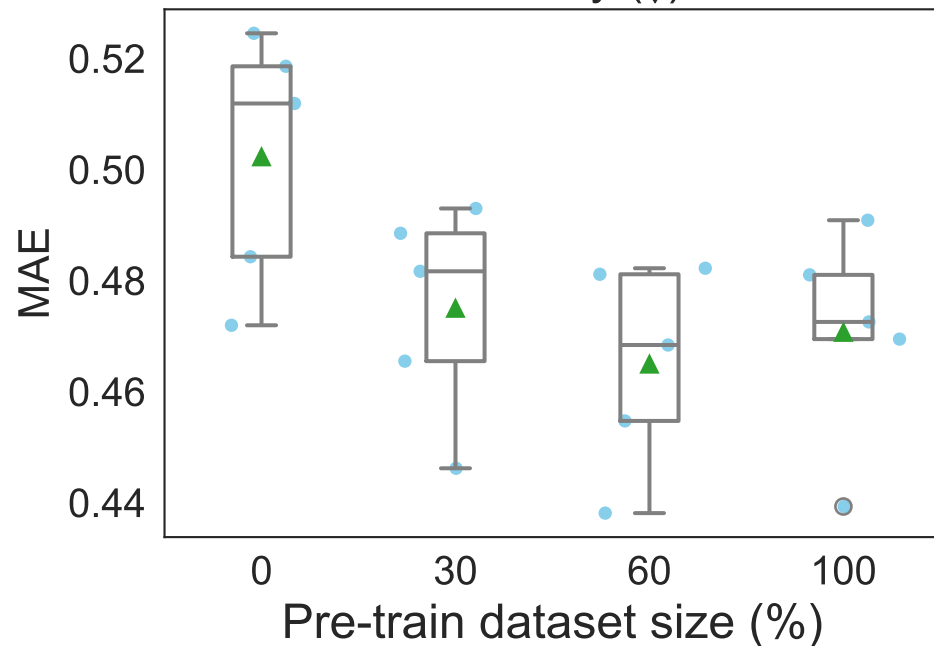
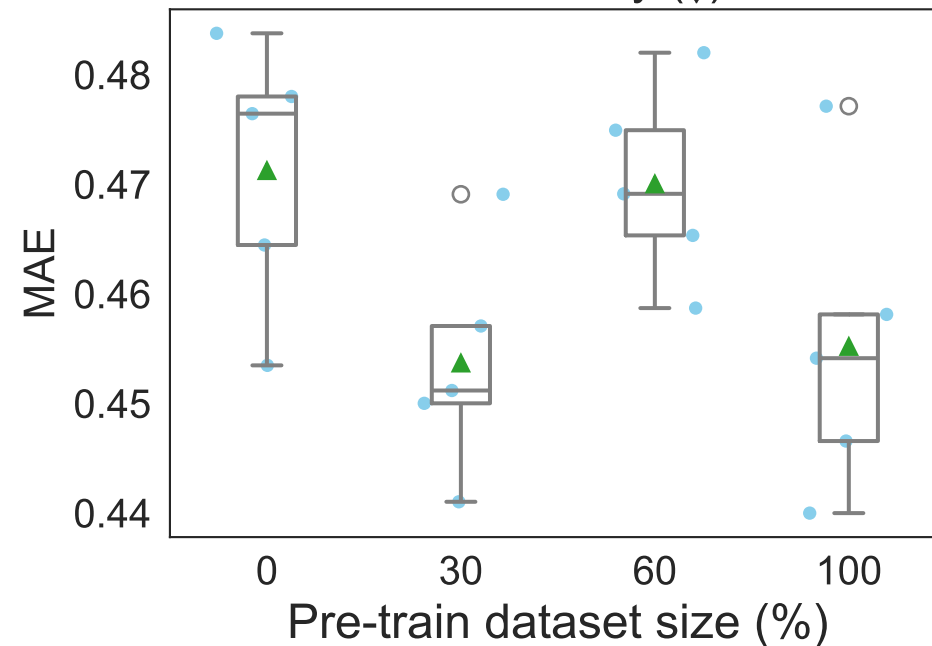


Comparing pre-training dataset sizes / pre-training objective = MLM / split = random

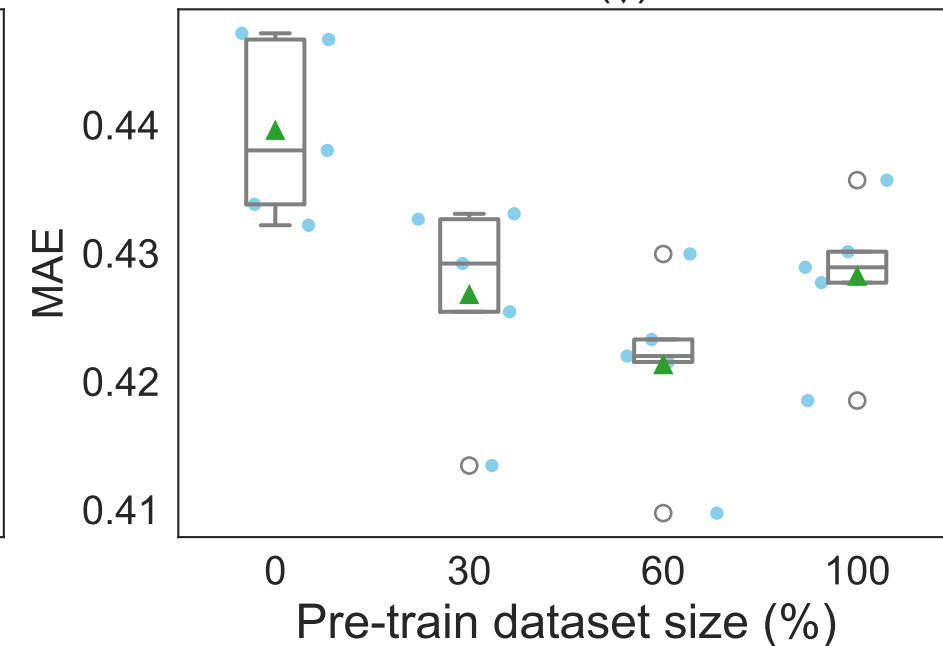
Solubility ( $\downarrow$ )



