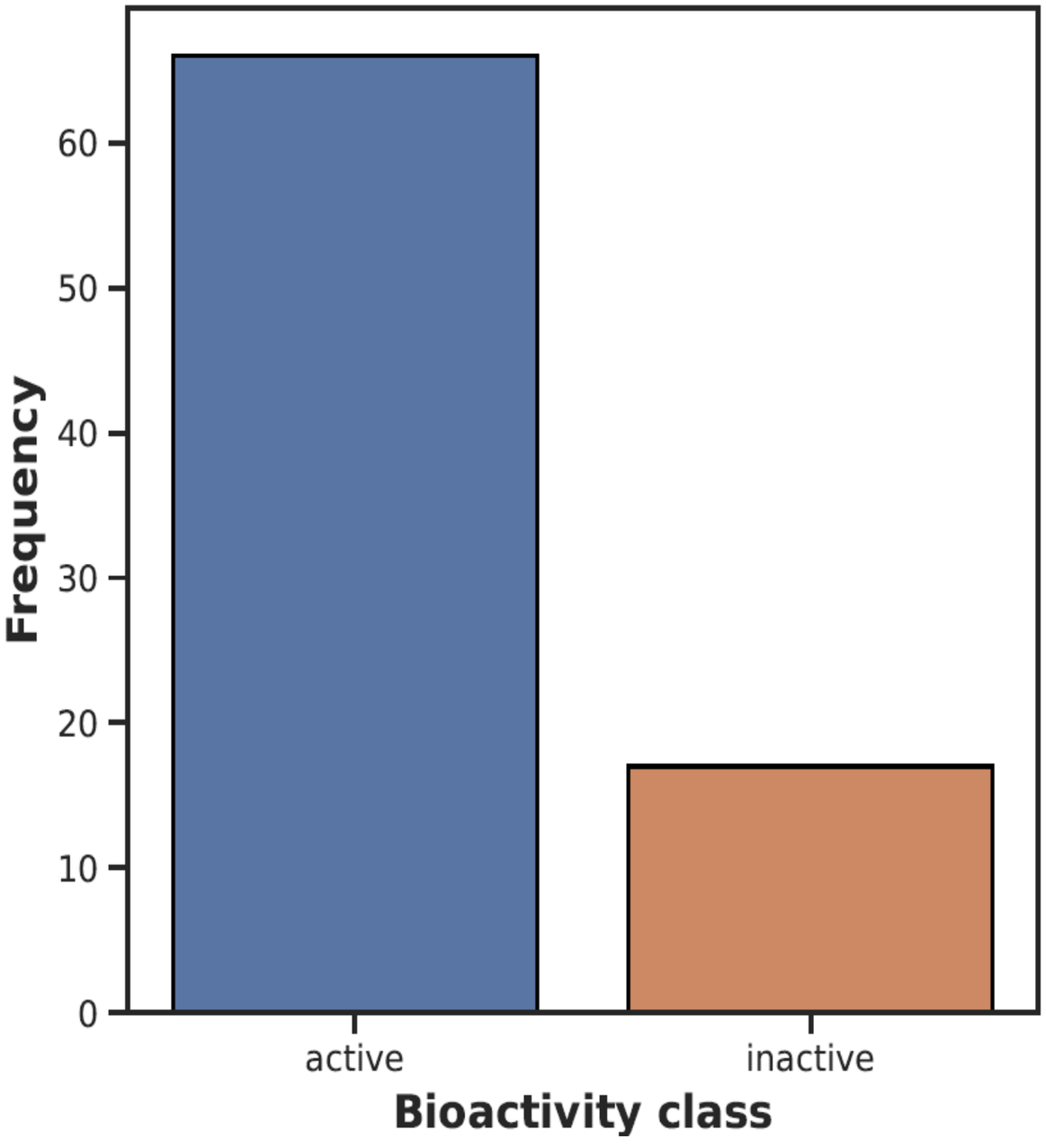
Decoding the inhibition selectivity and physicochemical multiverse of beta-carbonic anhydrase inhibitors through random forest and neural network-assisted QSAR/QSAAR modeling: A novel approach for the rational