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

[†]Email: kcarp@stanford.edu

Background

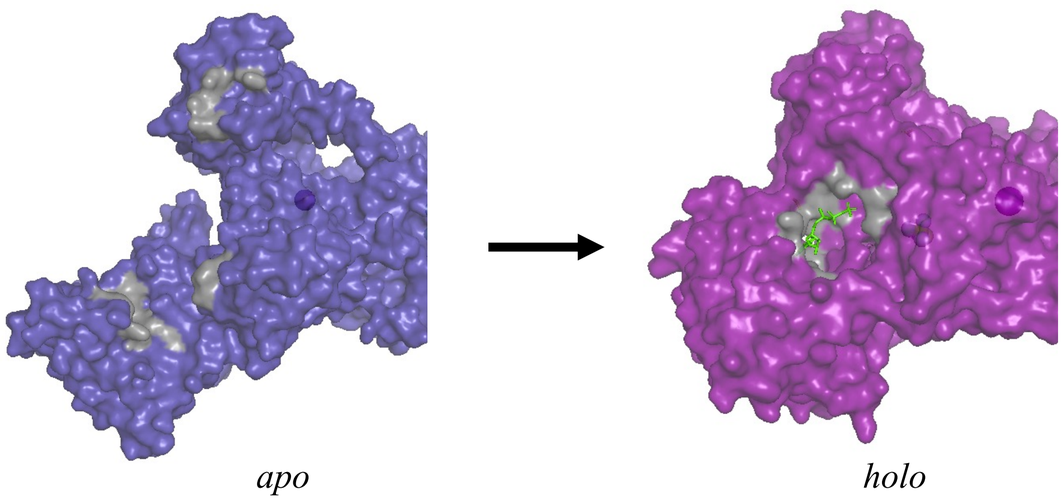

Fig. 1. SR calcium ATPase in the apo (left; PDB ID: 1SU4) and holo (right; PDB ID: 3FGO) forms. Ligand (ACP) shown on holo form in green. Residues within 5Å of the ligand in the holo form colored in gray. Dramatic structural change in binding site between apo and holo forms indicates the existence of a cryptic pocket.

- Identification of ligand binding pockets is an essential early step for structure-based drug discovery
- Most pocket-finding algorithms (e.g. Fpocket [1], ConCavity [2]) use geometric approaches to identify cavities on the protein surface
- Cryptic pockets undergo substantial structural change upon binding a ligand and are therefore undetectable by geometric pocket-finding methods in the apo form
- Previous work [3] used molecular dynamics (MD) information to identify cryptic pockets, but MD simulations are very computationally expensive and time-consuming