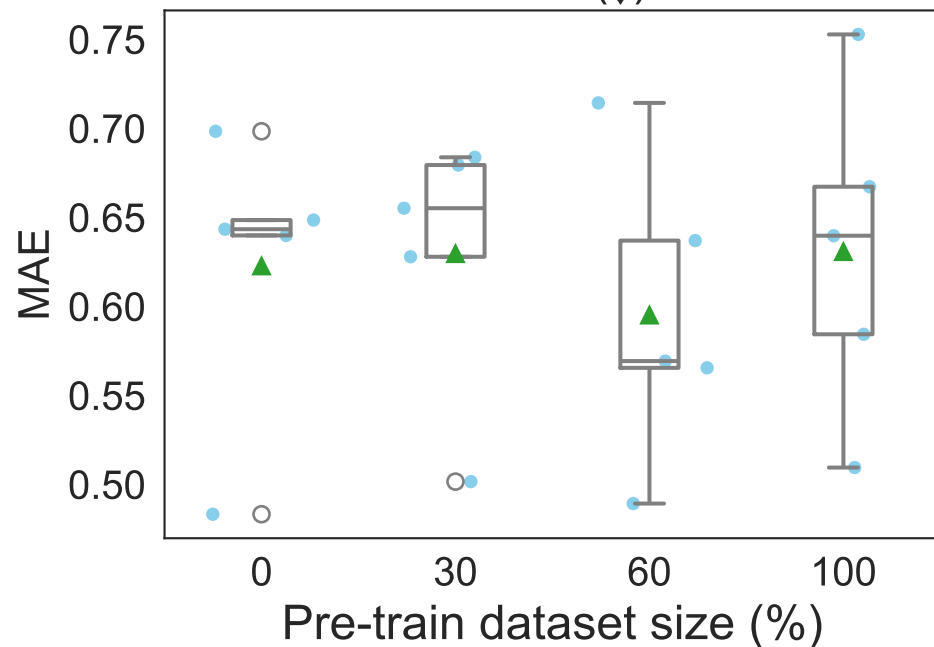
Permeability ( $\downarrow$ )



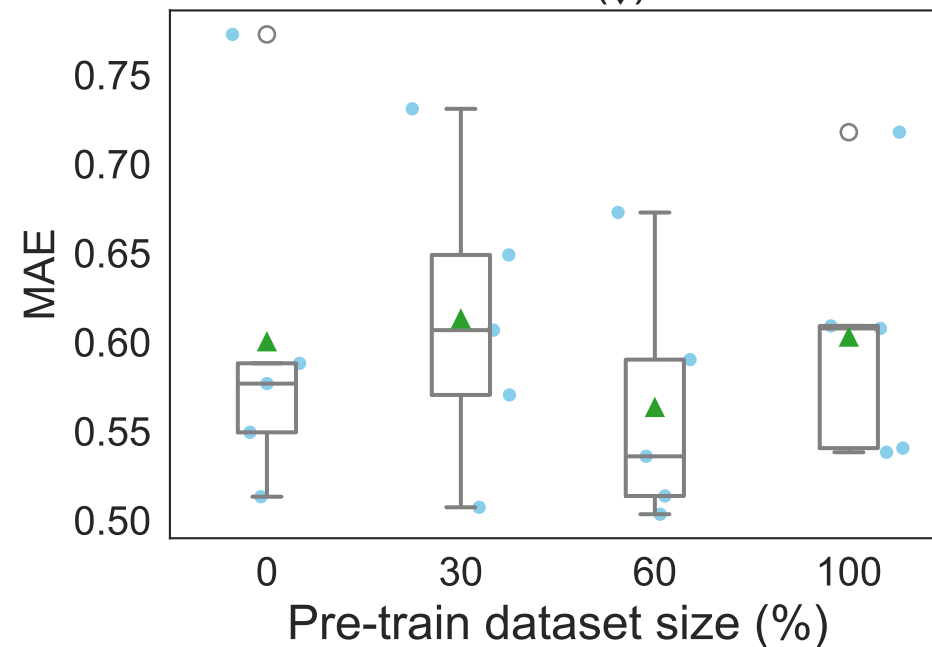
HLM ( $\downarrow$ )



hPPB ( $\downarrow$ )



rPPB ( $\downarrow$ )



RLM ( $\downarrow$ )

