reach beyond protein-protein binding optimization. The model can be easily adapted to tackle both protein-protein and protein-ligand docking challenges, thereby fostering potential implementation in minibinder and enzyme design. Nonetheless, we note that the method for mutant structure generation, a prerequisite for structure-based $\Delta\Delta G_{\text{bind}}$ prediction, needs to be further improved to reduce the time cost of the overall pipeline. Future efforts can focus on training a faster (but still accurate) model to predict the mutant structure, or alternatively bypassing the mutant structure modeling step and directly predicting the $\Delta\Delta G_{\text{bind}}$.

In conclusion, our work represents a substantial advancement in protein binding affinity change prediction using geometric deep learning and pretraining techniques. By developing GearBind, we have shown how the integration of atom and edge-level message passing along with a novel pretraining strategy could improve in protein $\Delta\Delta G_{\text{bind}}$ prediction. The experimentally validated success in the CR3022 case study underscores its ability to identify mutations that significantly enhance antibody-antigen binding affinity. Furthermore, the adaptability of GearBind to a broad range of biological tasks reinforces its potential as an indispensable tool for the future.

4 Methods

4.1 Datasets

We used the SKEMPI v2 [18] dataset for training and validation. The dataset contains 7,085 $\Delta\Delta G_{\text{bind}}$ measurements on 348 complexes. We performed preprocessing following [12, 17], discarding data with ambiguous $\Delta\Delta G_{\text{bind}}$ values or high $\Delta\Delta G_{\text{bind}}$ variance across multiple measurements, and arrived at 5,747 distinct mutations, with their $\Delta\Delta G_{\text{bind}}$ measurements, on 340 complexes. For each mutation, we sampled the mutant structure with FoldX 4 [10] based on the wild-type PDB crystal structure. We used PDBFixer v1.8 [28] to fix the PDB structures beforehand if the raw structure could not be processed by FoldX. Mutant structures that could not be read by torchdrug [29] were discarded for fair comparison. The resulting dataset was split into five subsets by PDB complexes with roughly the same size. As in cross validation, we performed inference on each data point from a certain fold with the model trained on the other four folds. We report results on all 4060 single-point mutations of the processed dataset.

For pretraining, we retrieved a dataset of 123,505 experimentally-determined protein structures from the Protein Data Bank (PDB) [30]. To ensure data quality, we retained structures with resolutions ranging from 0.0 to 2.5 angstroms and excluded those with lower resolutions. Both single-chain and multi-chain proteins were utilized for pretraining. Regarding the multi-chain proteins, we randomly selected two chains for each modeling iteration to capture their interaction upon mutations.

4.2 Implementation details of GearBind

Given a pair of wild-type and mutant structures, GearBind predicts the binding free energy change $\Delta\Delta G_{\text{bind}}$ by building a geometric encoder on a multi-relational graph, which is further enhanced by self-supervised pre-training. Note that the key feature that makes the neural network *geometric* is that it considers the spatial relationship between entities, *i.e.*, nodes in a graph. In the following sections, we will discuss the construction of multi-relational graphs, multi-level message passing and pre-training methods.

4.2.1 Constructing relational graphs for protein complex structures

Given a protein-protein complex, we construct a multi-relational graph for its interface and discard all other atoms. Here a residue is considered on the interface if its Euclidean distance to the nearest residue from the binding partner is no more than 6Å. Each atom on the interface is regarded as a node in the graph. We add three types of edges to represent different interactions between these atoms. For two atoms with a sequential distance lower than 3, we add a **sequential edge** between them, the type of which is determined by their relative position in the protein sequence. For two atoms with spatial distance

lower than 5Å, we add a **radial edge** between them. Besides, each atom is also linked to its **10-nearest neighbors** to guarantee the connectivity of the graph. Spatial edges that connect two atoms adjacent in the protein sequence are not interesting and thus discarded. The relational graph is denoted as $(\mathcal{V}, \mathcal{E}, \mathcal{R})$ with $\mathcal{V}$, $\mathcal{E}$, $\mathcal{R}$ denoting the sets of nodes, edges and relation types, respectively. We use the tuple (i, j, r) to denote the edge between atom i and j with type r. We use one-hot vectors of residues types and atom types as node features for each atom and further include sequential and spatial distances in edge features for each edge.

4.2.2 Building geometric encoder by multi-level message passing

On top of the constructed interface graph, we now perform multi-level message passing to model interactions between connected atoms, edges and residues. We use $\boldsymbol{a}_i^{(l)}$ and $\boldsymbol{e}_{(j,i,r)}^{(l)}$ to denote the representations of node i and edge (j, i, r) at the l-th layer. Specially, we use $\boldsymbol{a}_i^{(0)}$ to denote the node feature for atom i and $\boldsymbol{e}_{(j,i,r)}^{(0)}$ to denote the edge feature for edge (j, i, r). Then, the representations are updated by the following procedures:

$$\boldsymbol{a}_i^{(l)} \leftarrow \mathrm{AtomMP}(\boldsymbol{a}_i^{(l-1)}), \tag{1}$$

$$\boldsymbol{e}_{(j,i,r)}^{(l)} \leftarrow \mathrm{EdgeMP}(\boldsymbol{e}_{(j,i,r)}^{(l-1)}), \tag{2}$$

$$\boldsymbol{a}_i^{(l)} \leftarrow \boldsymbol{a}_i^{(l)} + \mathrm{AGGR}(\boldsymbol{e}_{(j,i,r)}^{(l)}), \tag{3}$$

$$\boldsymbol{a}_{\mathrm{C}\alpha(i)}^{(l)} \leftarrow \boldsymbol{a}_{\mathrm{C}\alpha(i)}^{(l)} + \mathrm{ResAttn}(\boldsymbol{a}_{\mathrm{C}\alpha(i)}^{(l)}). \tag{4}$$

First, we perform atom-level message passing (AtomMP) on the atom graph. Then, a line graph is constructed for the message passing between edges (EdgeMP) so as to learn effective representations between atom pairs. The edge representations are used to update atom representations via an aggregation function (AGGR). Finally, we take the representations $\boldsymbol{a}_{\mathrm{C}\alpha(i)}^{(l)}$ of the alpha carbon as residue representation and perform a residue-level attention mechanism (ResAttn), which can be seen as a special kind of message passing on a fully-connected graph. In the following paragraphs, we will discuss these components in details.

