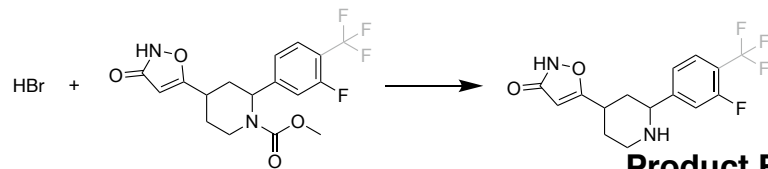


Near identical reactions from USPTO 50:50 random train / test split

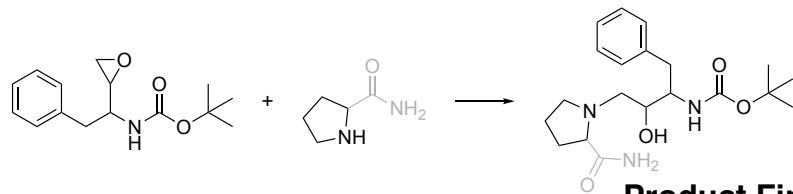
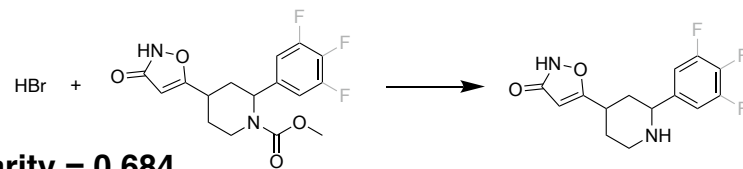
Training Set



Product Fingerprint similarity = 0.684

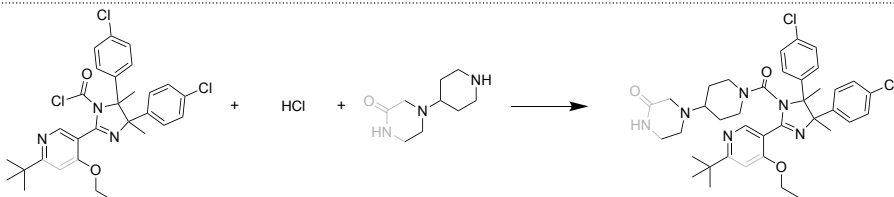
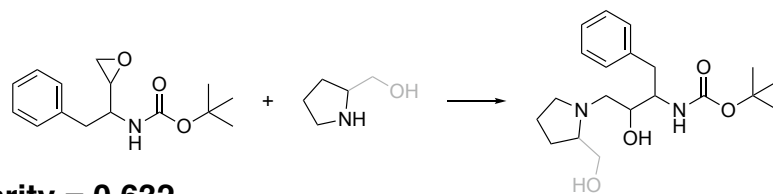
Reaction Fingerprint similarity = 0.699

Test Set



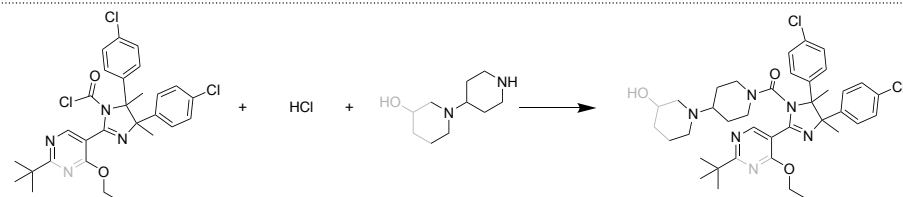
Product Fingerprint similarity = 0.632

Reaction Fingerprint similarity = 0.582



Product Fingerprint similarity = 0.538

Reaction Fingerprint similarity = 0.739



⋮

**~78% of test set reactions have Tanimoto
similar 'sibling' in training set**

⋮