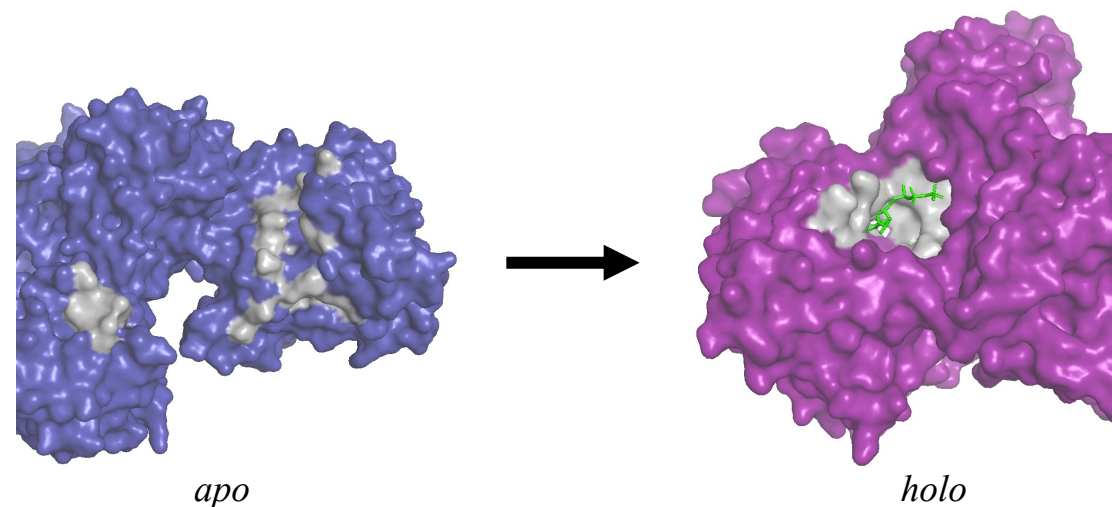# SiteFEATURE: Using the Protein Microenvironment for Cryptic Pocket Prediction
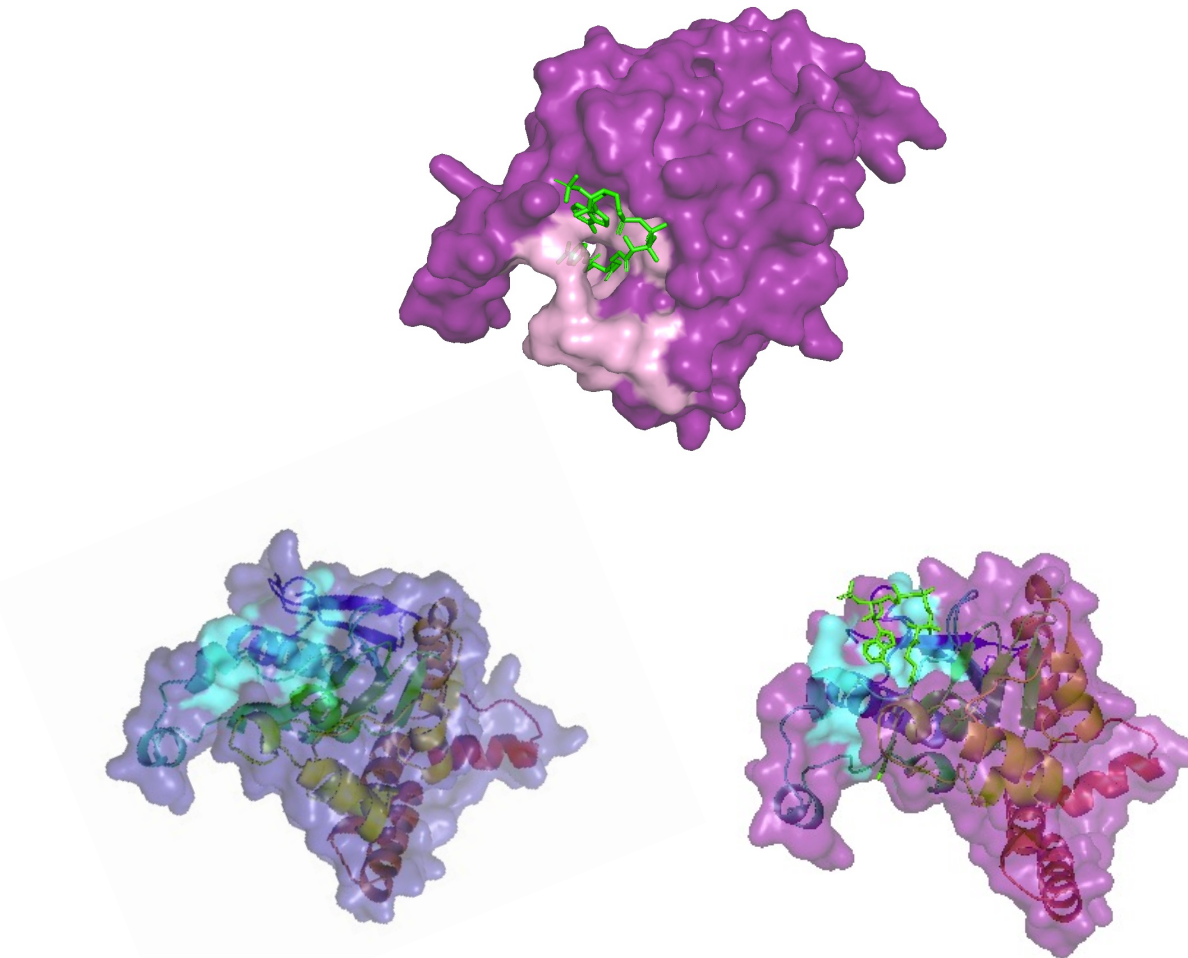
Kristy A. Carpenter[1,†], Yu-Chen Lo[2], and Russ B. Altman[1,3]

