Loop-Diffusion: an equivariant diffusion model for designing and scoring protein loops

Kevin Borisiak
Department of Physics, University of Washington
kborisia@uw.edu

Gian Marco Visani*
Paul G. Allen School of Computer Science and Engineering, University of Washington
gvisan01@cs.washington.edu

Armita Nourmohammad*
Department of Physics, University of Washington
Department of Applied Mathematics, University of Washington
Paul G. Allen School of Computer Science and Engineerings, University of Washington
Fred Hutch Cancer Research Center, Seattle, WA
armita@uw.edu

Abstract

Predicting protein functional characteristics from structure remains a central problem in protein science, with broad implications from understanding the mechanisms of disease to designing novel therapeutics. Unfortunately, current machine learning methods are limited by scarce and biased experimental data, and physics-based methods are either too slow to be useful, or too simplified to be accurate. In this work, we present Loop-Diffusion, an energy based diffusion model which leverages a dataset of general protein loops from the entire protein universe to learn an energy function that generalizes to functional prediction tasks. We evaluate Loop-Diffusion's performance on scoring TCR-pMHC interfaces and demonstrate state-of-the-art results in recognizing binding-enhancing mutations.

1 Introduction and Related Work

Predicting the functional characteristics of proteins has been a central goal in protein science. Within structural biology, the sequence-structure-function paradigm is qualitatively well understood, but rigorous quantitative predictions remain out of grasp. Understanding the relationship between a protein's structure and its function is particularly elusive due to the large number of degrees of freedom involved in determining a protein's structure and its complex dynamics over a wide range of time scales. This is further exacerbated by the scarcity of experimental data for functional measurements, which is expensive to generate, and is often noisy, biased, and riddled with batch effects. An important challenge is modeling the interaction between T-cell receptors (TCRs) and peptide-MHC (pMHC) antigens, which could enable the design of novel antigen-specific immune receptors for cancer immunotherapy (Cappell and Kochenderfer 2023; Cameron et al. 2013) and autoimmune disease prevention (Bjornevik et al. 2022). However, computational models in this domain are limited by data quality. For example, sequence data for paired TCR-pMHC complexes

*Correspondence should be addressed to Gian Marco Visani (gvisan01@cs.washington.edu) and Armita Nourmohammad (armita@uw.edu).

Machine Learning for Structural Biology Workshop, NeurIPS 2024.

is limited, highly biased, and skewed toward specific subsets (Shugay et al. 2018). Additionally, structural information is available for only a few hundred experimentally resolved TCR-pMHC complexes, and computational protein-folding models often lack reliability in this domain, making it difficult to create robust and generalizable models.

Traditionally, molecular dynamics simulations (Ayaz et al. 2023) and empirical energy functions (Chaudhury et al. 2010) have been used to model protein-protein interactions in a data-free way, but the former are too slow to be useful at scale, and the latter often lack accuracy. Machine learning is an attractive alternative, but currently available data is unsuitable for supervised learning. To ameliorate the data-scarcity issue, a growing body of work has attempted to use unsupervised learning and zero-shot protocols to infer biophysical energy functions for proteins. For example, Roney and Ovchinnikov (2022) showed that the output confidence score of a structure prediction model such as AlphaFold (Jumper et al. 2021) can be used to distinguish real protein structures from “decoy" structures, suggesting that the model has learned some physical potential around the equilibrium structure. Unfortunately, this approach has been less successful on more nuanced tasks: (Pak et al. 2023) found that AlphaFold output metrics do not correlate with change of protein stability upon mutations. Other work has sought to augment AlphaFold with domain-specific knowledge to improve its structure predictions and scoring ability, for example for TCR-pMHC complexes (Bradley 2023). Recently, generative models have exploded in popularity owing to their success in natural language processing and computer vision. At their core, generative models seek to learn and sample from the probability distribution of training examples. In equilibrium physical systems, this distribution is intricately tied to the energy of a system through the Boltzmann distribution. Indeed, several works have shown that generative models can be used as zero-shot estimators of energy-based functional quantities like protein stability (Pun et al. 2024; Visani et al. 2024; Meier et al. 2021) and protein-protein binding affinity (Visani et al. 2024; Jin et al. 2023). Design choices of the input space and careful curation of the training data are crucial factors that determine the types of quantities a generative model can capture and its overall performance. For example, (Meier et al. 2021) used the log-likelihood of a model trained to predict masked amino-acid identities given contextual protein sequence to infer mutation effects on protein function, noting that using training sequences at a higher similarity cutoff improved the zero-shot performance of mutation effects. Similarly, (Pun et al. 2024; Visani et al. 2024) trained models to predict masked amino-acid identities given a contextual atomic structure, and used its log-likelihood to infer mutations' effects on protein stability as well as protein-protein binding. (Jin et al. 2023) instead trained an energy-based model with score matching to score protein-protein interfaces, finding that the learned scores correlate well with the experimentally measured affinity between the binding partners.

