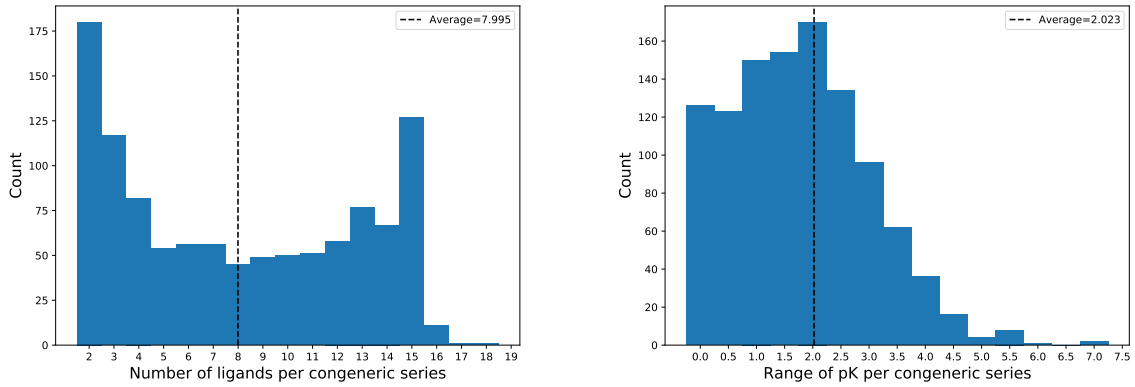
Supporting Information:

Improving $\Delta\Delta G$ predictions with a multi-task convolutional Siamese network

Andrew T. McNutt and David Ryan Koes*

Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA

E-mail: dkoes@pitt.edu



(a) Number of ligands per congeneric series (b) The pK range of each of the congeneric series

Figure S1: Statistics on the final, filtered BindingDB Congeneric Series dataset that was used for training and evaluation

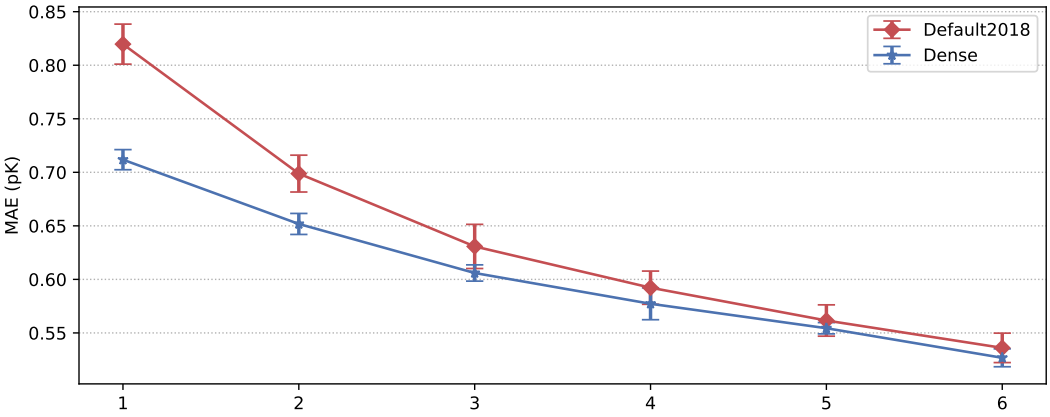


Figure S2: Evaluating mean absolute error (MAE) on the BindingDB congeneric series dataset when adding more information about each congeneric series to the training set. Error bars indicate ±1 standard deviation of 25 individual models (only 5 for Dense).

Table S1: Ablations and their impact on absolute affinity prediction. Parentheses indicate the ±1 standard deviation of the 25 train/test versions. **Bold** indicates that it is not significantly different from the Standard model ($p > 0.005$).

