Electronic Supplementary Material (ESI) for Digital Discovery.
This journal is © The Royal Society of Chemistry 2024

Supplementary Information

FSL-CP: A Benchmark for Small Molecule Activity Few-Shot Prediction using Cell Microscopy Images

Son V. Ha, Lucas Leuschner, and Paul Czodrowski*

Department of Chemistry, Johannes Gutenberg University Mainz, Germany

E-mail: czodpaul@uni-mainz.de,ORCID:0000-0002-7390-879

List of Figures

List of Tables

This document provides supplementary information for the publication "FSL-CP: A Benchmark for Small Molecule Activity Few-Shot Prediction using Cell Microscopy Images". We report details about the 201 tasks of the dataset, and additional performance metrics.

The dataset, model codes and performances of all models (including those not reported in the publication) are all publicly available on Github:

`https://github.com/czodrowskilab/FSL_CP`.

Model Hyperparameters

Here we include more details about the hyperparameters of benchmark models. If the reader wants to reproduce the result of the paper, we would encourage running the codes from our GitHub repository.

protonet_cp+:

num_episodes _train = 50000

num_episodes_val = 100

loss function: $nn.CrossEntropyLoss()$

optimizer: $optim.Adam(model.parameters(), lr = 0.0001)$

learning rate scheduler: $StepLR(optimizer, step_size = 20000, gamma = 0.1)$

Backbone model: 3-hidden-layer Fully-connected Neural Network

Size of output layer of backbone model = 256

Distance = ’Euclidean’

protonet_cp:

num_episodes _train = 50000

num_episodes_val = 100

loss function: $nn.CrossEntropyLoss()$

optimizer: $optim.Adam(model.parameters(), lr = 0.0001)$

learning rate scheduler: $StepLR(optimizer, gamma = 0.1)$

Backbone model: 3-hidden-layer Fully-connected Neural Network

Size of output layer of backbone model = 512

Distance = ’Euclidean’

protonet_img:

num_episodes _train = 30000

num_episodes_val = 100

loss function: $nn.CrossEntropyLoss()$

optimizer: $optim.Adam(model.parameters(), weight\_decay = 1e - 4)$

learning rate scheduler: $StepLR(optimizer, step_size = 10000, gamma = 0.1)$

Image transformation: $RandomCrop(300), Resize(200)$

Backbone model: ResNet50

Size of output layer of backbone model = 1600

Distance = 'CosineSimilarity'

maml_cp+:

num_episodes _train = 32000

num_episodes_val = 100

adaptation_steps=3

loss function: $nn.BCEWithLogitsLoss()$

optimizer: $optim.Adam(lr = 0.001)$

Image transformation: $RandomCrop(100), Resize(85)$

Model: ResNet50

MAML Fast adaptation learning rate = 0.01

maml_img:

num_episodes _train = 32000

num_episodes_val = 100

adaptation_steps=3

loss function: $nn.BCEWithLogitsLoss()$

optimizer: $optim.Adam(lr = 0.001)$

Model: 3-hidden-layer Fully-connected Neural Network

MAML Fast adaptation learning rate = 0.01

singletask_cp:

max_epochs = 50

loss function: $nn.BCEWithLogitsLoss()$

optimizer: $optim.Adam(lr = 0.0001)$

learning rate scheduler: $StepLR(optimizer, step_size = 100, gamma = 0.1)$

Model: 3-hidden-layer Fully-connected Neural Network

multitask_cp:

pretrain max_epochs = 50

pretrain loss function: $multitask\_bce$ (Binary Cross Entropy)

pretrain optimizer: $optim.SGD(lr = 1e - 2, momentum = 0.9, weight\_decay = 1e - 4)$

pretrain learning rate scheduler: $StepLR(optimizer, step\_size = 20, gamma = 0.1)$

Model: 3-hidden-layer Fully-connected Neural Network

inference max_epochs = 50

inference loss function: $nn.BCEWithLogitsLoss()$

inference optimizer: $optim.Adam(lr = 0.0001)$

inference learning rate scheduler: $StepLR(optimizer, step\_size = 100, gamma = 0.1)$

logistic_cp+:

We did a RandomizedSearchCV on these hyperparameters:

'C' : [0.01, 0.1, 1.0, 10.0, 100.0]

Test Tasks Similarity

In this section, we include some statistics about the similarity between 18 test tasks.

In Figure 7, we measured the Jaccard Index for the unique InChiKeys from every task pair in D_test. We observed that the majority of tasks shares very few common InChiKeys. Exceptions are tasks 737826, 737824_1 and 737825 whose targets resemble Cytochrome P450.