Protein structures are multi-scaled. Contiguous chunks of amino-acids within the protein's sequence form well defined structural motifs that are conserved across proteins. Between these motifs lie loops, regions with particularly high levels of thermal motion. While the more ordered motifs form the topological structure of the protein, the active regions responsible for protein function often contain disordered loops, such as the CDR3 loops of immune receptors and the peptide antigens within TCR-pMHC complexes, and the CDR3 loops of antibodies. In this work, we model protein loops to capture the biophysical interactions determining the activity and affinity of loops. We hypothesize that irrespective of their activities, data on loops in their structural contexts should inform the biophysical interactions that sustain a loop in the structure, and ultimately determine its function. Therefore, we propose to leverage the large set of general loops in proteins to learn a model that can score and design active loops, such as CDR3s, peptides, and more. To do so, we present Loop-Diffusion, an energy-based diffusion model trained on 433k atomic neighborhoods surrounding loops of various lengths, extracted from 20k non-redundant protein structures. Loop-Diffusion is trained to generate valid atom configurations for loops within a fixed local environment. With its energy-based architecture, Loop-Diffusion can be easily used to score loop configurations within their environment. We evaluate the ability of Loop-Diffusion to score mutations on peptides and CDR3 loops within TCR-pMHC interfaces, demonstrating that it achieves state-of-the-art results at recognizing binding-enhancing mutations compared to other unsupervised models from the literature.

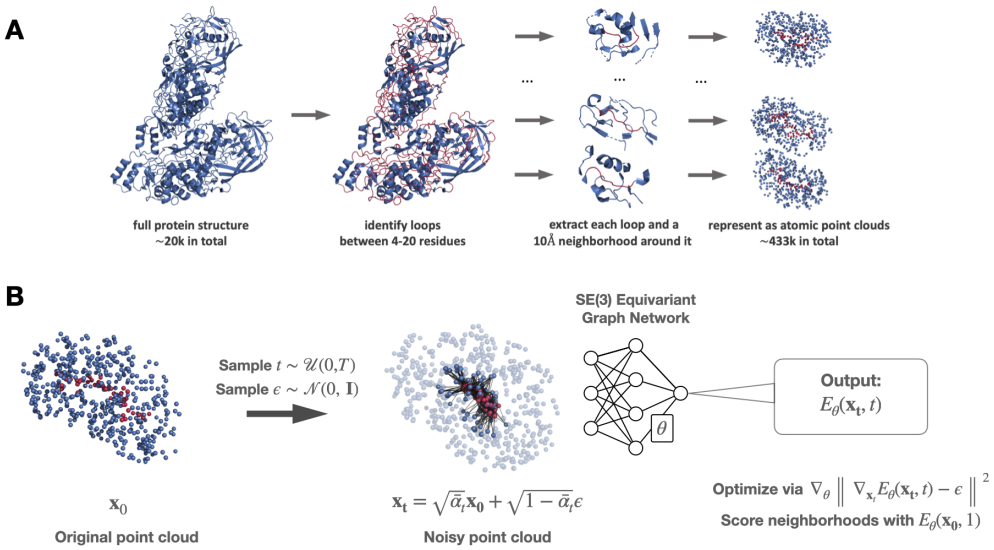

Figure 1: **Schematic of Loop-Diffusion. A)** Loop extraction pipeline. We consider all loops of length between 4 and 20 residues, and atoms in their 10 Å neighborhoods. **B)** Loop-Diffusion is an energy-based model trained with the DDPM objective.

2 Methods

2.1 Structure Preprocessing and Loop Extraction

We use 20k protein structures from the ProteinNet split of CASP12 at 30% similarity cut-off (AlQuraishi 2019). From a protein structure file, we extract a set of neighborhoods, which are a subset of the full protein structure. We define the neighborhoods as follows. We identify loop residues using the DSSP algorithm within PyRosetta (Chaudhury et al. 2010), and identify loops as contiguous sets of loop residues with length ranging from 4 to 20 residues; we show the distribution of loop lengths in our training set in Figure S1. We then define a loop's neighborhood as all atoms within the 10 Å radius of the loop residues' alpha carbons (α-C's). Atoms that belong to the loop are marked as loop atoms, while the rest of the atoms are marked as the environment (Figure 1). In addition to atomic coordinates, we save each atom's element type as well as its partial charge computed by PyRosetta. We omit hydrogen to save compute.

2.2 Energy Based Diffusion Model

Our model aims to leverage the information contained within the distribution of loop conformations within protein structures to learn a useful energy function for downstream tasks. We assume that the loop conformations we observe in crystal structures lie at a local minima of some energy landscape, and that the probability of observing a loop conformation $\mathbf{x}$ is given by a Boltzmann Distribution: $p(\mathbf{x}) = e^{-E(\mathbf{x})/kT}/Z$; where $E(\mathbf{x})$ is some unknown energy function containing physical interactions between the constituents of the loop and its environment, Z is an unknown normalizing constant, k is the Boltzmann constant, and T is temperature (assumed constant throughout this work). We train a neural network to estimate the energy function using the Denoising Diffusion Probabilistic Model (DDPM) objective, described by Ho et al. (2020).

2.2.1 The Objective of the Denoising Diffusion Probabilistic Model (DDPM)

The diffusion framework described by Ho et al. (2020) has three key components: (i) the forward process, (ii) the reverse process, and (iii) the optimization objective. In this work, we focus on the forward process and the optimization objective, as these are the components required to train an energy-based model. The forward process is defined by a fixed Markov chain that gradually adds

noise to the data according to a fixed variance schedule $\beta_1, ..., \beta_T$:

$$p(\mathbf{x}_{1:T}|\mathbf{x}_0) := \prod_{t=1}^{T} p(\mathbf{x}_t|\mathbf{x}_{t-1}), \qquad p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\,\mathbf{x}_{t-1},\, \beta_t\mathbf{I}). \tag{1}$$

where $p(x_t|x_{t-1})$ represents the conditional probability distribution of the state at time t given the state at the previous time step, and has a Gaussian form with mean $\sqrt{1-\beta_t}\,\mathbf{x}_{t-1}$ and variance $\beta_t\mathbf{I}$. Using the notation $\alpha_t := 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, we can sample a noisy $\mathbf{x}_t$ starting from data $\mathbf{x}_0 \sim p_{data}$ using $p_{\alpha_t}(\mathbf{x}_t|\mathbf{x_0}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0,\, (1-\bar{\alpha}_t)\mathbf{I})$, or equivalently generating the data at time t as $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$, with $\epsilon \sim \mathcal{N}(0, \mathbf{I})$.

