DeeplyTough: Learning Structural Comparison of Protein Binding Sites (Supporting Information)

Martin Simonovsky[*,†,‡,¶] and Joshua Meyers[*,†]

†BenevolentAI, London W1T 5HD, United Kingdom

‡École des Ponts ParisTech, 77455 Champs sur Marne, France

¶Université Paris-Est, 77455 Champs sur Marne, France

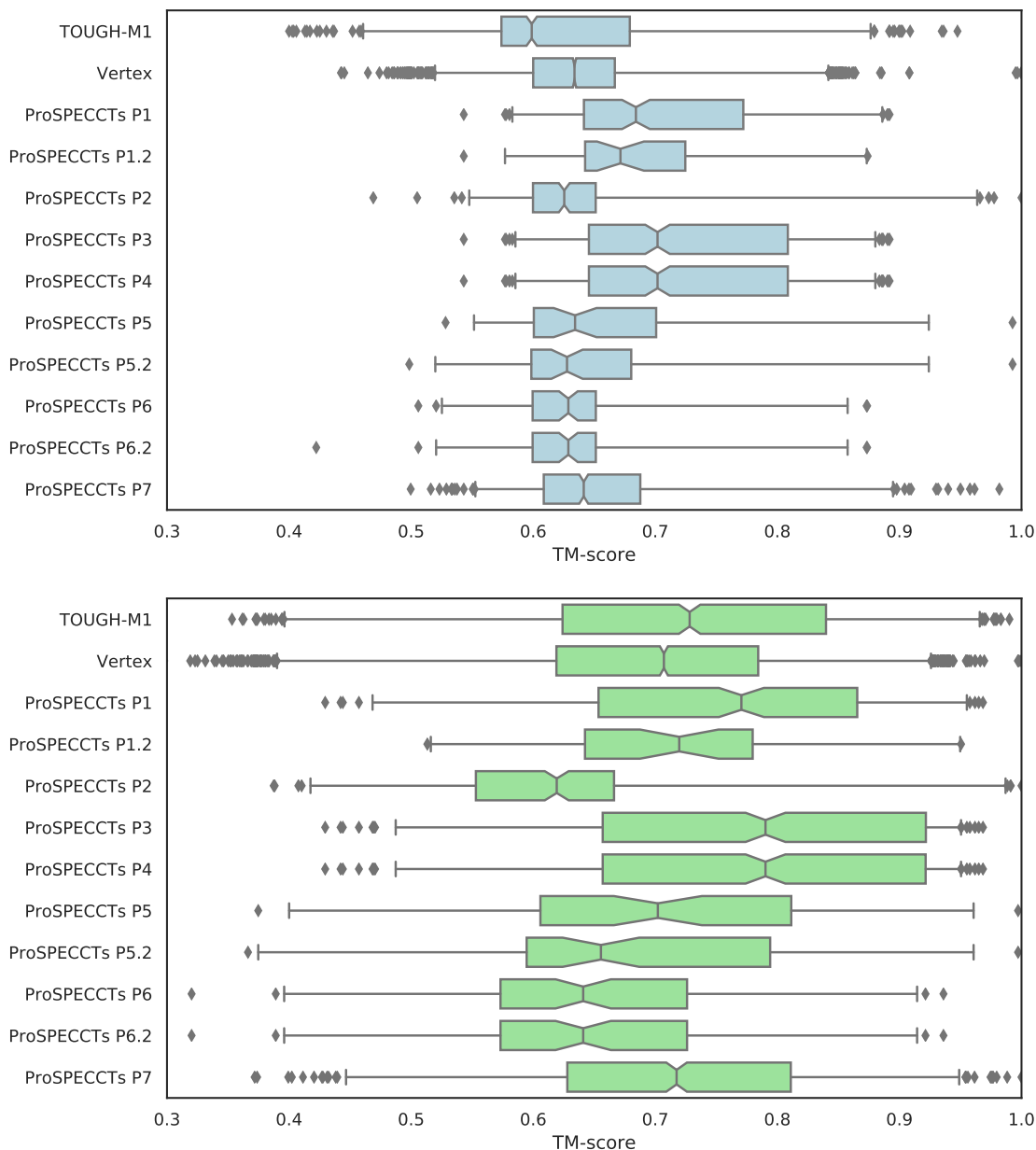E-mail: martin.simonovsky@enpc.fr; joshua.meyers@benevolent.ai