Ablation	Pearson's R	RMSE (pK)	MAE (pK)
Standard	0.864(±0.0124)	0.841(±0.0377)	0.556(±0.0303)
No $L_{\Delta\Delta G}$	0.873(±0.0057)	0.814(±0.018)	0.517(±0.0136)
No $L_{\Delta G}$	0.0278(±0.107)	101(±34.4)	101(±34.5)
No L_{Rotation}	**0.866(±0.0059)**	**0.833(±0.0169)**	**0.553(±0.0141)**
No $L_{\text{Consistency}}$	**0.866(±0.00559)**	**0.833(±0.0171)**	**0.559(±0.0158)**
No $L_{\Delta\Delta G}$, $L_{\text{Consistency}}$	0.873(±0.00425)	0.814(±0.0132)	0.514(±0.0121)
No $L_{\Delta G}$, $L_{\text{Consistency}}$	−0.0118(±0.0494)	6.65(±0.465)	6.44(±0.48)
Concatenation	**0.87(±0.00477)**	**0.824(±0.0155)**	0.531(±0.0158)
No Siamese Network	0.846(±0.0175)	0.883(±0.0479)	0.616(±0.0442)
Subtraction, Single-order	0.834(±0.0111)	0.919(±0.0236)	0.666(±0.027)
Concatentation, Single-order	0.842(±0.00876)	0.9(±0.0221)	0.64(±0.0241)

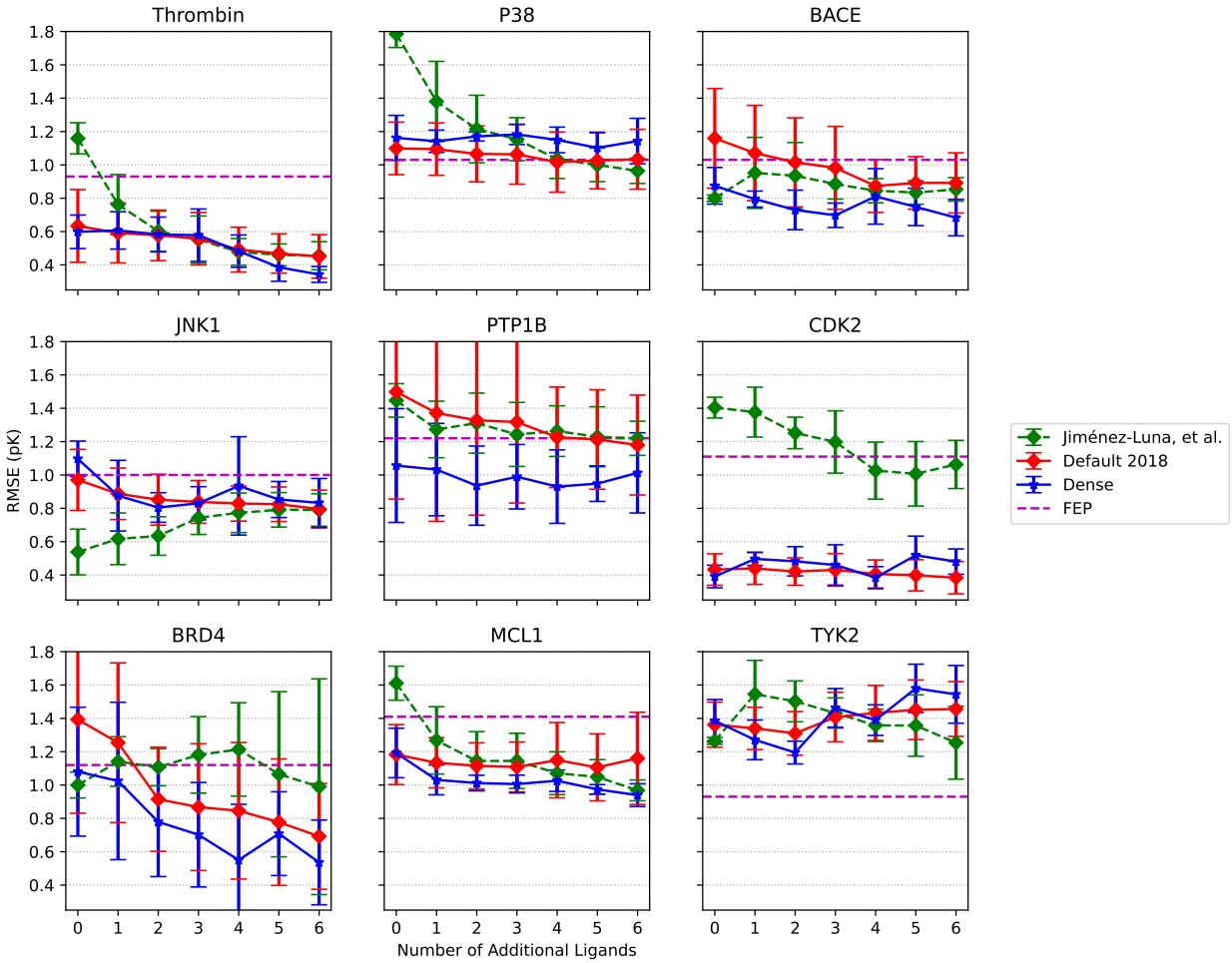


Figure S3: RMSE on both the Mobley et al.[1] (BRD4) and Wang et al.[2] external datasets evaluated in terms of both no-shot (0) and few-shot performance. FEP performance as reported in Jiménez-Luna et al.[3].

Table S2: Effect of seeing both orderings of ligand pairs vs only one ordering of ligand pairs during training. We sum the prediction of $\Delta\Delta G$ for each ordering of the test set pairs and take the average across all ligand pairs, noting that the sum should be zero for each ligand pair. Subtraction means latent space subtraction for RBFE prediction. Concatenation means latent space concatenation for RBFE prediction. Combinations means only one ordering of a ligand pair is seen during training, otherwise both orderings of a ligand pair are seen during training.

Ablation	Difference from Reverse Prediction (pK)
Normal Model	0.0004 ($\pm$ 0.0011)
Subtraction & Combinations	0.1909 ($\pm$ 0.0471)
Concatenation	0.0039 ($\pm$ 0.0078)
Concatenation & Combinations	0.2675 ($\pm$ 0.0595)

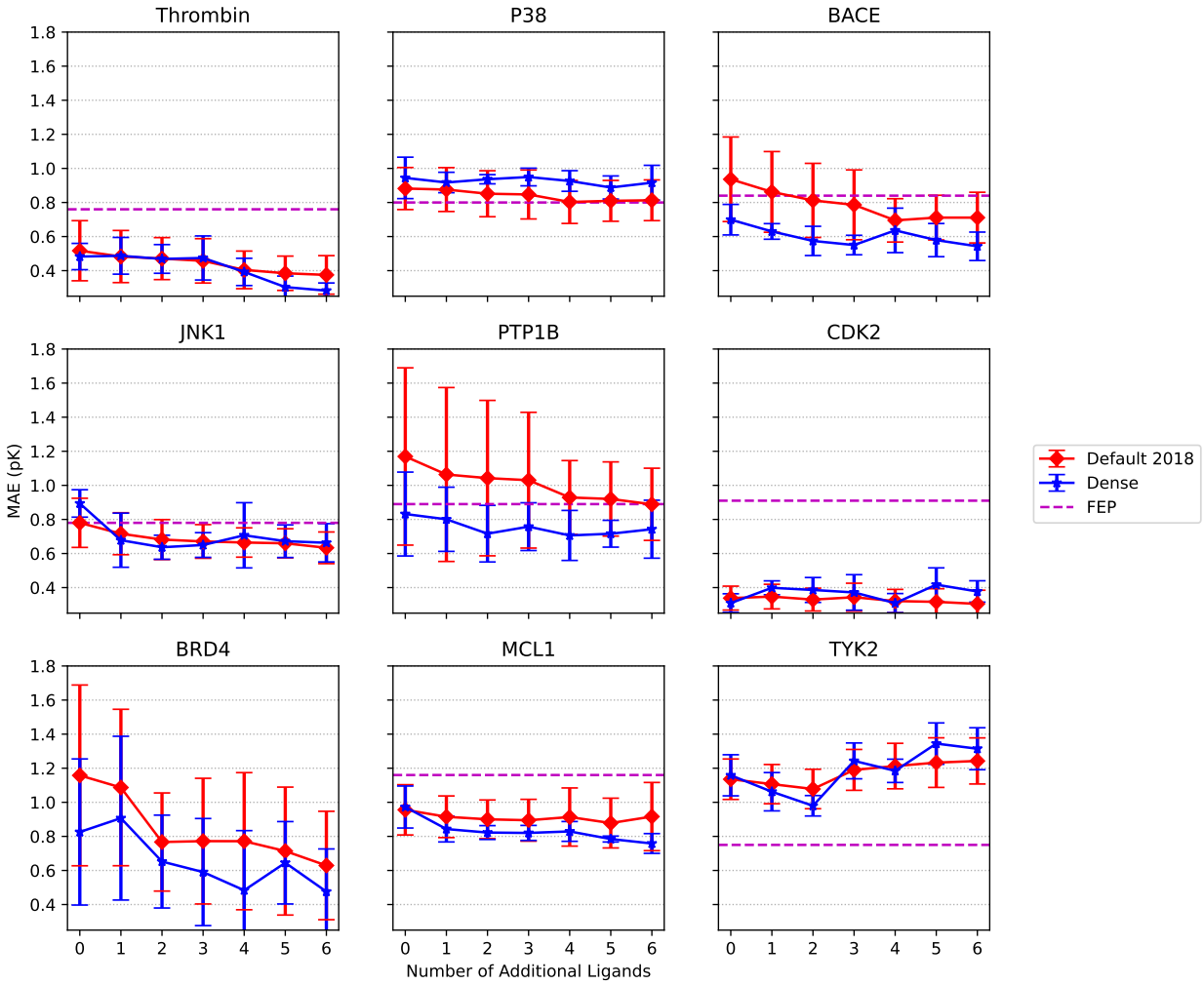


Figure S4: MAE on both the Mobley *et al.*[1] (BRD4) and Wang et al.[2] external datasets in terms of both no-shot (0) and few-shot performance

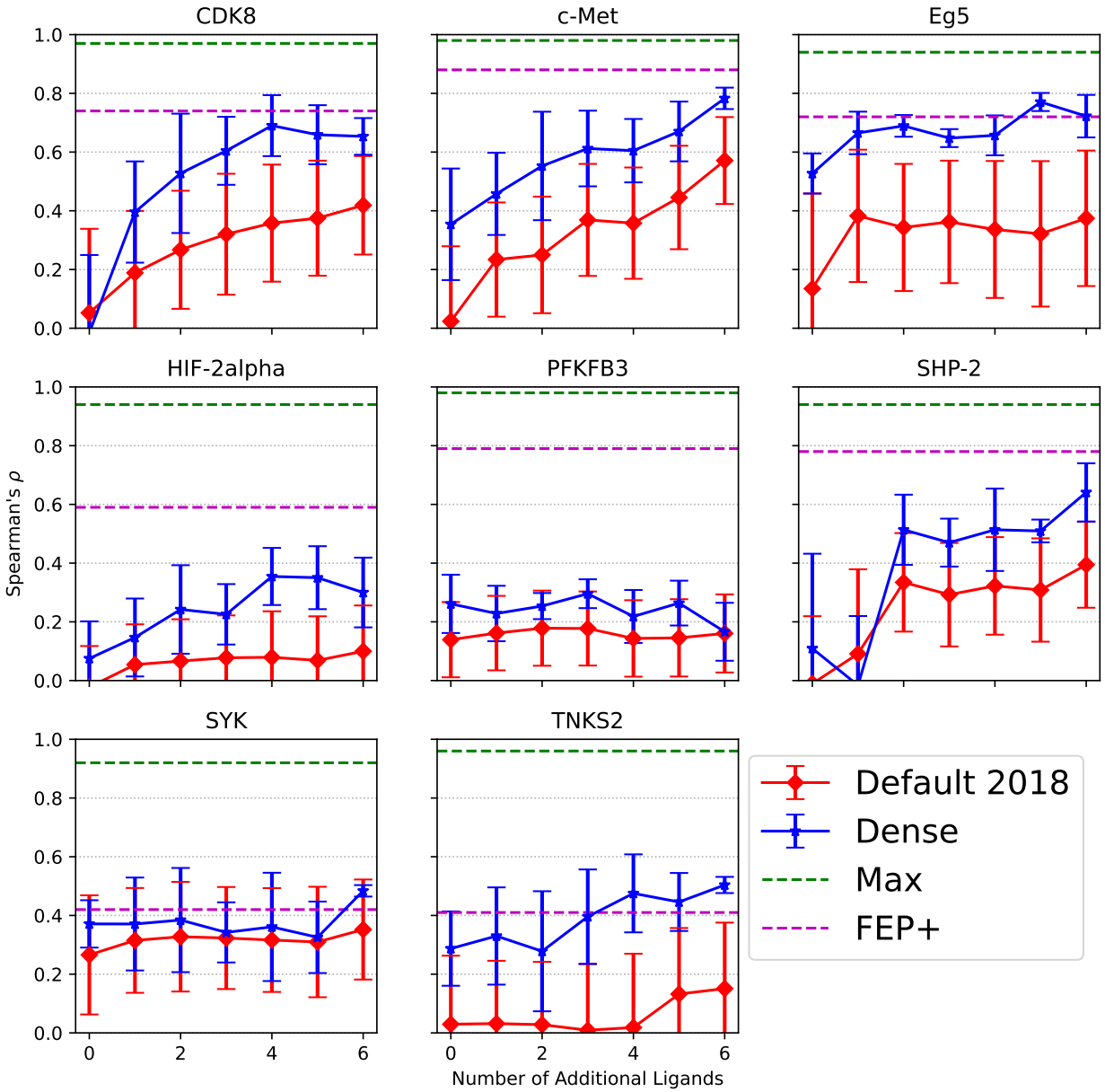


Figure S5: Spearman's (ρ) correlation on the Schindler et al.[4] external dataset in terms of both no-shot (0) and few-shot performance. FEP+ performance given by Schindler et al.[4]. `Max` is the maximum possible correlation given the error in the experimental assay.

