Supporting Information For
GPCRLigNet: Rapid Screening for GPCR Active Ligands Using Machine Learning

Jacob M Remington,[a] Kyle McKay,[a] Noah B Beckage,[a] Jonathon B Ferrell, [a] Severin T. Schneebeli,[a,b,c] and Jianing Li[a,c,d]*

a. Department of Chemistry, University of Vermont, Burlington, VT 05405.
b. Department of Industrial and Physical Pharmacy and Department of Chemistry, Purdue University, West Lafayette, IN 47906.
c. Department of Pathology, University of Vermont, Burlington, VT 05405
d. Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 47906.

* jianing-li@purdue.edu

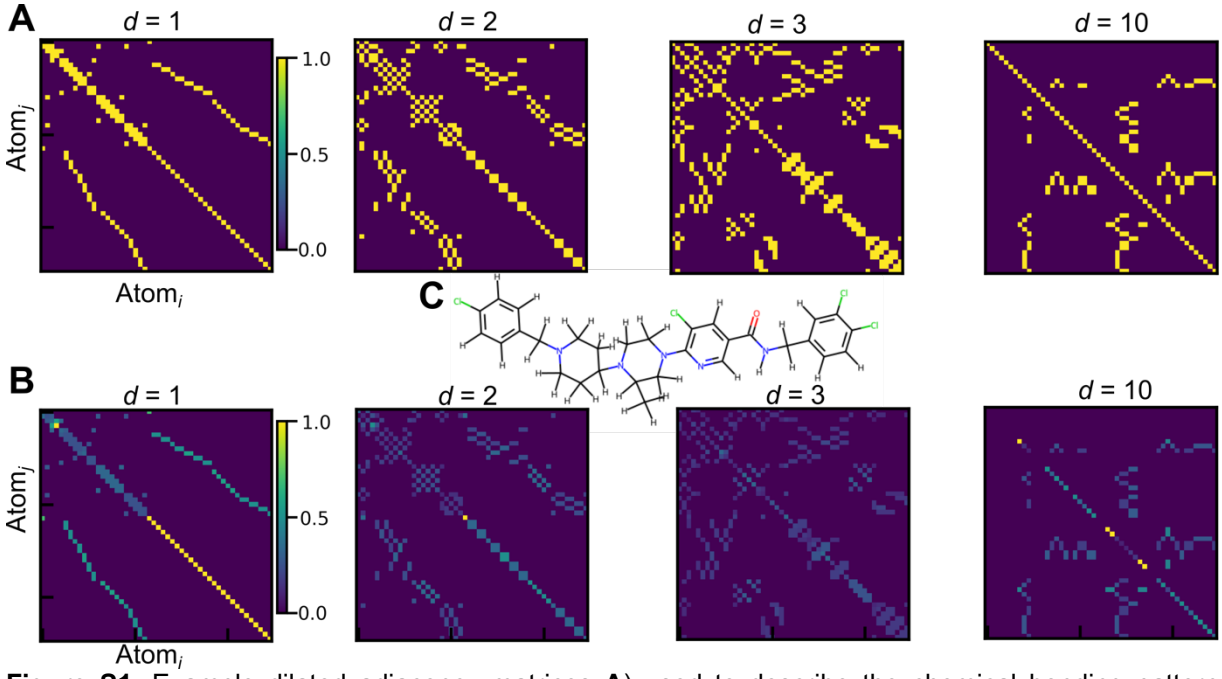

Figure S1. Example dilated adjacency matrices **A**) used to describe the chemical bonding pattern of chemicals with increasing dilation (*d* = 1, 2, 3, and 10) from left to right, and the corresponding edge-weighted adjacency matrices **B**. The molecule that these graphs represent is shown in **C**. Each dilated adjacency matrix encodes different scales of chemical bonding information such that the larger dilations allow a direct transfer of information from distant regions of the molecule.