Methods

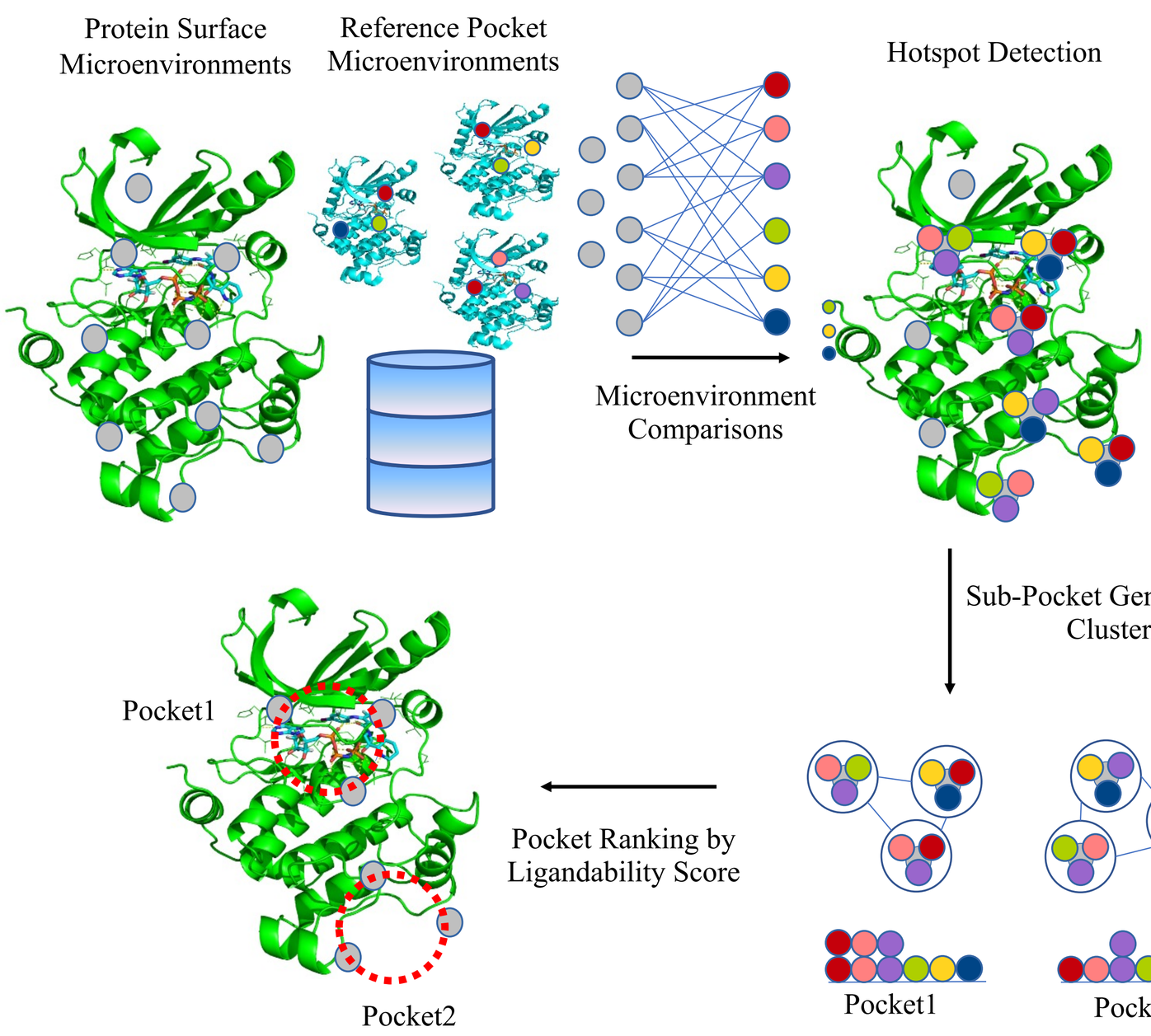

Fig. 2. SiteFEATURE framework.
FEATURE vectors are computed for every residue on the target protein surface, then compared those of a reference library of known binding pockets. Residues with high similarity scores (as defined by PocketFEATURE [4]) to the reference library are termed "hotspots". Hotspots and their surrounding residues are clustered into pockets using the CAST algorithm, then ranked by ligandability score and presented to the user for manual review.

Results

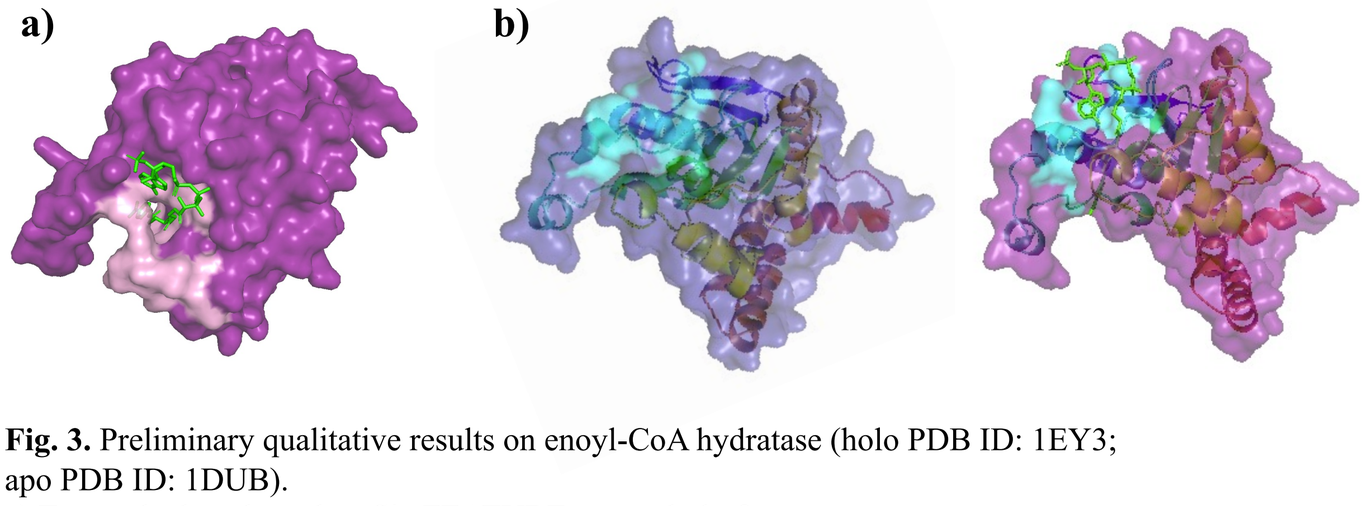

Fig. 3. Preliminary qualitative results on enoyl-CoA hydratase (holo PDB ID: 1EY3; apo PDB ID: 1DUB).
a) Top-ranked pocket when SiteFEATURE run on holo form.
b) Top-ranked pocket when SiteFEATURE run on apo form, visualized on both apo (blue) and holo (purple) forms.

- 55.41% of targets in holo form have at least one correct pocket predicted (mean hits@5 = 0.854, mean hits@10 = 0.951, mean reciprocal rank = 0.597)
- 32.39% of targets in apo form have at least one correct pocket predicted (mean hits@5 = 0.826, mean hits@10 = 0.913, mean reciprocal rank = 0.666)

Discussion

- Visual analysis of randomly-selected cases suggests that the algorithm can identify true pockets from both holo and apo forms
- Percentage of targets with correct pockets predicted can be improved by refining microenvironment feature calculation and clustering algorithm
- Improving the rank of correctly predicted pockets can be accomplished through further work on the scoring function
- Machine learning may be incorporated through the microenvironment feature calculation and scoring function

References

[1] V. Le Guilloux, P. Schmidtke, & P. Tuffery, *BMC Bioinf.* ***10*** (2009).
[2] J.A. Capra, R.A. Laskowski *et al., PLoS Comput. Biol.* ***23*** (2009).
[3] P. Cimermancic, P. Weinkam *et al., J. Mol. Biol.* ***428*** (2016).
[4] T. Liu & R.B. Altman, *PLoS Comput. Biol.* ***7*** (2011).

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship.