Electronic Supplementary Material (ESI) for Digital Discovery.
This journal is © The Royal Society of Chemistry 2023

# Supporting Information for Machine Learning-Augmented Docking 1. CYP inhibition prediction

Benjamin Weiser, Jerome Genzling, Mihai Burai Patrascu, Ophélie Rostaing and Nicolas Moitessier*

## Code Workflow Procedure

1. Cleaned and combined test and train Pei sets with CYP_clean_files.ipnyb. Sets combined and then clustered to create new train and test sets
2. Dock each ligand 5 times to its respective isoform using FITTED. Docked data can be found here: (to be inserted)
3. Create analogue sets using FITTED. Create max train-test similarity using CYP_TC_DataSets.py
4. Run RF with Feature Importances using max train-test similarity of 0.8 using ML_over_Tanimoto.py which calls CYP_inhibition_functions.py and Do_ML2.py
5. Using these selected features run all ML models on all datasets using ML_over_Tanimoto.py which calls CYP_inhibition_functions.py and Do_ML2.py
6. Use CYP_evaluate_and_ensemble.py which calls CYP_evaluate_and_ensemble_functions.py to make ensembles and evaluate and graph model performance

## 1 List of input Ligand Features

- Energy
- van der Waals
- Electrostatics
- Hydrogen Bonds
- Hydrogen Bond-like Metal interaction
- Metal Coordination
- Water van der Waals
- Water Electrostatics
- Water Hydrogen Bonds
- Number of Water Molecules
- Number of Rotatable Bonds
- Rotatable Bonds Entropy
- Non-polar Solvation Energy
- Polar Ligand Solvation Energy
- Polar Protein Solvation Energy
- RMSD
- RMSD (MBG)
- Strain Energy
- RankScore
- MatchScore
- Molecular Weight
- Net Charge
- Hydrogen Bond Donors
- Hydrogen Bond Acceptors
- Rotatable Bonds
- Rings
- Ionizable Centers
- Oxygens
- Nitrogens
- Sulfurs
- Heteroatoms
- Heavy Atoms
- Michael Acceptors
- tPSA
- logP
- Fsp3
- Stereogenic Centers Ratio
- logS
- Stereogenic Centers
- Aromatic Rings
- Aromatic Proportion
- Halogens
- McGowan Molecular Volume
- Molecular van der Waals Surface Area
- Polar SASA
- non Polar SASA
- Molecular Density
- Molecular Polarity
- Blood-Brain Barrier
- Molecular Softness
- Molecular Electronegativity
- Negatively Charged Oxygens
- Positively Charged Nitrogen
- Molecular Hardness
- Molecular Polarizability
- 3D-Wiener Index
- Geometric Radius
- Geometric Diameter
- Geometrical Shape Coefficient
- Span
- Radius of Gyration
- Ovality Index
- Globularity Factor

- Dipole moment
- Badapple pscore
- Quantitative Estimate of Druglikeness
- aromatic
- carboxylic acid
- carboxylate
- ester aromatic
- ester aliphatic
- ester conjugated
- ester
- lactone
- aldehyde aromatic
- aldehyde aliphatic
- aldehyde conjugated
- aldehyde
- acyl chloride
- acyl bromide
- amide ter
- amide
- lactame
- anhydride
- ketone
- sulfonamide
- thiol
- thiolate
- carbamate
- primary amine
- secondary amine
- tertiary amine
- quat ammonium
- amine
- aniline
- nitrile
- imine
- oxime
- nitro
- nitroso
- azide
- isocyanate
- hydroxamic acid
- hydroxamate
- alkyl chloride
- alkyl bromide
- alkyl iodide
- aryl chloride
- aryl bromide
- aryl iodide
- boronic acid aromatic
- boronic ester aromatic
- boronic acid aliphatic
- boronic ester aliphatic
- boronic acid vinylic
- boronic ester vinylic
- boronic acid
- sulfonyl chloride
- sulfonyl bromide
- vinyl chloride
- vinyl bromide
- vinyl iodide
- alkene
- ketone O
- formyl
- alcohol
- chloride
- bromide
- fluoride
- iodide
- silicon
- primary amine neutral
- secondary amine neutral
- tertiary amine neutral
- alpha fluoroketone
- alpha chloroketone
- epoxide
- aziridine
- beta-lactam
- primary alcohol
- disulfide
- azo
- isothiocyanate
- hydrazine
- ketene
- peroxide
- silyl fluoride
- silyl bromide
- silyl chloride
- silyl iodide
- thiocarbonyl
- thioester
- betaKeto

# Figures and Tables

Table S1: Dataset sizes for 5 CYP P450 isoforms

| | 1A2 | 2C19 | 2C9 | 2D6 | 3A4 |
|---|---|---|---|---|---|
| Before Cleaning Data Size | 10072 | 10769 | 10050 | 10851 | 11436 |
| After Cleaning Data Size | 9773 | 10636 | 9935 | 10717 | 11401 |
| # Actives | 3598 | 4564 | 2603 | 1406 | 3602 |
| # Inactives | 6175 | 6072 | 7332 | 9311 | 7799 |

# Protein structures (PDB IDs)

CYP1A2: 2HI4, CYP2C9: 1R9O, CYP2C19: 4GQS, CYP2D6: 3QM4, CYP3A4: 3NXU

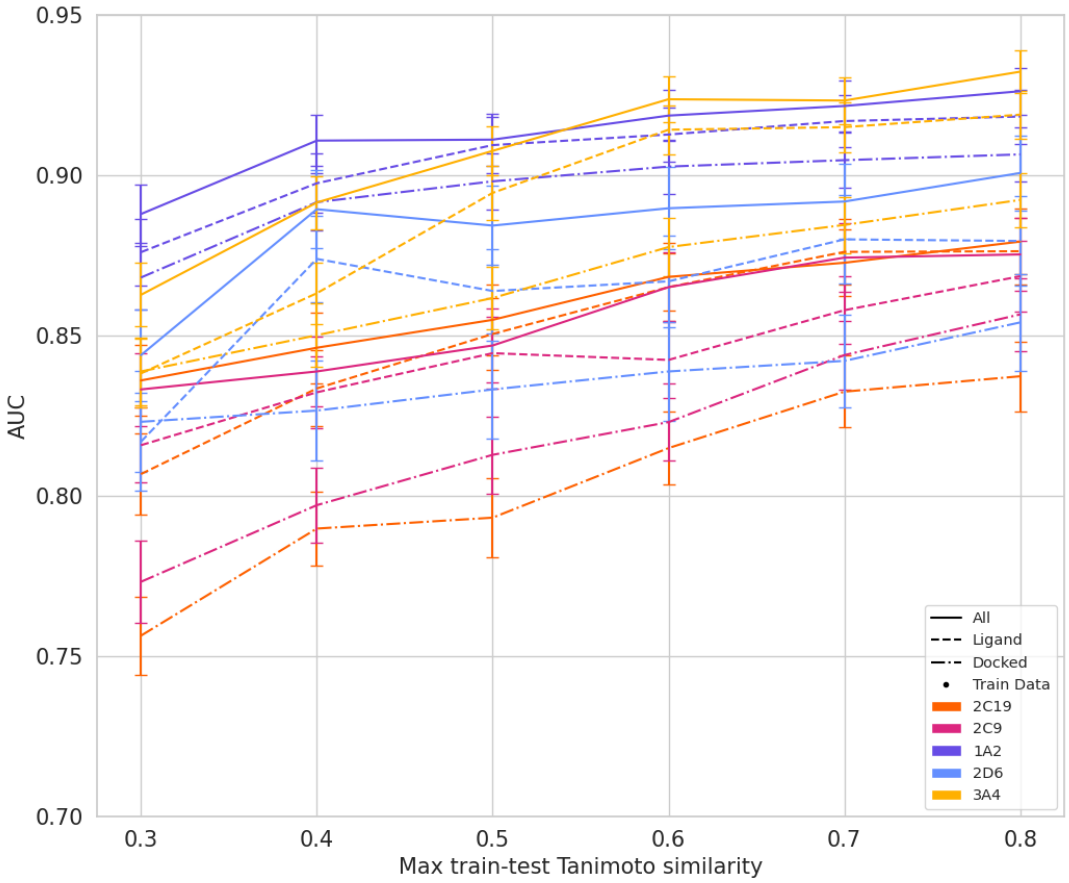


Figure S1: AUC of ML augmented docking using ligand only features, docking features, and all features over max train-test Tanimoto similarity

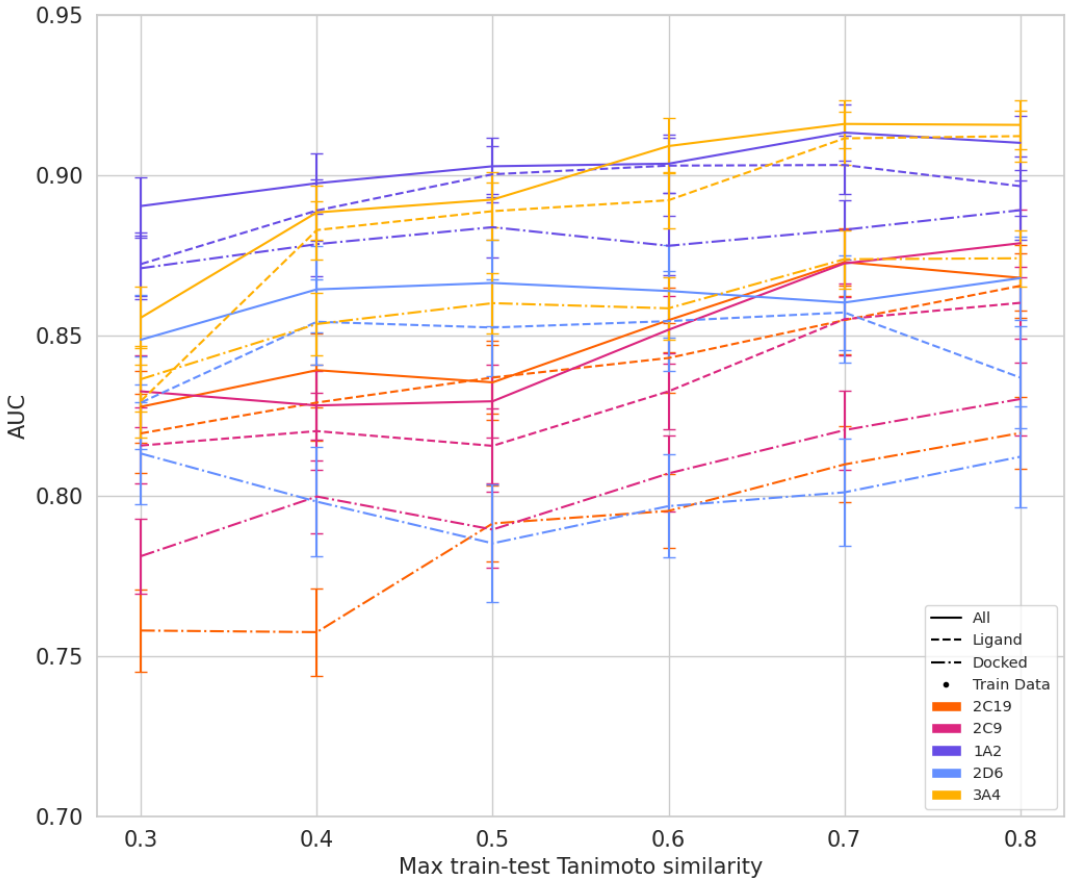


Figure S2: AUC of ML augmented docking using ligand-only features, docking features, and all features over max train-test Tanimoto similarity using same data sizes. Data size capped at data size of 0.3. Data size used; 1A2: 2406, 2C9: 2622, 2C19: 2698, 2D6: 2789, 3A4: 2649. Using all features become more predictive than ligand featurization at low Tanimoto similarity