Taking a closer look at these three tasks 737826, 737824_1 and 737825, they have 800, 840 and 779 unique InChiKeys, respectively. 2 datapoints from 2 tasks is similar if they have the same (InChiKey, labels) pair. Tasks (737826 and 737824_1) have 580 (InChiKey, labels) pair in common. For tasks (737826, 737825) and (737824_1, 737825) the number is 450 and 469. So around 60%-70% of the (InChiKeys, labels) pair is similar between these tasks.

In simpler words, these 3 tasks are 60%-70% 'similar' to each other. In our opinion, they are still different enough to be different tasks in the test set. However, we acknowledge that future data curation effort should take notice of similar tasks like this.

Table 1: **Test tasks**: Details of the assays in the test set and their targets in ChEMBL.

TASK_ID	assay_chembl_id	target_chembl_id	target_type
688267	CHEMBL1614530	CHEMBL2026	SINGLE PROTEIN
600886	CHEMBL1040692	CHEMBL364	ORGANISM
737826	CHEMBL1741325	CHEMBL3397	SINGLE PROTEIN
737824_1	CHEMBL1741323	CHEMBL3622	SINGLE PROTEIN
737825	CHEMBL1741324	CHEMBL340	SINGLE PROTEIN
1495405	CHEMBL3562136	CHEMBL612545	UNCHECKED
737053	CHEMBL1738598	CHEMBL612545	UNCHECKED
737400	CHEMBL1738606	CHEMBL5501	SINGLE PROTEIN
736947	CHEMBL1738312	CHEMBL1741220	SINGLE PROTEIN
752347	CHEMBL1794311	CHEMBL1977	SINGLE PROTEIN
752496	CHEMBL1794486	CHEMBL5027	SINGLE PROTEIN
752509	CHEMBL1794499	CHEMBL1795091	SINGLE PROTEIN
752594	CHEMBL1794584	CHEMBL1293258	SINGLE PROTEIN
809095	CHEMBL1964081	CHEMBL2007624	SINGLE PROTEIN
845173	CHEMBL2114784	CHEMBL1795085	SINGLE PROTEIN
845196	CHEMBL2114807	CHEMBL5990	SINGLE PROTEIN
954338	CHEMBL2354287	CHEMBL2362981	SINGLE PROTEIN
845206	CHEMBL2114817	CHEMBL4377	SINGLE PROTEIN