design of new anti-tuberculosis drugs

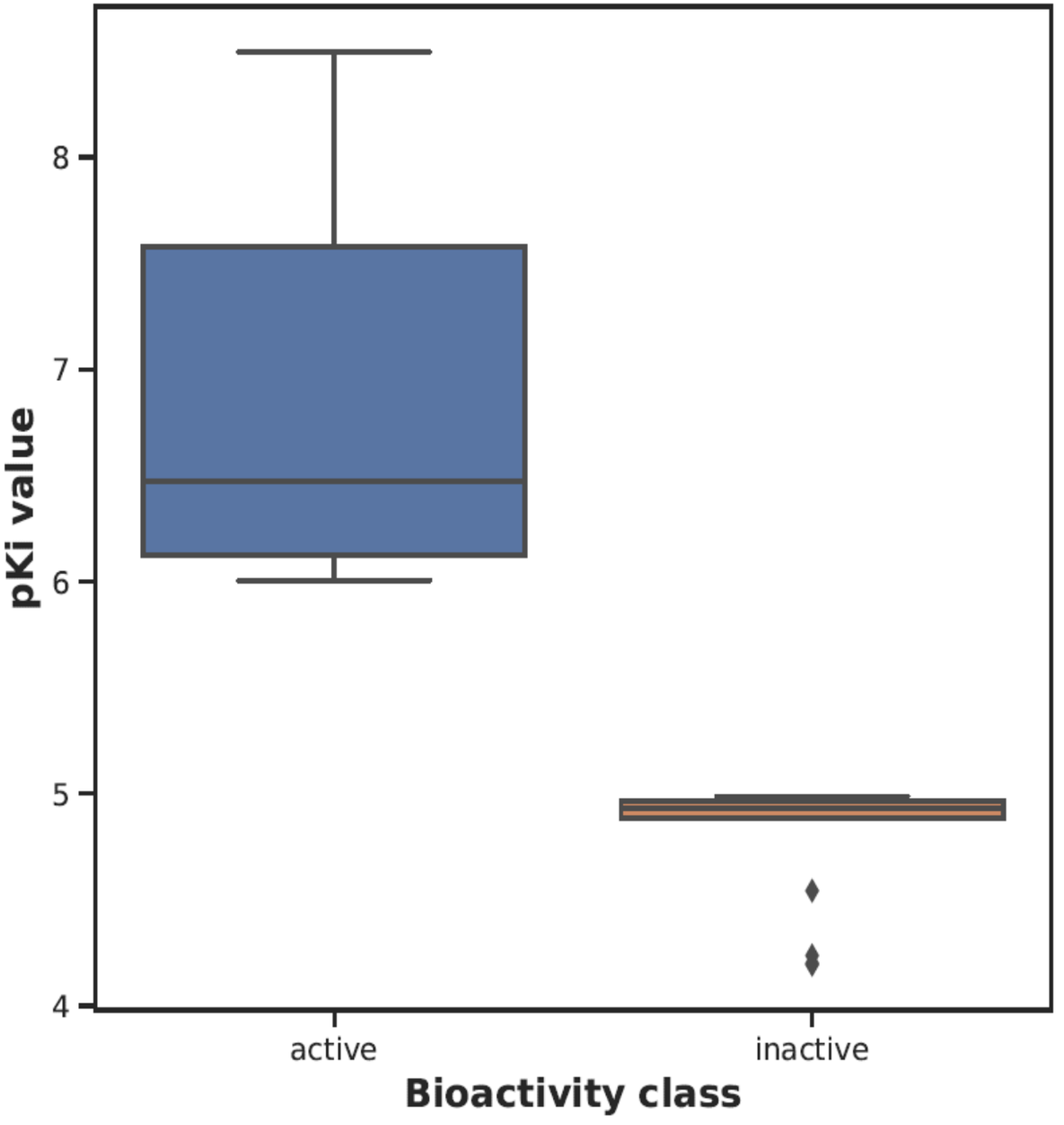
- **Principle Investigator:** Dr. Ashok Aspatwar
- **ML Model Generation and Optimization/Software Development/Chemical Data Curation/ Molecular Modeling Studies:** Ratul Bhowmik