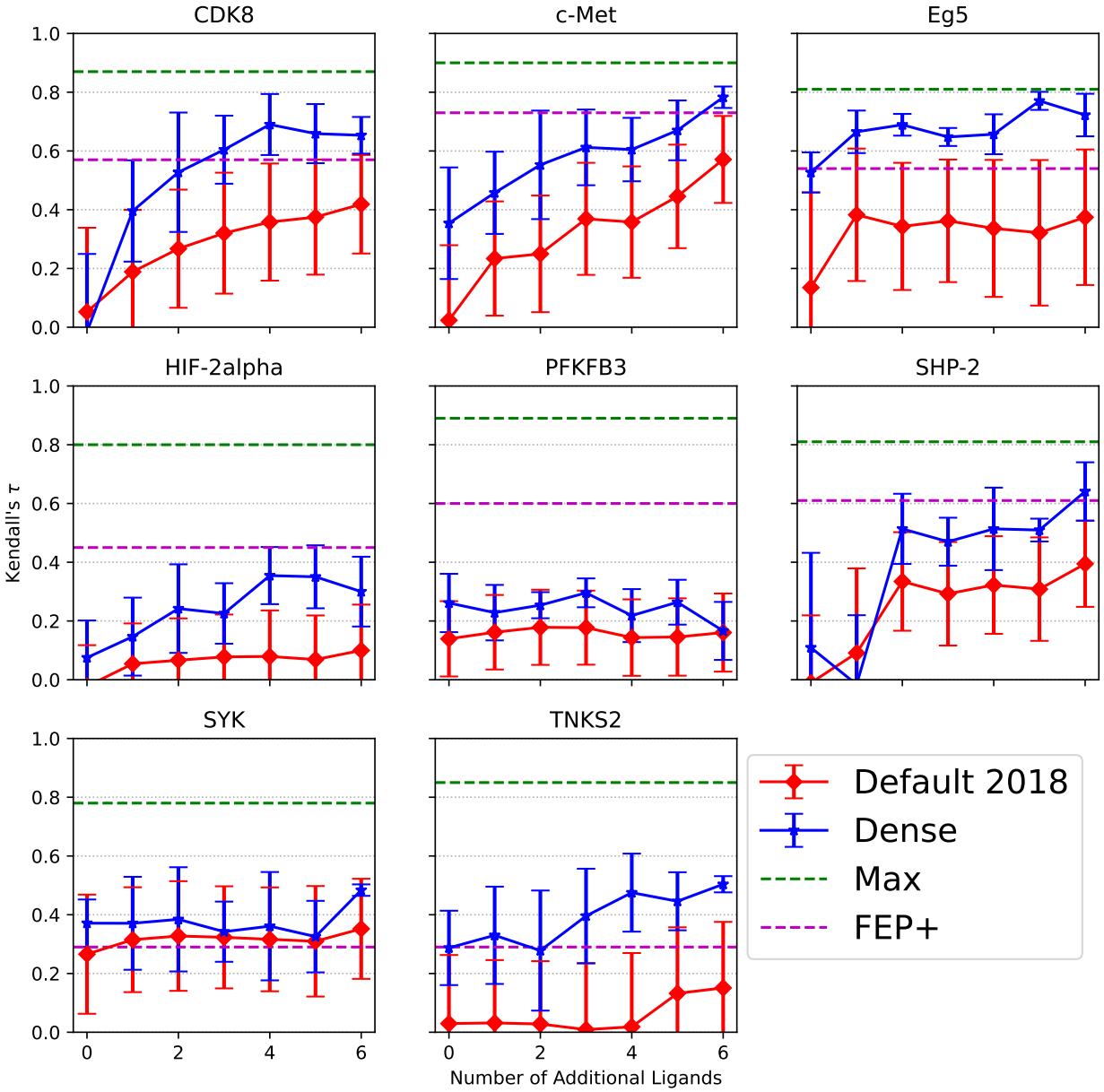


Figure S6: Kendall's τ coefficient on the Schindler et al.[4] external dataset in terms of both no-shot (0) and few-shot performance. FEP+ performance given by Schindler et al.[4]. `Max` is the maximum possible correlation given the error in the experimental assay.