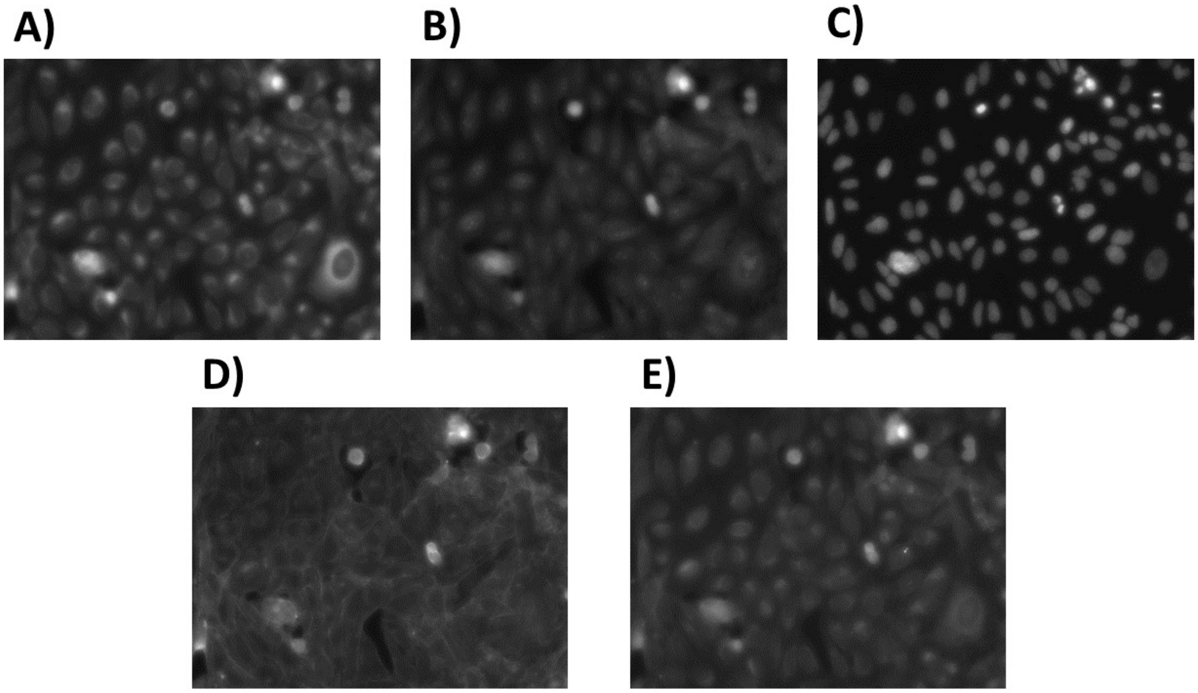


Figure 1: **Sample Cell Painting images from file 24294-I23-2.npz.** Each of these image are from the same view, but with different dyes: a)Mito, b)Ph_Golgi, c)Hoechst, d)ERSytoBleed, e)ERSyto. For modelling, we stack these 5 views to create one 5-channel image.

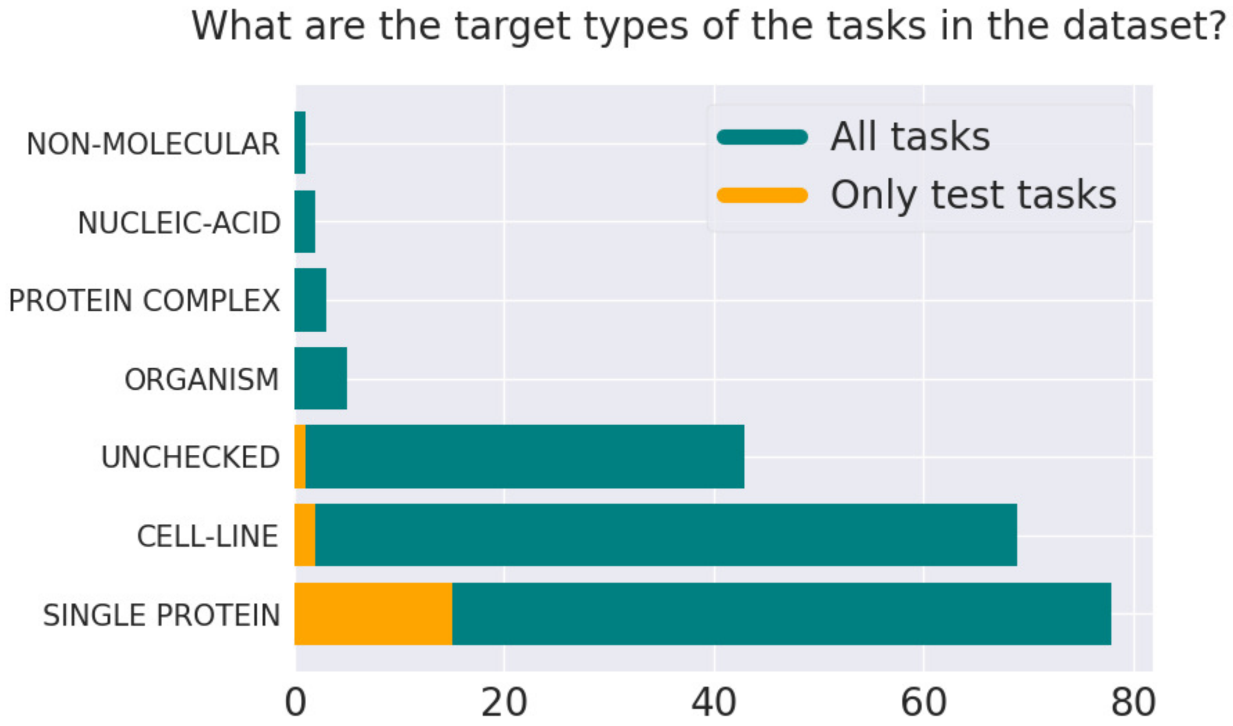


Figure 2: **What are the target types of the tasks in the dataset?** The majority are single proteins and cell lines, both in the entire set of tasks and the test tasks only.