Atom-level message passing

Following GearNet [21], we use a relational graph neural network (RGCN) [31] to pass messages between atoms. In a message passing step, each node aggregates messages from its neighbors to update its own representation. The message is computed as the output of a relation (edge type)-specific linear layer when applied to the neighbor representation. Formally, the message passing

step is defined as:

$$\mathrm{AtomMP}(\boldsymbol{a}_i^{(l-1)}) = \boldsymbol{a}_i^{(l-1)} + \sigma\left(\mathrm{BN}\left(\sum\nolimits_{r\in\mathcal{R}} \boldsymbol{W}_r^{(a)} \sum\nolimits_{(j,i,r)\in\mathcal{E}} \boldsymbol{a}_j^{(l-1)}\right)\right),$$

where $\mathrm{BN}(\cdot)$ denotes batch norm and $\sigma(\cdot)$ is the ReLU activation function.

Edge-level message passing and aggregation

Modeling sequential proximity or spatial distance alone is not enough for capturing the complex protein-protein interactions (PPI) contributing to binding. Multiple works have demonstrated the benefits of incorporating angular information using edge-level message passing [19, 21, 32]. Here we construct a line graph [33], *i.e.* a relational graph among all edges of the above atom-level graph. Two edges are connected if and only if they share a common end node. The relations, or edge types, are defined as the angle between the atom-level edge pair, discretized into 8 bins. We use $(\mathcal{V}', \mathcal{E}', \mathcal{R}')$ to denote the constructed line graph. Then, relational message passing is used on the line graph:

$$\mathrm{EdgeMP}(\boldsymbol{e}_x^{(l-1)}) = \sigma\left(\mathrm{BN}\left(\sum\nolimits_{r'\in\mathcal{R}'} \boldsymbol{W}_{r'}^{(e)} \sum\nolimits_{(y,x,r')\in\mathcal{E}'} \boldsymbol{e}_y^{(l-1)}\right)\right),$$

where x and y denote edge tuples in the original graph for abbreviation.

Once we updated the edge representations, we aggregate them into its end nodes. These representations are fed into a linear layer and multiplied with the edge type-specific kernel matrix $\boldsymbol{W}_r^{(a)}$ in AtomMP:

$$\mathrm{AGGR}(\boldsymbol{e}_{(j,i,r)}^{(l)}) = \sigma\left(\mathrm{BN}\left(\sum\nolimits_{r\in\mathcal{R}} \boldsymbol{W}_r^{(a)} \sum\nolimits_{(j,i,r)\in\mathcal{E}} \mathrm{Linear}(\boldsymbol{e}_{(j,i,r)}^{(l)})\right)\right),$$

which will be used to update the representation for atom i as in equation 3.

Residue-level message passing

Constrained by the computational complexity, atom and edge-level message passing only consider sparse interactions while ignoring global interactions between all pairs of residues. By modeling a coarse-grained view of the interface at the residue level, we are able to perform message passing between all pairs of residues. To do this, we design a geometric graph attention mechanism, which takes the representations of the alpha carbon of residues as input and updates their representations with the output as in equation 4. Here we follow the typical definition of self-attention to calculate attention logits with query and key vectors and apply the probability on the value vectors to get residue representations $\boldsymbol{r}_i$:

$$\alpha_{ij} = \mathrm{Softmax}_j\left(\frac{1}{d}\cdot \mathrm{Linear}_q(\boldsymbol{a}_{\mathrm{C}\alpha(i)}^{(l)}) \cdot \mathrm{Linear}_k(\boldsymbol{a}_{\mathrm{C}\alpha(j)}^{(l)})\right),$$

$$\boldsymbol{r}_i^{(l)} = \sum\nolimits_j \alpha_{ij} \cdot \mathrm{Linear}_v(\boldsymbol{a}_{\mathrm{C}\alpha(i)}^{(l)}),$$

where d is the hidden dimension of the representation $\boldsymbol{a}_{\mathrm{C}\alpha(i)}^{(l)}$ and the Softmax function is taken over all j.

Besides traditional self-attention, we also include geometric information in the attention mechanism, which should be invariant to roto-translational transformations on the global complex structure. Therefore, we construct a local frame for each residue with coordinates of its Nitrogen, Carbon and alpha Carbon atoms:

$$\begin{aligned}
&\boldsymbol{v}_{i1} = \boldsymbol{x}_{\mathrm{N}(i)} - \boldsymbol{x}_{\mathrm{C}\alpha(i)}, \\
&\boldsymbol{v}_{i2} = \boldsymbol{x}_{\mathrm{C}(i)} - \boldsymbol{x}_{\mathrm{C}\alpha(i)}, \\
&\boldsymbol{u}_{i1}, \boldsymbol{u}_{i2} = \mathrm{GramSchmidt}(\boldsymbol{v}_{i1}, \boldsymbol{v}_{i2}) \\
&\boldsymbol{R}_i = \left[\frac{\boldsymbol{u}_{i1}}{\|\boldsymbol{u}_{i1}\|}, \frac{\boldsymbol{u}_{i2}}{\|\boldsymbol{u}_{i2}\|}, \frac{\boldsymbol{u}_{i1}}{\|\boldsymbol{u}_{i1}\|} \times \frac{\boldsymbol{u}_{i2}}{\|\boldsymbol{u}_{i2}\|} \right],
\end{aligned}$$

where we use $\boldsymbol{x}$ to denote the coordinate of an atom and $\mathrm{GramSchmidt}(\cdot)$ refers to the Gram–Schmidt process for orthogonalization. Then, the geometric attention is designed to model the relative position of beta carbons of all residues j in the local frame of residue i:

$$\boldsymbol{p}_i^{(l)} = \sum\nolimits_j \alpha_{ij} \boldsymbol{R}_i^\top (\boldsymbol{x}_{\mathrm{C}\alpha(i)} - \boldsymbol{x}_{\mathrm{C}\beta(j)}),$$

where $\boldsymbol{p}_i^{(l)}$ is the spatial representations for the residue i. When the complex structure is rotated, the frame $\boldsymbol{R}_i$ and relative position $\boldsymbol{x}_{\mathrm{C}\alpha(i)} - \boldsymbol{x}_{\mathrm{C}\beta(j)}$ are rotated accordingly and the effects will be counteracted, which guarantees the rotation invariance of our model.