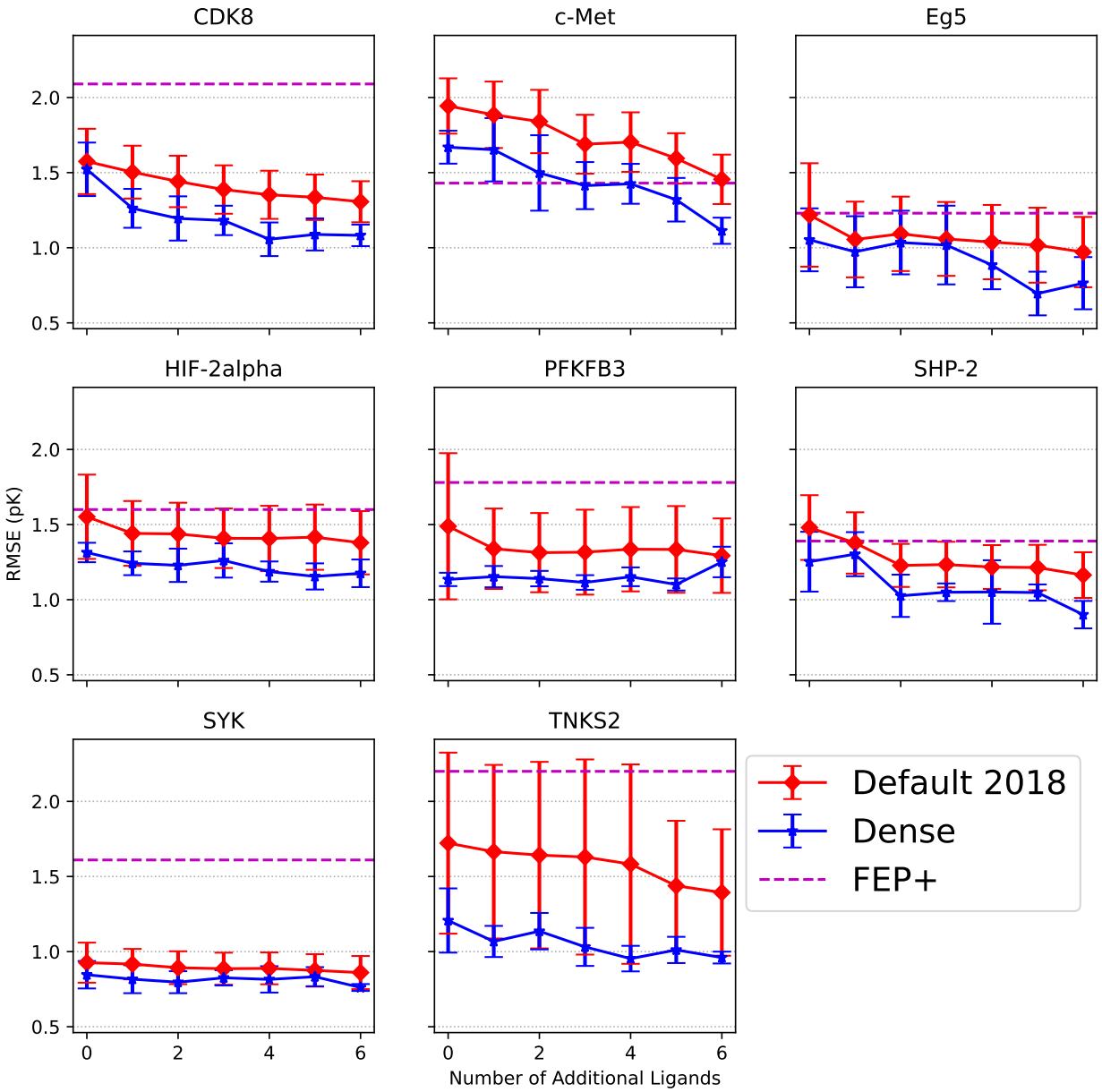


Figure S7: RMSE on the Schindler et al.[4] external dataset in terms of both no-shot (0) and few-shot performance. FEP+ performance given by Schindler et al.[4].