- **Co-authors:** Ajay Manaithiya, Rajarshi Ray
- **Affiliation:** Tampere University, Finland

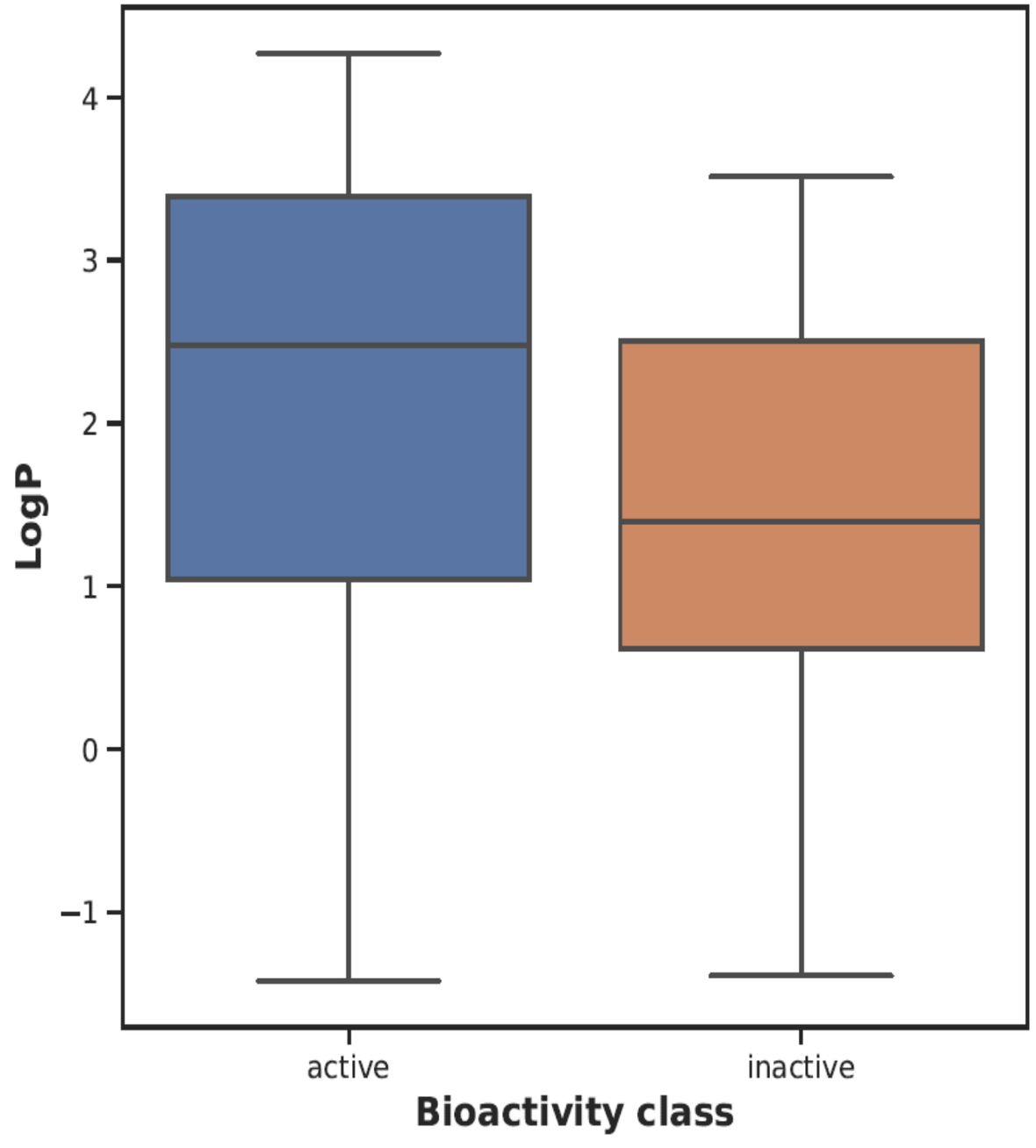
PRIMARY OBJECTIVES

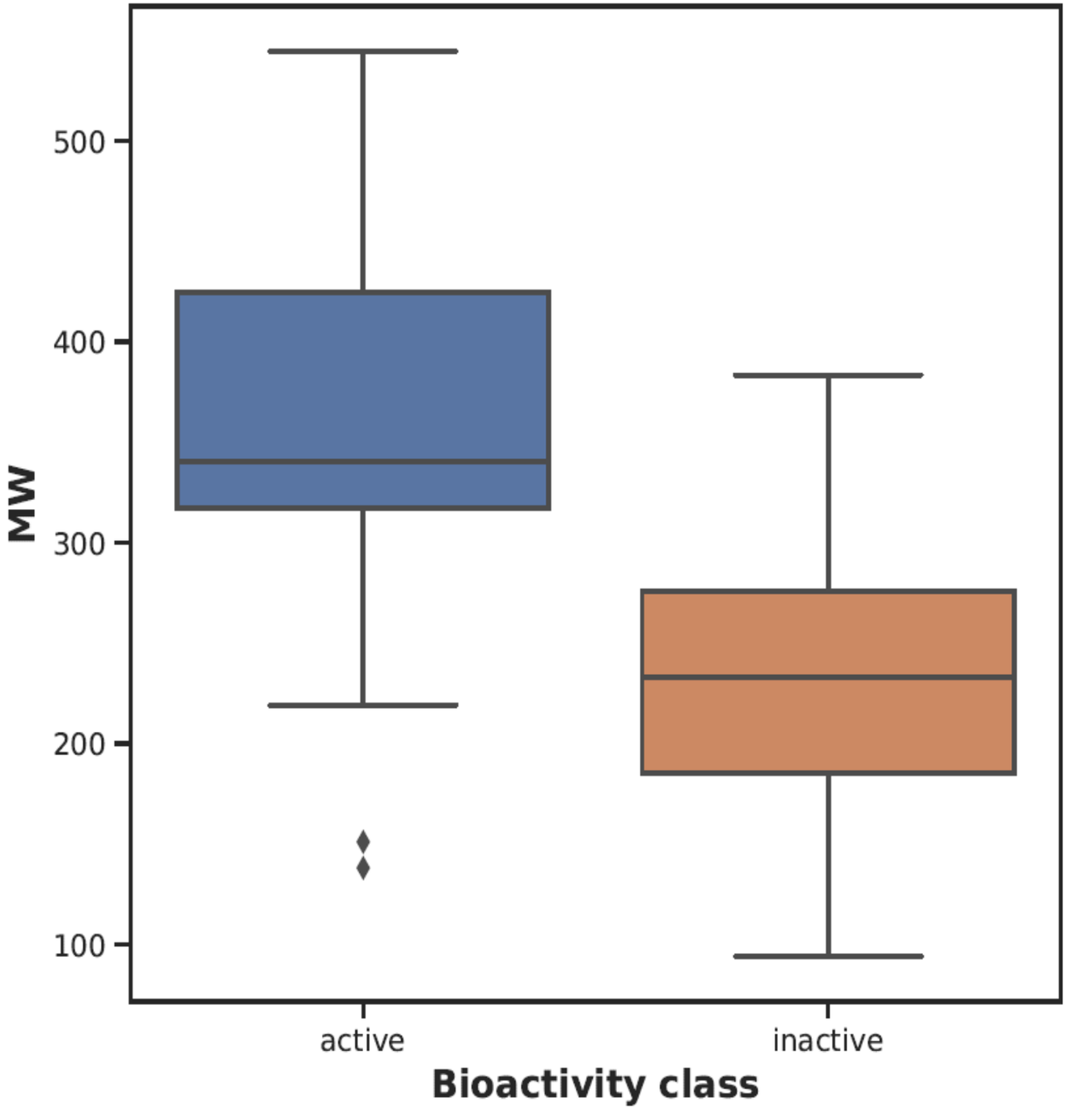
- Generation of selective robust multi-molecular