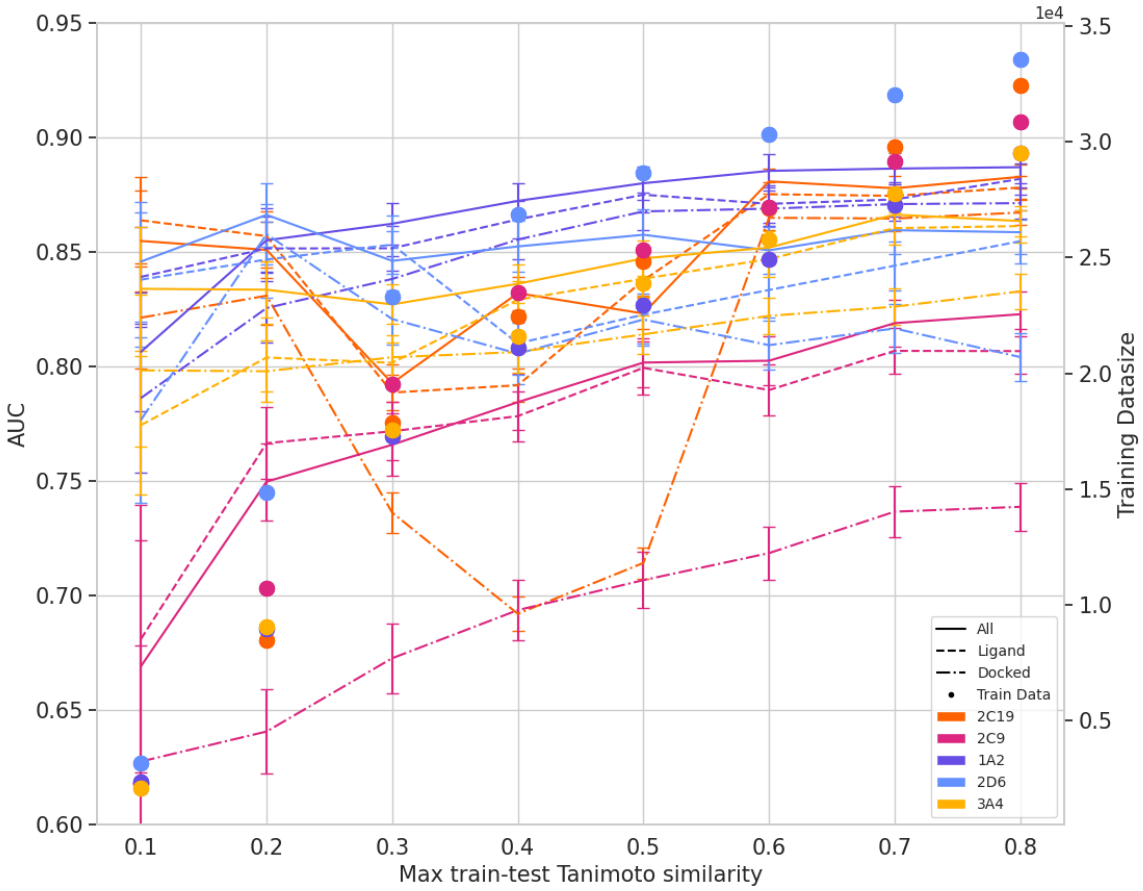


Figure S3: AUC of ML augmented docking using ligand only features, docking features, and all features over max train-test Tanimoto similarity using analog datasets

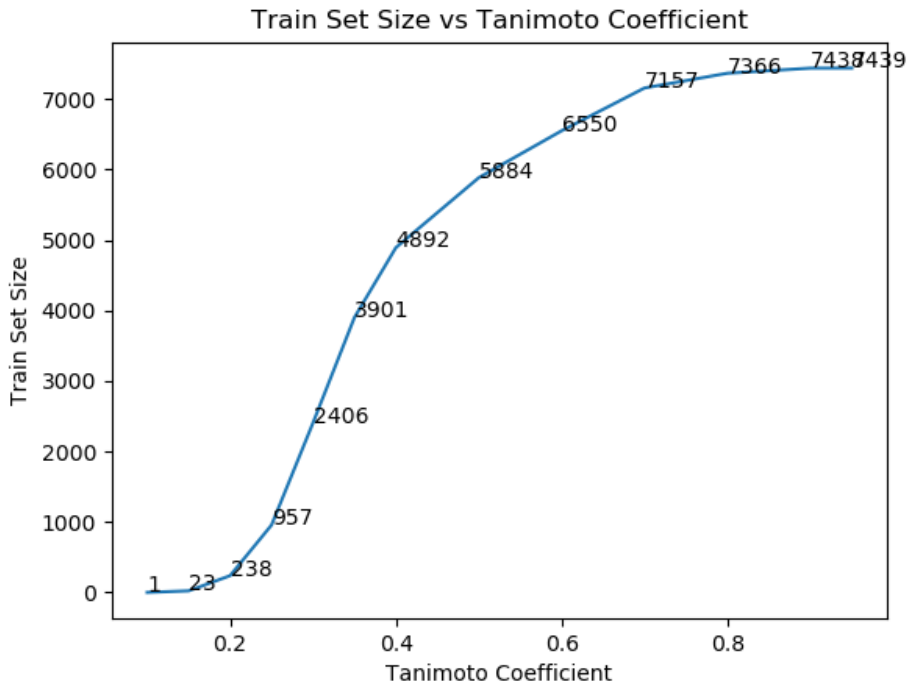


Figure S4: 1A2 CYP inhibition data set size after filtering by max similarity between test and train set using a test set of 100 clusters of 5 molecules.

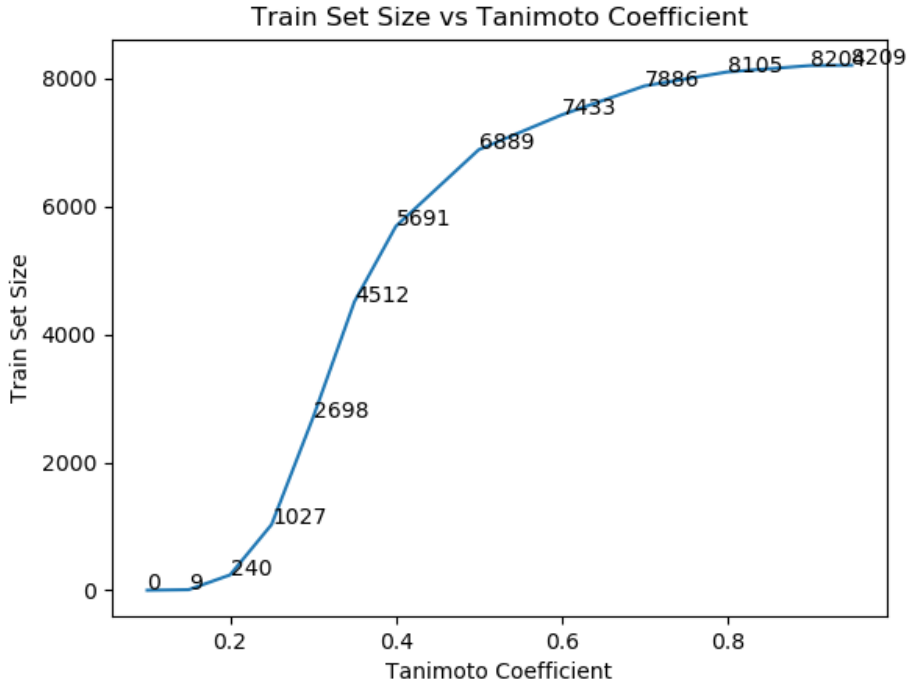


Figure S5: 2C19 CYP inhibition data set size after filtering by max similarity between test and train set using a test set of 100 clusters of 5 molecules.

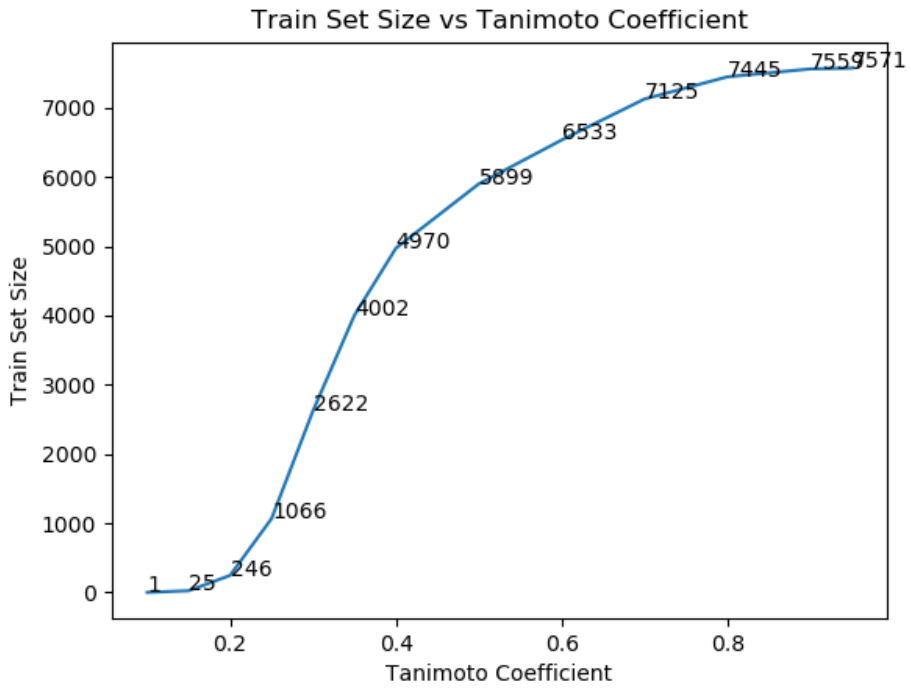


Figure S6: 2C9 CYP inhibition data set size after filtering by max similarity between test and train set using a test set of 100 clusters of 5 molecules.