The goal of DDPM is to learn a parametric model of the inverse process $p_\theta(\mathbf{x_{t-1}}|\mathbf{x_t})$ so that, starting from pure noise $\mathbf{x}_T$, one can generate samples from the true data distribution by running the inverse of 1. Such a model can be trained by optimizing the variational bound on the negative log likelihood. For the author's Ho et al. (2020) choice of parametrization of model, this yields the following objective function:

$$\mathcal{L}_{DDPM} = \mathcal{L}_0 + \sum_{t=1}^{T-1} \mathbb{E}_{\mathbf{x}_0,\epsilon}\left[\frac{\beta_t^2}{2\sigma_t^2\alpha_t}\left\|\epsilon_\theta(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t}\log p_{\alpha_t}(\mathbf{x}_t|\mathbf{x}_0)\right\|^2\right] + \mathcal{L}_T \tag{2}$$

The optimal network $\epsilon_\theta(\mathbf{x}_t, t)$ will approximate the score of the true data distribution, perturbed by some Gaussian kernel $p_{\alpha_t}(\mathbf{x}_t) := \int p_{data}(\mathbf{x}) p_{\alpha_t}(\mathbf{x}_t|\mathbf{x}) d\mathbf{x}$ (Song and Ermon 2020)(Ho et al. 2020)(Vincent 2011). In practice, we parameterize $\epsilon_\theta(\mathbf{x}_t, t)$ as the negative gradient of our energy model:

$$\epsilon_\theta(\mathbf{x}_t, t) = -\nabla_{\mathbf{x}_t} E_\theta(\mathbf{x}_t, t) \tag{3}$$

and adopt $\mathcal{L}_{simple}$ from Ho et al. (2020), which drops the scaling coefficient on 2 and replaces the sum over t with a sample from $\mathcal{U}(0, T)$ A.1. Noting that $\nabla_{\mathbf{x}_t}\log p_{\alpha_t}(\mathbf{x}_t|\mathbf{x}_0) = -\epsilon/\sqrt{1-\bar{\alpha}_t}$, and $1/\sqrt{1-\bar{\alpha}_t}$ is dropped when adopting $\mathcal{L}_{simple}$, our training objective becomes:

$$\mathcal{L}_{\text{Loop-Diffusion}} = \mathbb{E}_{t,\mathbf{x}_0,\epsilon}\left[\left\|\nabla_{\mathbf{x}_t} E_\theta(\mathbf{x}_t, t) - \epsilon\right\|^2\right] \tag{4}$$

We assume that early in the diffusion process, i.e. small t, the distribution $p_{\alpha_t}(\mathbf{x})$ approximates the Boltzmann distribution underlying our data. Thus, we expect that when evaluated on real neighborhoods at the smallest time step $t = 1$, our learned energy model will approximate the true energy $E(\mathbf{x})$ up to a scaling constant:

$$\begin{aligned} -\nabla_{\mathbf{x}} E_\theta(\mathbf{x}, 1) &\simeq \nabla_{\mathbf{x}}\log p_{\alpha_1}(\mathbf{x}) \\ &= \nabla_{\mathbf{x}}\log e^{\frac{-E(\mathbf{x})}{kT}} = -kT\nabla_{\mathbf{x}}E(\mathbf{x}) \end{aligned} \tag{5}$$

We implement E_θ using an equivariant graph convolutional network architecture built with the `e3nn` library (Geiger and Smidt 2022). For more details on model implementation and training, see A.1.

It is worth noting that Jin et al. (2023) follows a similar approach, employing a simpler Denoising Score Matching (DSM) objective (Vincent 2011) to learn $E(\mathbf{x})$. Rather than conditioning on t and learning the full Markov chain, they predict the noise added in a single step without conditioning on the noise level. In theory, this should be equally capable of learning the target Boltzmann distribution, however, DSM models are less effective at sampling from the learned distribution. We believe sampling may be useful, as it would allow us to relax protein structures using our learned energy function, which motivated us to pursue the DDPM approach.

3 Results

We evaluated the performance of Loop-Diffusion in predicting the effect of mutations on the binding affinity $\Delta\Delta G$ of the TCR-pMHC complexes. For this task, we leverage the ATLAS dataset (Borrman