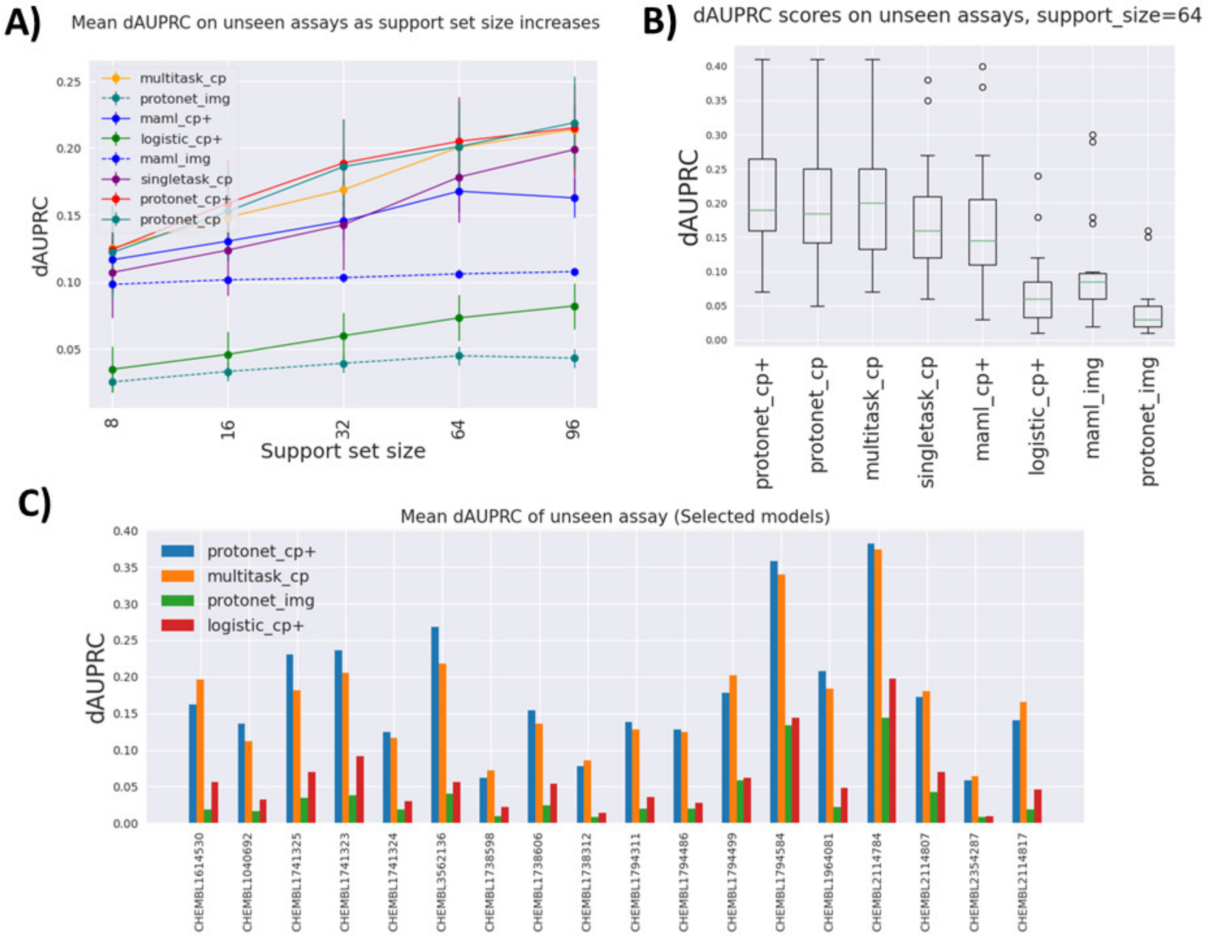


Figure 3: **Result reported using dAUPRC.** Figure A): Mean dAUPRC on test tasks as support set size increases. Figure B): Distribution of dAUPRC across all test tasks at support set size 64. Figure C): Mean dAUPRC of selected models for each task across all support set sizes.