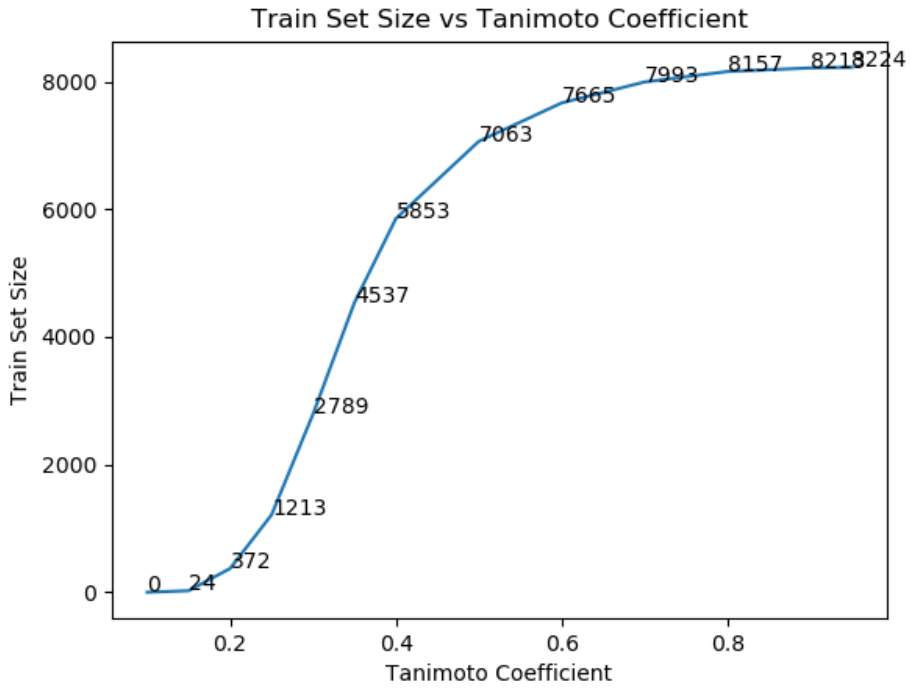


Figure S7: 2D6 CYP inhibition data set size after filtering by max similarity between test and train set using a test set of 100 clusters of 5 molecules.

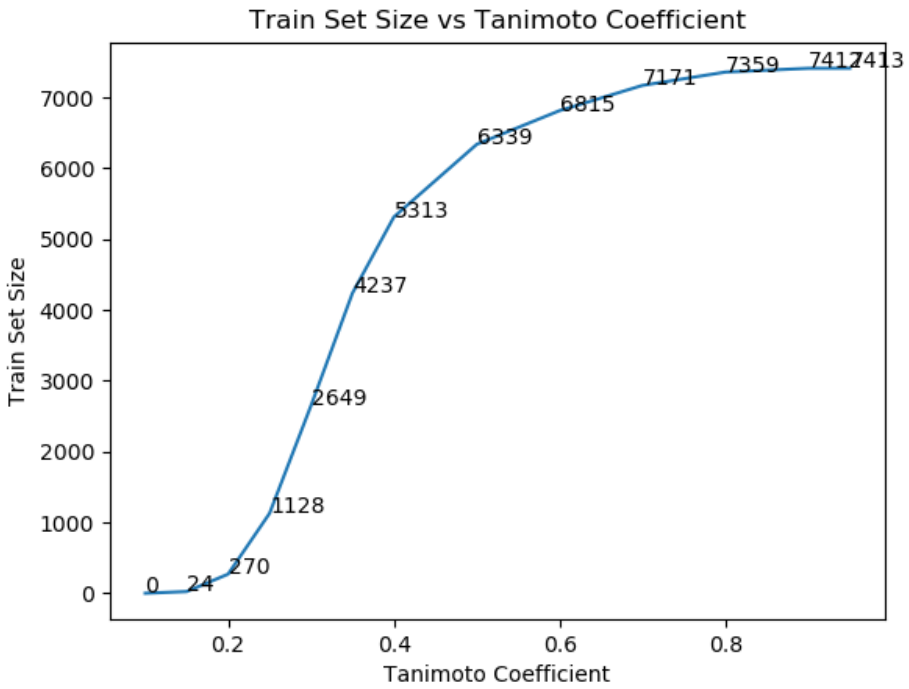


Figure S8: 3A4 CYP inhibition data set size after filtering by max similarity between test and train set using a test set of 100 clusters of 5 molecules.

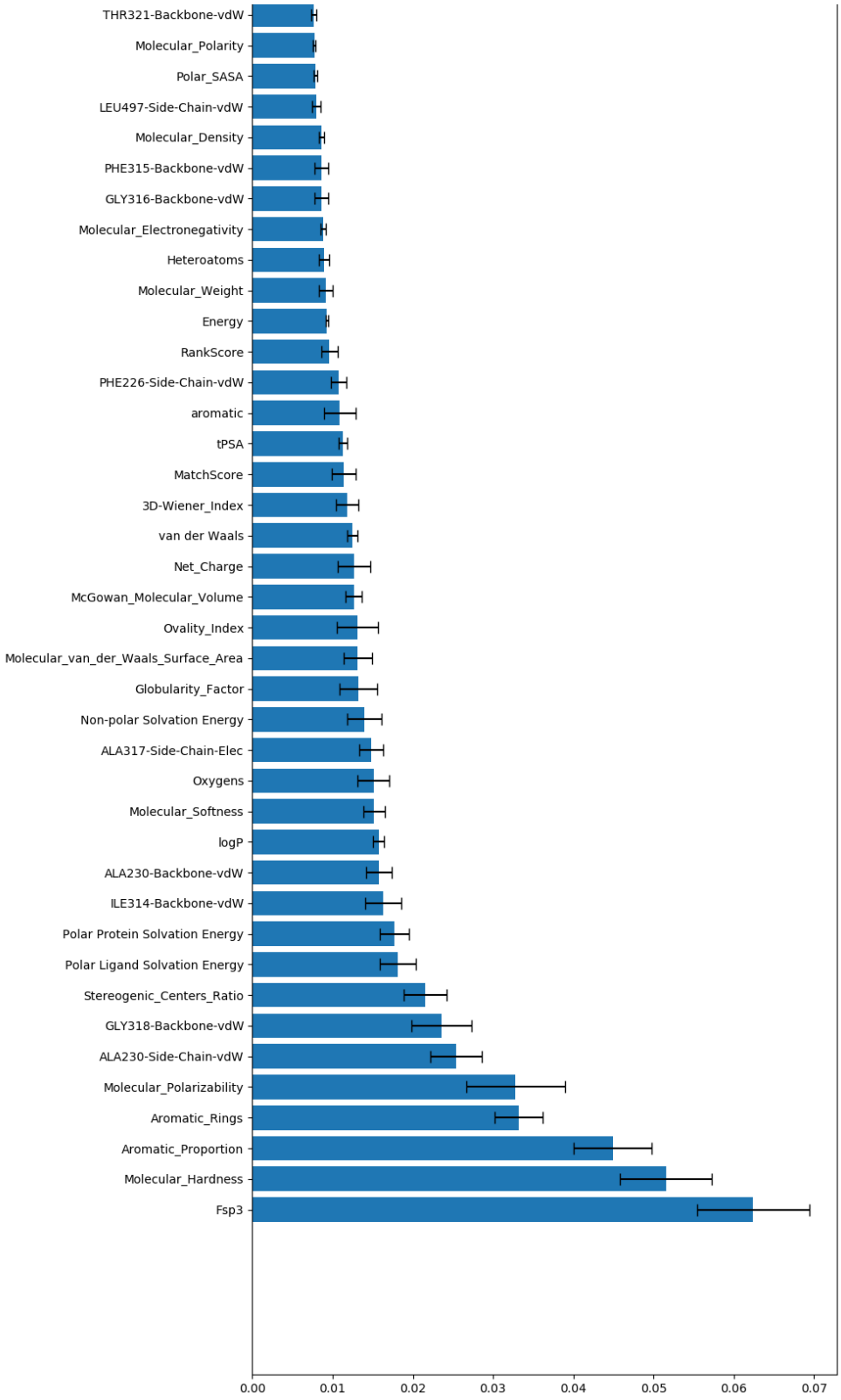


Figure S9: Feature importance of random forest model using selected top 100 features for isoform 1A2.

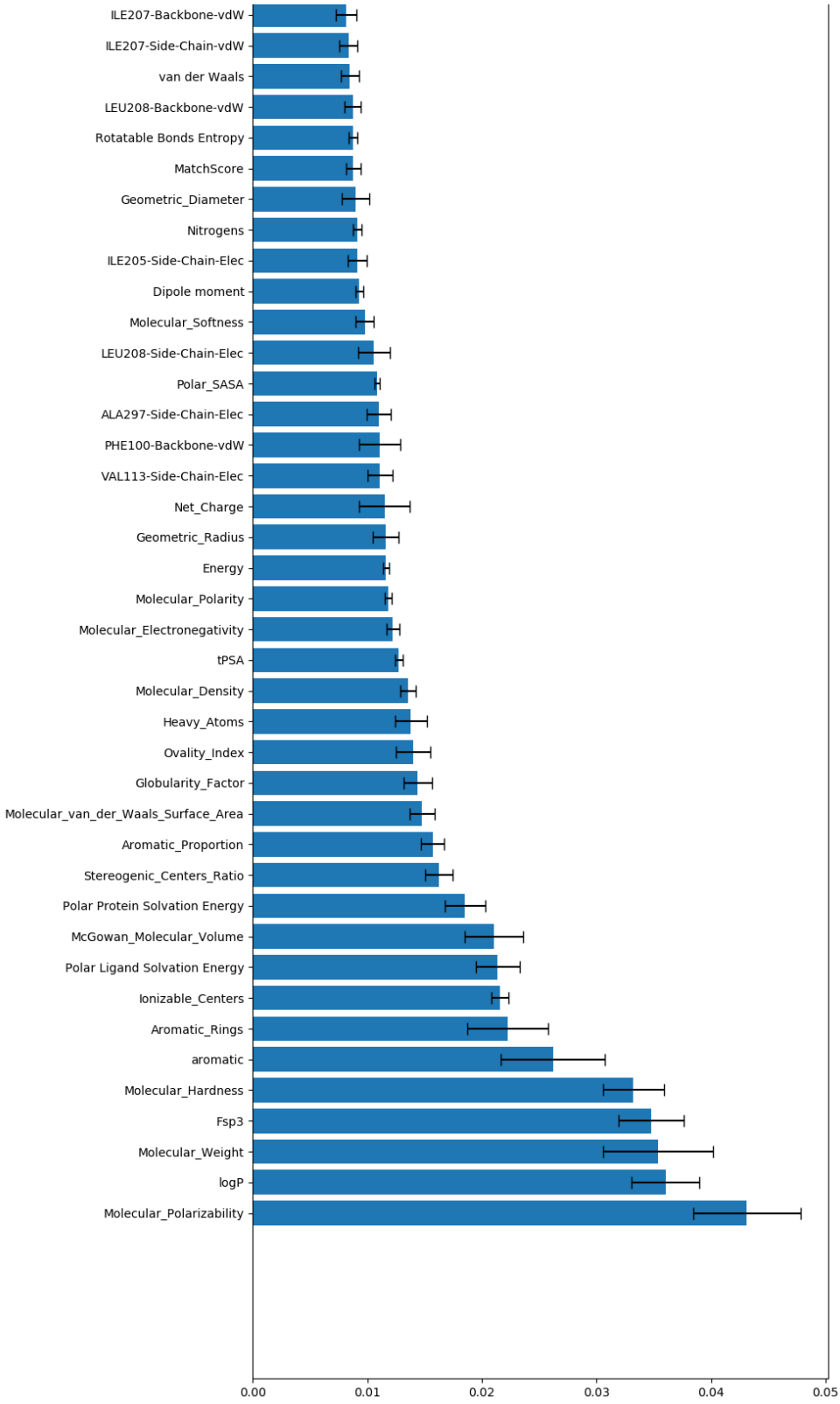


Figure S10: Feature importance of random forest model using selected top 100 features for isoform 2C9.

