

---

# Does Structural Information Improve ESM3 for Protein Binding Affinity Prediction?

---

Thomas Loux<sup>1,3</sup>, Dianzhuo Wang<sup>2,3</sup>, Eugene I. Shakhnovich<sup>3,†</sup>

<sup>1</sup>Ecole Polytechnique, Institut Polytechnique de Paris

<sup>2</sup>John A. Paulson School of Engineering and Applied Sciences, Harvard University

<sup>3</sup>Department of Chemistry and Chemical Biology, Harvard University

†shakhnovich@chemistry.harvard.edu.

## Abstract

This paper investigates the impact of incorporating structural information into the protein-protein interaction predictions made by ESM3, a multimodal protein language model (pLM). We utilized various structural variants as inputs and compared three widely used structure acquisition pipelines—EvoEF2, Gromacs, and Rosetta Relax—to assess their effects on ESM3’s performance. Our findings reveal that the use of a consistent identical structure, regardless of whether it is relaxed or variant, consistently enhances model performance across various datasets. This improvement is striking in few-shot learning. However, performance deteriorates when different relaxed mutant structures are used for each variant. Based on these results, we advise caution when integrating distinct mutant structures into ESM3 and similar models. This study highlights the critical need for careful consideration of structural inputs in protein binding affinity prediction.

## 1 Introduction

The global COVID-19 pandemic has highlighted the critical role of protein-protein interactions in virus behavior, particularly through mutations such as amino acid substitutions. These mutations can disrupt existing interactions and alter binding energies  $\Delta\Delta G = \Delta G_{MUT} - \Delta G_{WT}$ , impacting the structures and functions of proteins. The understanding of these dynamics, especially the interactions between the receptor binding domain (RBD) and antibodies or Angiotensin-converting enzyme 2 (ACE2), is crucial due to their implications in viral infectivity and evasion mechanisms [1].

Traditional experimental methods, including high-throughput experiments like deep mutational scan [2], while reliable, are often limited by their time-intensive and costly nature, making them impractical in urgent pandemic responses. This has led to an increased reliance on computational methods, which have shown to be invaluable not only in identifying potential viral threats but also in facilitating the rapid development and optimization of therapeutic antibodies [3] [4].

Computational approaches to studying protein dynamics can be broadly categorized into traditional biophysical methods and machine learning or deep learning techniques. Traditional biophysical methods, such as those utilizing force fields in tools like FoldX [5] and Rosetta [6], relying on empirical interatomic interactions to predict folding stability or binding affinity. While effective, the highest accuracy in these methods often demands extensive computational resources, such as those required for Free Energy Perturbation [7].

Machine learning models and pLMs become a popular alternative to these resource-heavy physical simulations. Despite the inability to validate variant structures, the integration of 3D structural data into machine learning models is transforming the landscape of PPI prediction. Different models use different relaxation methods to obtain variant structure, in order to predict free energy difference

$\Delta\Delta G$  for protein-protein complexes. DDAffinity[8], a message passing neural network, uses FoldX [5] BuildMutant for both the wildtype and variant structures. DDGPred [9], a deep neural network, uses Rosetta Relax [10] to obtain mutated structures to predict. EGRAL[11], a graph neural network, uses Rosetta Relax and Backrub [12], in a pipeline close to the conformation sampling of flex ddg [6] method. It allows them to obtain variant structure for SKEMPI2.0 dataset [13] which only provides wildtype (wildtype) structures in order to predict mutational property changes. Unibind,[14] a Graph Neural Network, uses EvoEF2 BuildMutant [15] to obtain the structure graph.

While these papers concentrate on refining the architecture of their models to enhance performance on pretraining tasks and subtasks, comparatively little attention is given to evaluating the impact of these structural acquisition pipelines. These set the stage for our investigation into the ESM3 model [16], a multimodal pLM that, unlike previous ESM models [17] which is trained on sequence only, natively incorporates sequence data, all-atom coordinates, and functional annotations. This capability renders ESM3 exceptionally versatile, allowing it to generate embeddings from either sequence-only inputs or combined sequence and structure inputs. Importantly, there are no stringent requirements for the structures provided, nor is there a need for preprocessing these structures. This flexibility means the method to obtain protein structures can be modified easily and evaluated as part of the end-to-end pipeline. In the following work, we demonstrate that (i) identical protein structures significantly enhances the binding affinity predictions of ESM3, and (ii) the performance of the model is degraded by using a relaxed structure for each variant, compared to using the one identical structure of any variant.

## 2 Methodology

### 2.1 ESM3 and model inputs

The model we benchmarked is composed of ESM3 and a regression head. We have extracted the hidden representation used by output heads after the attention mechanism. Additionally, it is crucial to highlight that the geometric encoder is invariant to rotations and translations, so that alignment is not needed. ESM weights are kept frozen. The embeddings are mean-pooled to obtain a complex embedding, instead of a tensor of residue embeddings. We renormalized the embeddings and binding affinities  $\Delta\Delta G$  by mean and variance. In our results, we report Spearman correlations on several high-throughput datasets: the Bloom dataset[2], which consists of a deep mutational scan of the SARS-CoV-2 RBD binding to ACE2, covering 3,803 single-mutation variants of the RBD. The Desai dataset [18] [19], which includes  $2^{15} = 32,768$  combinatorial mutational data of RBD binding with ACE2 and LY-CoV555, CB6, REGN10987 and S309 antibodies. It is crucial to note that for these datasets, all mutations are on the RBD.

Our comparison identifies the SVM, implemented with default settings in scikit-learn [20], as the most effective regression head for these datasets, as detailed in SI Figure 7. Consequently, we will utilize the SVM for subsequent analyses. However, it is important to note that our findings should not depend on the choice of the regression model.

We evaluate three distinct types of inputs: sequence-only, sequence paired with identical structure, and sequence paired with variant structures. For sequence and identical structure analyses, we use the ProteinChain class from ESM3 repository to load the desired PDB file, which reads both the sequence and atom coordinates. We can then modify the returned object to change the sequence while keeping structure preserved. For sequence and variant structures, we have the flexibility to directly load the PDB file or adjust the sequence to match the desired variant, depending on whether the sequence aligns with the target variant.