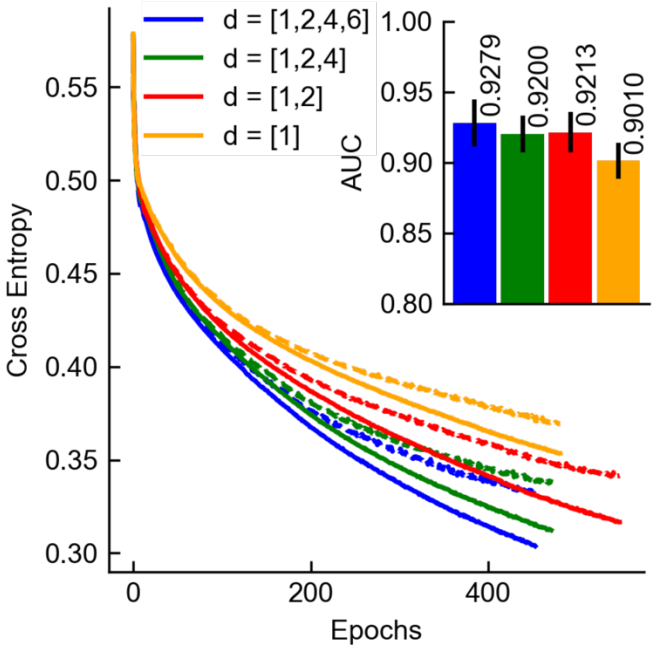

Figure S2. Training of dilated graph convolution neural networks for GPCR activity classification. The cross entropy loss as a function of training epoch is plotted for the training (solid) and validation (dashed) datasets. Models with decreasing number of dilations from four *d* values of 1, 2, 4, and 6 to only a single dilation with *d* of 1 (corresponding to the standard chemical adjacency matrix) are shown from blue to orange. The area under the ROC curve for the models evaluated on the validation set is shown in the inset. Error bars are 1σ values obtained from training on three datasets shuffled differently. Early stopping with a patience of 10 epochs was used to determine when to halt training. Increasing number of dilations improved the quality of the graph convolution networks.

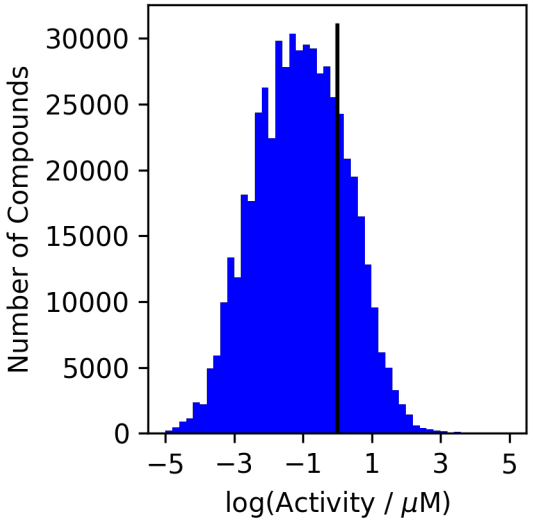

Figure S3. GPCR activity ($\log_{10}$ μM units) of the GLASS compounds is shown. The 1 μM activity cut-off is shown with a solid black line. While the GLASS compounds were imbalanced towards active molecules, the addition of DUD-E alleviated the issue. In addition, the 1 μM cutoff represented a practical goal for initial high throughput virtual screening protocols.

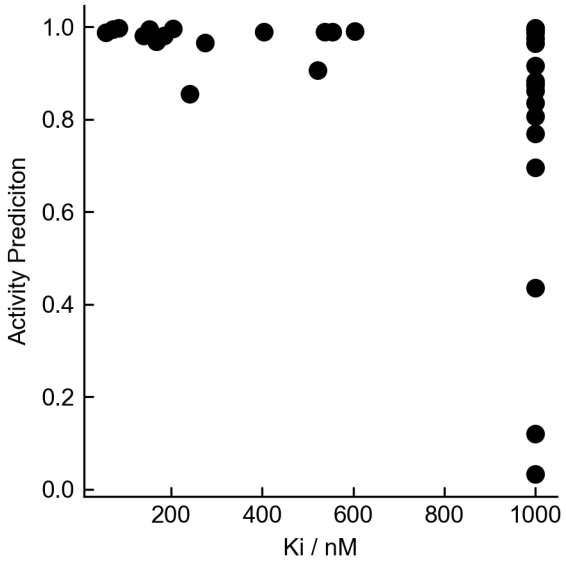

Figure S4. GPCR Activity prediction scores for the currently known antagonists of PAC1R. The molecules without exact Ki values are plotted with Ki = 1000 nM.