signature implemented machine learning-assisted quantitative structural activity relationship **(ML-QSAR)** models for predicting bioactivity (**Ki**) against ***MtbCA1*** and ***MtbCA2***
- Structural interpretation of active and inactive molecules concerning ***MtbCA1*** and ***MtbCA2*** inhibition through variance importance plot, correlation matrix, and feature plot analysis
- Development of a web application using the generated ML-QSAR models
- Generation of neural network-assisted quantitative structural activity-activity relationship **(ML-QSAAR)** models to infer the structural-functional correlation between ***MtbCA1*** and ***MtbCA2*** inhibitor's binding mechanism
- Validation of key molecular signatures extracted from **ML-QSAR** and **ML-QSAAR** analysis through molecular docking approach

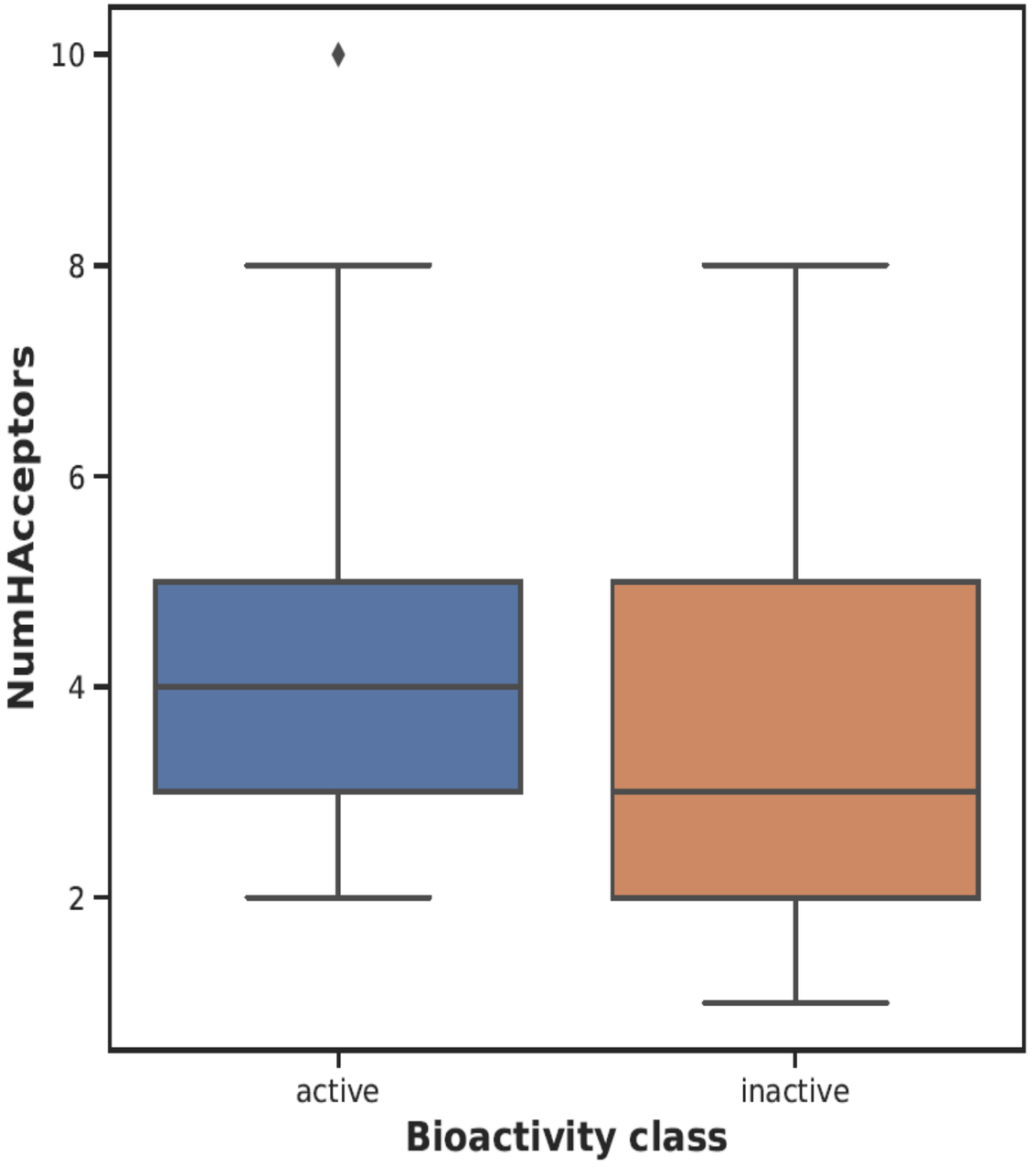
MtbCA1 Exploratory Data Analysis

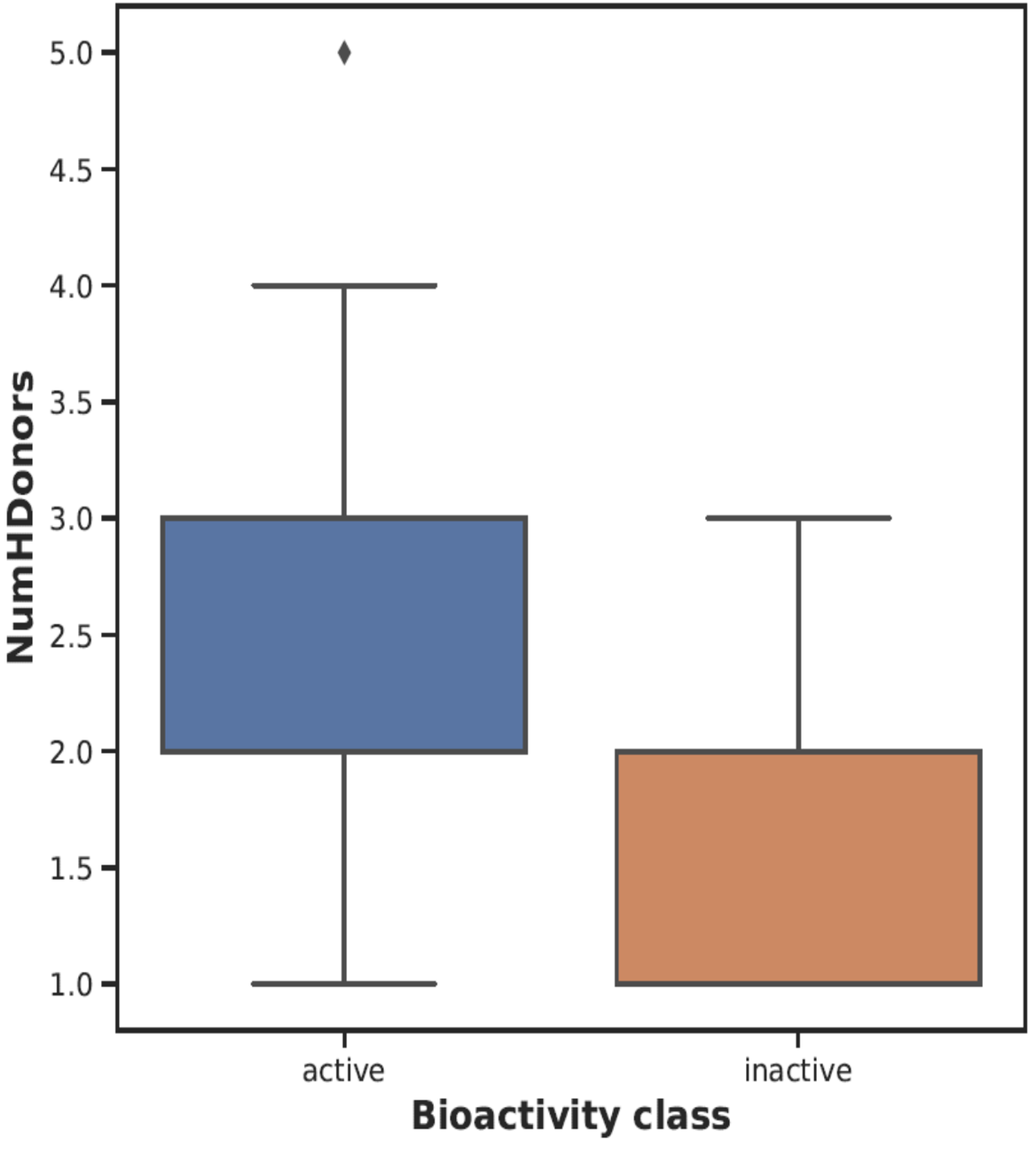

Frequency
60
50
40
30
20
10
0
active
inactive
Bioactivity class