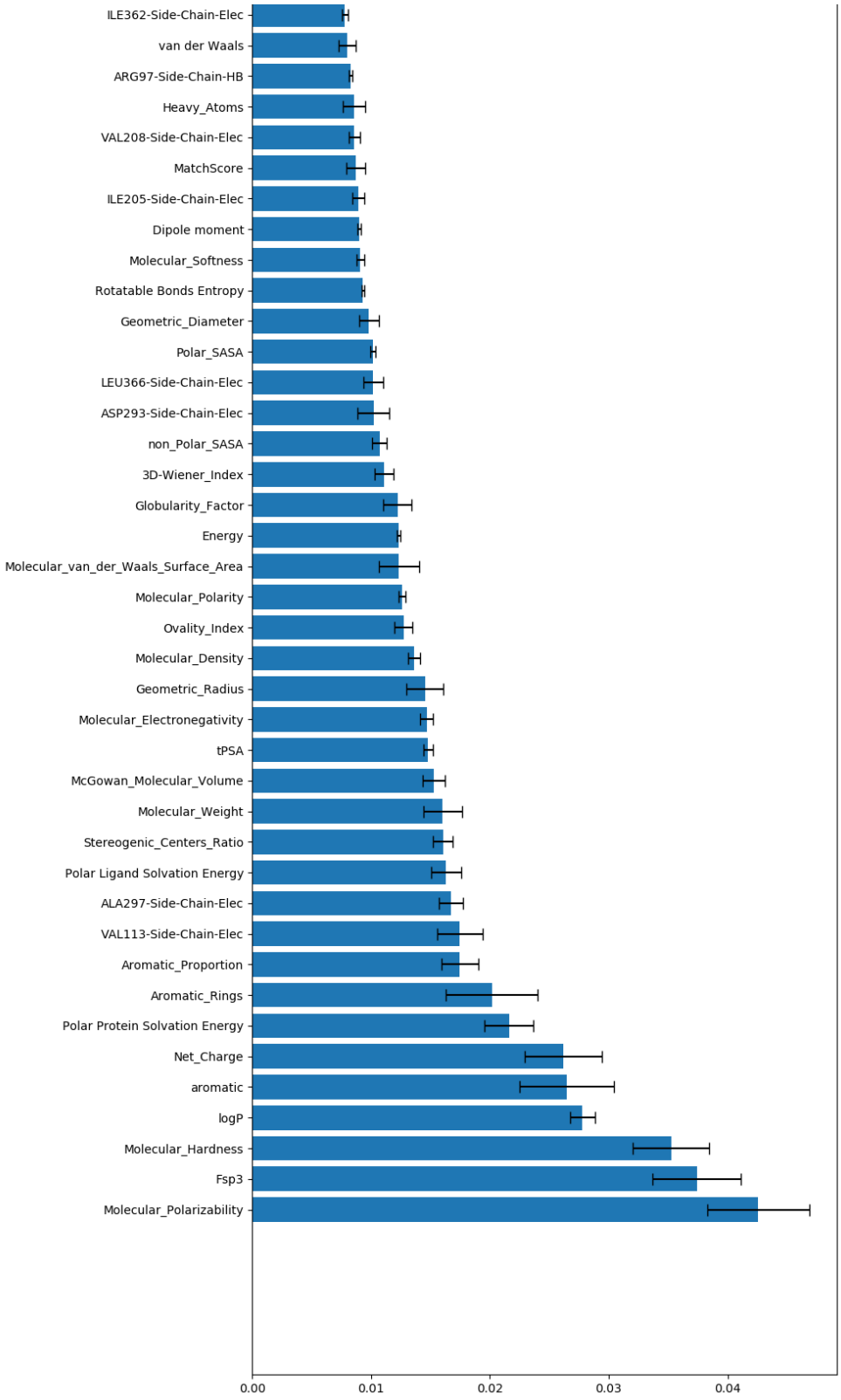


Figure S11: Feature importance of random forest model using selected top 100 features for isoform 2C19.

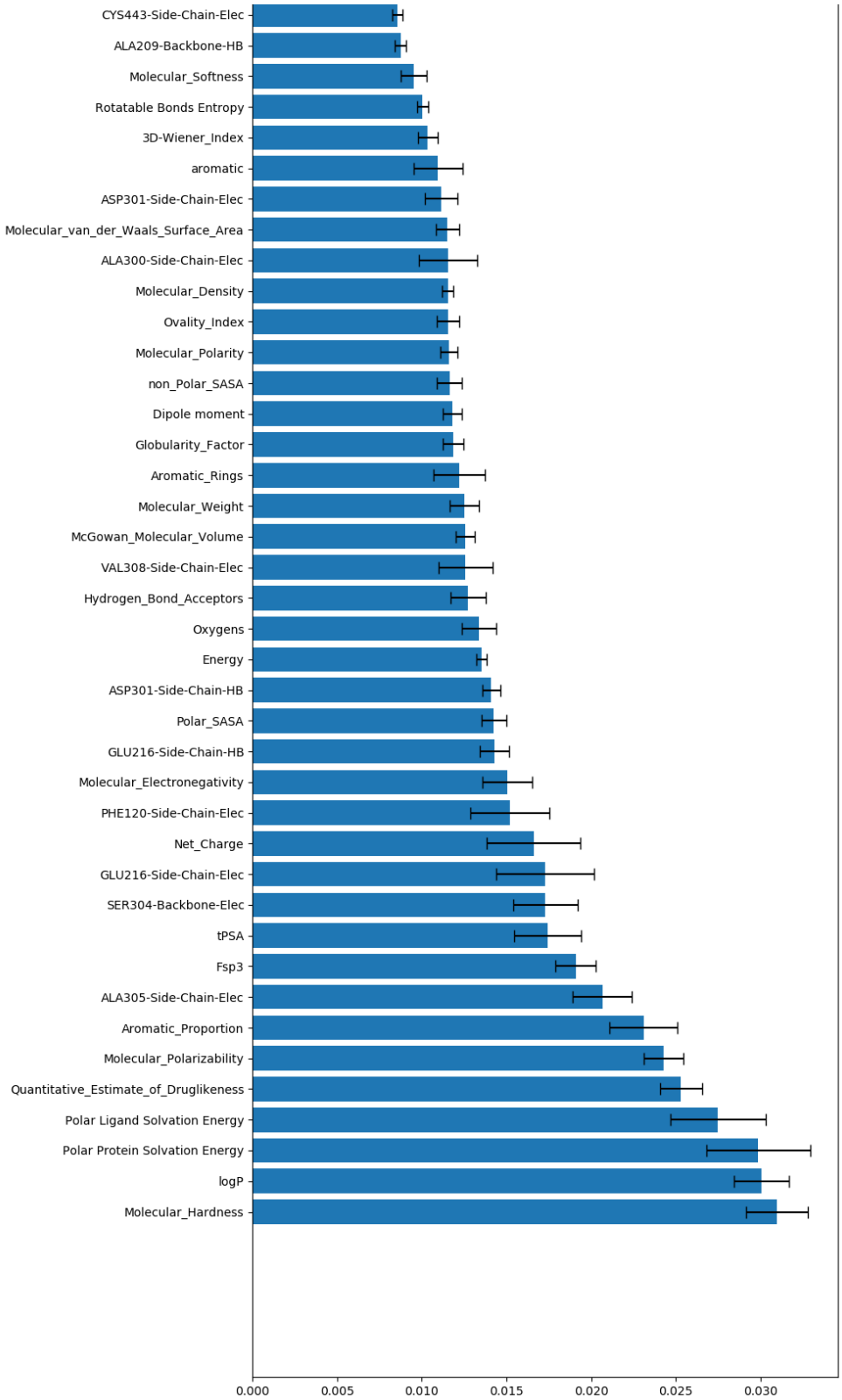


Figure S12: Feature importance of random forest model using selected top 100 features for isoform 2D6.

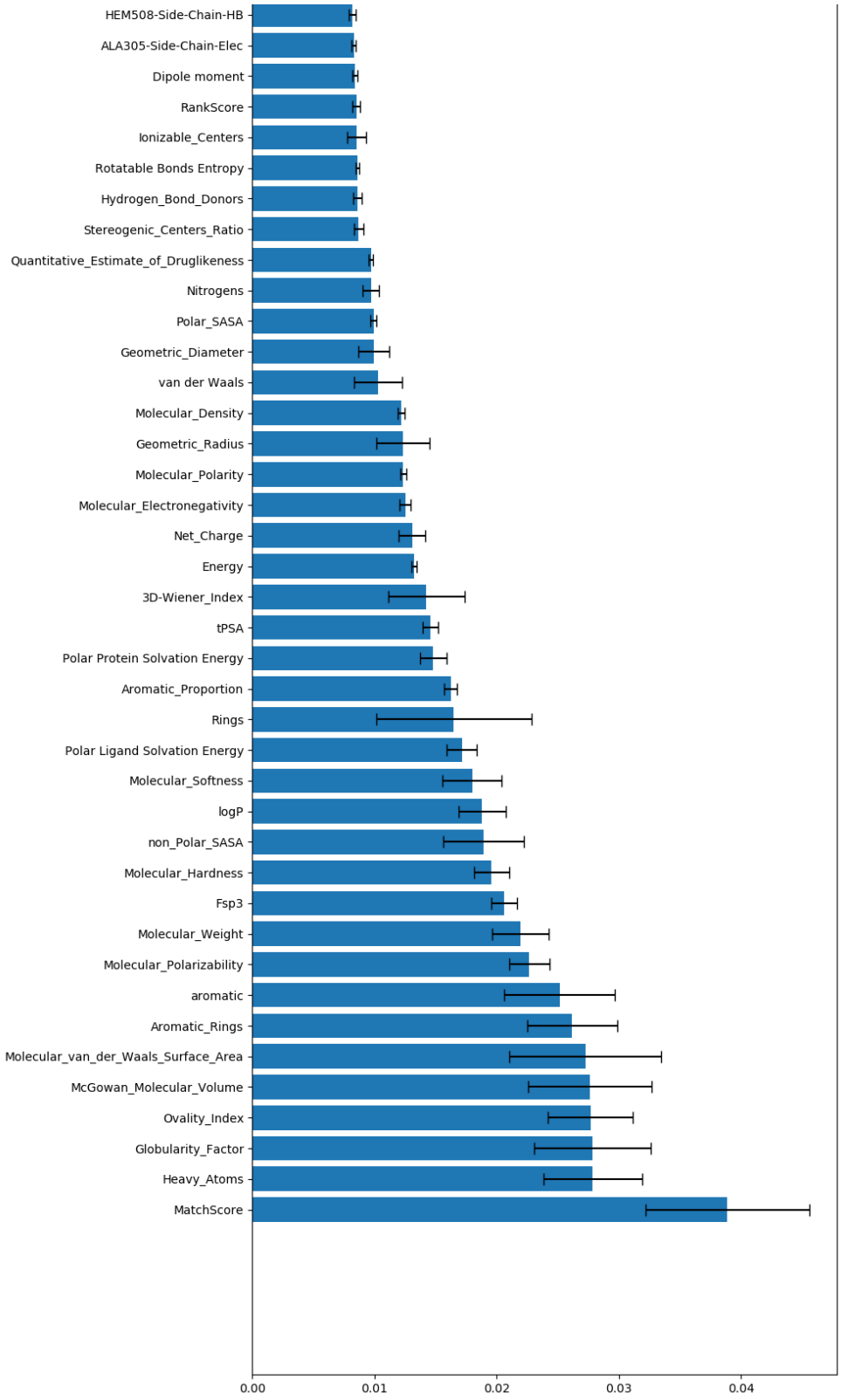


Figure S13: Feature importance of random forest model using selected top 100 features for isoform 3A4.