### 2.2 Acquiring variants structures

Computational techniques such as in silico mutations, rotamer optimization, and energy minimization are typically used to generate variant structures. These methods are based on the physical principle that the most stable structures are also the most probable, due to the conformation space following a Boltzmann distribution. At low temperatures, the minimum energy state serves as an approximation for the minimum of free energy, allowing for the representation of a protein by a single structure.

In line with these principles, our initial step involves performing an in silico mutation using EvoEF2. The EvoEF2 BuildMutant commandt[15], which performs rotamer optimization to mutate specific

residues using pre-defined libraries of side chain conformations. After mutation, there are two popular protocols: (i) Rosetta Relaxation protocol [10] is used to obtain variant structure with alternating side chain and backbone optimization. (ii) Gromacs energy minimization uses a steep descent energy minimization method with force field CHARMM36.[21].

In the category of sequence paired with variant structures, three widely used relaxation protocols for obtaining variant structures were evaluated : (i) EVoEF2 alone (ii) EvoEF2 with Rosetta Relax, and (iii) EvoEF2 with Gromacs energy minimization. The wildtype RBD structure is extracted from PDB file 6XF5[22], a trimer of the isolated spike protein. In the latter experiments, unless specifically mentioned, the training size is 100 random samples.

Additionally, as a control, we conduct Molecular Dynamics(MD) simulations using Gromacs and the CHARMM36 force field to generate structures under thermodynamic fluctuations for the wildtype. Structures are extracted every 2 femtoseconds and assigned to specific variants. Furthermore, we introduce noisy structure variants by applying Gaussian noise with a 0.1 Ångströms(Å) variance to all atoms in an EvoEF2-relaxed structure. This method helps us assess the impact of structural perturbations on model performance.

We used Intel Xeon Sapphire Rapids CPUs and A100 GPUs. Rosetta Relax required approximately 10 minutes per sample, while Gromacs took about 20 minutes per sample, each using a single CPU core. The MD simulation for the wildtype was completed in 24 hours using a single A100 GPU and 15 CPU cores. Meanwhile, ESM3 generated embeddings within an hour on a single GPU, without any inference optimization. Notably, the EvoEF2 BuildMutant was quite efficient, processing roughly 30,000 variants serially in just about one hour.

### 3 Results

Firstly, we evaluated the performance of ESM3 using only sequence data, as well as when paired with identical structures. As shown in Figure 1, while the performance of ESM1, ESM2 and ESM3 varies when sequence is the only input, the performance of ESM3 is consistently superior when the sequence is paired with an identical structure. This enhancement is also significant across various datasets we tested, as shown in SI Figure 12.

One might anticipate that using RBD structures from RBD-ACE2 binding interactions would yield better results when predicting the binding between RBD mutants and ACE2. As there are minor structural change on the binding surface due to PPI. However, our findings show no significant difference in predictive accuracy. For instance, the structure of 7W9I [23], which is the SARS-CoV-2 RBD bound to ACE2, did not demonstrate enhanced predictive accuracy when predicting on Bloom RBD-ACE2 dataset. Similarly, 7KMG [24], which contains RBD binding with the LY-CoV555 antibody, and 6XF5 [25], the SARS-CoV-2 spike protein with RBD in the down position, also did not exhibit a significant difference in predictive performance in either task shown in Figure 1. We have also tested other PDB structures containing various variants and antibodies, from which we use only the RBD chain, as shown in SI Figure 6, and found no significant differences. This suggests that, although there are minor structural differences in these PDB structures, the ESM3 model could not effectively utilize these variations to enhance prediction accuracy.

Secondly, we evaluate the performance of ESM3 when it is provided with sequences and structures for each mutant, sourced from various pipelines. As indicated previously, using the wildtype structure—referred to here as 'Wildtype'—enhances the performance of the ESM3 model compared to using sequence data alone.

Interestingly, as shown in Figure 2, utilizing the EvoEF2 generated mutant structure does not show any observable effects on the predictive power of ESM3 across all datasets we tested. Furthermore, after using EvoEF2 to introduce mutations, we applied two popular relaxation methods. Surprisingly, the minimized mutated structures produced via Rosetta or Gromacs exhibited lower Spearman correlations, aligning their performance with that of sequence-only inputs.

Consequently, we tested two additional structural pipelines: the 'MD Frame' in which each mutant is provided with a structure extracted every 2 femtoseconds during MD simulations, and the 'Noisy 0.1 Å', which involves a Gaussian noise with mean 0 and variance 0.1 Å applied to all atoms in an EvoEF2-generated configuration. Interestingly, despite the minimal perturbation in the 'MD Frame' and 'Noisy 0.1Å', there is a significant reduction in accuracy, even falling below that of sequence-only

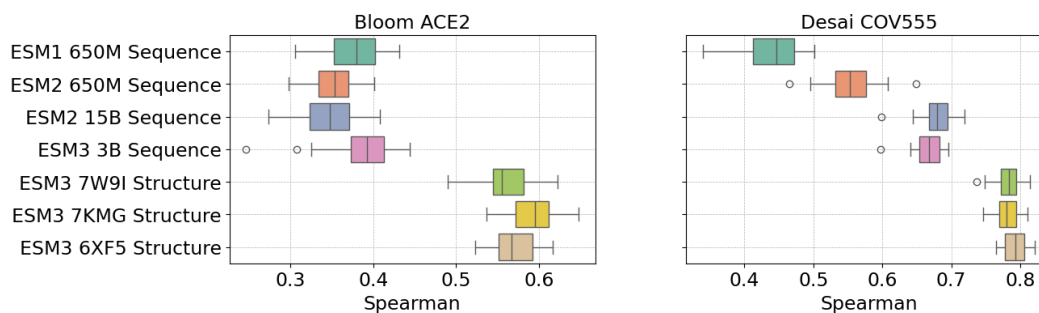


Figure 1: **Left:** Spearman correlation for Models: ESM1, ESM2, and ESM3 sequence only and sequence plus selected structures, evaluated on Bloom dataset for ACE2 binding affinity. 7W9I is the SARS-CoV-2 RBD bound to ACE2, 7KMG is RBD binding with the LY-CoV555 antibody, 6XF5 is the SARS-CoV-2 spike protein with RBD in the down position **Right:** Same analysis on the Desai dataset for RBD and LY-CoV555 antibody binding affinity.

embeddings. This indicates that the embeddings are highly sensitive to noise introduced by processing pipelines. These results are consistent on other datasets we tested, all summarized in SI Figure 11. Further specifics for each dataset can be found in (LY-CoV555: 14, CB6: SI Figure 15, REGN10987: SI Figure 16, S309: SI Figure 17 and ACE2: SI Figure 13).