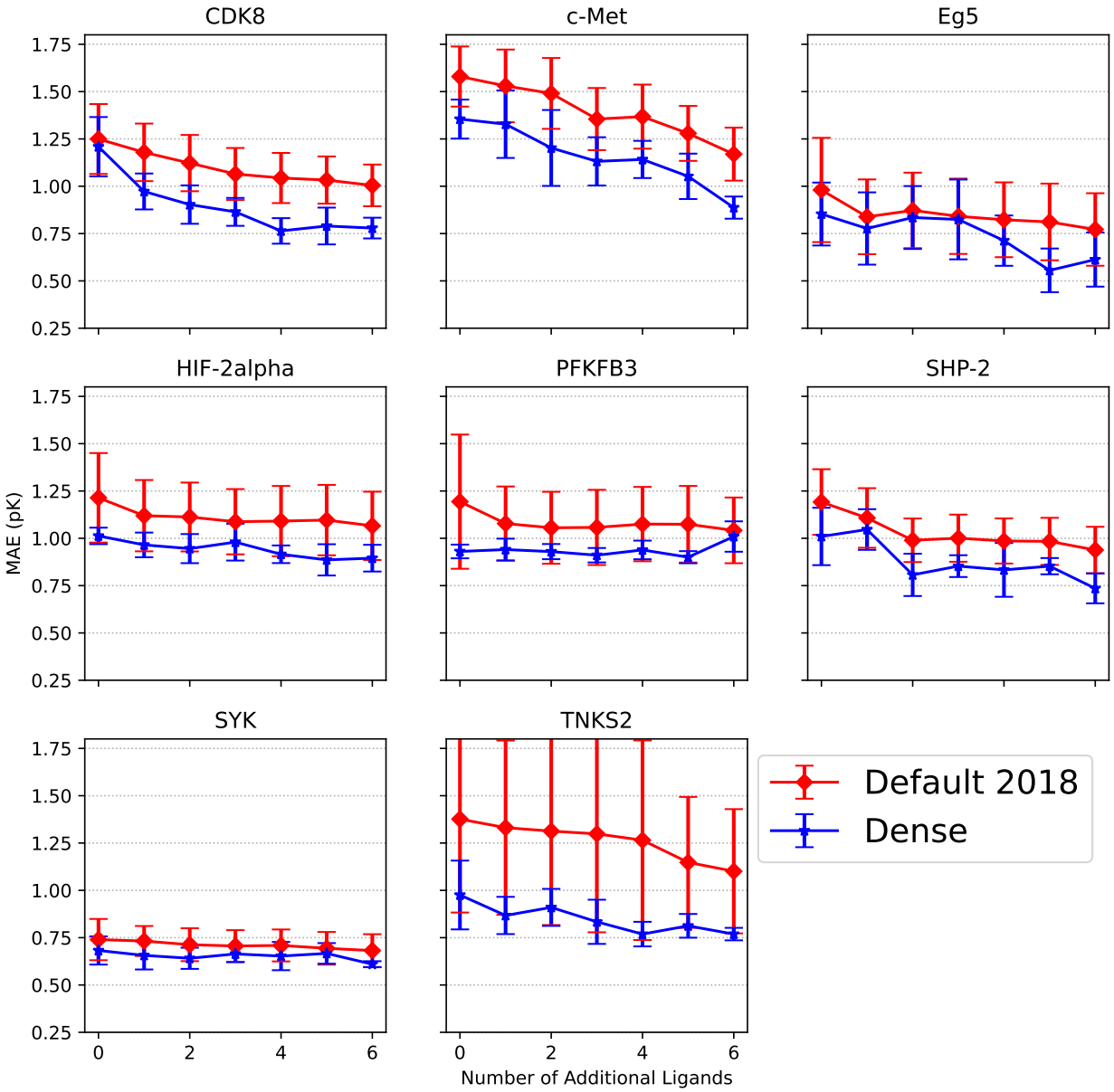


Figure S8: MAE on the Schindler et al.[4] external dataset in terms of both no-shot (0) and few-shot performance.