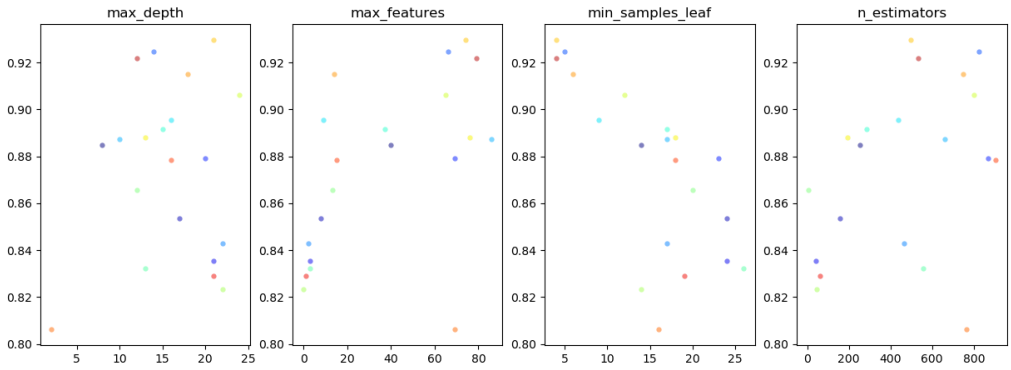


Figure S14: Example of hyperparameter tuning using hyperopt TPE for random forest.

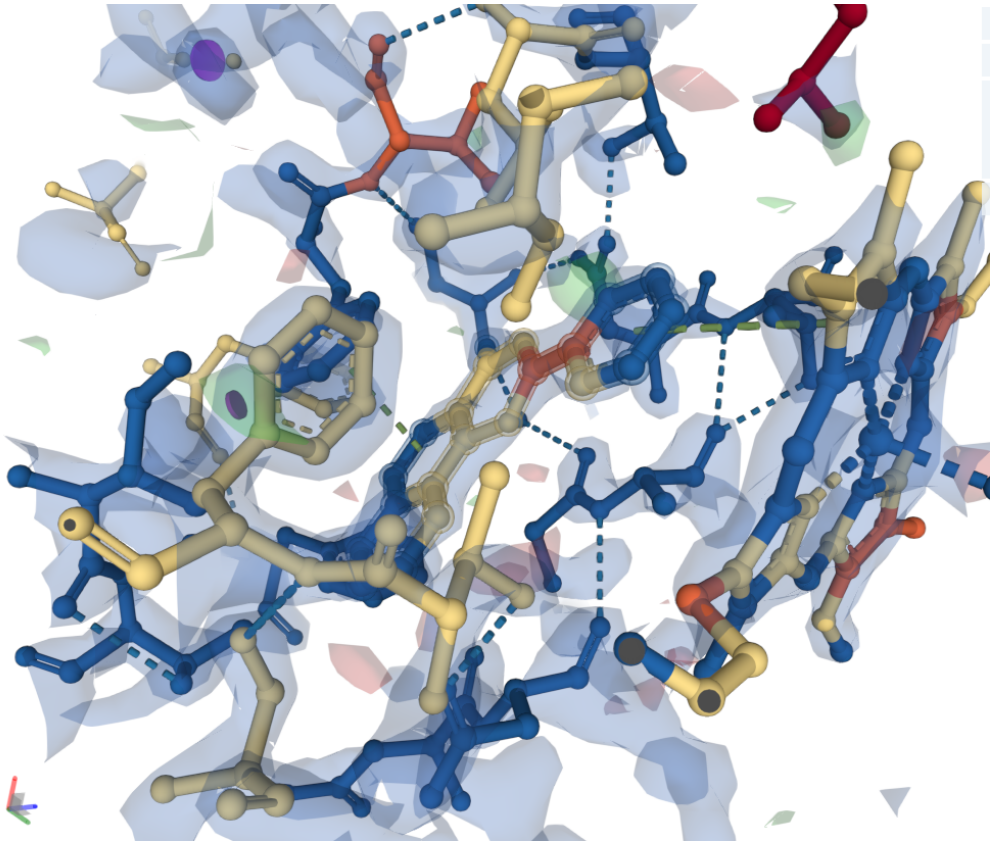

Figure S15: 2HI4: `https://www.rcsb.org/3d-view/2HI4?preset=electronDensityMaps`

Table S2: 1A2 Data

| | **30** | **80** | **70** | **40** | **20** | **50** | **60** |
|---|---|---|---|---|---|---|---|
| **roc_auc_all** | 0.888 | 0.926 | 0.922 | 0.911 | 0.849 | 0.911 | 0.919 |
| **bootstrapped_scores_mean_all** | 0.888 | 0.927 | 0.922 | 0.911 | 0.849 | 0.911 | 0.918 |
| **bootstrapped_scores_std_all** | 0.009 | 0.007 | 0.008 | 0.008 | 0.010 | 0.008 | 0.008 |
| **roc_auc_ligand** | 0.876 | 0.918 | 0.917 | 0.898 | 0.830 | 0.909 | 0.913 |
| **bootstrapped_scores_mean_ligand** | 0.876 | 0.918 | 0.917 | 0.897 | 0.831 | 0.910 | 0.913 |
| **bootstrapped_scores_std_ligand** | 0.010 | 0.008 | 0.008 | 0.009 | 0.012 | 0.009 | 0.009 |
| **roc_auc_docked** | 0.868 | 0.907 | 0.905 | 0.892 | 0.783 | 0.898 | 0.903 |
| **bootstrapped_scores_mean_docked** | 0.868 | 0.906 | 0.905 | 0.892 | 0.783 | 0.898 | 0.903 |
| **bootstrapped_scores_std_docked** | 0.010 | 0.008 | 0.009 | 0.009 | 0.012 | 0.009 | 0.009 |
| **fitted_auc_docked** | 0.677 | 0.677 | 0.677 | 0.677 | 0.677 | 0.677 | 0.677 |
| **auc_improvement_docked** | 19.120 | 22.960 | 22.780 | 21.470 | 10.640 | 22.120 | 22.580 |
| **optimal_threshold_docked** | 0.505 | 0.394 | 0.404 | 0.374 | 0.081 | 0.242 | 0.343 |
| **accuracy_docked** | 0.786 | 0.831 | 0.830 | 0.813 | 0.704 | 0.819 | 0.821 |
| **active_pos_rate_docked** | 0.783 | 0.873 | 0.856 | 0.876 | 0.823 | 0.872 | 0.875 |
| **inactive_neg_rate_docked** | 0.789 | 0.798 | 0.810 | 0.764 | 0.609 | 0.777 | 0.779 |
| **threshold_95_docked** | 0.152 | 0.172 | 0.091 | 0.182 | 0.030 | 0.061 | 0.162 |
| **accuracy_95_docked** | 0.731 | 0.769 | 0.735 | 0.752 | 0.639 | 0.755 | 0.767 |
| **active_acc_95_docked** | 0.930 | 0.937 | 0.945 | 0.939 | 0.932 | 0.939 | 0.946 |
| **inactive_acc_95_docked** | 0.574 | 0.637 | 0.569 | 0.605 | 0.408 | 0.609 | 0.626 |
| **fitted_auc_ligand** | 0.677 | 0.677 | 0.677 | 0.677 | 0.677 | 0.677 | 0.677 |
| **auc_improvement_ligand** | 19.910 | 24.140 | 24.000 | 22.060 | 15.350 | 23.250 | 23.580 |
| **optimal_threshold_ligand** | 0.192 | 0.475 | 0.343 | 0.455 | 0.626 | 0.465 | 0.495 |
| **accuracy_ligand** | 0.778 | 0.846 | 0.839 | 0.824 | 0.741 | 0.840 | 0.841 |
| **active_pos_rate_ligand** | 0.899 | 0.864 | 0.879 | 0.872 | 0.836 | 0.840 | 0.842 |
| **inactive_neg_rate_ligand** | 0.682 | 0.832 | 0.808 | 0.787 | 0.667 | 0.841 | 0.841 |
| **threshold_95_ligand** | 0.051 | 0.172 | 0.071 | 0.152 | 0.162 | 0.061 | 0.182 |
| **accuracy_95_ligand** | 0.710 | 0.770 | 0.761 | 0.760 | 0.668 | 0.757 | 0.763 |
| **active_acc_95_ligand** | 0.948 | 0.950 | 0.948 | 0.947 | 0.948 | 0.945 | 0.945 |
| **inactive_acc_95_ligand** | 0.522 | 0.628 | 0.612 | 0.613 | 0.447 | 0.609 | 0.619 |
| **fitted_auc_all** | 0.677 | 0.677 | 0.677 | 0.677 | 0.677 | 0.677 | 0.677 |
| **auc_improvement_all** | 21.100 | 24.930 | 24.470 | 23.390 | 17.170 | 23.420 | 24.170 |
| **optimal_threshold_all** | 0.263 | 0.192 | 0.303 | 0.192 | 0.333 | 0.485 | 0.232 |
| **accuracy_all** | 0.796 | 0.830 | 0.833 | 0.798 | 0.761 | 0.835 | 0.816 |
| **active_pos_rate_all** | 0.880 | 0.915 | 0.902 | 0.930 | 0.842 | 0.826 | 0.922 |
| **inactive_neg_rate_all** | 0.729 | 0.762 | 0.777 | 0.694 | 0.697 | 0.842 | 0.732 |
| **threshold_95_all** | 0.101 | 0.111 | 0.162 | 0.182 | 0.061 | 0.232 | 0.172 |
| **accuracy_95_all** | 0.762 | 0.814 | 0.798 | 0.797 | 0.689 | 0.799 | 0.799 |
| **active_acc_95_all** | 0.934 | 0.943 | 0.950 | 0.934 | 0.964 | 0.940 | 0.944 |
| **inactive_acc_95_all** | 0.625 | 0.713 | 0.679 | 0.689 | 0.471 | 0.687 | 0.684 |

