

How are virtual screening programs assessed?

- Binding poses
 - Similarity of predicted poses to solved structures
 - Root mean square deviation (RMSD) $< 2 \text{ \AA}$ considered good

- Docking scores
 - Correlation between scores and measured affinities
 - Active molecules should have lower scores than
 - inactive compounds
 - decoys, which have similar chemical properties but different connectivity than active molecules
 - Quantified by receiver operating characteristic (ROC) curve, area under the ROC curve (AUC), enrichment factor

- Reported values can be biased
 - Computational chemists, especially methods developers, tempted to tweak approach (e.g. parameters) until experiments are reproduced.
 - Blinded challenges (e.g. D3R Grand Challenge) reduce bias



How are virtual screening program assessed?

- Binding poses
 - Similarity of predicted poses to solved structures
 - Root mean square deviation (RMSD) < 2 Å considered good
- Docking scores
 - Correlation between scores and measured affinities
 - Active molecules should have lower scores than
 - inactive compounds
 - decoys, which have similar chemical properties but different connectivity than active molecules
 - Quantified by receiver operating characteristic (ROC) curve, area under the ROC curve (AUC), enrichment factor
- Reported values can be biased
 - Computational chemists, especially methods developers, tempted to tweak approach (e.g. parameters) until experiments are reproduced.
 - Blinded challenges (e.g. D3R Grand Challenge) reduce bias

The Receiver Operating Characteristic (ROC) Curve

score	fraction of actives	fraction of inactives
-53.4	0.2	0
-50.2	0.2	0.2
-49.2	0.4	0.2
-45.7	0.6	0.2
-42.1	0.6	0.4
-35.2	0.6	0.6
-30.0	0.6	0.8
-21.3	0.8	0.8
-20.7	1.0	0.8
-4.2	1.0	1.0