Figure S1: Box plots of TM-scores indicating the structurally similarity of pockets in a test set to their nearest neighbor pockets in a training set (again, in terms of TM-score) for each training scenario in this work. For TOUGH-M1 dataset, a single random permutation out of ten is visualized. The scores are computed with TM-align[1], with 1.0 indicating identity and 0.0 no alignment. Whiskers mark 1% and 99% percentiles of a distribution. Pockets in TOUGH-M1 can be defined in two ways: i) computationally with Fpocket 2.0[2] as used for training DeeplyTough (blue plots; top), or ii) around their bound ligands as in Vertex and ProSPECCTs (green plots; bottom). In general, the distributions suggest that there is no strong contamination between training and test sets (especially with computationally defined pockets), with only several pocket pairs being near-identical.
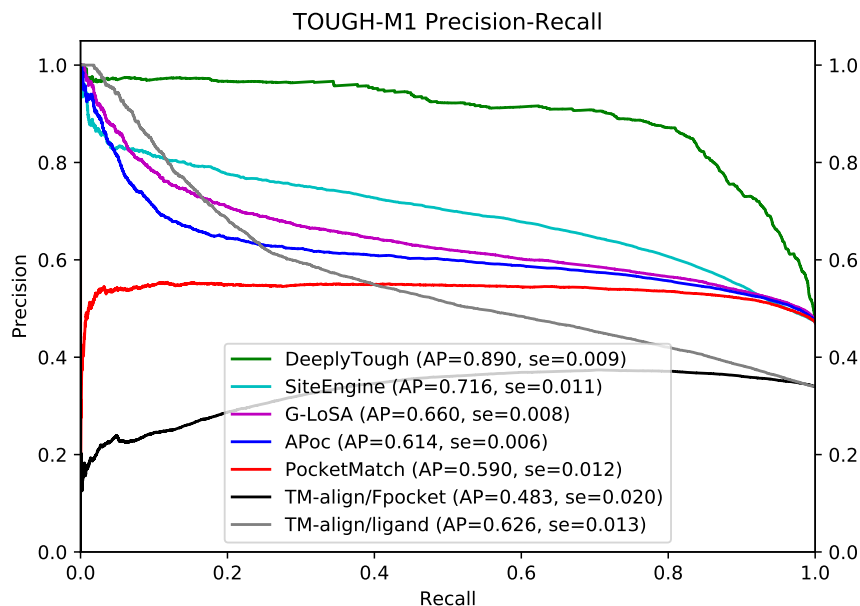
Figure S2: Precision-recall plot with associated average precision (AP) values evaluating the performance of pocket matching algorithms on TOUGH-M1 testing folds. Standard error, denoted as *se*, is measured over ten random splits.
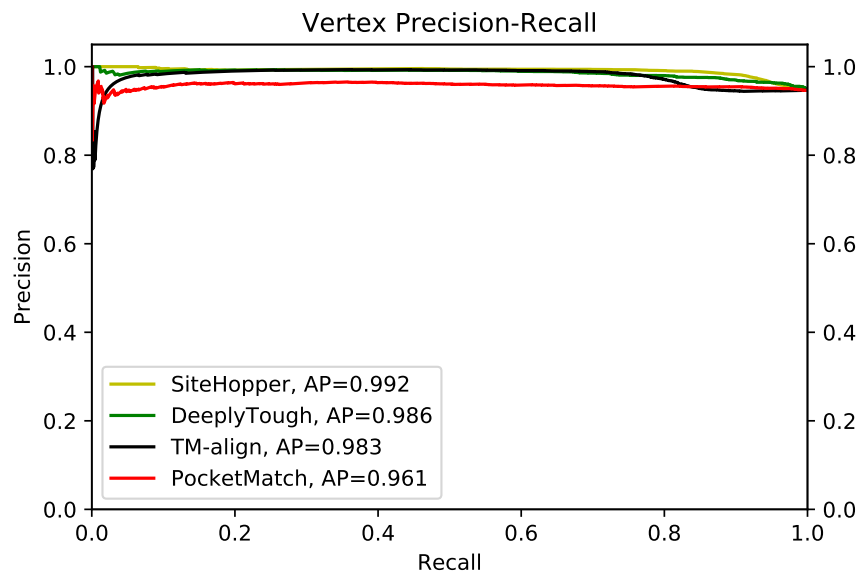

Figure S3: Precision-recall plot with associated average precision (AP) values evaluating the performance of pocket matching algorithms on the Vertex dataset (6,977 protein pairs).

Table S1: AP values for DeeplyTough on each of ten ProSPECCTs datasets.

	P1	P1.2	P2	P3	P4	P5	P5.2	P6	P6.2	P7
DeeplyTough	0.78	0.90	0.57	0.76	0.77	0.33	0.28	0.37	0.37	0.45

References

(1) Zhang, Y.; Skolnick, J. TM-align: A Protein Structure Alignment Algorithm Based on the TM-score. *Nucleic acids research* **2005**, *33*, 2302–2309.

(2) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinform.* **2009**, *10*.