Table S3: 3A4 Data

| | **60** | **20** | **30** | **80** | **40** | **50** | **70** |
|---|---|---|---|---|---|---|---|
| **Roc_auc_all** | 0.924 | 0.822 | 0.863 | 0.932 | 0.892 | 0.908 | 0.923 |
| **Bootstrapped_scores_mean_all** | 0.924 | 0.822 | 0.863 | 0.932 | 0.892 | 0.907 | 0.923 |
| **Bootstrapped_scores_std_all** | 0.007 | 0.011 | 0.010 | 0.007 | 0.008 | 0.008 | 0.007 |
| **Roc_auc_ligand** | 0.914 | 0.808 | 0.838 | 0.919 | 0.863 | 0.895 | 0.915 |
| **Bootstrapped_scores_mean_ligand** | 0.915 | 0.808 | 0.838 | 0.919 | 0.863 | 0.895 | 0.915 |
| **Bootstrapped_scores_std_ligand** | 0.008 | 0.012 | 0.011 | 0.008 | 0.010 | 0.008 | 0.008 |
| **Roc_auc_docked** | 0.878 | 0.762 | 0.839 | 0.892 | 0.850 | 0.862 | 0.885 |
| **Bootstrapped_scores_mean_docked** | 0.878 | 0.762 | 0.839 | 0.892 | 0.850 | 0.862 | 0.885 |
| **Bootstrapped_scores_std_docked** | 0.009 | 0.012 | 0.011 | 0.008 | 0.010 | 0.010 | 0.009 |
| **Fitted_auc_docked** | 0.742 | 0.742 | 0.742 | 0.742 | 0.742 | 0.742 | 0.742 |
| **Auc_improvement_docked** | 13.610 | 2.010 | 9.730 | 15.080 | 10.860 | 12.020 | 14.300 |
| **Optimal_threshold_docked** | 0.333 | 0.061 | 0.384 | 0.495 | 0.404 | 0.273 | 0.576 |
| **Accuracy_docked** | 0.773 | 0.634 | 0.763 | 0.804 | 0.776 | 0.760 | 0.801 |
| **Active_pos_rate_docked** | 0.862 | 0.896 | 0.805 | 0.790 | 0.793 | 0.867 | 0.737 |
| **Inactive_neg_rate_docked** | 0.714 | 0.459 | 0.735 | 0.813 | 0.765 | 0.689 | 0.844 |
| **Threshold_95_docked** | 0.192 | 0.040 | 0.141 | 0.192 | 0.152 | 0.182 | 0.263 |
| **Accuracy_95_docked** | 0.743 | 0.622 | 0.700 | 0.741 | 0.713 | 0.746 | 0.742 |
| **Active_acc_95_docked** | 0.932 | 0.925 | 0.936 | 0.934 | 0.939 | 0.918 | 0.927 |
| **Inactive_acc_95_docked** | 0.617 | 0.420 | 0.542 | 0.613 | 0.562 | 0.631 | 0.619 |
| **Fitted_auc_all** | 0.742 | 0.742 | 0.742 | 0.742 | 0.742 | 0.742 | 0.742 |
| **Auc_improvement_all** | 18.220 | 8.090 | 12.120 | 19.080 | 15.000 | 16.610 | 18.180 |
| **Optimal_threshold_all** | 0.384 | 0.404 | 0.263 | 0.535 | 0.222 | 0.354 | 0.232 |
| **Accuracy_all** | 0.850 | 0.730 | 0.783 | 0.862 | 0.817 | 0.829 | 0.840 |
| **Active_pos_rate_all** | 0.847 | 0.755 | 0.860 | 0.792 | 0.857 | 0.820 | 0.899 |
| **Inactive_neg_rate_all** | 0.852 | 0.714 | 0.732 | 0.908 | 0.791 | 0.836 | 0.801 |
| **Threshold_95_all** | 0.111 | 0.051 | 0.131 | 0.111 | 0.081 | 0.101 | 0.101 |
| **Accuracy_95_all** | 0.797 | 0.647 | 0.734 | 0.798 | 0.771 | 0.799 | 0.805 |
| **Active_acc_95_all** | 0.938 | 0.949 | 0.954 | 0.948 | 0.931 | 0.946 | 0.947 |
| **Inactive_acc_95_all** | 0.702 | 0.445 | 0.587 | 0.699 | 0.664 | 0.701 | 0.711 |
| **Fitted_auc_ligand** | 0.742 | 0.742 | 0.742 | 0.742 | 0.742 | 0.742 | 0.742 |
| **Auc_improvement_ligand** | 17.270 | 6.690 | 9.670 | 17.740 | 12.170 | 15.300 | 17.350 |
| **Optimal_threshold_ligand** | 0.354 | 0.212 | 0.202 | 0.263 | 0.111 | 0.202 | 0.485 |
| **Accuracy_ligand** | 0.831 | 0.723 | 0.753 | 0.834 | 0.763 | 0.803 | 0.830 |
| **Active_pos_rate_ligand** | 0.883 | 0.795 | 0.814 | 0.877 | 0.898 | 0.901 | 0.792 |
| **Inactive_neg_rate_ligand** | 0.796 | 0.675 | 0.713 | 0.805 | 0.674 | 0.738 | 0.855 |
| **Threshold_95_ligand** | 0.131 | 0.030 | 0.051 | 0.111 | 0.030 | 0.101 | 0.141 |
| **Accuracy_95_ligand** | 0.797 | 0.680 | 0.709 | 0.808 | 0.713 | 0.763 | 0.802 |
| **Active_acc_95_ligand** | 0.939 | 0.928 | 0.953 | 0.933 | 0.954 | 0.943 | 0.936 |
| **Inactive_acc_95_ligand** | 0.702 | 0.514 | 0.546 | 0.724 | 0.552 | 0.642 | 0.712 |

Table S4: 2C9 Data

| | **30** | **40** | **60** | **20** | **50** | **80** | **70** |
|---|---|---|---|---|---|---|---|
| **Roc_auc_all** | 0.833 | 0.839 | 0.865 | 0.773 | 0.847 | 0.875 | 0.874 |
| **Bootstrapped_scores_mean_all** | 0.833 | 0.839 | 0.865 | 0.773 | 0.847 | 0.875 | 0.874 |
| **Bootstrapped_scores_std_all** | 0.011 | 0.011 | 0.011 | 0.013 | 0.012 | 0.011 | 0.011 |
| **Roc_auc_ligand** | 0.816 | 0.832 | 0.842 | 0.776 | 0.845 | 0.869 | 0.858 |
| **Bootstrapped_scores_mean_ligand** | 0.816 | 0.832 | 0.842 | 0.775 | 0.845 | 0.868 | 0.858 |
| **Bootstrapped_scores_std_ligand** | 0.012 | 0.011 | 0.012 | 0.011 | 0.011 | 0.011 | 0.011 |
| **Roc_auc_docked** | 0.773 | 0.797 | 0.823 | 0.711 | 0.813 | 0.857 | 0.844 |
| **Bootstrapped_scores_mean_docked** | 0.773 | 0.797 | 0.822 | 0.711 | 0.813 | 0.856 | 0.844 |
| **Bootstrapped_scores_std_docked** | 0.013 | 0.012 | 0.012 | 0.013 | 0.012 | 0.011 | 0.011 |
| **Fitted_auc_all** | 0.687 | 0.687 | 0.687 | 0.687 | 0.687 | 0.687 | 0.687 |
| **Auc_improvement_all** | 14.630 | 15.190 | 17.820 | 8.650 | 16.000 | 18.840 | 18.740 |
| **Optimal_threshold_all** | 0.273 | 0.646 | 0.495 | 0.758 | 0.636 | 0.333 | 0.485 |
| **Accuracy_all** | 0.736 | 0.786 | 0.818 | 0.740 | 0.810 | 0.832 | 0.832 |
| **Active_pos_rate_all** | 0.769 | 0.632 | 0.695 | 0.606 | 0.612 | 0.734 | 0.707 |
| **Inactive_neg_rate_all** | 0.725 | 0.837 | 0.860 | 0.785 | 0.877 | 0.865 | 0.875 |
| **Threshold_95_all** | 0.020 | 0.040 | 0.020 | 0.010 | 0.030 | 0.010 | 0.040 |
| **Accuracy_95_all** | 0.608 | 0.602 | 0.613 | 0.433 | 0.595 | 0.649 | 0.626 |
| **Active_acc_95_all** | 0.946 | 0.931 | 0.933 | 0.952 | 0.929 | 0.933 | 0.937 |
| **Inactive_acc_95_all** | 0.494 | 0.492 | 0.507 | 0.259 | 0.483 | 0.554 | 0.522 |
| **Fitted_auc_docked** | 0.687 | 0.687 | 0.687 | 0.687 | 0.687 | 0.687 | 0.687 |
| **Auc_improvement_docked** | 8.630 | 11.020 | 13.610 | 2.430 | 12.590 | 16.970 | 15.700 |
| **Optimal_threshold_docked** | 0.222 | 0.394 | 0.535 | 0.313 | 0.414 | 0.455 | 0.566 |
| **Accuracy_docked** | 0.631 | 0.697 | 0.781 | 0.689 | 0.723 | 0.808 | 0.806 |
| **Active_pos_rate_docked** | 0.844 | 0.774 | 0.646 | 0.533 | 0.768 | 0.717 | 0.630 |
| **Inactive_neg_rate_docked** | 0.560 | 0.671 | 0.827 | 0.741 | 0.709 | 0.838 | 0.865 |
| **Threshold_95_docked** | 0.061 | 0.121 | 0.030 | 0.030 | 0.101 | 0.071 | 0.061 |
| **Accuracy_95_docked** | 0.516 | 0.570 | 0.555 | 0.465 | 0.584 | 0.596 | 0.584 |
| **Active_acc_95_docked** | 0.944 | 0.934 | 0.944 | 0.923 | 0.933 | 0.948 | 0.941 |
| **Inactive_acc_95_docked** | 0.372 | 0.449 | 0.425 | 0.312 | 0.468 | 0.479 | 0.464 |
| **Fitted_auc_ligand** | 0.687 | 0.687 | 0.687 | 0.687 | 0.687 | 0.687 | 0.687 |
| **Auc_improvement_ligand** | 12.890 | 14.540 | 15.550 | 8.860 | 15.760 | 18.160 | 17.100 |
| **Optimal_threshold_ligand** | 0.586 | 0.556 | 0.404 | 0.869 | 0.475 | 0.636 | 0.253 |
| **Accuracy_ligand** | 0.745 | 0.779 | 0.794 | 0.753 | 0.786 | 0.834 | 0.797 |
| **Active_pos_rate_ligand** | 0.658 | 0.676 | 0.738 | 0.593 | 0.709 | 0.646 | 0.771 |
| **Inactive_neg_rate_ligand** | 0.774 | 0.813 | 0.813 | 0.807 | 0.812 | 0.896 | 0.806 |
| **Threshold_95_ligand** | 0.061 | 0.020 | 0.020 | 0.040 | 0.051 | 0.010 | 0.020 |
| **Accuracy_95_ligand** | 0.581 | 0.577 | 0.573 | 0.423 | 0.610 | 0.678 | 0.623 |
| **Active_acc_95_ligand** | 0.948 | 0.925 | 0.939 | 0.944 | 0.919 | 0.911 | 0.922 |
| **Inactive_acc_95_ligand** | 0.458 | 0.461 | 0.450 | 0.248 | 0.507 | 0.601 | 0.523 |