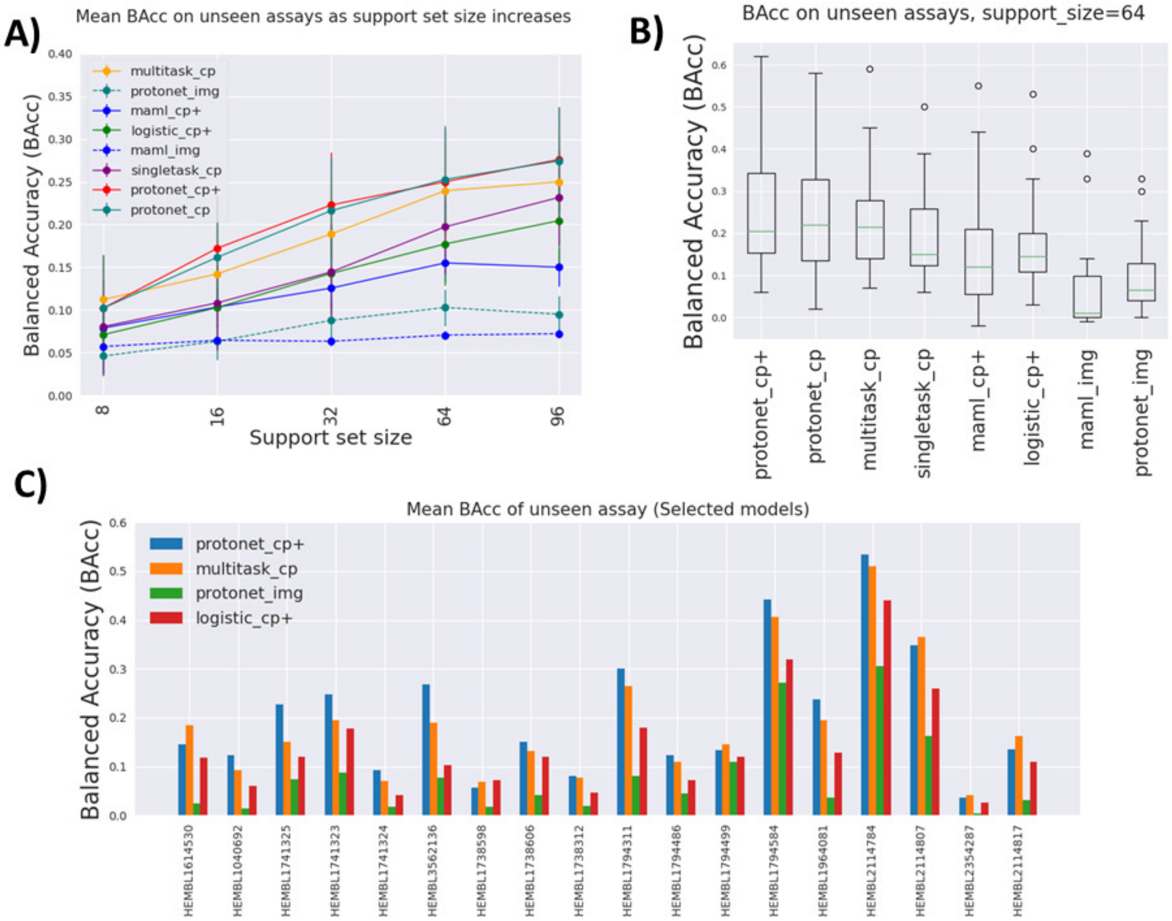


Figure 4: **Result reported using Balanced Accuracy (BAcc).** Figure A): Mean BAcc on test tasks as support set size increases. Figure B): Distribution of BAcc across all test tasks at support set size 64. Figure C): Mean BAcc of selected models for each task across all support set sizes.