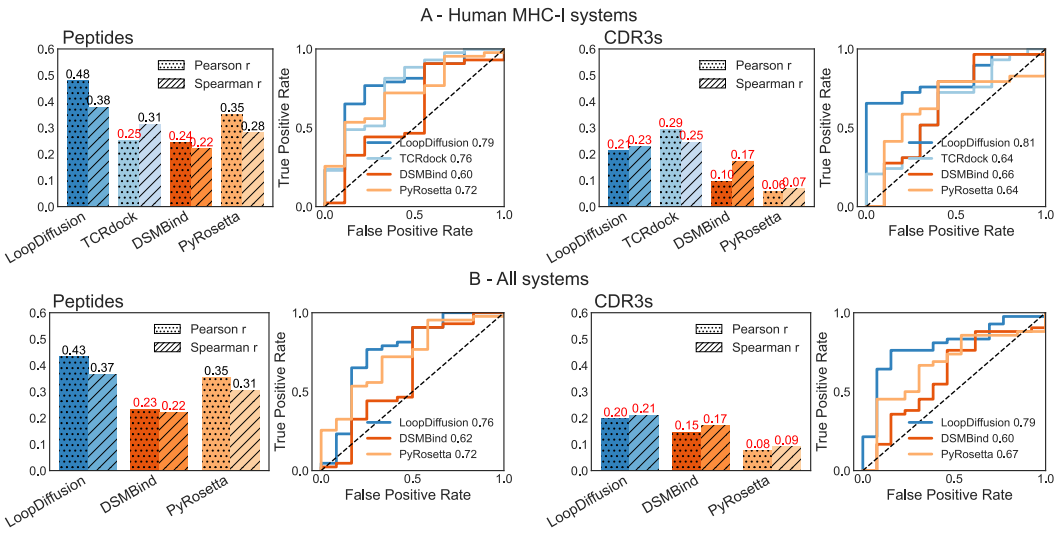

Figure 2: **Peptides and CDR3 mutational effects on the binding affinity of TCR-pMHC complexes.** The predictions for the effect of mutations on binding affinity in peptides (left) and in the CDR3 loop of TCRs (right) is compared across models for TCR-pMHC complexes of **(A)** Human MHC-I system (n=52 for peptides, n=37 for CDR3), and **(B)** all the TCR-pMHC complexes in the ATLAS database with a mutation in the peptide or the CDR3 Loop (n=55 for both peptides and CDR3's). All panels show correlation coefficients between the experimental data and model predictions (left), and insignificant correlations (p-value>0.05) are indicated in red. All panels show ROC curves and the corresponding AUROC for all models to classify between favorable ($\Delta\Delta G_{binding} \geq 0$) and unfavorable ($\Delta\Delta G_{binding} < 0$) mutations in each set (right).

et al. 2017). Due to the nature of our pipeline, we focus specifically on mutations that occur within the loops, specifically on the peptide antigen or the two CDR3 loops of a TCR in the complex. CDR3 loop locations were identified via alignment against annotated TCR sequences, using code from the TCRdock Github repository (Bradley 2023). To score a mutation on a loop (either CRD3 or peptide), we extract both the wild-type and the mutant loop neighborhoods from the respective structures; conveniently, ATLAS provides *in-silico* generated mutant structures. We evaluate the energy of each structure as the sum of the model node energies evaluated at $t = 1$ within the diffusion framework, since at $t = 1$ we expect to capture the statistics of the true Boltzmann distribution as closely as possible. We define the model's predicted mutational effect on the binding affinity $\Delta\Delta G$ to be the difference between the predicted energies of the mutant and the wild-type.

We compared our approach to three other methods, one traditional physics-based energy function and two unsupervised machine learning methods, which similar to ours, were not expressively trained to predict mutational effects on binding $\Delta\Delta G$. These methods are: (i) PyRosetta binding $\Delta\Delta G$ (Park et al. 2016), computed using our implementation of the "cartesian-ddG" protocol, (ii) TCRdock (Bradley 2023), which is a protocol enhancing AlphaFold-Multimer's (Evans et al. 2022) ability to predict the structure of TCR-pMHC complexes; it can be used to score TCR-pMHC binding via its predicted alignment error of the interface, and (iii) DSMBind (Jin et al. 2023), which is an energy-based model trained with score matching on protein-protein interfaces. See Section A.2 for further details on these models.

For TCR-pMHC proteins associated with Human MHC Class I, Loop-Diffusion achieves best correlations to the experimental $\Delta\Delta G$ values for mutations on peptide antigens, and second-best for mutations on CDR3s. In both regions, Loop-Diffusion shows the best classification accuracy between favorable ($\Delta\Delta G \geq 0$) and unfavorable ($\Delta\Delta G < 0$) mutations, measured by the Area under the Receiver Operating Curve (AUROC); see Fig. 2A, and Fig. S2 for the corresponding scatter plots. Moreover, Loop-Diffusion shows best correlations with the experimental data and best classification accuracies for other MHC systems (human MHC II and mouse systems), compared to all other models, except for TCRdock, which cannot be evaluated on this data; see Figs. 2B, S3.

4 Discussion

In this work we have enhanced the power of generative models for zero-shot prediction of protein-protein binding energy by carefully selecting the training data to reflect the distribution of targets of interest. We presented Loop-Diffusion, an energy-based model trained as a DDPM to denoise loops from the protein universe, and applied it to score mutations within loops at protein's functional interfaces. We tested Loop-Diffusion specifically on CDR3 loops and peptides within TCR-pMHC systems, finding that it is stronger than comparable unsupervised models at identifying binding-enhancing mutations. In future work, we plan on using Loop-Diffusion to score loops within other functional contexts, such as the CDR3 loops of antibodies. Furthermore, we plan on further training Loop-Diffusion with emphasis on the generative task, which can be used for example to generate mutant structures prior to scoring them. On a similar note, we plan on exploring the use of correlated noise structures, so that we can more easily generate valid loop conformations (Jing et al. 2023; Jin et al. 2023).

5 Acknowledgments

This work has been supported by the National Institutes of Health MIRA award (R35 GM142795), the CAREER award from the National Science Foundation (grant No: 2045054), A&S PhD Fellowship Support from the UW Provost, and the Allen School Computer Science & Engineering Research Fellowship from the Paul G. Allen School of Computer Science & Engineering at the University of Washington. This work is also supported, in part, through the Departments of Physics and Computer Science and Engineering, and the College of Arts and Sciences at the University of Washington.