Table S5: 2C19

| | **30** | **50** | **70** | **80** | **40** | **60** | **20** |
|---|---|---|---|---|---|---|---|
| **Roc_auc_all** | 0.836 | 0.855 | 0.873 | 0.879 | 0.846 | 0.868 | 0.720 |
| **Bootstrapped_scores_mean_all** | 0.836 | 0.855 | 0.873 | 0.879 | 0.846 | 0.868 | 0.720 |
| **Bootstrapped_scores_std_all** | 0.011 | 0.011 | 0.011 | 0.010 | 0.011 | 0.011 | 0.012 |
| **Roc_auc_ligand** | 0.807 | 0.851 | 0.876 | 0.876 | 0.834 | 0.865 | 0.730 |
| **Bootstrapped_scores_mean_ligand** | 0.807 | 0.851 | 0.877 | 0.877 | 0.834 | 0.866 | 0.730 |
| **Bootstrapped_scores_std_ligand** | 0.013 | 0.011 | 0.010 | 0.010 | 0.012 | 0.011 | 0.013 |
| **Roc_auc_docked** | 0.756 | 0.793 | 0.833 | 0.837 | 0.790 | 0.815 | 0.659 |
| **Bootstrapped_scores_mean_docked** | 0.756 | 0.793 | 0.832 | 0.838 | 0.790 | 0.814 | 0.658 |
| **Bootstrapped_scores_std_docked** | 0.012 | 0.012 | 0.011 | 0.011 | 0.012 | 0.011 | 0.012 |
| **Fitted_auc_ligand** | 0.616 | 0.616 | 0.616 | 0.616 | 0.616 | 0.616 | 0.616 |
| **Auc_improvement_ligand** | 19.110 | 23.480 | 26.040 | 26.060 | 21.780 | 24.940 | 11.420 |
| **Optimal_threshold_ligand** | 0.364 | 0.354 | 0.404 | 0.434 | 0.263 | 0.606 | 0.323 |
| **Accuracy_ligand** | 0.716 | 0.757 | 0.810 | 0.813 | 0.745 | 0.809 | 0.665 |
| **Active_pos_rate_ligand** | 0.810 | 0.874 | 0.838 | 0.820 | 0.830 | 0.727 | 0.750 |
| **Inactive_neg_rate_ligand** | 0.657 | 0.684 | 0.793 | 0.809 | 0.693 | 0.860 | 0.611 |
| **Threshold_95_ligand** | 0.061 | 0.141 | 0.152 | 0.172 | 0.030 | 0.162 | 0.010 |
| **Accuracy_95_ligand** | 0.628 | 0.649 | 0.720 | 0.715 | 0.651 | 0.680 | 0.537 |
| **Active_acc_95_ligand** | 0.934 | 0.939 | 0.944 | 0.941 | 0.941 | 0.941 | 0.937 |
| **Inactive_acc_95_ligand** | 0.439 | 0.470 | 0.581 | 0.575 | 0.470 | 0.517 | 0.290 |
| **Fitted_auc_all** | 0.616 | 0.616 | 0.616 | 0.616 | 0.616 | 0.616 | 0.616 |
| **Auc_improvement_all** | 22.030 | 23.920 | 25.700 | 26.360 | 23.050 | 25.260 | 10.430 |
| **Optimal_threshold_all** | 0.535 | 0.475 | 0.303 | 0.384 | 0.293 | 0.404 | 0.303 |
| **Accuracy_all** | 0.762 | 0.785 | 0.790 | 0.808 | 0.744 | 0.802 | 0.656 |
| **Active_pos_rate_all** | 0.758 | 0.805 | 0.851 | 0.838 | 0.868 | 0.828 | 0.743 |
| **Inactive_neg_rate_all** | 0.765 | 0.772 | 0.752 | 0.790 | 0.667 | 0.785 | 0.601 |
| **Threshold_95_all** | 0.111 | 0.111 | 0.091 | 0.101 | 0.101 | 0.071 | 0.030 |
| **Accuracy_95_all** | 0.650 | 0.662 | 0.711 | 0.713 | 0.661 | 0.704 | 0.531 |
| **Active_acc_95_all** | 0.937 | 0.940 | 0.939 | 0.946 | 0.941 | 0.937 | 0.944 |
| **Inactive_acc_95_all** | 0.472 | 0.490 | 0.569 | 0.569 | 0.487 | 0.559 | 0.274 |
| **Fitted_auc_docked** | 0.616 | 0.616 | 0.616 | 0.616 | 0.616 | 0.616 | 0.616 |
| **Auc_improvement_docked** | 14.070 | 17.750 | 21.680 | 22.160 | 17.410 | 19.920 | 4.280 |
| **Optimal_threshold_docked** | 0.424 | 0.364 | 0.434 | 0.424 | 0.414 | 0.465 | 0.202 |
| **Accuracy_docked** | 0.661 | 0.680 | 0.747 | 0.756 | 0.696 | 0.720 | 0.580 |
| **Active_pos_rate_docked** | 0.813 | 0.876 | 0.807 | 0.783 | 0.783 | 0.778 | 0.778 |
| **Inactive_neg_rate_docked** | 0.567 | 0.559 | 0.709 | 0.739 | 0.642 | 0.684 | 0.457 |
| **Threshold_95_docked** | 0.202 | 0.242 | 0.182 | 0.162 | 0.061 | 0.202 | 0.040 |
| **Accuracy_95_docked** | 0.578 | 0.614 | 0.634 | 0.665 | 0.612 | 0.642 | 0.482 |
| **Active_acc_95_docked** | 0.944 | 0.936 | 0.949 | 0.939 | 0.936 | 0.950 | 0.938 |
| **Inactive_acc_95_docked** | 0.351 | 0.413 | 0.438 | 0.496 | 0.411 | 0.450 | 0.200 |

Table S6: 2D6

| | **30** | **70** | **60** | **80** | **50** | **20** | **40** |
|---|---|---|---|---|---|---|---|
| **Roc_auc_all** | 0.844 | 0.892 | 0.890 | 0.901 | 0.884 | 0.692 | 0.889 |
| **Bootstrapped_scores_mean_all** | 0.844 | 0.892 | 0.889 | 0.901 | 0.885 | 0.692 | 0.890 |
| **Bootstrapped_scores_std_all** | 0.014 | 0.012 | 0.013 | 0.012 | 0.012 | 0.015 | 0.012 |
| **Roc_auc_ligand** | 0.817 | 0.880 | 0.867 | 0.880 | 0.864 | 0.715 | 0.874 |
| **Bootstrapped_scores_mean_ligand** | 0.817 | 0.880 | 0.867 | 0.879 | 0.864 | 0.715 | 0.874 |
| **Bootstrapped_scores_std_ligand** | 0.015 | 0.014 | 0.014 | 0.014 | 0.013 | 0.017 | 0.014 |
| **Roc_auc_docked** | 0.823 | 0.842 | 0.839 | 0.854 | 0.833 | 0.624 | 0.827 |
| **Bootstrapped_scores_mean_docked** | 0.824 | 0.842 | 0.838 | 0.854 | 0.834 | 0.624 | 0.826 |
| **Bootstrapped_scores_std_docked** | 0.016 | 0.014 | 0.015 | 0.015 | 0.015 | 0.017 | 0.016 |
| **Fitted_auc_ligand** | 0.551 | 0.551 | 0.551 | 0.551 | 0.551 | 0.551 | 0.551 |
| **Auc_improvement_ligand** | 26.570 | 32.880 | 31.570 | 32.830 | 31.270 | 16.380 | 32.270 |
| **Optimal_threshold_ligand** | 0.162 | 0.192 | 0.313 | 0.697 | 0.253 | 0.051 | 0.545 |
| **Accuracy_ligand** | 0.847 | 0.873 | 0.893 | 0.915 | 0.874 | 0.516 | 0.899 |
| **Active_pos_rate_ligand** | 0.586 | 0.711 | 0.557 | 0.491 | 0.601 | 0.869 | 0.495 |
| **Inactive_neg_rate_ligand** | 0.886 | 0.897 | 0.944 | 0.980 | 0.916 | 0.462 | 0.961 |
| **Threshold_95_ligand** | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.020 | 0.020 |
| **Accuracy_95_ligand** | 0.564 | 0.575 | 0.801 | 0.570 | 0.571 | 0.381 | 0.502 |
| **Active_acc_95_ligand** | 0.886 | 0.938 | 0.775 | 0.938 | 0.944 | 0.960 | 0.963 |
| **Inactive_acc_95_ligand** | 0.515 | 0.519 | 0.805 | 0.514 | 0.514 | 0.293 | 0.432 |
| **Fitted_auc_docked** | 0.551 | 0.551 | 0.551 | 0.551 | 0.551 | 0.551 | 0.551 |
| **Auc_improvement_docked** | 27.190 | 29.090 | 28.760 | 30.290 | 28.200 | 7.290 | 27.540 |
| **Optimal_threshold_docked** | 0.202 | 0.253 | 0.323 | 0.212 | 0.283 | 0.010 | 0.293 |
| **Accuracy_docked** | 0.833 | 0.853 | 0.874 | 0.864 | 0.855 | 0.727 | 0.858 |
| **Active_pos_rate_docked** | 0.553 | 0.590 | 0.493 | 0.640 | 0.530 | 0.430 | 0.568 |
| **Inactive_neg_rate_docked** | 0.876 | 0.893 | 0.933 | 0.898 | 0.905 | 0.773 | 0.902 |
| **Threshold_95_docked** | 0.020 | 0.020 | 0.010 | 0.010 | 0.020 | 0.010 | 0.020 |
| **Accuracy_95_docked** | 0.462 | 0.547 | 0.558 | 0.551 | 0.526 | 0.727 | 0.468 |
| **Active_acc_95_docked** | 0.940 | 0.925 | 0.919 | 0.925 | 0.929 | 0.430 | 0.936 |
| **Inactive_acc_95_docked** | 0.389 | 0.489 | 0.503 | 0.494 | 0.465 | 0.773 | 0.397 |
| **Fitted_auc_all** | 0.551 | 0.551 | 0.551 | 0.551 | 0.551 | 0.551 | 0.551 |
| **Auc_improvement_all** | 29.270 | 34.060 | 33.850 | 34.950 | 33.310 | 14.110 | 33.820 |
| **Optimal_threshold_all** | 0.131 | 0.232 | 0.323 | 0.202 | 0.192 | 0.121 | 0.111 |
| **Accuracy_all** | 0.844 | 0.890 | 0.899 | 0.889 | 0.874 | 0.698 | 0.871 |
| **Active_pos_rate_all** | 0.678 | 0.617 | 0.565 | 0.680 | 0.653 | 0.526 | 0.713 |
| **Inactive_neg_rate_all** | 0.869 | 0.932 | 0.950 | 0.921 | 0.908 | 0.725 | 0.896 |
| **Threshold_95_all** | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| **Accuracy_95_all** | 0.593 | 0.677 | 0.717 | 0.624 | 0.609 | 0.419 | 0.694 |
| **Active_acc_95_all** | 0.898 | 0.909 | 0.898 | 0.946 | 0.940 | 0.869 | 0.900 |
| **Inactive_acc_95_all** | 0.547 | 0.642 | 0.690 | 0.575 | 0.558 | 0.350 | 0.662 |