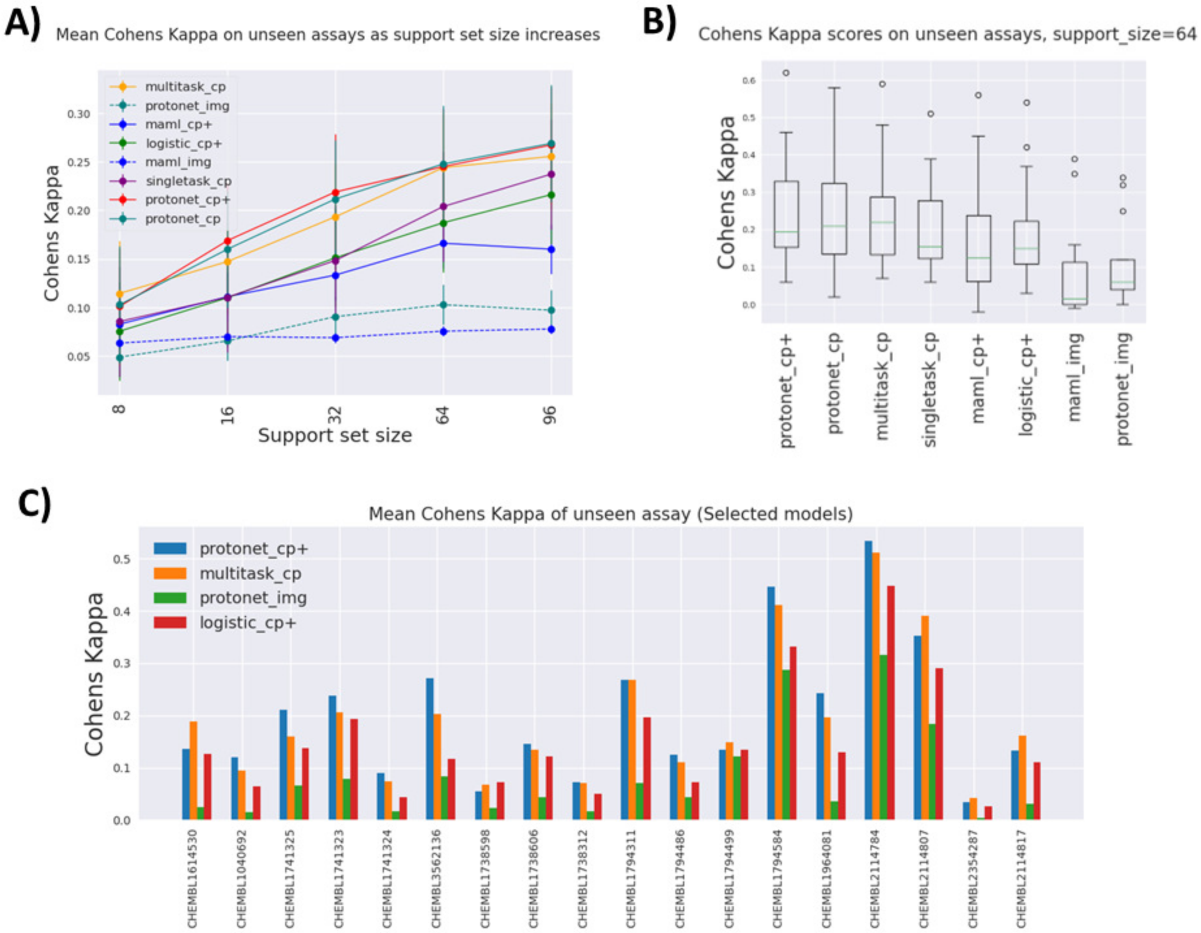


Figure 5: **Result reported using Cohens Kappa.** Figure A): Mean Cohens Kappa on test tasks as support set size increases. Figure B): Distribution of Cohens Kappa across all test tasks at support set size 64. Figure C): Mean Cohens Kappa of selected models for each task across all support set sizes.