The final output is the concatenation of residue representations $\boldsymbol{r}_i^{(l)}$ and spatial representations $\boldsymbol{p}_i^{(l)}$:

$$\mathrm{ResAttn}(\boldsymbol{a}_{\mathrm{C}\alpha(i)}^{(l)}) = \mathrm{Concat}(\boldsymbol{r}_i^{(l)}, \boldsymbol{p}_i^{(l)}).$$

After obtaining representations for each atom, we apply a mean pooling layer over representations of all alpha carbons $\boldsymbol{a}_{\mathrm{C}\alpha(i)}$ to get protein representations $\boldsymbol{h}$. An anti-symmetric prediction head is then applied to guarantee that back mutations would have the exact opposite predicted $\Delta\Delta G_{\mathrm{bind}}$ values:

$$\widetilde{\Delta\Delta G_{\mathrm{bind}}} = \mathrm{MLP}(\boldsymbol{h}^{(\mathrm{wt})}, \boldsymbol{h}^{(\mathrm{mt})}) - \mathrm{MLP}(\boldsymbol{h}^{(\mathrm{mt})}, \boldsymbol{h}^{(\mathrm{wt})}), \tag{5}$$

where $\boldsymbol{h}^{(\mathrm{wt})}$ and $\boldsymbol{h}^{(\mathrm{mt})}$ denote the representations for wild type and mutant complexes and $\widetilde{\Delta\Delta G_{\mathrm{bind}}}$ is the predicted $\Delta\Delta G_{\mathrm{bind}}$ from our GearBind model.

4.2.3 Modeling energy landscape of proteins via noise contrastive estimation.

As paired binding free energy change data is of relatively small size, it would be beneficial to pretrain GearBind with massive protein structural data. The high-level idea of our pretraining method is to model the distribution of native protein structures, which helps identify harmful mutations yielding unnatural structures. Denoting a protein structure as $\boldsymbol{x}$, its distribution can be modeled with Boltzmann distribution as:

$$p(\boldsymbol{x};\ \boldsymbol{\theta}) = \frac{\exp(-E(\boldsymbol{x};\ \boldsymbol{\theta}))}{A(\boldsymbol{\theta})},\ A(\boldsymbol{\theta}) = \int \exp(-E(\boldsymbol{x};\ \boldsymbol{\theta}))d\boldsymbol{x}, \tag{6}$$

where $\boldsymbol{\theta}$ denotes learnable parameters in our encoder, $E(\boldsymbol{x};\ \boldsymbol{\theta})$ denotes the energy function for the protein x and $A(\boldsymbol{\theta})$ is the partition function to normalize the distribution. The energy function is predicted by applying a linear layer on the GearBind representations $\boldsymbol{h}(\boldsymbol{x})$ of protein $\boldsymbol{x}$:

$$E(\boldsymbol{x};\ \boldsymbol{\theta}) = \text{Linear}(\boldsymbol{h}(\boldsymbol{x})). \tag{7}$$

Given the observed dataset $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_T\}$ from PDB, our objective is to maximize the probability of these samples:

$$\text{maximize}\ \ \frac{1}{2T}\sum\nolimits_t \log p(\boldsymbol{x}_t;\ \boldsymbol{\theta}). \tag{8}$$

However, direct optimization of this objective is intractable, since calculating the partition function requires integration over the whole protein structure space. To address this issue, we adopt a popular method for learning energy-based models called noise contrastive estimation [23]. For each observed structure $\boldsymbol{x}_t$, we sample a negative structure $\boldsymbol{y}_t$ and then the problem can be transformed to a binary classification task, *i.e.*, whether a sample is observed in the dataset or not.

$$\text{minimize}\ \ \frac{1}{2T}\sum\nolimits_t \log\left[\sigma(E(\boldsymbol{x}_t;\ \boldsymbol{\theta}) - E(\boldsymbol{y}_t;\ \boldsymbol{\theta}))\right], \tag{9}$$

where $\sigma(\cdot)$ denotes the sigmoid function for calculating the probability for a sample $\boldsymbol{x}_t$ belonging to the positive class. We could see that the above training objective tries to push down the energy of the positive examples (*i.e.* the observed structures) while pushing up the energy of the negative samples (*i.e.* the mutant structures.

For negative sampling, we perform random single-point mutations on the corresponding positive samples and then generate its conformation by keeping the backbone unchanged and sampling side-chain torsional angles at the mutation site from a backbone-dependent rotamer library [24].

After pretraining on the PDB database, we adopt the GearBind encoder for extracting protein complex representations and then train a gradient boosting tree on downstream datasets for prediction to avoid overfitting.

4.3 Training details.

We implement our model with the TorchDrug library [29]. For message passing, we employed a 4-layer GearBind model with a hidden dimension of 128. Regarding edge message passing, the connections between edges are categorized into 8 bins according to the angles between them. To predict the $\Delta\Delta G_{\text{bind}}$ value from graph representations, we utilized a 2-layer MLP.

The model was trained using the Adam optimizer with a learning rate of 1e-4 and a batch size of 8. The training process is performed on 1 A100 GPU for 40 epochs. For pretraining, we increased the hidden dimension to 512 to enhance the model's capacity. The pretraining was conducted using the Adam optimizer with a learning rate of 5e-4 and a batch size of 8, employing 4 A100 GPUs for 10 epochs.

4.4 Mutation tendency between different residue types

To estimate the tendency of mutation between different residue types, we randomly selected 5,000 proteins from the PDB database and introduced random mutations at various mutation sites. The average cosine similarity between residue representations before and after the mutations is utilized as a measure of the tendency between the two residue types. A high similarity indicates that the two residue types perform similar roles in that particular position, thereby reflecting distinct patterns associated with different residue types.

4.5 In silico affinity maturation of antibody CR3022

We chose PDB 6XC3 [34], in which chains H and L comprise antibody CR3022 and chain C is the SARS-CoV-2 RBD, as the starting complex. We note that potentially, structure prediction tools such as AlphaFold-Multimer [35] can be utilized to predict the complex structure. To better simulate the CR3022 interaction with Omicron RBD, We constructed the complex structures for BA.4 and BA.1.1 mutants with SWISS-MODEL [36]. We then performed saturation mutagenesis on the CDRs of CR3022 and generated mutant structures using FoldX [10] and Flex-ddG [11]. Specifically, residues 26-35, 50-66, 99-108 from the heavy chain and residues 24-40, 56-62, 95-103 from the light chain are mutated. This totals 1400 single-point mutations (if we count in the self-mutations). We use our ensemble model to rank the mutations and select the top-ranked mutants for synthesis and subsequent experimental validation. Mutations are ranked by the modified z-score (where values are subtracted by the median rather than the mean to be less sensitive to outliers) averaged across multiple $\Delta\Delta G_{\text{bind}}$ prediction methods.