[1]Department of Biomedical Data Science, [2]Department of Pediatrics, [3]Departments of Bioengineering, Genetics, and Medicine, Stanford University, Stanford, CA 94305, USA

Background

apo → *holo*

Methods

Protein Surface Microenvironment — Protein Pocket Microenvironment — Hotspot Detections

Microenvironment Comparisons

Sub-Pockets Generation and Scoring

Pocket1 — Pocket2

Binding Site Identification

Pocket1 Pocket2

Results

ROC Curve for Holo Cryptic Sites
max residue score
mean dist

- 5 (AUC = 0.806)
- 10 (AUC = 0.663)
- 15 (AUC = 0.562)
- 50 (AUC = 0.496)

ROC Curve for Apo Cryptic Sites
max residue score
mean dist

- 5 (AUC = 0.571)
- 10 (AUC = 0.585)
- 15 (AUC = 0.554)
- 50 (AUC = 0.553)

Discussion

- Preliminary results indicate strong prediction of pockets from holo forms but more challenges with apo forms
- Visual analysis of randomly-selected cases suggests that the algorithm can identify true pockets in some cases
- More appropriate evaluation metrics such as hits in top k or mean reciprocal rank may give better estimates of performance
- The clustering algorithm and scoring function are still undergoing refinement
- After further optimization, future work includes evaluation on held-out CryptoSite [1] test set, pipelining with FragFEATURE [2] for lead discovery, and deployment on the AlphaFold Protein Structure Database [3] for estimation of the size of the cryptic pocketome and druggable proteome
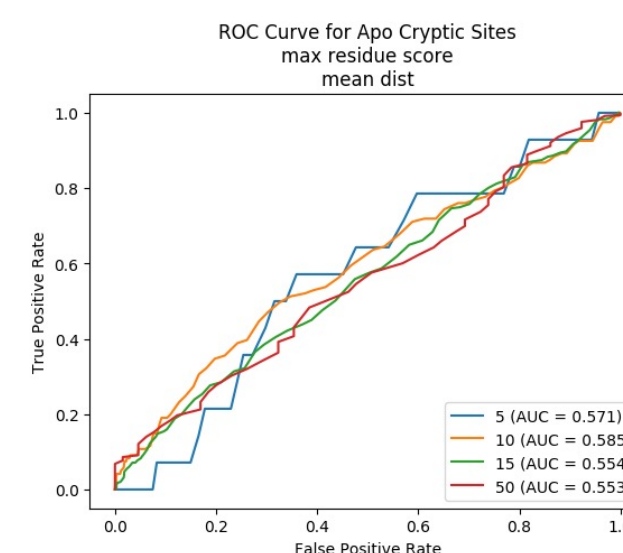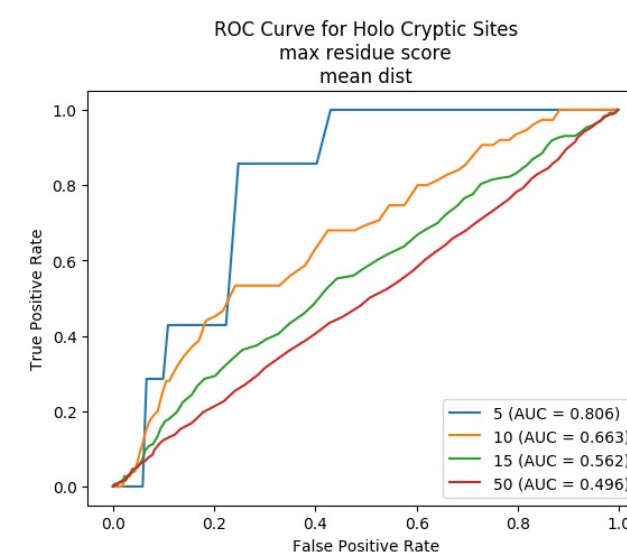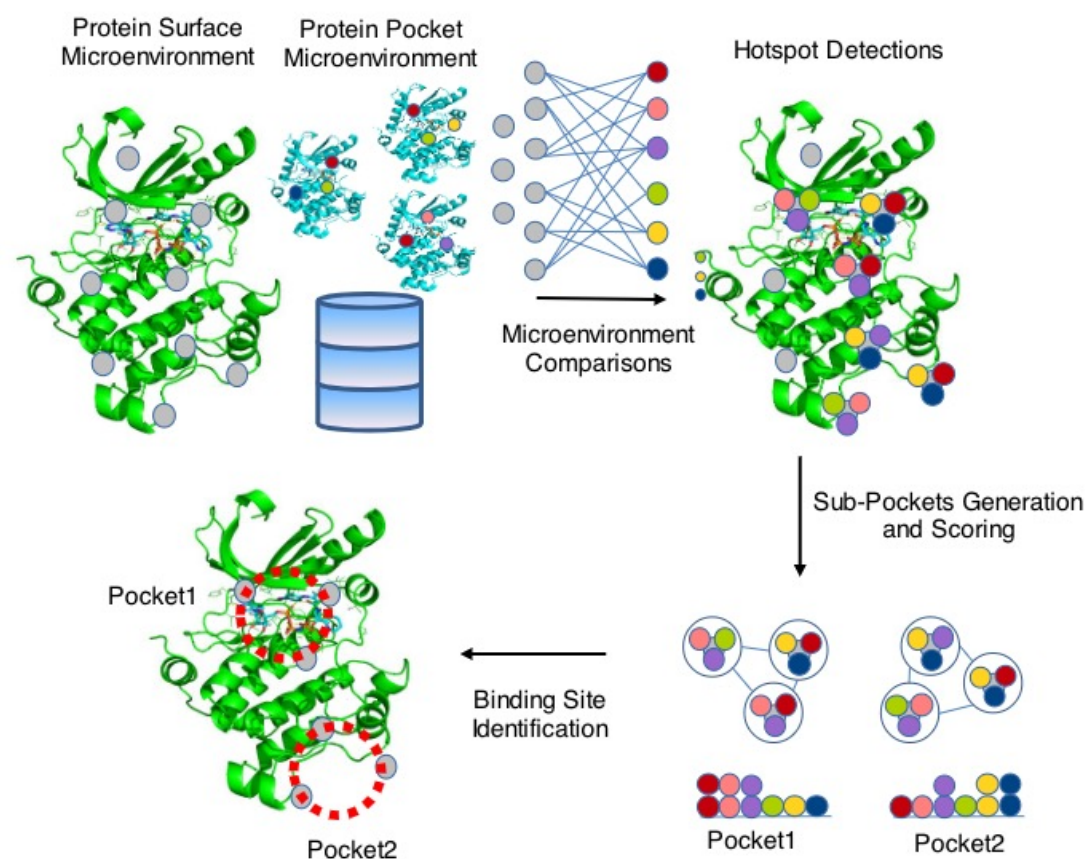
References

[1] P. Cimermancic, P. Weinkam *et al., J. Mol. Biol.* **428** (2016).
[2] G. W. Tang and R. B. Altman, *PLoS Comput. Biol.* **10** (2014).
[3] K. Tunyasuvunakool, J. Adler *et al., Nature.* **596** (2021).

†Email: kcarp@stanford.edu

Stanford MEDICINE

STANFORD BIOSCIENCES