

retrain procedure:

- ① write\_csv\_train\_new.sh → generate training and test csvs (11 paths for all proteins dataset, no prediction)  
model a:
- ② bash\_training\_greedy.sh
- ③ bash\_predict\_eval\_greedy.sh → save protein a w/ params in results-greedy for all protein dataset
- ④ write\_csv\_evaluation\_greedy.sh → generate evaluation csvs (12 paths, no signal) for different params
- ⑤ bash\_evaluate\_greedy.sh → evaluate training set for all protein dataset for all p
- ⑥ bash\_predict\_optimal.sh → find the optimal p with best  $\text{Exc}/\text{MSE}$  and update the predictions of model a with best p model in results/3d directory for both train and test
- ⑦ bash\_evaluate\_greedy.sh → evaluate test set for all proteins dataset with optimal p

In the greedy training loop for each organelle:

- ① bash\_training\_greedy
- ② bash\_predict\_eval\_greedy.
- ④ write\_csv\_evaluation\_greedy.
- ⑤ bash\_evaluate\_greedy (for all p for training set)
- ⑥ bash\_predict\_optimal. replace all the layer of model a in all datasets for all t
- ⑦ bash\_evaluate\_greedy. (for best p for test set)