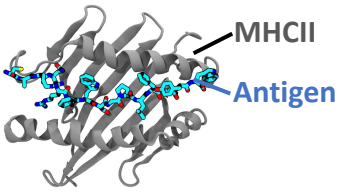
Design of enhanced binding antigen peptides using Gaussian process regression

David R. Bell[1] and Serena H. Chen[2]
[1]Frederick National Laboratory for Cancer Research
[2]Oak Ridge National Laboratory

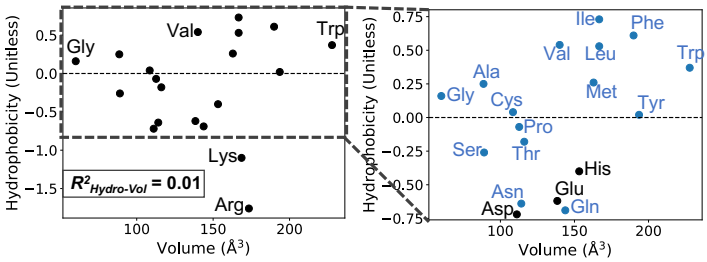
Frederick National Laboratory for Cancer Research
sponsored by the National Cancer Institute
OAK RIDGE National Laboratory
NIH National Cancer Institute · U.S. Department of Energy

Background: In biologics drug discovery, one objective is to design and optimize biological therapeutics such as antibodies and antigen-specific immunotherapies for enhanced binding. A naïve way of optimizing protein binding is to mutate protein residues and determine the difference in binding affinity. This mutagenesis technique is costly and often leads to decreased affinity mutants. Here, we develop a way to minimize this cost and predict enhanced affinity mutants from minimal prior data.

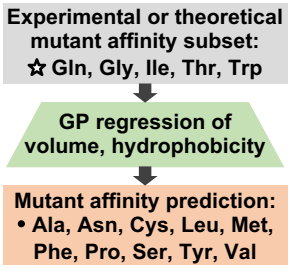
Methods: Use Gaussian process (GP) regression across residue volume and hydrophobicity to predict mutant peptide-MHCII binding affinities from a small residue subset



MHCII-antigen peptide system. We focus on predicting mutant antigen binding affinities



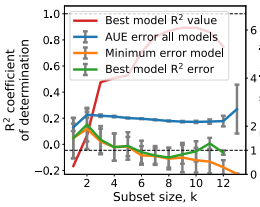
Residue volume and hydrophobicity feature space, with Pearson R^2 correlation



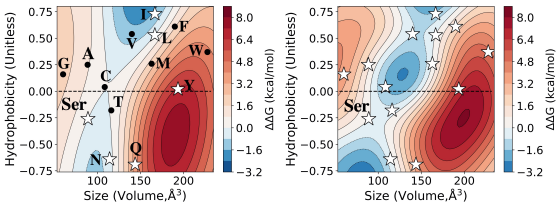
GP regression workflow, focusing on affinity prediction of neutral residues

GP regression accurately predicts mutant binding affinities

Computational Dataset from FEP calculations

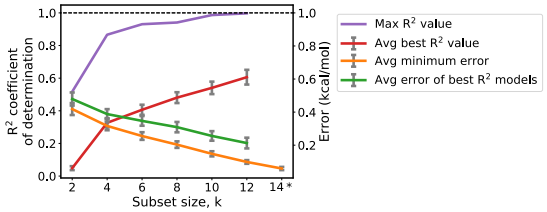


R^2 coefficient of determination and errors for prediction of Free Energy Perturbation (FEP) affinities



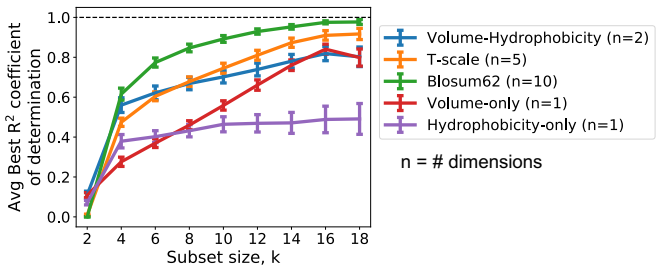
GP regression-predicted affinity landscapes for (left) a 6-residue subset and (right) all neutral residues based on an FEP-computed affinity dataset

Experimental Dataset from IEDB



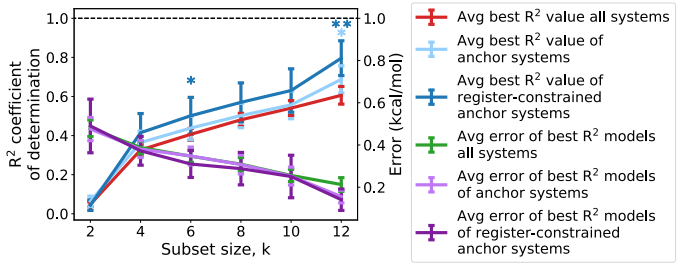
R^2 coefficient of determination and errors for prediction of experimental affinities from the Immune Epitope Database (IEDB) (n=167 antigen test set with all 20 residue mutations)

Residue volume and hydrophobicity are powerful 2-D descriptors, outperforming 5-D principal components



Residue volume and hydrophobicity features achieve similar accuracy to Blosum62 (n=10 dimensions) and outperform T-scale (n=5 dimensions) features for small residue subsets

GP regression accuracy increases for buried residues without competing conformations



GP regression accuracy improves when applied to point mutations of buried antigen 'anchor' residues and mutant systems that do not undergo MHC-antigen conformational changes, so-called 'register-constrained'

Discussion: This GP regression method predicts mutant binding affinities with uncertainty information in real-time across a visually- and physically-interpretable feature space. This approach is well-suited for active learning workflows where the optimal mutant residue is predicted and optimized from binding affinity knowledge of a small residue subset.

References:

1. Bell, D.R., Chen, S.H. https://doi.org/10.1021/acs.jcim.1c00458
2. Hie, B., et al. https://doi.org/10.1016/j.cels.2020.09.007

Contact Information: david.bell@nih.gov, chens@ornl.gov

Acknowledgements: The authors would like to thank Leili Zhang, Guojing Cong, Giacomo Domeniconi, Chih-Chieh Yang, Ruhong Zhou, Jeffrey K Weber, and Sangyun Lee for insightful discussions. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.