References

[1] Mohammed AlQuraishi. "ProteinNet: a standardized data set for machine learning of protein structure". In: *BMC Bioinformatics* 20.1 (June 2019), p. 311. ISSN: 1471-2105. DOI: 10.1186/s12859-019-2932-0. URL: https://doi.org/10.1186/s12859-019-2932-0 (visited on 09/23/2022).

[2] Pelin Ayaz et al. "Structural mechanism of a drug-binding process involving a large conformational change of the protein target". en. In: *Nature Communications* 14.1 (Apr. 2023). Publisher: Nature Publishing Group, p. 1885. ISSN: 2041-1723. DOI: 10.1038/s41467-023-36956-5. URL: https://www.nature.com/articles/s41467-023-36956-5 (visited on 09/20/2024).

[3] Kjetil Bjornevik et al. "Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis". In: *Science* 375.6578 (Jan. 2022). Publisher: American Association for the Advancement of Science, pp. 296–301. DOI: 10.1126/science.abj8222. URL: https://www.science.org/doi/10.1126/science.abj8222 (visited on 09/07/2024).

[4] Tyler Borrman et al. "ATLAS: A database linking binding affinities with structures for wild-type and mutant TCR-pMHC complexes". en. In: *Proteins: Structure, Function, and Bioinformatics* 85.5 (2017). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.25260, pp. 908–916. ISSN: 1097-0134. DOI: 10.1002/prot.25260. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.25260 (visited on 09/18/2024).

[5] Philip Bradley. "Structure-based prediction of T cell receptor:peptide-MHC interactions". In: *eLife* 12 (Jan. 2023). Ed. by Michael L Dustin, Tadatsugu Taniguchi, and Michael L Dustin. Publisher: eLife Sciences Publications, Ltd, e82813. ISSN: 2050-084X. DOI: 10.7554/eLife.82813. URL: https://doi.org/10.7554/eLife.82813 (visited on 01/16/2024).

[6] Brian J. Cameron et al. "Identification of a Titin-Derived HLA-A1–Presented Peptide as a Cross-Reactive Target for Engineered MAGE A3–Directed T Cells". In: *Science translational medicine* 5.197 (Aug. 2013), 197ra103. ISSN: 1946-6234. DOI: 10.1126/scitranslmed.3006034. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6002776/ (visited on 09/07/2024).

[7] Kathryn M. Cappell and James N. Kochenderfer. "Long-term outcomes following CAR T cell therapy: what we know so far". en. In: *Nature Reviews Clinical Oncology* 20.6 (June 2023). Publisher: Nature Publishing Group, pp. 359–371. ISSN: 1759-4782. DOI: 10.1038/s41571-023-00754-1. URL: https://www.nature.com/articles/s41571-023-00754-1 (visited on 09/07/2024).

[8] Sidhartha Chaudhury, Sergey Lyskov, and Jeffrey J. Gray. "PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta". eng. In: *Bioinformatics (Oxford, England)* 26.5 (Mar. 2010), pp. 689–691. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btq007.

[9] Richard Evans et al. *Protein complex prediction with AlphaFold-Multimer*. en. Pages: 2021.10.04.463034 Section: New Results. Mar. 2022. DOI: 10.1101/2021.10.04.463034. URL: https://www.biorxiv.org/content/10.1101/2021.10.04.463034v2 (visited on 09/18/2024).

[10] Mario Geiger and Tess Smidt. *e3nn: Euclidean Neural Networks*. en. arXiv:2207.09453 [cs]. July 2022. URL: http://arxiv.org/abs/2207.09453 (visited on 09/12/2024).

[11] Jonathan Ho, Ajay Jain, and Pieter Abbeel. *Denoising Diffusion Probabilistic Models*. en. arXiv:2006.11239 [cs, stat]. Dec. 2020. URL: http://arxiv.org/abs/2006.11239 (visited on 09/18/2024).

[12] Wengong Jin et al. *DSMBind: SE(3) denoising score matching for unsupervised binding energy prediction and nanobody design*. en. Pages: 2023.12.10.570461 Section: New Results. Dec. 2023. DOI: 10.1101/2023.12.10.570461. URL: https://www.biorxiv.org/content/10.1101/2023.12.10.570461v1 (visited on 09/18/2024).

[13] Bowen Jing et al. *EigenFold: Generative Protein Structure Prediction with Diffusion Models*. arXiv:2304.02198 [physics, q-bio]. Apr. 2023. DOI: 10.48550/arXiv.2304.02198. URL: http://arxiv.org/abs/2304.02198 (visited on 09/18/2024).

[14] John Jumper et al. "Highly accurate protein structure prediction with AlphaFold". en. In: *Nature* 596.7873 (Aug. 2021). Publisher: Nature Publishing Group, pp. 583–589. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03819-2. URL: https://www.nature.com/articles/s41586-021-03819-2 (visited on 09/18/2024).

[15] Risi Kondor, Zhen Lin, and Shubhendu Trivedi. *Clebsch-Gordan Nets: a Fully Fourier Space Spherical Convolutional Neural Network*. arXiv:1806.09231 [cs, stat]. Nov. 2018. URL: http://arxiv.org/abs/1806.09231 (visited on 08/24/2023).

[16] Zeming Lin et al. "Evolutionary-scale prediction of atomic-level protein structure with a language model". In: *Science* 379.6637 (Mar. 2023). Publisher: American Association for the Advancement of Science, pp. 1123–1130. DOI: 10.1126/science.ade2574. URL: https://www.science.org/doi/10.1126/science.ade2574 (visited on 09/07/2024).