Lastly, Gromacs minimization and Rosetta Relax were run on the wildtype structure and provided as structure to ESM3. These embeddings with the same structure gave all equal performance with Wildtype pipeline, as shown with "Noisy Wildtype" in Figure 2 and in SI Figure 13 - 17. This effectively suggests that small perturbation is not an issue provided that the same structure is used for all variants. Hence, ESM3 seems to be sensitive to RMSD when provided with protein structure variants.

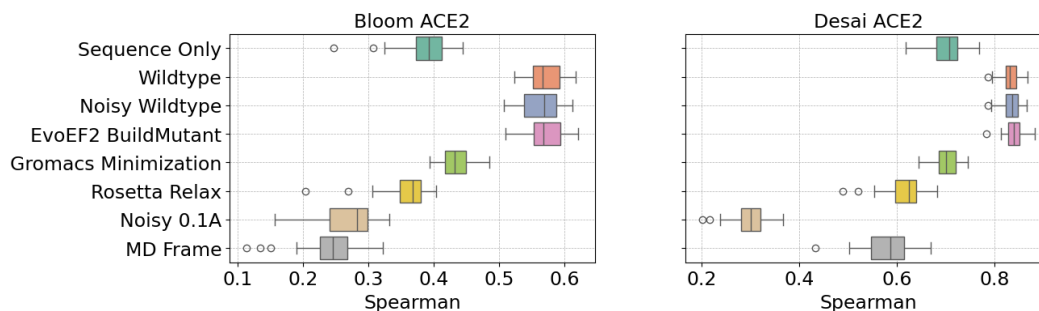


Figure 2: **Left:**Spearman correlation for sequence plus variant structures obtained through different structure acquisition pipelines, evaluated using ESM3 on the Bloom ACE2 **Right:** Same analysis on the Desai dataset for ACE2 binding affinity. Example of aligned structures are shown in SI Figure 4 and 5

It is important to note that the structural variations induced by these pipelines are extremely subtle. To accurately measure the differences between the structures generated by these pipelines, we analyzed the Root Mean Square Deviation (RMSD) distribution by comparing the structures associated with the wildtype variant to those relaxed using the same method. Across all structures, the RMSD consistently remains below 2Å, as shown in Figure 3 which is still within the limits of PDB measurement resolution.

Nonetheless, we can observe that the ranking of methods by mean RMSD—Minimized Gromacs, Relax Rosetta, and MD Frame—correlates inversely with their Spearman correlation rankings, discounting the noisy structure, which is unphysical. The MD Frame, showing the highest RMSD differences, performed the worst, while the Gromacs Minimization performed the best with the lowest RMSD. This is counter-intuitive to the physical understanding that under thermodynamic fluctuations, MD and relaxed structures should represent the actual protein structure and thus be treated the same by the model.

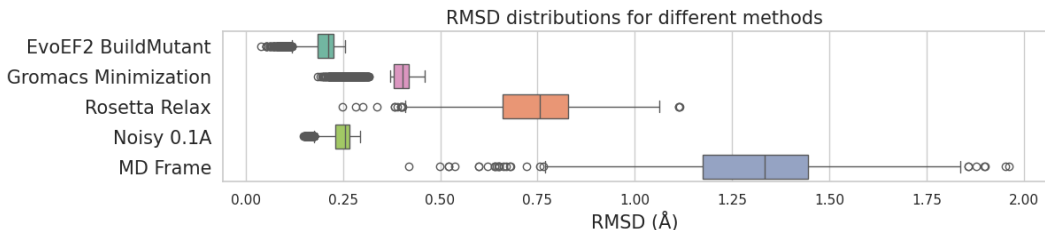


Figure 3: RMSD distribution for structures acquired using different methods.

### 3.1 Embeddings similarities

A further analysis of the embeddings before the renormalization shows a very high similarity of the embeddings whose structures are generated from the same pipeline. The cosine similarity of these embeddings compared to its own wildtype embedding is at least 0.99, as shown in Table 1. This indicates that the generated embeddings is contained in a very small subspace of ESM3 hidden representation space. This may explain the sensitivity of ESM3 given minimal change on the variant structures: it probably reaches the accuracy limit of this model pretrained on a very diverse dataset of proteins.

## 4 Discussions

Various studies utilize mutant structures generated by computational methods as inputs to their models for predicting PPI. In our work, we aim to benchmark these pipelines. We have adapted ESM3 for use as an encoder for PPI prediction. This adaptation was tested across several protein binding datasets with various mutant structures and different structure minimization pipelines. Initially, we observed that although using identical PDB structures enhances the model’s performance in predicting variant binding affinities, using identical RBD structures from different RBD-protein complexes does not significantly impact results. This finding supports the use of identical structures as a priori information in ESM3 for PPI prediction, even if these structures do not perfectly represent the actual protein configuration during binding.

Additionally, providing ESM3 with mutant structures generated by EVoEF2 instead of the structure does not lead to an increase in model accuracy. Furthermore, we noted that ESM3 exhibits sensitivity when provided with structural variations, whether they arise from relaxed structures, thermal noise in MD simulations, or artificially introduced Gaussian noise. To mitigate such sensitivity, constraints could be applied, for example, on amino acids far from mutation sites. This would allow for exploration of relevant parts of the conformation space while limiting displacement in the unaffected regions of the protein, leading to a smaller RMSD.