4.6 Protein expression and purification

The gene encoding SARS-CoV RBD was synthesized by Genscript (Nanjing, China) and subcloned into pSectag 2B vector with C-terminal human IgG1 Fc fragment and AviTag. The recombinant vector was transfected to Expi 293 cells and cultured in 37 ℃ for 5 days, followed by centrifugation at 4,000 rpm for 20 minutes. The supernatant was harvested and filtered through a 0.22 µm vacuum filter. The protein G resin (Genscript) was loaded into column, washed by PBS and flow the supernatant through to fully combine the resin. Then the targeted protein was eluted by 0.1 M glycine (pH 3.0) and neutralized with 1 M Tris-HCL (pH 9.0), followed by buffer-exchanged and concentrated with PBS using an Amicon ultra centrifugal concentrator (Millipore) with a molecular weight cut-off of 3 kDa. Protein concentration was measured using the NanoDrop 2000 spectrophotometer (Thermo Fisher), and protein purity was examined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The Delta RBD protein were purchased from Vazyme (Nanjing, China) and Omicron S protein were purchased from ACROBiosystems (Beijing, China).

4.7 Preparation for mutant and wild-type CR3022 antibodies

The heavy chain and light chain genes of different CR3022 antibodies were synthesized and subcloned into expression vector pcDNA 3.4 in IgG1 format. These constructed vectors were transfected into CHO cells and purified by Protein A. All antibodies were produced by Biointron Biological Inc (Shanghai, China).

4.8 Enzyme-linked immunosorbent assay (ELISA)

The receptor-binding domain (RBD) of Delta (B.1.617.2) strain and spike (S) protein of Omicron (B.1.1.529) strain at 100 ng per well was coated in 96 wells half area microplate (Corning #3690) overnight at 4 ℃. The antigen coated plate was washed by three times of PBST (PBS with 0.05% Tween-20) and blocked with 3% MPBS (PBS with 3% skim milk) in 37 ℃ for 1 h. Following by three times washing with PBST, 50 µL of three-fold serially diluted antibody in 1% MPBS was added and incubated at 37 ℃ for 1.5 h. The HRP-conjugated anti-Fab and anti-Fc (Sigma-Aldrich) secondary antibodies was used for detection of different tested antibodies. After washing with PBST for 5 times, the enzyme activity was measured after addition of ABTS substrate (Invitrogen) for 15 min. The data was acquired by measuring the absorbance at 405 nm using a Microplate Spectrophotometer (Biotek) and the EC50 (concentration for 50% of maximal effect) was calculated by Graphpad Prism8.0 software.

4.9 Protein structure analysis

Protein structure analysis is conducted by python scripts. The antibody-antigen complex structure after mutation was obtained from Rosetta relaxation. The relaxed protein structure can provide more accurate side chain conformations, which are critical for accurate contact and conformational analysis. The improved accuracy of such analyses enables a deeper understanding of the underlying binding mechanisms and can facilitate the identification of key characteristics involved in protein-protein interactions. The attribution scores are derived by using Integrated Gradients (IG) [37], a model-agnostic attribution method, on our model to obtain residue-level interpretation following [21].

4.10 Data & Code Availability

See the "Methods > Datasets" section for detail on data collection and preprocessing. We plan to make our code publicly available on GitHub upon acceptance of this paper.

Appendix A Appendix

References

[1] Victora, G.D., Nussenzweig, M.C.: Germinal centers. Annual review of immunology **40**, 413–442 (2022)

[2] Gram, H., Marconi, L.-A., Barbas 3rd, C., Collet, T.A., Lerner, R.A., Kang, A.S.: In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. Proceedings of the National Academy of Sciences **89**(8), 3576–3580 (1992)

[3] Jackson, J.R., Sathe, G., Rosenberg, M., Sweet, R.: In vitro antibody maturation. improvement of a high affinity, neutralizing antibody against il-1 beta. Journal of immunology (Baltimore, Md.: 1950) **154**(7), 3310–3319 (1995)

[4] Li, B., Fouts, A.E., Stengel, K., Luan, P., Dillon, M., Liang, W.-C., Feierbach, B., Kelley, R.F., Hötzel, I.: In vitro affinity maturation of a natural human antibody overcomes a barrier to in vivo affinity maturation. In: MAbs, vol. 6, pp. 437–445 (2014). Taylor & Francis

[5] Wang, J., An, L., Zhao, Y., Zhang, C., Li, S., Ye, C., Jing, S., Hang, H.: In vitro affinity maturation of antibody against membrane-bound gpcr molecules. Applied microbiology and biotechnology **103**, 7703–7717 (2019)

[6] Tabasinezhad, M., Talebkhan, Y., Wenzel, W., Rahimi, H., Omidinia, E., Mahboudi, F.: Trends in therapeutic antibody affinity maturation: From in-vitro towards next-generation sequencing approaches. Immunology Letters **212**, 106–113 (2019)

[7] Murphy, K., Weaver, C.: Janeway's Immunobiology. Garland science, ??? (2016)

[8] Sulea, T., Hussack, G., Ryan, S., Tanha, J., Purisima, E.O.: Application of assisted design of antibody and protein therapeutics (adapt) improves efficacy of a clostridium difficile toxin a single-domain antibody. Scientific reports **8**(1), 2260 (2018)

[9] Shan, S., Luo, S., Yang, Z., Hong, J., Su, Y., Ding, F., Fu, L., Li, C., Chen, P., Ma, J., *et al.*: Deep learning guided optimization of human antibody against sars-cov-2 variants with broad neutralization. Proceedings of the National Academy of Sciences **119**(11), 2122954119 (2022)

[10] Delgado, J., Radusky, L.G., Cianferoni, D., Serrano, L.: Foldx 5.0: working with rna, small molecules and a new graphical interface. Bioinformatics

35(20), 4168–4169 (2019)