[17] Joshua Meier et al. "Language models enable zero-shot prediction of the effects of mutations on protein function". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 29287–29303. URL: https://proceedings.neurips.cc/paper/2021/hash/f51338d736f95dd42427296047067694-Abstract.html (visited on 06/28/2024).

[18] Marina A. Pak et al. "Using AlphaFold to predict the impact of single mutations on protein stability and function". en. In: *PLOS ONE* 18.3 (Mar. 2023). Publisher: Public Library of Science, e0282689. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0282689. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0282689 (visited on 09/10/2024).

[19] Hahnbeom Park et al. "Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules". In: *Journal of Chemical Theory and Computation* 12.12 (Dec. 2016). Publisher: American Chemical Society, pp. 6201–6212. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.6b00819. URL: https://doi.org/10.1021/acs.jctc.6b00819 (visited on 09/18/2024).

[20] Michael N. Pun et al. "Learning the shape of protein microenvironments with a holographic convolutional neural network". In: *Proceedings of the National Academy of Sciences* 121.6 (Feb. 2024). Publisher: Proceedings of the National Academy of Sciences. DOI: 10.1073/pnas.2300838121. URL: https://www.pnas.org/doi/10.1073/pnas.2300838121 (visited on 09/16/2024).

[21] James P. Roney and Sergey Ovchinnikov. "State-of-the-Art Estimation of Protein Model Accuracy Using AlphaFold". en. In: *Physical Review Letters* 129.23 (Nov. 2022), p. 238101. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.129.238101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.129.238101 (visited on 09/10/2024).

[22] Mikhail Shugay et al. "VDJdb: a curated database of T-cell receptor sequences with known antigen specificity". In: *Nucleic Acids Research* 46.D1 (Jan. 2018), pp. D419–D427. ISSN: 0305-1048. DOI: 10.1093/nar/gkx760. URL: https://doi.org/10.1093/nar/gkx760 (visited on 09/18/2024).

[23] Yang Song and Stefano Ermon. *Generative Modeling by Estimating Gradients of the Data Distribution*. arXiv:1907.05600 [cs, stat]. Oct. 2020. URL: http://arxiv.org/abs/1907.05600 (visited on 11/01/2023).

[24] Ashish Vaswani et al. "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017. URL: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (visited on 09/19/2024).

[25] Pascal Vincent. "A Connection Between Score Matching and Denoising Autoencoders". en. In: *Neural Computation* 23.7 (July 2011), pp. 1661–1674. ISSN: 0899-7667, 1530-888X. DOI: 10.1162/NECO_a_00142. URL: https://direct.mit.edu/neco/article/23/7/1661-1674/7677 (visited on 09/18/2024).

[26] Gian Marco Visani et al. *HERMES: Holographic Equivariant neuRal network model for Mutational Effect and Stability prediction*. en. Pages: 2024.07.09.602403 Section: New Results. July 2024. DOI: 10.1101/2024.07.09.602403. URL: https://www.biorxiv.org/content/10.1101/2024.07.09.602403v1 (visited on 09/19/2024).

A Appendix

A.1 Implementation Details

As input, our model takes a PyTorch Geometric graph of the loop neighborhood, with each atom receiving being assigned to a node with a position value $\mathbf{x}_{pos} \in \mathbb{R}^3$, a feature vector $\mathbf{x}_{feat} \in \mathbb{R}^6$ containing one-hot encoded vector of the 5 atom types it may encounter (N, S, H, C, O) and a single scalar value for the charge, and a one-dimensional attribute vector $\mathbf{z}$ containing a binary value indicating whether the atom belongs to the loop or the environment. During training, we use auto-differentiation to take the gradient of the network energy with respect to the loop node coordinates.

During training, we append to the node features a 10-dimensional sinusoidal time-embedding, expanding the feature dimension to 16.

For the network, we use the basic graph convolutional network implementation provided within the `e3nn` package. We experimented with other architectures on a simple n-body force prediction task and found that graph networks outperformed the transformers (Vaswani et al. 2017) and Clebsch-Gordan nets (Kondor et al. 2018) we tried. Additionally, we believe the geometric pairwise interactions encoded by a graph network are more reflective of the underlying physics of the inter-atomic interactions in this problem. `e3nn` assigns node and edge features according to the Irreducible Representation (irrep) they transform by under SO(3) (Geiger and Smidt 2022). The irrep is identified by the degree value l. $l = 0$ corresponds to scalar values, $l = 1$ corresponds to vector values, and so on. Within the network, the node and edge features increase up to $l = 4$, with a multiplicity of 8 for each type i.e. each node receives 8 scalars, 8 vectors, 8 traceless symmetric tensors, etc. We found that these values performed well on our baselines during model selection. We use a network depth of 3, 3 radial basis functions for the edge embedding, and 100 radial neurons.

For our implementation of the diffusion protocol, we choose a linear β_t schedule interpolating between $\beta_0 = 0.0001$ and $\beta_T = 0.002$ with $T = 2000$. Additionally during training we only sample times from $[0, \frac{T}{2}]$ so the model could better focus on learning the early distribution. We note that omitting the scaling coefficient in 2 is intended to improve sampling quality, which makes it easier to monitor the models learning, however it down-weights the loss at early time steps, which is when the model should be learning the Boltzmann distribution. In future experiments, we would like to train with the weighting coefficient to see if it improves performance on scoring.

