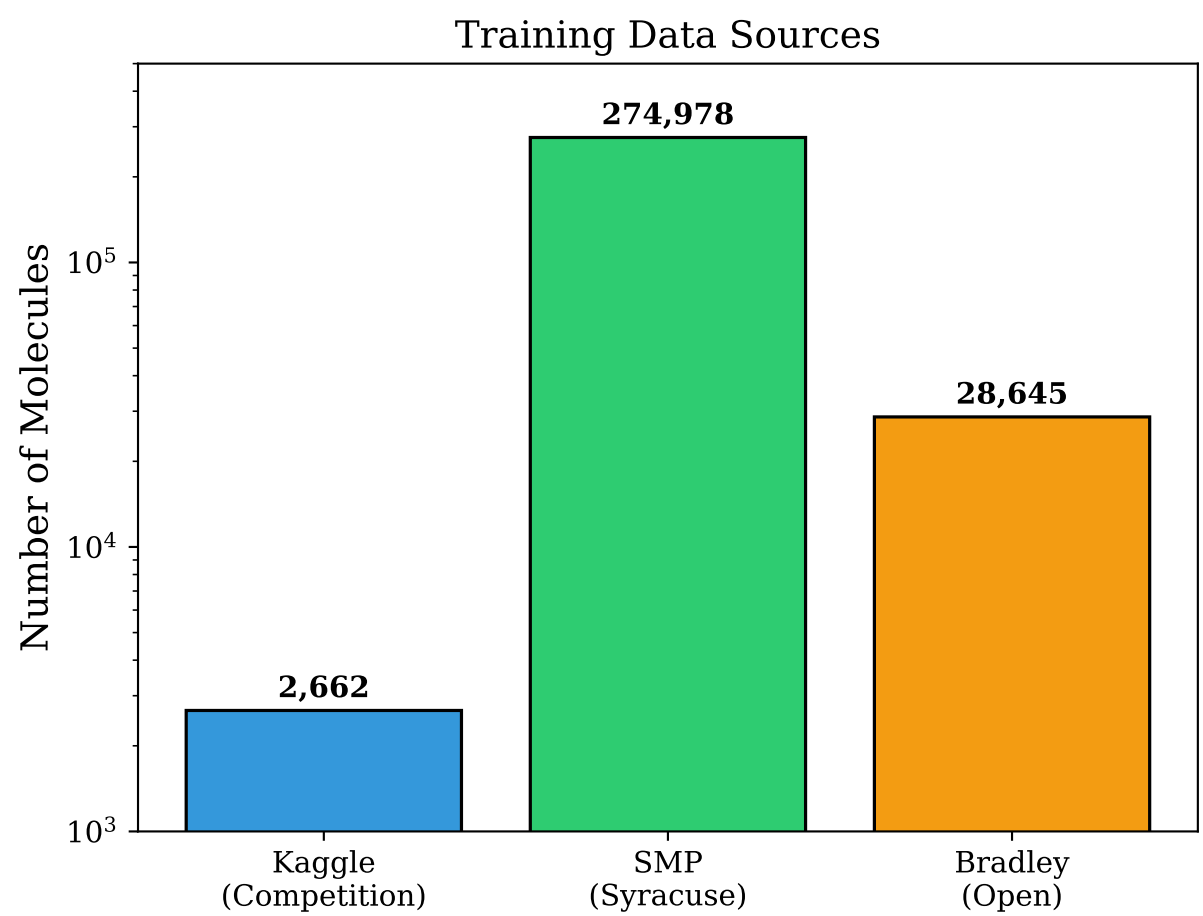
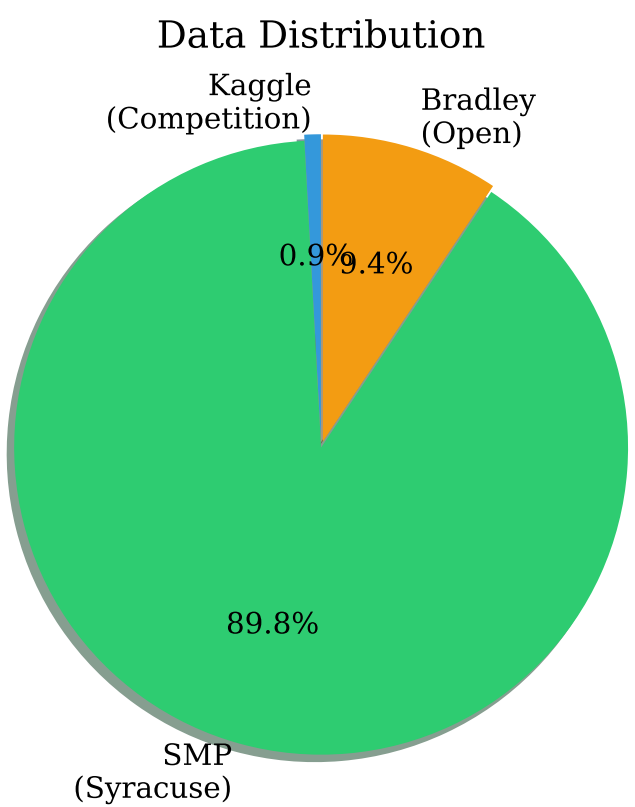

Training Data Sources
Number of Molecules
10^5
10^4
10^3
274,978
28,645
2,662
Kaggle
(Competition)
SMP
(Syracuse)
Bradley
(Open)


Data Distribution
Kaggle
(Competition)
Bradley
(Open)
0.9%
9.4%
89.8%
SMP
(Syracuse)