[11] Barlow, K.A., Ó Conchúir, S., Thompson, S., Suresh, P., Lucas, J.E., Heinonen, M., Kortemme, T.: Flex ddg: Rosetta ensemble-based estimation of changes in protein–protein binding affinity upon mutation. The Journal of Physical Chemistry B **122**(21), 5389–5399 (2018)

[12] Rodrigues, C.H., Myung, Y., Pires, D.E., Ascher, D.B.: mcsm-ppi2: predicting the effects of mutations on protein–protein interactions. Nucleic acids research **47**(W1), 338–344 (2019)

[13] Myung, Y., Rodrigues, C.H., Ascher, D.B., Pires, D.E.: mcsm-ab2: guiding rational antibody design using graph-based signatures. Bioinformatics **36**(5), 1453–1459 (2020)

[14] Petukh, M., Dai, L., Alexov, E.: Saambe: webserver to predict the charge of binding free energy caused by amino acids mutations. International journal of molecular sciences **17**(4), 547 (2016)

[15] Huang, X., Zheng, W., Pearce, R., Zhang, Y.: Ssipe: accurately estimating protein–protein binding affinity change upon mutations using evolutionary profiles in combination with an optimized physical energy function. Bioinformatics **36**(8), 2429–2437 (2020)

[16] Jiang, Y., Quan, L., Li, K., Li, Y., Zhou, Y., Wu, T., Lyu, Q.: Dgcddg: Deep graph convolution for predicting protein-protein binding affinity changes upon mutations. IEEE/ACM Transactions on Computational Biology and Bioinformatics (2023)

[17] Liu, X., Luo, Y., Li, P., Song, S., Peng, J.: Deep geometric representations for modeling effects of mutations on protein-protein binding affinity. PLoS computational biology **17**(8), 1009284 (2021)

[18] Jankauskaitė, J., Jiménez-García, B., Dapkūnas, J., Fernández-Recio, J., Moal, I.H.: SKEMPI 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. Bioinformatics **35**(3), 462–469 (2018) https://academic.oup.com/bioinformatics/article-pdf/35/3/462/27700009/bty635.pdf. https://doi.org/10.1093/bioinformatics/bty635

[19] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., *et al.*: Highly accurate protein structure prediction with alphafold. Nature **596**(7873), 583–589 (2021)

[20] Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil,

R., Kabeli, O., Shmueli, Y., *et al.*: Evolutionary-scale prediction of atomic-level protein structure with a language model. Science **379**(6637), 1123–1130 (2023)

[21] Zhang, Z., Xu, M., Jamasb, A., Chenthamarakshan, V., Lozano, A., Das, P., Tang, J.: Protein representation learning by geometric structure pretraining. arXiv preprint arXiv:2203.06125 (2022)

[22] Zhang, Z., Xu, M., Lozano, A., Chenthamarakshan, V., Das, P., Tang, J.: Physics-inspired protein encoder pre-training via siamese sequence-structure diffusion trajectory prediction. arXiv preprint arXiv:2301.12068 (2023)

[23] Gutmann, M., Hyvärinen, A.: Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 297–304 (2010). JMLR Workshop and Conference Proceedings

[24] Shapovalov, M.V., Dunbrack, R.L.: A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure **19**(6), 844–858 (2011)

[25] Ter Meulen, J., Van Den Brink, E.N., Poon, L.L.M., Marissen, W.E., Leung, C.S.W., Cox, F., Cheung, C.Y., Bakker, A.Q., Bogaards, J.A., Van Deventer, E., *et al.*: Human monoclonal antibody combination against sars coronavirus: synergy and coverage of escape mutants. PLoS medicine **3**(7), 237 (2006)

[26] Tian, X., Li, C., Huang, A., Xia, S., Lu, S., Shi, Z., Lu, L., Jiang, S., Yang, Z., Wu, Y., Ying, T.: Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. Emerg Microbes Infect **9**(1), 382–385 (2020)

[27] Yuan, M., Wu, N.C., Zhu, X., Lee, C.-C.D., So, R.T.Y., Lv, H., Mok, C.K.P., Wilson, I.A.: A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science **368**(6491), 630–633 (2020)

[28] Eastman, P.: PDBFixer. GitHub (2013)

[29] Zhu, Z., Shi, C., Zhang, Z., Liu, S., Xu, M., Yuan, X., Zhang, Y., Chen, J., Cai, H., Lu, J., et al.: Torchdrug: A powerful and flexible machine learning platform for drug discovery. arXiv preprint arXiv:2202.08320 (2022)

[30] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The protein data bank. Nucleic acids

research **28**(1), 235–242 (2000)

[31] Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I., Welling, M.: Modeling relational data with graph convolutional networks. In: European Semantic Web Conference, pp. 593–607 (2018). Springer

[32] Gasteiger, J., Groß, J., Günnemann, S.: Directional message passing for molecular graphs. In: International Conference on Learning Representations (ICLR) (2020)

[33] Harary, F., Norman, R.Z.: Some properties of line digraphs. Rendiconti del Circolo Matematico di Palermo **9**(2), 161–168 (1960). https://doi.org/10.1007/BF02854581

[34] Yuan, M., Liu, H., Wu, N.C., Lee, C.-C.D., Zhu, X., Zhao, F., Huang, D., Yu, W., Hua, Y., Tien, H., *et al.*: Structural basis of a shared antibody response to sars-cov-2. Science **369**(6507), 1119–1123 (2020)

[35] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al.: Protein complex prediction with alphafold-multimer. BioRxiv, 2021–10 (2021)

[36] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., *et al.*: Swiss-model: homology modelling of protein structures and complexes. Nucleic acids research **46**(W1), 296–303 (2018)

[37] Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. In: International Conference on Machine Learning, pp. 3319–3328 (2017). PMLR