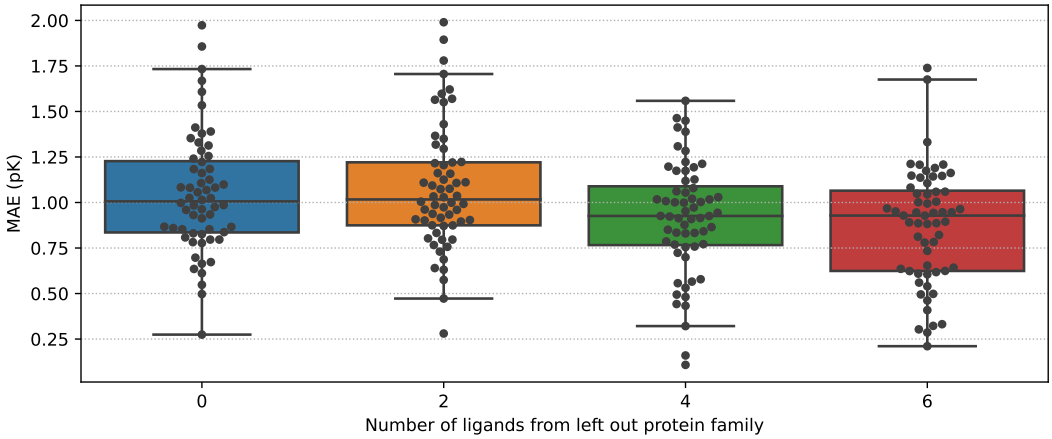


Figure S9: Average MAE on left out protein family for the protein families in the BindingDB dataset. The MAE of the model decreases as we include information from the left out test set.

Table S3: Minimum protein distance (`Biopython.align.globalxx`[5]) between each external test set and the BindingDB Congeneric Series set, with the PDB ID of the protein in the BindingDB set. Maximum ligand similarity (`rdkit.DataStructs.FingerprintSimilarity`[6]) of the ligands in each external test set and the BindingDB Congeneric Series set with the number of ligands in the external set that share the maximal ligand similarity value. Pearson's R and RMSE are the no-shot performance of our model on each congeneric series trained on all of the BindingDB data. The Default2018 performance is given first and the Dense performance is given in parentheses for both Pearon's R and RMSE.

External Dataset	Minimum Protein Distance (PDBID)	Max Ligand Similarity (% of ligands at max similarity)	Pearson's R	RMSE (pK)
PTP1B	0.431034 (1BZJ)	1.0 (95.7)	0.382 (0.626)	1.498 (1.056)
MCL1	0.606667 (1A29)	0.670551 (2.4)	0.292 (0.219)	1.183 (1.192)
JNK1	0.478261 (MCL1)	0.540767 (4.8)	0.309 (-0.099)	0.970 (1.097)
TYK2	0.462069 (3E63)	0.587719 (6.2)	0.327 (0.248)	1.362 (1.386)
BRD4	0.669421 (1TFQ)	0.657881 (10.0)	0.475 (0.693)	1.393 (1.080)
Thrombin	0.269481 (1ETS)	0.757866 (9.1)	0.270 (0.011)	0.634 (0.599)
CDK2	0.463063 (3PY1)	1.0 (100.0)	0.949 (0.955)	0.433 (0.391)
P38	0.465714 (3ERK)	0.557272 (2.9)	0.338 (0.135)	1.099 (1.162)
BACE	0.080488 (2OHM)	0.586625 (2.8)	0.141 (0.374)	1.159 (0.875)

References

(1) Mobley, D. L.; Gilson, M. K. Predicting binding free energies: frontiers and benchmarks. *Annual review of biophysics* **2017**, *46*, 531–558.

(2) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J., et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *Journal of the American Chemical Society* **2015**, *137*, 2695–2703.

(3) Jiménez-Luna, J.; Pérez-Benito, L.; Martínez-Rosell, G.; Sciabola, S.; Torella, R.; Tresadern, G.; De Fabritiis, G. DeltaDelta neural networks for lead optimization of small molecule potency. *Chemical science* **2019**, *10*, 10911–10918.

(4) Schindler, C. E.; Baumann, H.; Blum, A.; Böse, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D., et al. Large-scale assessment of binding free energy calculations in active drug discovery projects. *Journal of Chemical Information and Modeling* **2020**, *60*, 5457–5474.

(5) Cock, P. J.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B., et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423.

(6) RDKit: Open-source cheminformatics. `http://www.rdkit.org`, [Online; accessed 11-April-2013].