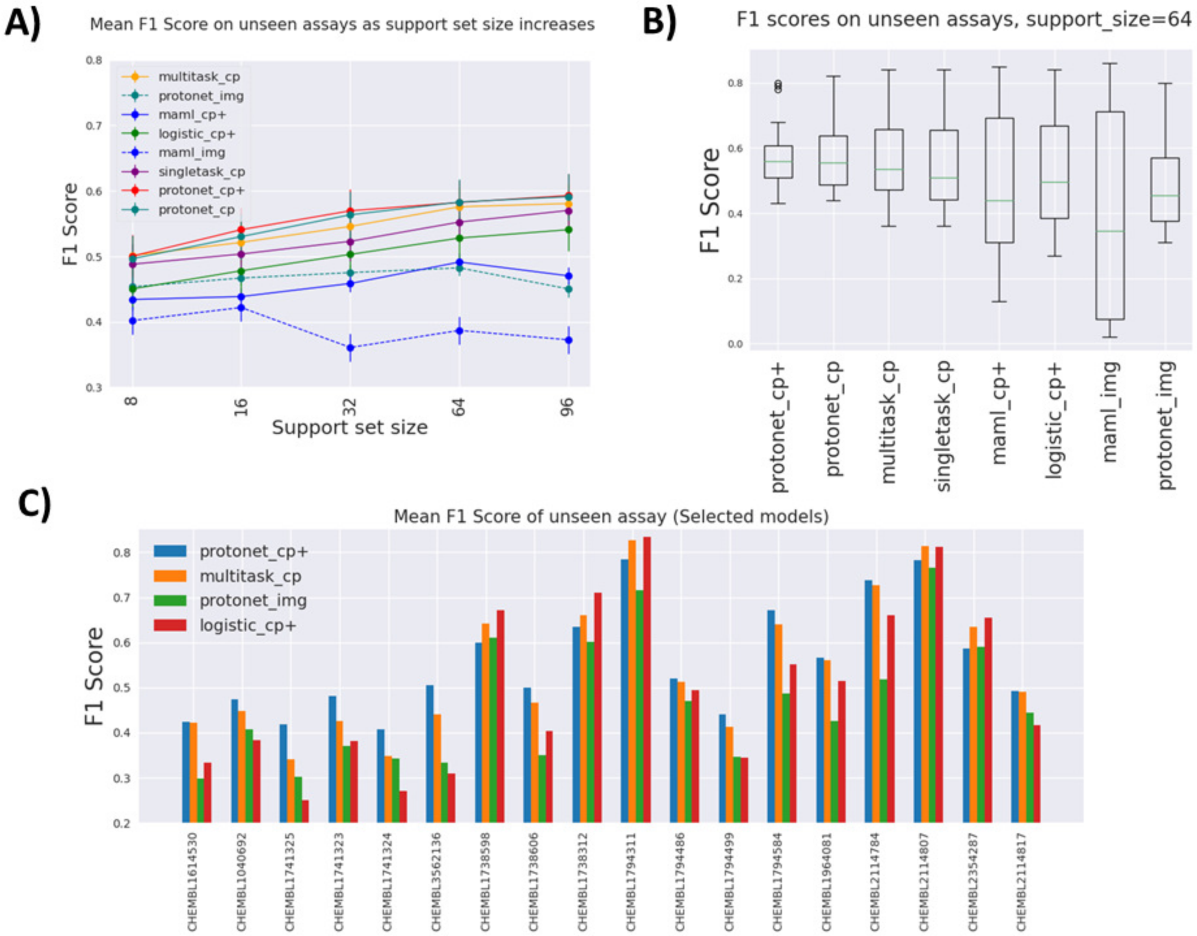
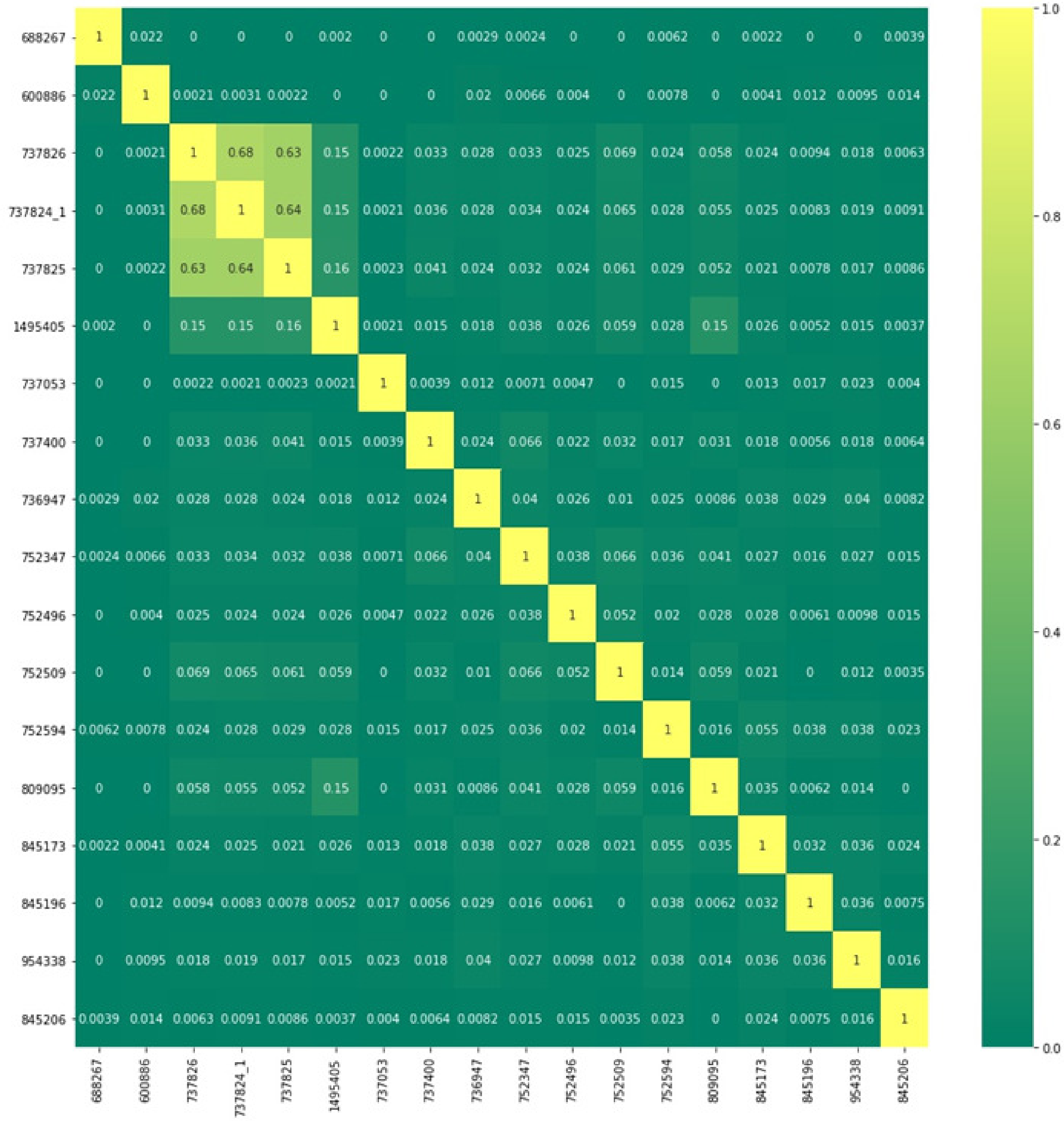