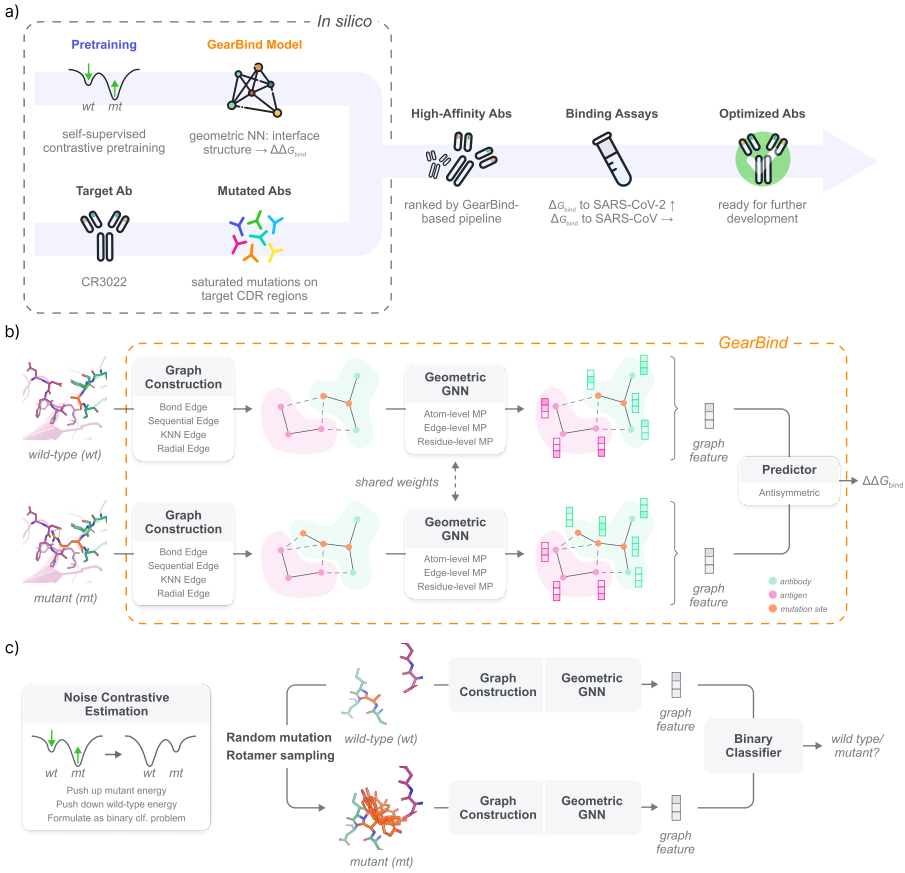

Fig. 1: **Geometric deep learning-based platform for antibody affinity maturation.** (a) Pipeline Overview: The pipeline utilizes the geometric encoder, GearBind, which undergoes pretraining on PDB and supervised learning on SKEMPIv2. It is employed to select mutations on a target antibody that can enhance binding affinity with the antigen. Multiple rounds of antibody optimization, guided by the model's predictions, result in antibodies with improved binding affinity to SARS-CoV-2 and conserved binding affinity to SARS-CoV, ready for further development. NN: neural network. (b) GearBind Model: The model employs a shared graph neural network to encode both the wild-type and mutant complex structures. This network facilitates multi-relational and multi-level message passing to capture interactions between atoms at different levels of granularity. The mutational effect $\Delta\Delta G_{\text{bind}}$ is predicted by combining the representations of the two complexes. (c) Self-supervised Pretraining on Unlabeled Protein Structures: The model undergoes self-supervised pretraining using unlabeled protein structures. A contrastive learning loss function is introduced, which models the energy landscape of native protein structures. This loss function maximizes the probability of native structures while minimizing the probability of randomly mutated structures.

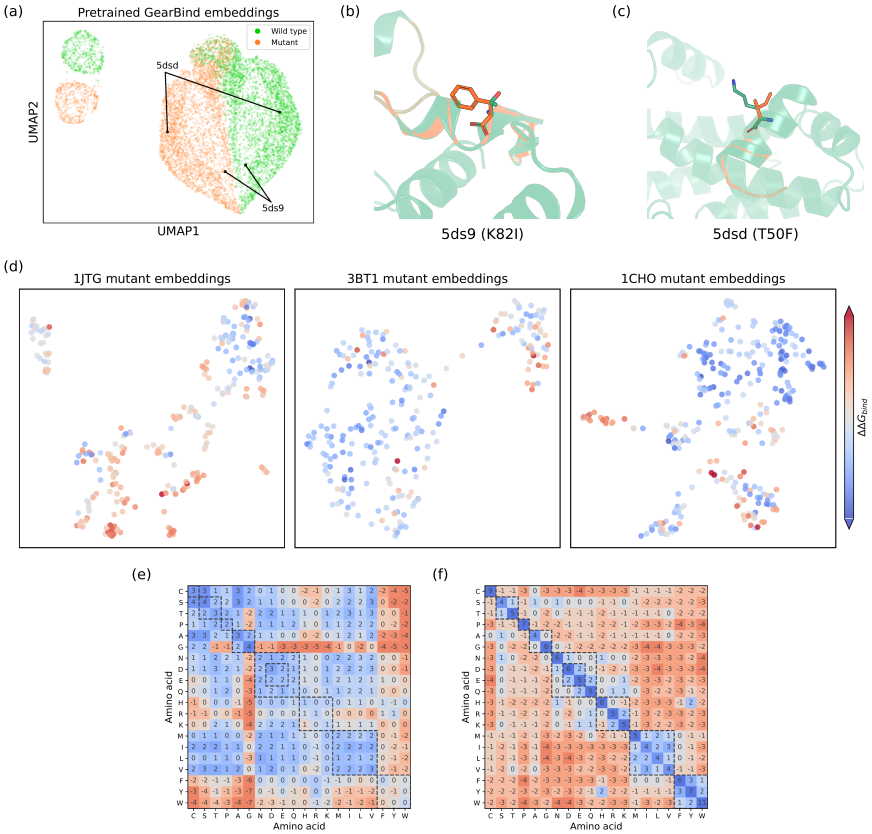

Fig. 2: **Evaluation of the pretrained GearBind method.** (a) Visualization of pretrained GearBind protein representations on a set of wild-type structures (green, $n = 3919$) and their random mutants (orange) in PDB. (b-c) show the structural changes upon mutations of two examples, 5ds9 (b) and 5dsd (c). (d) Visualization of pretrained GearBind representations of mutants of three proteins (1jtg, 3bt1 and 1cho) in SKEMPI colored by $\Delta\Delta G_{\text{bind}}$. (e-f) Similarity matrix for (e) BLOSUM62 and (f) GearBind. Colors signify acceptability of mutations with blue indicating a higher acceptability and red signifying lower acceptability. The black dashed rectangles delineate clusters of amino acids exhibiting similar chemical properties, namely acidic, basic, hydrophobic, and aromatic characteristics.