Our findings highlight the critical need for enhanced computational pipelines that better accommodate variants with measurable structural changes, given that none of the mutant-generating pipelines we tested improved model performance. Additionally, our results highlight the importance of further evaluating the effects of mutant structure pipelines on predicting PPIs and other task, both in pLMs and other models that require structural inputs.

Based on our findings, researchers using ESM3 to predict mutant PPIs might consider the following: (i) Researchers could potentially bypass MD relaxation or energy minimization, opting to use the PDB file directly from the PDB databank. (ii) Caution should be exercised when using relaxed structures generated by EvoEF2. Furthermore, while not tested in this paper, we encourage the evaluation of the impact of relaxation for pretrained or end-to-end models, to ensure that the use of the relaxation pipelines effectively improves the performance of the model while not overfitting the generated mutant structures.

## 5 Code Availability

The code for reproducing our experiments is available at <https://github.com/Dianzhuo-Wang/esm3-structural-inputs>

## References

- [1] Dianzhuo Wang, Marian Huot, Vaibhav Mohanty, and Eugene I. Shakhnovich. Biophysical principles predict fitness of sars-cov-2 variants. *Proceedings of the National Academy of Sciences*, 121(23):e2314518121, 2024.
- [2] Tyler N. Starr, Allison J. Greaney, Sarah K. Hilton, Daniel Ellis, Katharine H.D. Crawford, Adam S. Dingens, Mary Jane Navarro, John E. Bowen, M. Alejandra Tortorici, Alexandra C. Walls, Neil P. King, David Vesler, and Jesse D. Bloom. Deep mutational scanning of sars-cov-2 receptor binding domain reveals constraints on folding and ace2 binding. *Cell*, 182(5):1295–1310.e20, 2020.
- [3] Thomas Desautels, Adam Zemla, Edmond Lau, Magdalena Franco, and Daniel Faissol. Rapid in silico design of antibodies targeting sars-cov-2 using machine learning and supercomputing. *BioRxiv*, pages 2020–04, 2020.
- [4] Théau Debroise, Rose Hoste, Quentin Chamayou, Hervé Minoux, Bruno Filoche-Rommé, Marc Bianciotto, Jean-Philippe Rameau, Laurent Schio, and Maximilien Levesque. In silico drug repositioning for covid-19 using absolute binding free energy calculations. *arXiv preprint arXiv:2109.03565*, 2021.
- [5] Javier Delgado, Leandro G Radusky, Damiano Cianferoni, and Luis Serrano. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics*, 35(20):4168–4169, 03 2019.
- [6] Kyle A. Barlow, Shane Ó Conchúir, Samuel Thompson, Pooja Suresh, James E. Lucas, Markus Heinonen, and Tanja Kortemme. 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, May 2018.
- [7] Jared M. Sampson, Daniel A. Cannon, Jianxin Duan, Jordan C. K. Epstein, Alina P. Sergeeva, Phinikoula S. Katsamba, Seetha M. Mannepalli, Fabiana A. Bahna, H el ene Adihou, St ephanie M. Gu eret, Ranganath Gopalakrishnan, Stefan Geschwindner, D. Gareth Rees, Anna Sigurdardottir, Trevor Wilkinson, Roger B. Dodd, Leonardo De Maria, Juan Carlos Mobarec, Lawrence Shapiro, Barry Honig, Andrew Buchanan, Richard A. Friesner, and Lingle Wang. Robust Prediction of Relative Binding Energies for Protein–Protein Complex Mutations Using Free Energy Perturbation Calculations. *Journal of Molecular Biology*, 436(16):168640, August 2024.
- [8] Guanglei Yu, Qichang Zhao, Xuehua Bi, and Jianxin Wang. DDAffinity: predicting the changes in binding affinity of multiple point mutations using protein 3D structure. *Bioinformatics*, 40(Supplement\_1):i418–i427, July 2024.
- [9] Sisi Shan, Shitong Luo, Ziqing Yang, Junxian Hong, Yufeng Su, Fan Ding, Lili Fu, Chenyu Li, Peng Chen, Jianzhu Ma, 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):e2122954119, 2022.
- [10] Firas Khatib, Seth Cooper, Michael D Tyka, Kefan Xu, Ilya Makedon, Zoran Popovi c, David Baker, and Foldit Players. Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences*, 108(47):18949–18953, 2011.
- [11] Arturo Fiorellini-Bernardis, Sebastien Boyer, Christoph Brunken, Bakary Diallo, Karim Beguir, Nicolas Lopez-Carranza, and Oliver Bent. Protein binding affinity prediction under multiple substitutions applying eGNNs on Residue and Atomic graphs combined with Language model information: eGRAL, May 2024. arXiv:2405.02374 [cs, q-bio].
- [12] Ian W Davis, W Bryan Arendall, David C Richardson, and Jane S Richardson. The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure*, 14(2):265–274, 2006.
- [13] Justina Jankauskait e, Brian Jim enez-Garc ıa, Justas Dapk unas, Juan Fern andez-Recio, and Iain H Moal. SKEMPI 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 35(3):462–469, 07 2018.