Figure 6: **Result reported using F1 Score.** Figure A): Mean F1 score on test tasks as support set size increases. Figure B): Distribution of F1 score across all test tasks at support set size 64. Figure C): Mean F1 score of selected models for each task across all support set sizes.



	688267	600886	737826	737824_1	737825	1495405	737053	737400	736947	752347	752496	752509	752594	809095	845173	845196	954338	845206
688267	1	0.022	0	0	0	0.002	0	0	0.0029	0.0024	0	0	0.0062	0	0.0022	0	0	0.0039
600886	0.022	1	0.0021	0.0031	0.0022	0	0	0	0.02	0.0066	0.004	0	0.0078	0	0.0041	0.012	0.0095	0.014
737826	0	0.0021	1	0.68	0.63	0.15	0.0022	0.033	0.028	0.033	0.025	0.069	0.024	0.058	0.024	0.0094	0.018	0.0063
737824_1	0	0.0031	0.68	1	0.64	0.15	0.0021	0.036	0.028	0.034	0.024	0.065	0.028	0.055	0.025	0.0083	0.019	0.0091
737825	0	0.0022	0.63	0.64	1	0.16	0.0023	0.041	0.024	0.032	0.024	0.061	0.029	0.052	0.021	0.0078	0.017	0.0086
1495405	0.002	0	0.15	0.15	0.16	1	0.0021	0.015	0.018	0.038	0.026	0.059	0.028	0.15	0.026	0.0052	0.015	0.0037
737053	0	0	0.0022	0.0021	0.0023	0.0021	1	0.0039	0.012	0.0071	0.0047	0	0.015	0	0.013	0.017	0.023	0.004
737400	0	0	0.033	0.036	0.041	0.015	0.0039	1	0.024	0.066	0.022	0.032	0.017	0.031	0.018	0.0056	0.018	0.0064
736947	0.0029	0.02	0.028	0.028	0.024	0.018	0.012	0.024	1	0.04	0.026	0.01	0.025	0.0086	0.038	0.029	0.04	0.0082
752347	0.0024	0.0066	0.033	0.034	0.032	0.038	0.0071	0.066	0.04	1	0.038	0.066	0.036	0.041	0.027	0.016	0.027	0.015
752496	0	0.004	0.025	0.024	0.024	0.026	0.0047	0.022	0.026	0.038	1	0.052	0.02	0.028	0.028	0.0061	0.0098	0.015
752509	0	0	0.069	0.065	0.061	0.059	0	0.032	0.01	0.066	0.052	1	0.014	0.059	0.021	0	0.012	0.0035
752594	0.0062	0.0078	0.024	0.028	0.029	0.028	0.015	0.017	0.025	0.036	0.02	0.014	1	0.016	0.055	0.038	0.038	0.023
809095	0	0	0.058	0.055	0.052	0.15	0	0.031	0.0086	0.041	0.028	0.059	0.016	1	0.035	0.0062	0.014	0
845173	0.0022	0.0041	0.024	0.025	0.021	0.026	0.013	0.018	0.038	0.027	0.028	0.021	0.055	0.035	1	0.032	0.036	0.024
845196	0	0.012	0.0094	0.0083	0.0078	0.0052	0.017	0.0056	0.029	0.016	0.0061	0	0.038	0.0062	0.032	1	0.036	0.0075
954338	0	0.0095	0.018	0.019	0.017	0.015	0.023	0.018	0.04	0.027	0.0098	0.012	0.038	0.014	0.036	0.036	1	0.016
845206	0.0039	0.014	0.0063	0.0091	0.0086	0.0037	0.004	0.0064	0.0082	0.015	0.015	0.0035	0.023	0	0.024	0.0075	0.016	1

Figure 7: **Figure 1: Heatmap of Jaccard Index between the unique InChiKeys of 18 tasks in D_test.** The majority of tasks share very few common InChiKeys. Outliers are tasks 737826, 737824_1 and 737825 whose targets resemble Cytochrome P450. But we believe they are still different enough to be separate tasks in the test set.