Algorithm 1: Training

Data: Loop Neighborhood
$\mathbf{x}_0 := [\mathbf{x}_{0,loop}, \mathbf{x}_{0,env}]$
repeat
  sample $\mathbf{x}_0$ from data
  sample $t \in [0, \frac{T}{2}]$
  sample $\epsilon \in \mathcal{N}(0, \mathbf{I})$
  add noise to loop coordinates only;
    $\mathbf{x}_{t,loop} = \sqrt{\bar{\alpha}_t}\mathbf{x}_{0,loop} + \sqrt{1-\bar{\alpha}_t}\epsilon$
    $\mathbf{x}_{t,env} = \mathbf{x}_{0,env}$
    $\mathbf{x}_t := [\mathbf{x}_{t,loop}, \mathbf{x}_{t,env}]$
  take gradient descent step on:
    $\nabla_\theta \left\| \nabla_{\mathbf{x}_t} E_\theta(\mathbf{x}_t, t) - \epsilon \right\|^2$
until *converged*;

Algorithm 2: Inference (specifically mutation scoring)

Data:
Wild Type Neighborhood $\mathbf{x}_{wt}$
Mutant Neighborhood $\mathbf{x}_{mt}$
Result:
Compute model energies at $t = 0$:
  $E_{wt} = E_\theta(\mathbf{x}_{wt}, 0)$
  $E_{mt} = E_\theta(\mathbf{x}_{mt}, 0)$
$\Delta\Delta G_{pred} = E_{mt} - E_{wt}$

A.2 Baselines details

PyRosetta (Park et al. 2016). We use our implementation of the "cartesian-ddG" protocol. Specifically, we compute the binding ΔG of a TCR-pMHC complex as $\Delta G = E_{\text{TCR-pMHC}} - (E_{\text{TCR}} - E_{\text{pMHC}})$, where each energy term E is computed using pyrosetta's cartesian scoring function. We then compute binding $\Delta\Delta G$ simply as $\Delta G_{\text{mt}} - \Delta G_{\text{wt}}$.

TCRdock (Bradley 2023). This is an AlphaFold-based algorithm that uses carefully-selected

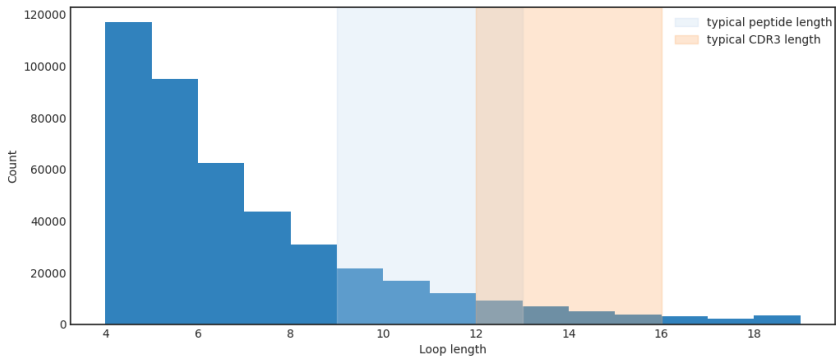

Figure S1: **Distribution of loop lengths within our dataset extracted from CASP12.** The peptides in our test dataset range from 9-13 residues in length, which is relatively well represented in our training dataset. CDR3's loops, however, are typically in the range of 12-16 residues in length, which is a more data scarce regime. Future work may attempt to crop the CDR3 loop around the mutation to see if performance is improved.

structural templates, alongside considerations about TCR-pMHC's common docking geometries, to enhance AlphaFold-Multimer's (Evans et al. 2022) capabilities on TCR-pMHC structure prediction. TCRdock's PAE score of the TCR-pMHC interface has been shown to have some discriminatory power of correct TCR-pMHC pairings. We thus treat the TCR-pMHC PAE as a binding score, and the difference between mutant and wildtype scores as a predictor of binding $\Delta\Delta G$. As we encountered issues when using TCRdock on complexes with Class II MHCs, we only use TCRdock for predictions with Class I MHCs, and leave the analysis to of MHC-Class II complexes to future work. Notably, as TCRdock is effectively a protein-folding algorithm, it does not rely on the availability of accurate structures, though its performance does deteriorate for TCR-pMHC systems that have low similarity matches among those that have structures in TCRdock's database (Bradley 2023). As all of the TCR-pMHC systems in ATLAS have a wildtype structure in TCRdock's database, the TCRdock scores we show are as good as they can get.

DSMBind (Jin et al. 2023). Similar to Loop-Diffusion, DSMBind is an energy-based model trained with score matching; we use the version of the model trained to score protein-protein interfaces. DSMBind adds noise by randomly roto-translating one of the two binding partners about its center of mass, as well as randomly rotating all the side-chains' orientations. DSMBind also uses ESM2 embeddings (Lin et al. 2023) as features to enhance their predictions, which Loop-Diffusion currently does not.