- [14] Guangyu Wang, Xiaohong Liu, Kai Wang, Yuanxu Gao, Gen Li, Daniel T Baptista-Hon, Xiaohong Helena Yang, Kanmin Xue, Wa Hou Tai, Zeyu Jiang, et al. Deep-learning-enabled protein–protein interaction analysis for prediction of sars-cov-2 infectivity and variant evolution. *Nature Medicine*, 29(8):2007–2018, 2023.
- [15] Xiaoqiang Huang, Robin Pearce, and Yang Zhang. Evoef2: accurate and fast energy function for computational protein design. *Bioinformatics*, 36(4):1135–1142, 2020.
- [16] Tomas Hayes, Roshan Rao, Halil Akin, Nicholas J Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil, Vincent Q Tran, Jonathan Deaton, Marius Wiggert, et al. Simulating 500 million years of evolution with a language model. *bioRxiv*, pages 2024–07, 2024.
- [17] Joshua Meier, Roshan Rao, Robert Verkuil, Jason Liu, Tom Sercu, and Alex Rives. Language models enable zero-shot prediction of the effects of mutations on protein function. *Advances in neural information processing systems*, 34:29287–29303, 2021.
- [18] Alief Moulana, Thomas Dupic, Angela M Phillips, Jeffrey Chang, Serafina Nieves, Anne A Roffler, Allison J Greaney, Tyler N Starr, Jesse D Bloom, and Michael M Desai. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 omicron BA.1. *Nat. Commun.*, 13(1):7011, November 2022.
- [19] Alief Moulana, Thomas Dupic, Angela M Phillips, Jeffrey Chang, Anne A Roffler, Allison J Greaney, Tyler N Starr, Jesse D Bloom, and Michael M Desai. The landscape of antibody binding affinity in sars-cov-2 omicron ba. 1 evolution. *Elife*, 12:e83442, 2023.
- [20] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- [21] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L De Groot, Helmut Grubmüller, and Alexander D MacKerell Jr. Charmm36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods*, 14(1):71–73, 2017.
- [22] Tongqing Zhou, I-Ting Teng, Adam S Olia, Gabriele Cerutti, Jason Gorman, Alexandra Nazzari, Wei Shi, Yaroslav Tsybovsky, Lingshu Wang, Shuishu Wang, et al. Structure-based design with tag-based purification and in-process biotinylation enable streamlined development of sars-cov-2 spike molecular probes. *Cell reports*, 33(4), 2020.
- [23] Yifan Wang, Caixuan Liu, Chao Zhang, Yanxing Wang, Qin Hong, Shiqi Xu, Zuyang Li, Yong Yang, Zhong Huang, and Yao Cong. Structural basis for sars-cov-2 delta variant recognition of ace2 receptor and broadly neutralizing antibodies. *Nature communications*, 13(1):871, 2022.
- [24] Bryan E Jones, Patricia L Brown-Augsburger, Kizzmekia S Corbett, Kathryn Westendorf, Julian Davies, Thomas P Cujec, Christopher M Wiethoff, Jamie L Blackbourne, Beverly A Heinz, Denisa Foster, et al. The neutralizing antibody, ly-cov555, protects against sars-cov-2 infection in nonhuman primates. *Science translational medicine*, 13(593):eabf1906, 2021.
- [25] Tongqing Zhou, I-Ting Teng, Adam S Olia, Gabriele Cerutti, Jason Gorman, Alexandra Nazzari, Wei Shi, Yaroslav Tsybovsky, Lingshu Wang, Shuishu Wang, et al. Structure-based design with tag-based purification and in-process biotinylation enable streamlined development of sars-cov-2 spike molecular probes. *Cell reports*, 33(4), 2020.

## A Appendix

### A.1 Similarity of embeddings

Method	Mean cosine similarity	Mean Euclidean distance
Wildtype	$0.99994 \pm 2.2 \times 10^{-5}$	$96.7 \pm 21.3$
Relax Rosetta	$0.99963 \pm 9.35 \times 10^{-5}$	$257 \pm 35$
Minimized Gromacs	$0.99962 \pm 1.15 \times 10^{-4}$	$260 \pm 40$
MD Frame	$0.99933 \pm 1.40 \times 10^{-4}$	$350 \pm 38$
Noisy 0.1Å	$0.99976 \pm 9.51 \times 10^{-5}$	$213 \pm 44$

Table 1: Comparison of methods based on cosine similarity and Euclidean distance

### A.2 Aligned PDB structures



Figure 4: 40 randomly selected RBD structures from the Bloom dataset, obtained via the "MD Frame" pipeline.

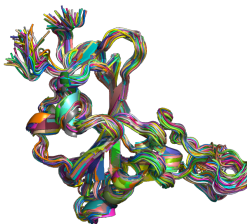


Figure 5: 40 randomly selected RBD structures from the Bloom dataset, obtained via the "Gromacs Minimization" pipeline.



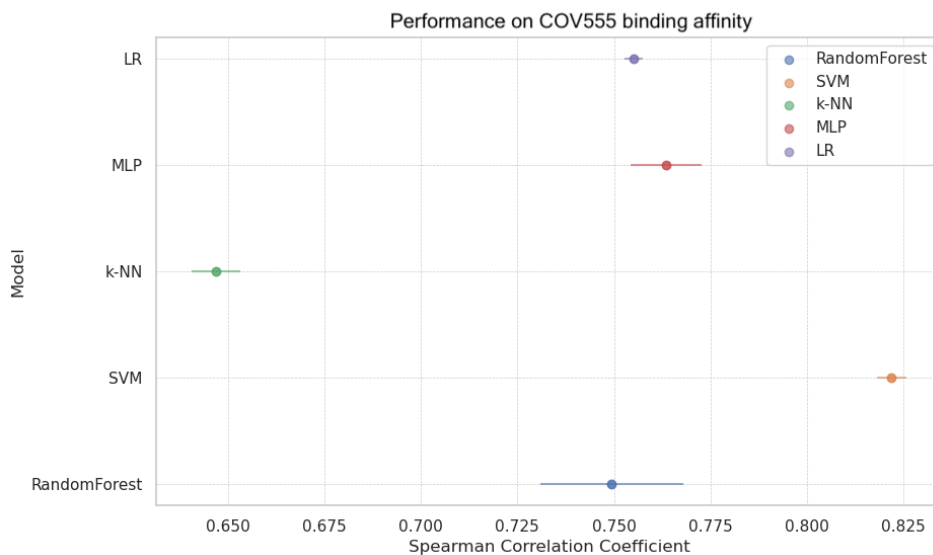


Figure 7: Choice of regression head for RBD binding affinities with COV555 antibody on Desai Dataset