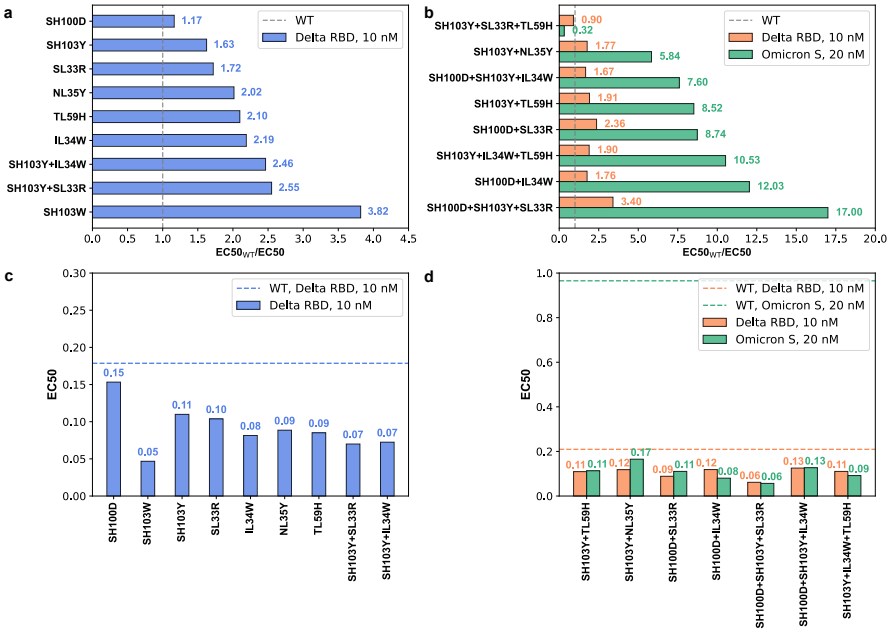

Fig. 3: **ELISA binding assay results for GearBind designed mutations of the CR3022 antibody.** Lower EC50 values indicate stronger antigen-antibody interactions. EC50 are denoted next to the respective bars. We present the binding affinity fold increase calculated by $EC50_{WT}/EC50$ for first round (a) and second round design (b), wild type is denoted in gray dashed line. For (a) and (b), fold increase change are denoted next to the respective bars. (c) illustrates the EC50 for selected mutants in first round with the 10 nM Delta RBD, wild type is denoted in blue dashed line. (d) explores the EC50 for selected second round double and triple site mutations with the 10 nM Delta RBD and 20 nM Omicron S; results are differentiated by blue (10 nM Delta RBD) and green (20 nM Omicron S), wild type is denoted in dashed line with corresponding color. For (c) and (d), EC50 values are displayed next to the respective bars.

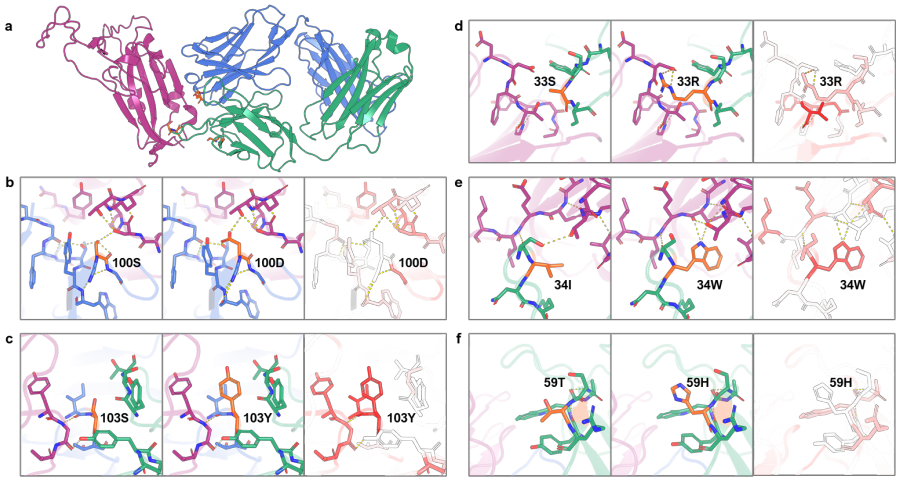

Fig. 4: **Comparative analysis of wild type and mutant protein structures in antibody CR3022 mutation sites.** Target RBD protein is represented in magenta, with the heavy chain in blue, and light chain in green, and mutations marked in orange. (a) Overview of all 5 effective mutations sites predicted by GearBind. Figures (b-f) present triptych comparisons between wild type structures (left), mutant structures (middle), and the GearBind predicted contribution of each residue in mutant structure (right). High contributions are denoted in red, while lower contributions are marked white. Specific mutations include: (b) S100D, (c) S103Y in the heavy chain, and (d) SL33R, (e) IL34W, (f) TL59H in the light chain.

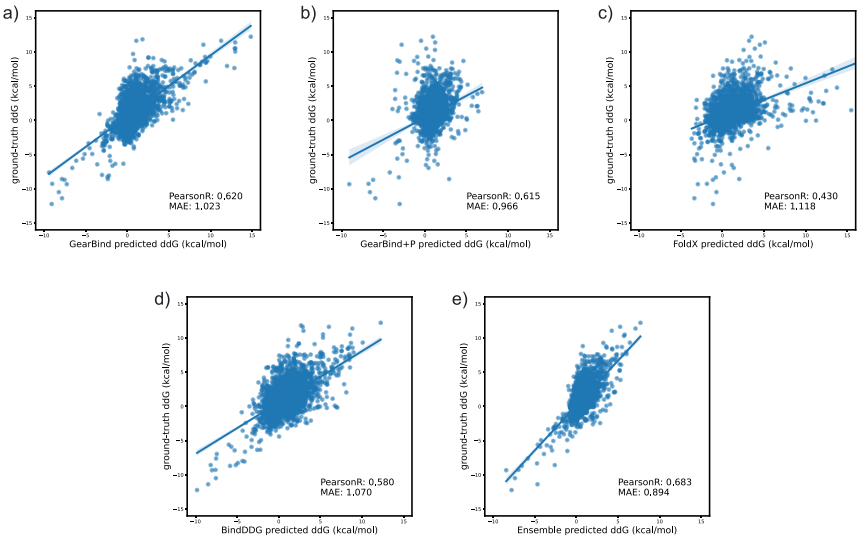

Fig. A1: Scatter plots of ground truth $\Delta\Delta G_{\text{bind}}$'s versus predicted $\Delta\Delta G_{\text{bind}}$'s of different models trained on SKEMPIv2 and tested on its single-mutant subset with split-by-complex cross validation ($n = 4060$). Linear regression model fits are plotted with the data, with the light blue region around the line representing the 95% confidence interval. Pearson correlation and mean absolute error (MAE) values are annotated in each plot. The prediction of the ensemble model aligns best with the experimental data.

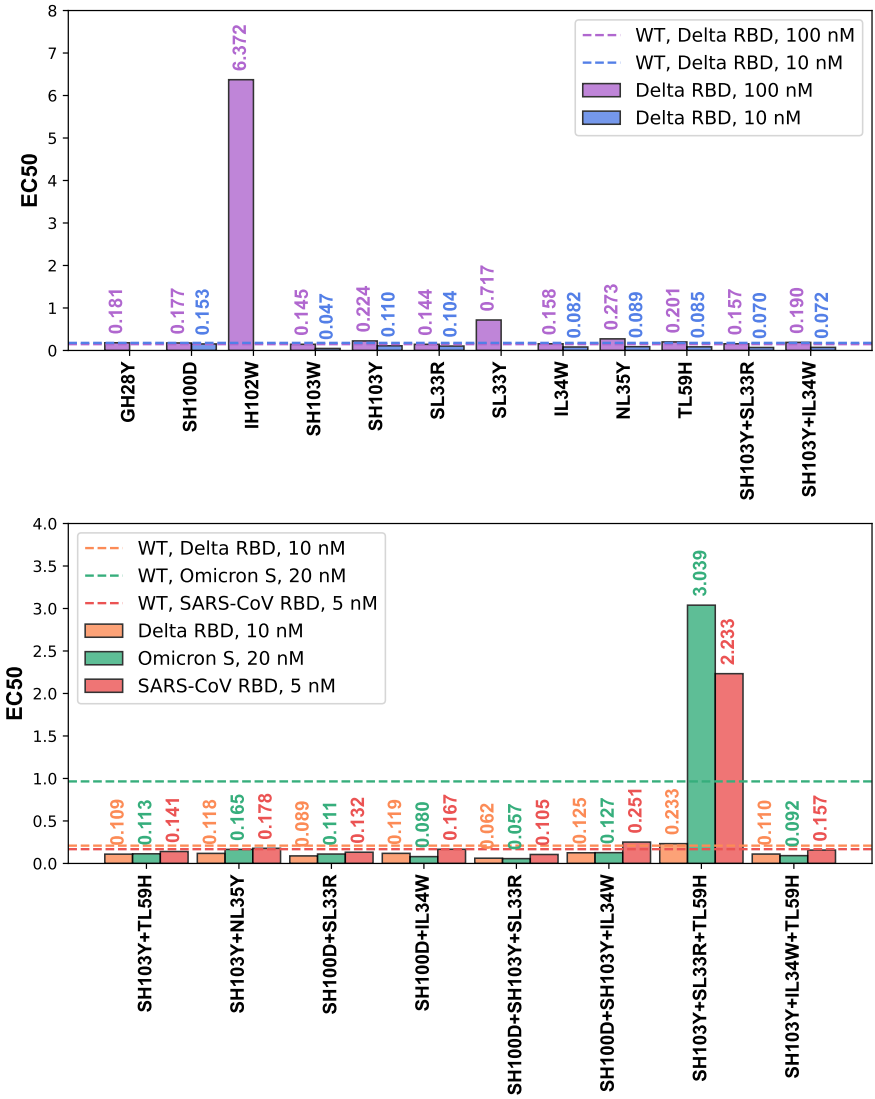

Fig. A2: **ELISA binding assay EC50 results for all GearBind designed mutations of the CR3022 antibody.** Lower EC50 values indicate stronger antigen-antibody interactions. EC50 values are denoted next to the respective bars. (a) illustrates the EC50 for selected mutants in the first round with the 100 nM Delta RBD (purple) and 10 nM Delta RBD (blue). (b) explores the EC50 for all second-round double and triple mutations with the 10 nM Delta RBD (orange), 20 nM Omicron S (green) and 5 nM SARS-CoV RBD (red), wild type is denoted in dashed line with corresponding color.

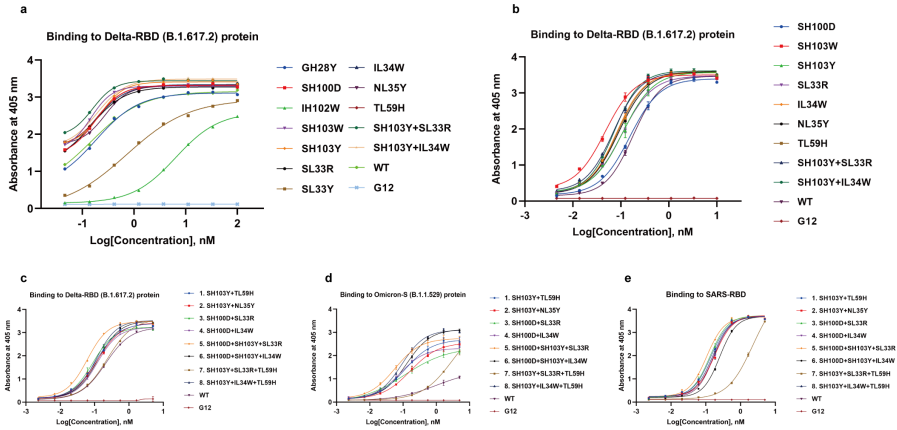

Fig. A3: **Concentration-response curve of GearBind designed antibody binding to the target antigens.** (a) Enzyme-linked immunosorbent assay test of the predictive CR3022 mutants binding with SARS-CoV-2 delta RBD. The initial concentration of each tested antibody was 100 nM. G12 was a HBsAg-reactive neutralizing antibody applied as the negative control in this experiment. (b) Further confirmation of the binding affinity of different mutants tested in a. The initial concentration was further decreased to 10 nM. (c) Binding validation of the second batch of CR3022 mutants with double or triple mutations predicted by GearBind to SARS-CoV-2 delta RBD. (d) Binding validation of the second batch of CR3022 mutants with SARS-CoV-2 omicron spike (S) protein. The initial concentration of each antibody was 20 nM. (e) Evaluation of the optimization specificity of GearBind by targeting SARS-CoV RBD using the second batch of CR3022 mutants. The initial concentration was 5 nM.