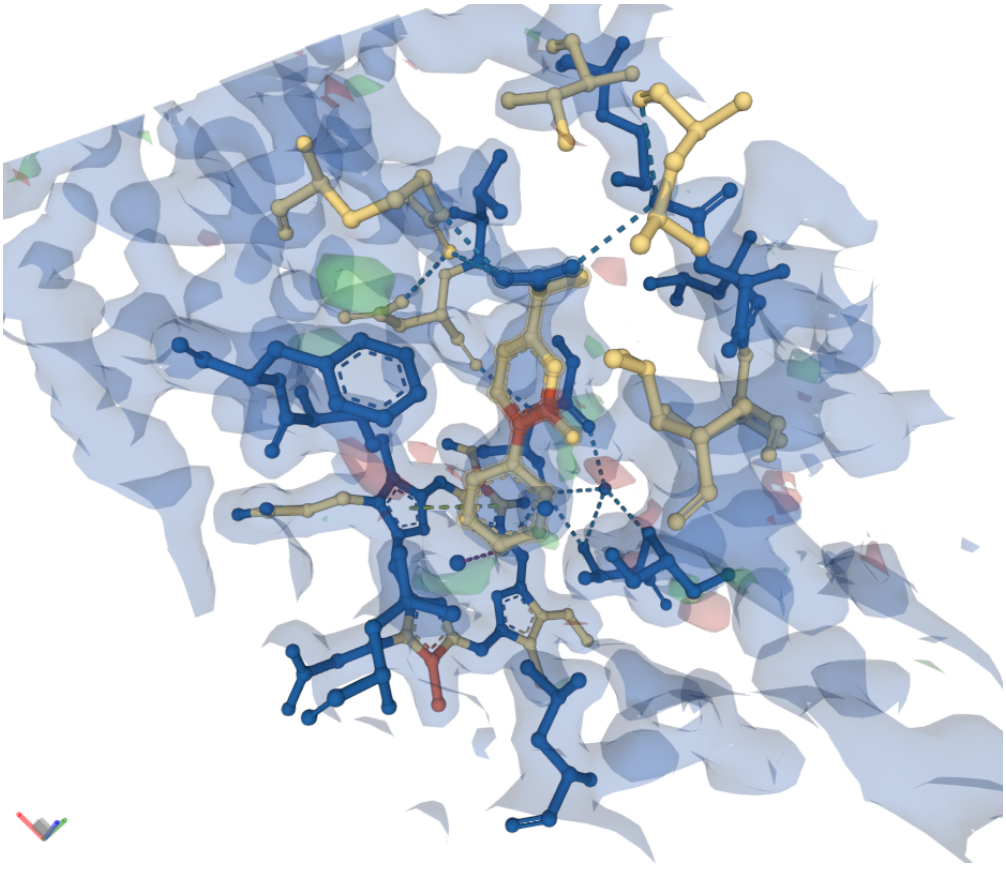

Figure S16: 1R9O: `https://www.rcsb.org/3d-view/1R9O?preset=electronDensityMaps`

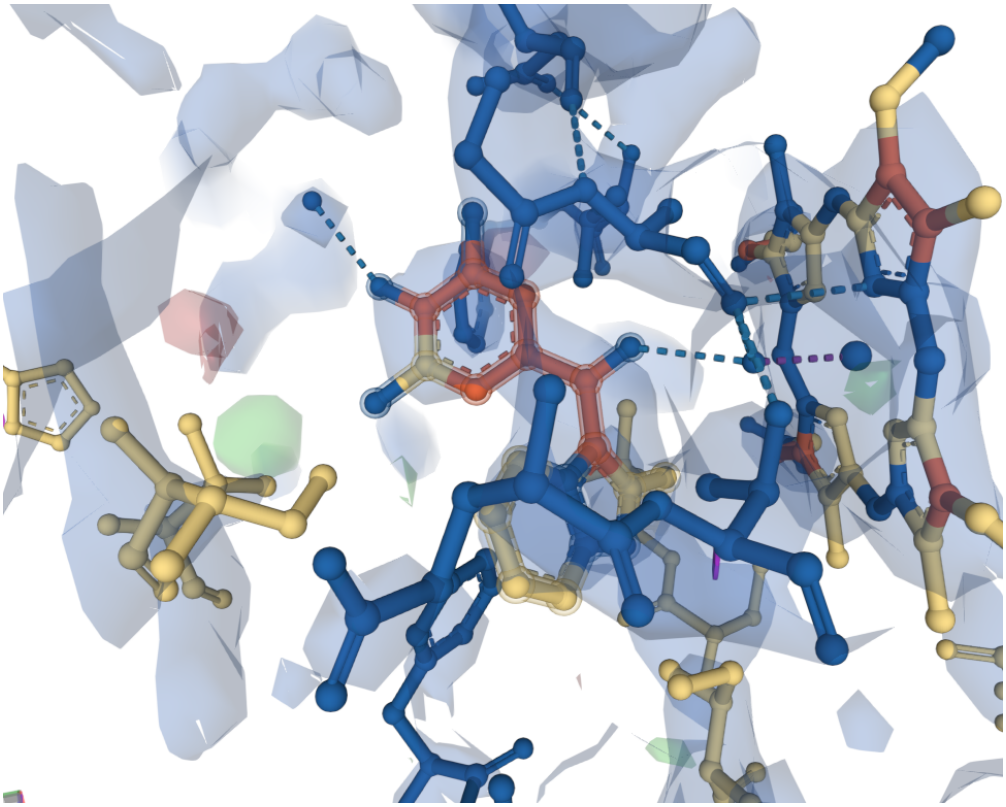

Figure S17: 4GQS: `https://www.rcsb.org/3d-view/4GQS?preset=electronDensityMaps`

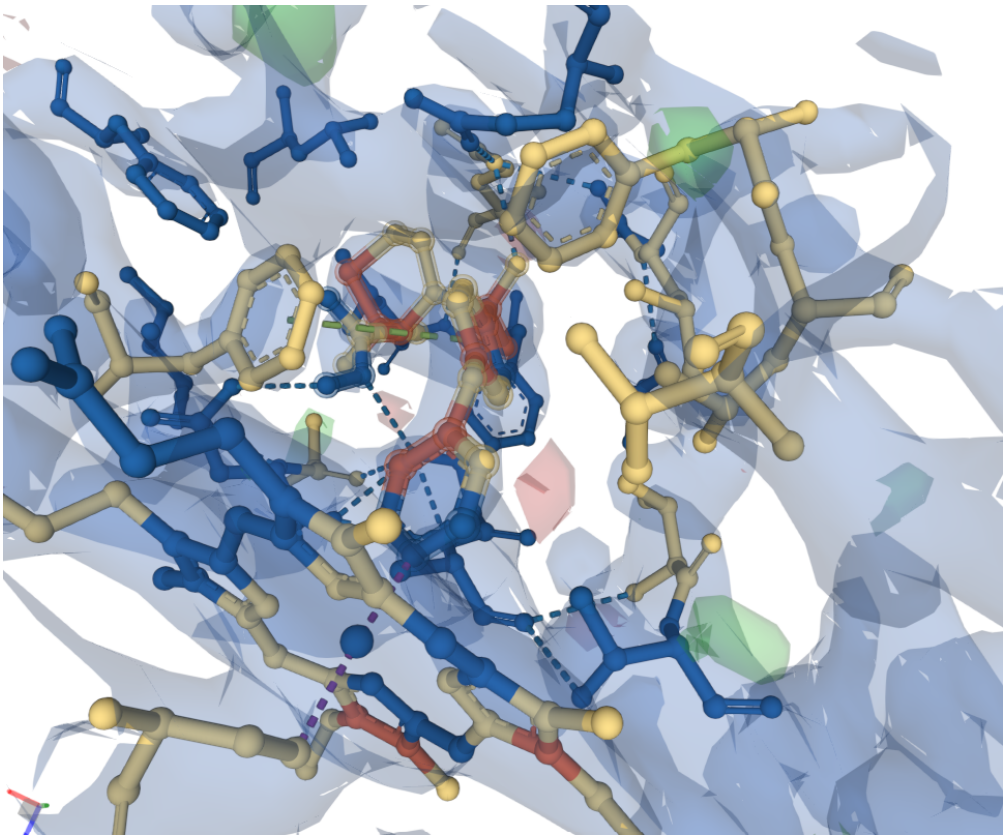

Figure S18: 3QM4: `https://www.rcsb.org/3d-view/3QM4?preset=electronDensityMaps`

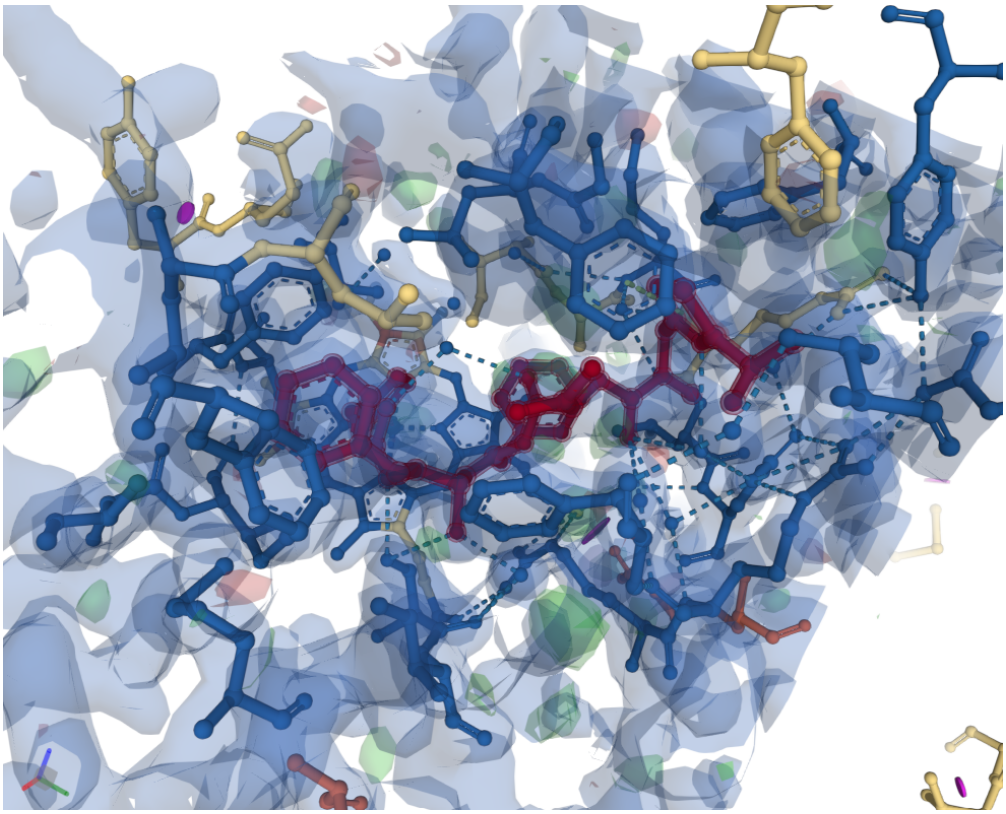

Figure S19: 3NXU: `https://www.rcsb.org/3d-view/3NXU?preset=electronDensityMaps`