### A.3 RBD structures extracted from PDB file perform similarly

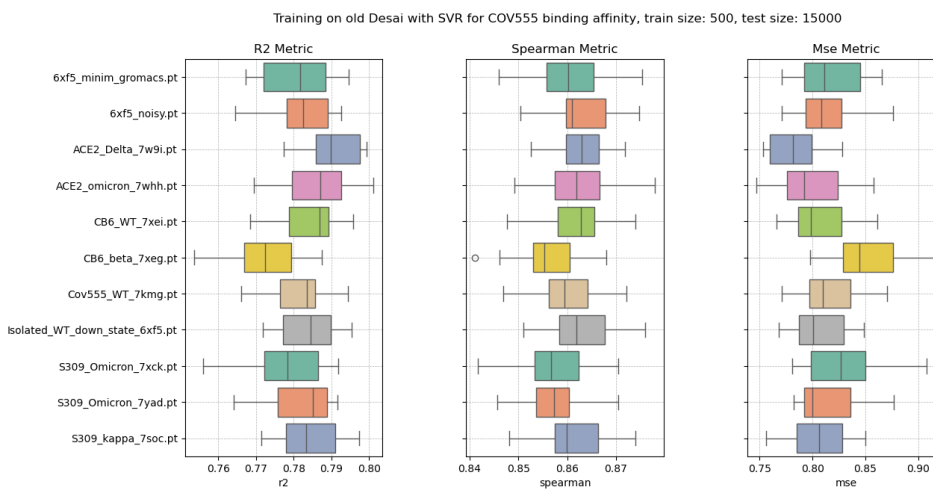


Figure 6: Performance for binding affinity with COV555 for RBD structures extracted from different PDB files

### A.4 Choice of regression head

A comparison over all antibodies binding affinities in Desai Dataset suggest that SVM regressor performs the best.

### A.5 ESM1 vs ESM2 vs ESM3

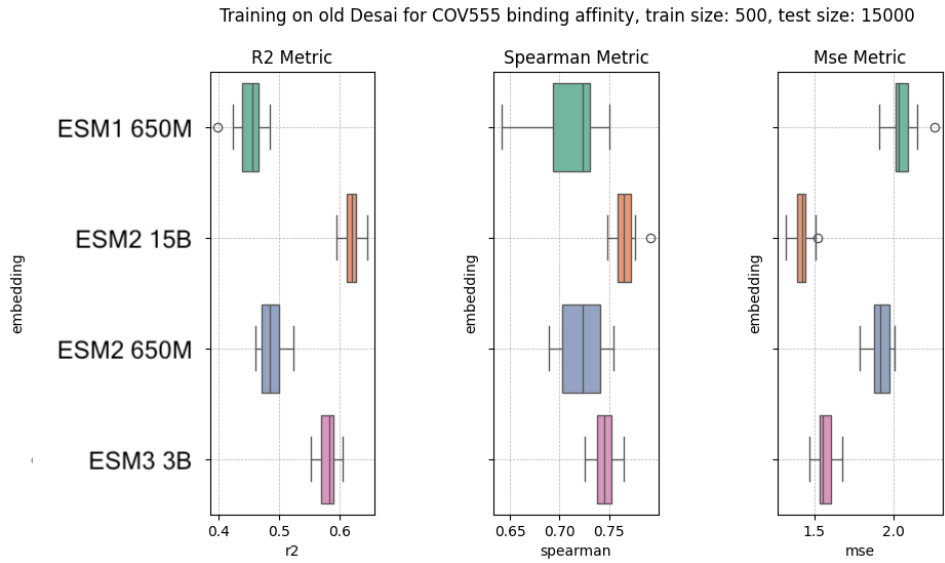


Figure 8: Performance of sequence-only for ESM1 650m, ESM2 650m, ESM2 15b, ESM3

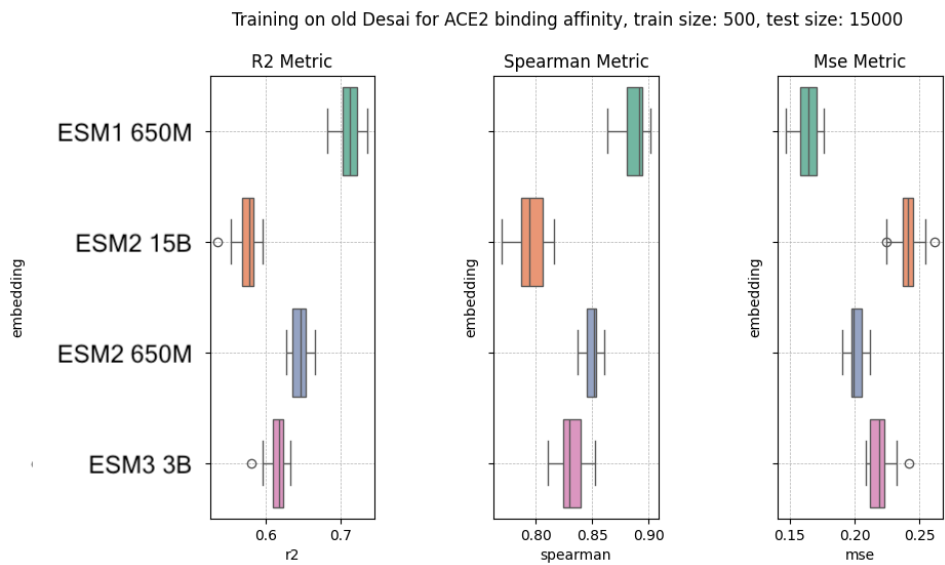


Figure 9: Spearman Correlation for RBD binding affinity with ACE2

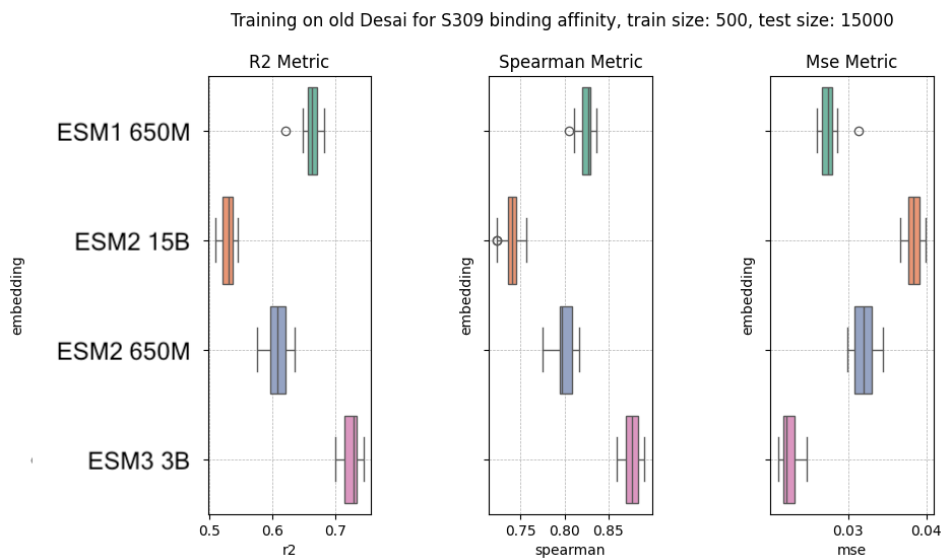


Figure 10: Spearman Correlation for RBD binding affinity with ACE2

## A.6 Performance on binding affinity prediction with antibodies on Desai Dataset

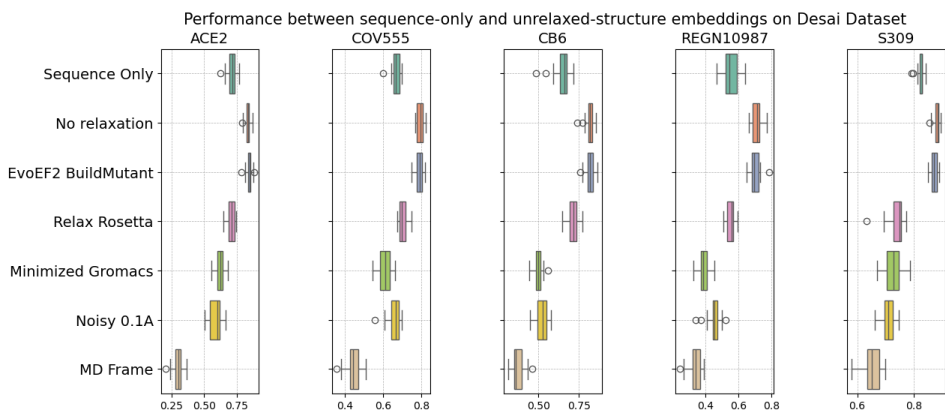


Figure 11: Spearman Correlation for RBD binding affinity with other proteins. Training size is 100 random samples.

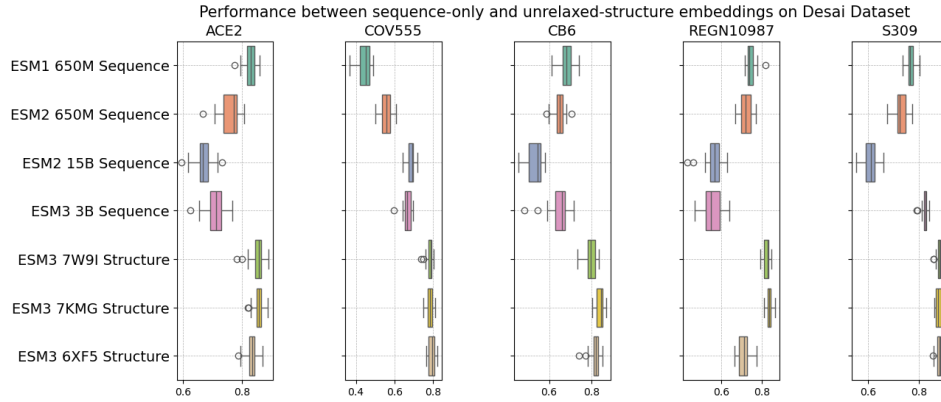


Figure 12: Spearman Correlation for RBD binding affinity with other proteins comparing sequence-only embeddings and sequence and structure. Training size is 100 random samples.