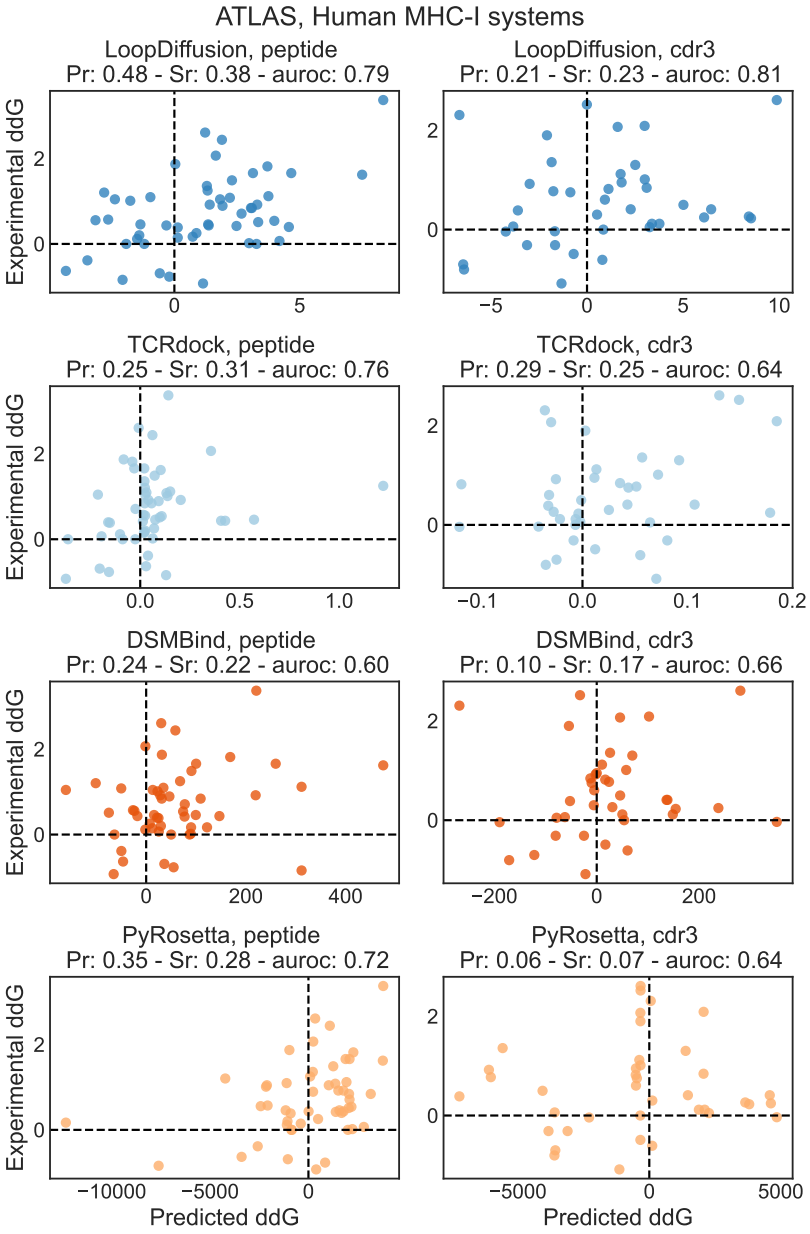

Figure S2: **Scatterplots of predicted vs. experimental binding $\Delta\Delta G$ on mutations occurring on peptides only (left columns) or one CDR3 only (right column), from the subset of the ATLAS dataset containing only Human MHC Class-I systems.**

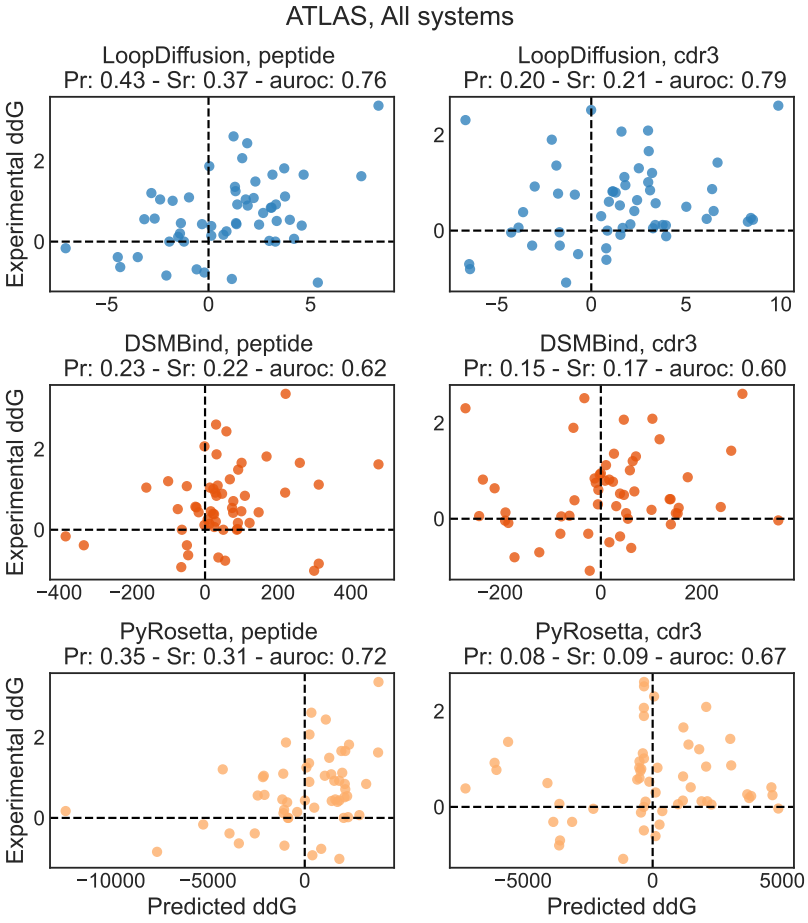

Figure S3: **Scatterplots of predicted vs. experimental binding $\Delta\Delta G$ on mutations occurring on peptides only (left columns) or one CDR3 only (right column), from the ATLAS dataset.**