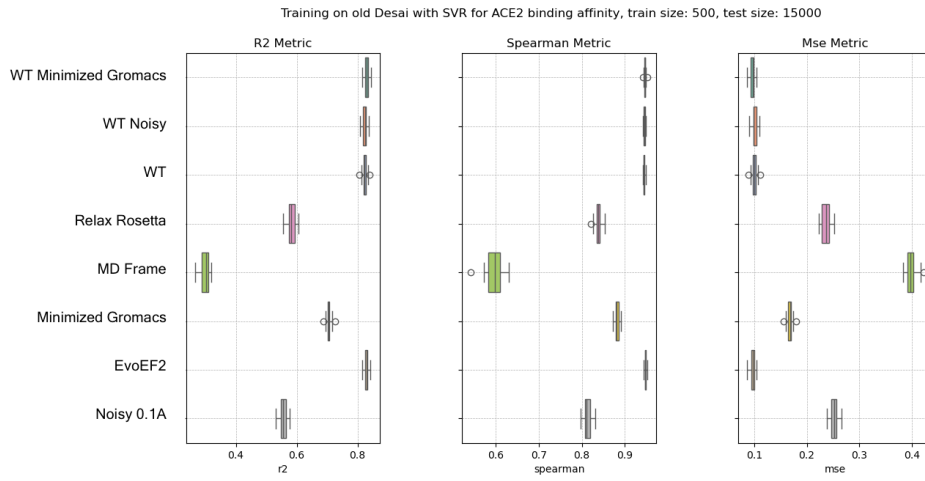


Figure 13: Spearman Correlation for RBD binding affinity with ACE2

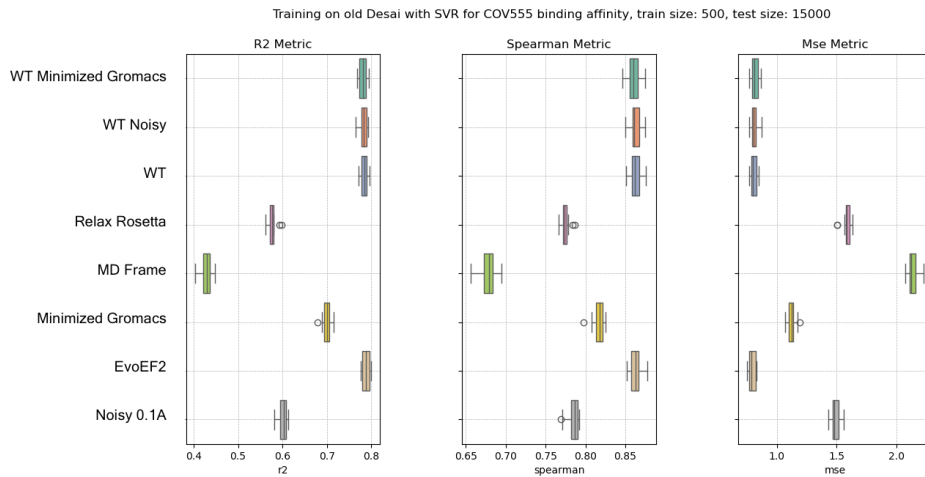


Figure 14: Spearman Correlation for RBD binding affinity with COV555

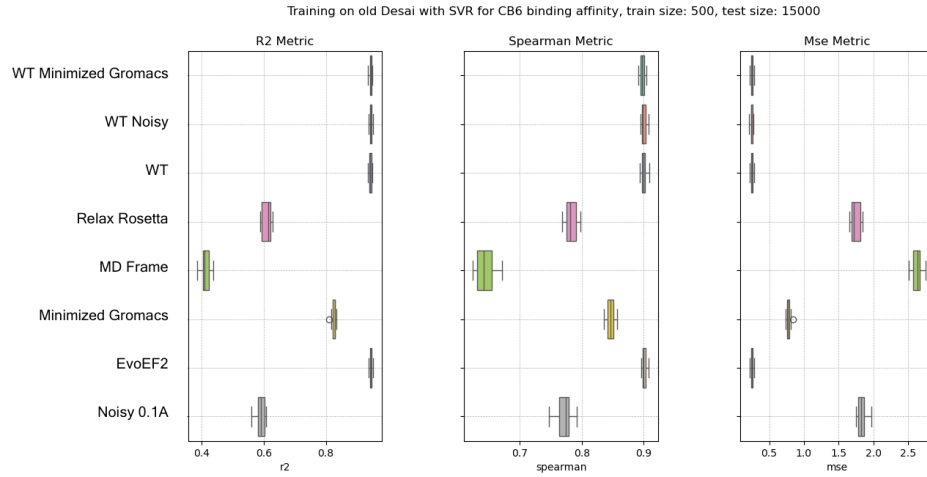


Figure 15: Spearman Correlation for RBD binding affinity with CB6

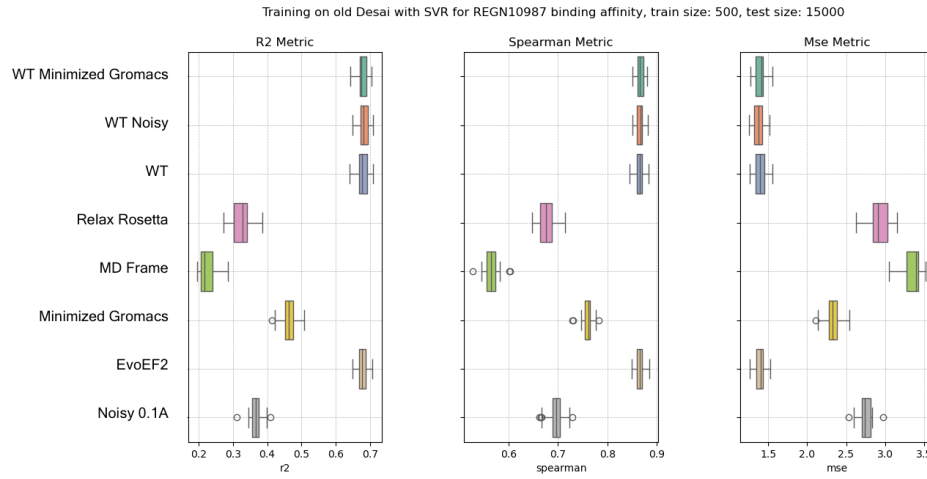


Figure 16: Spearman Correlation for RBD binding affinity with REGN10987

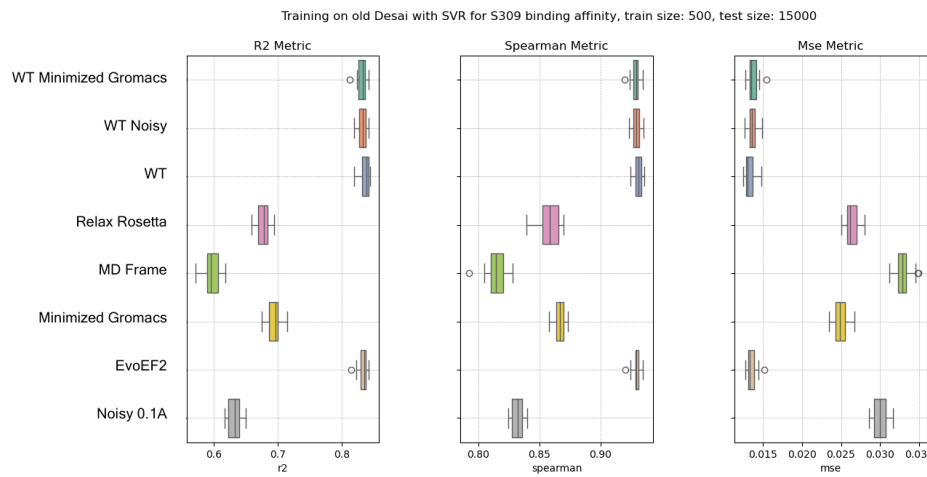


Figure 17: Spearman Correlation for RBD binding affinity with S309