pKi value
8
7
6
5
4
active
inactive
Bioactivity class

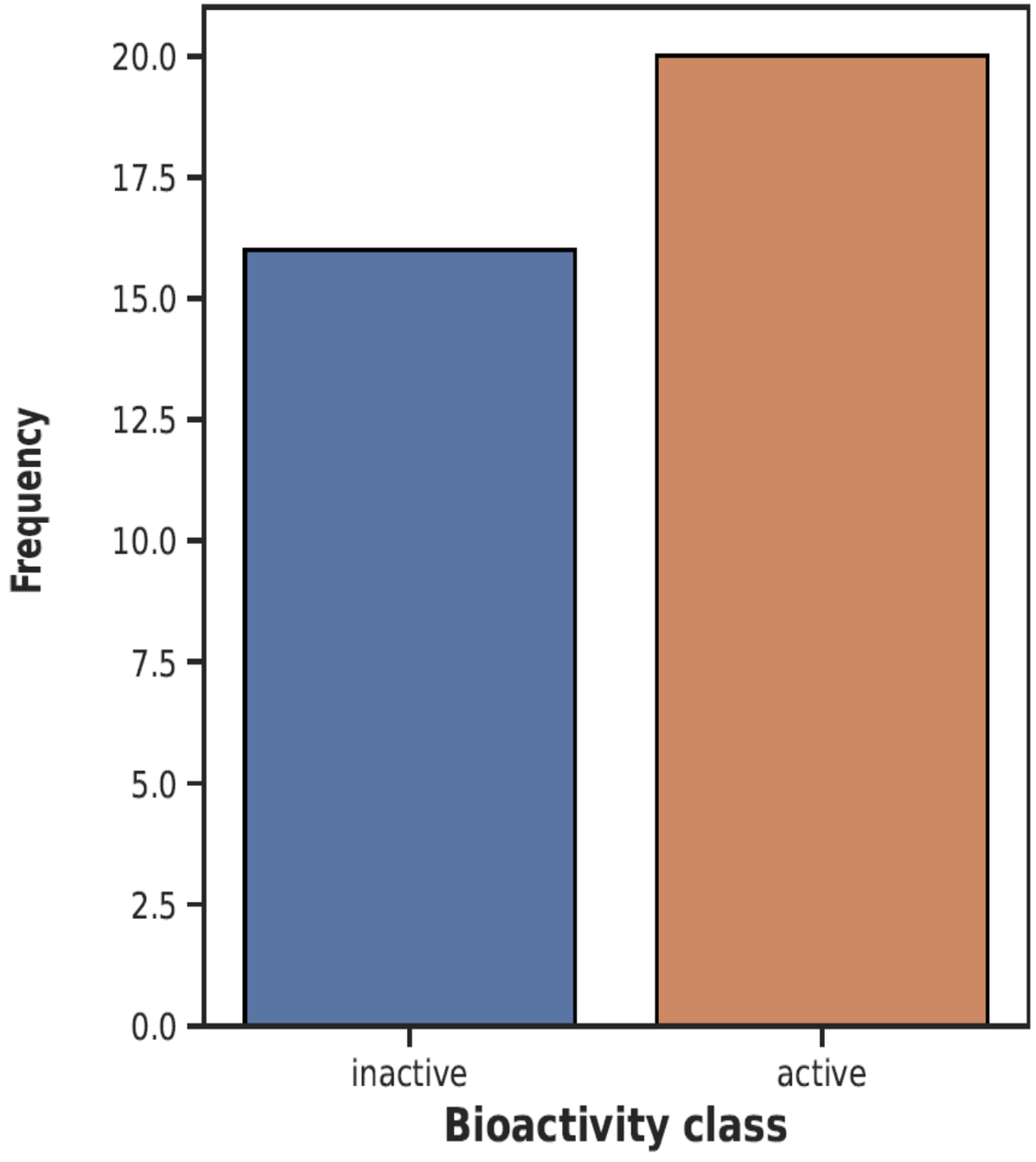

LogP
4
3
2
1
0
−1
active
inactive
Bioactivity class

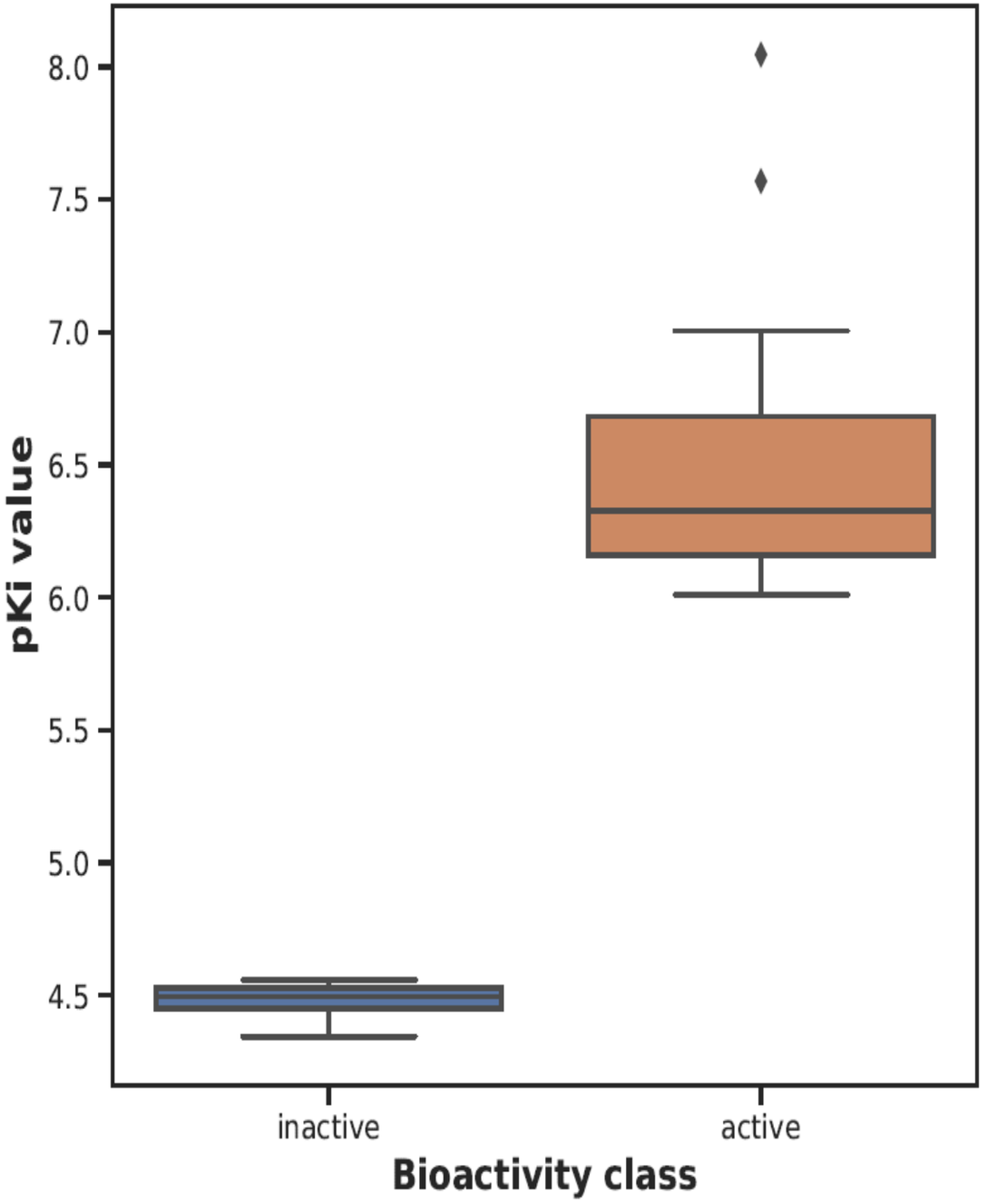

MW
500
400
300
200
100
active
inactive
Bioactivity class

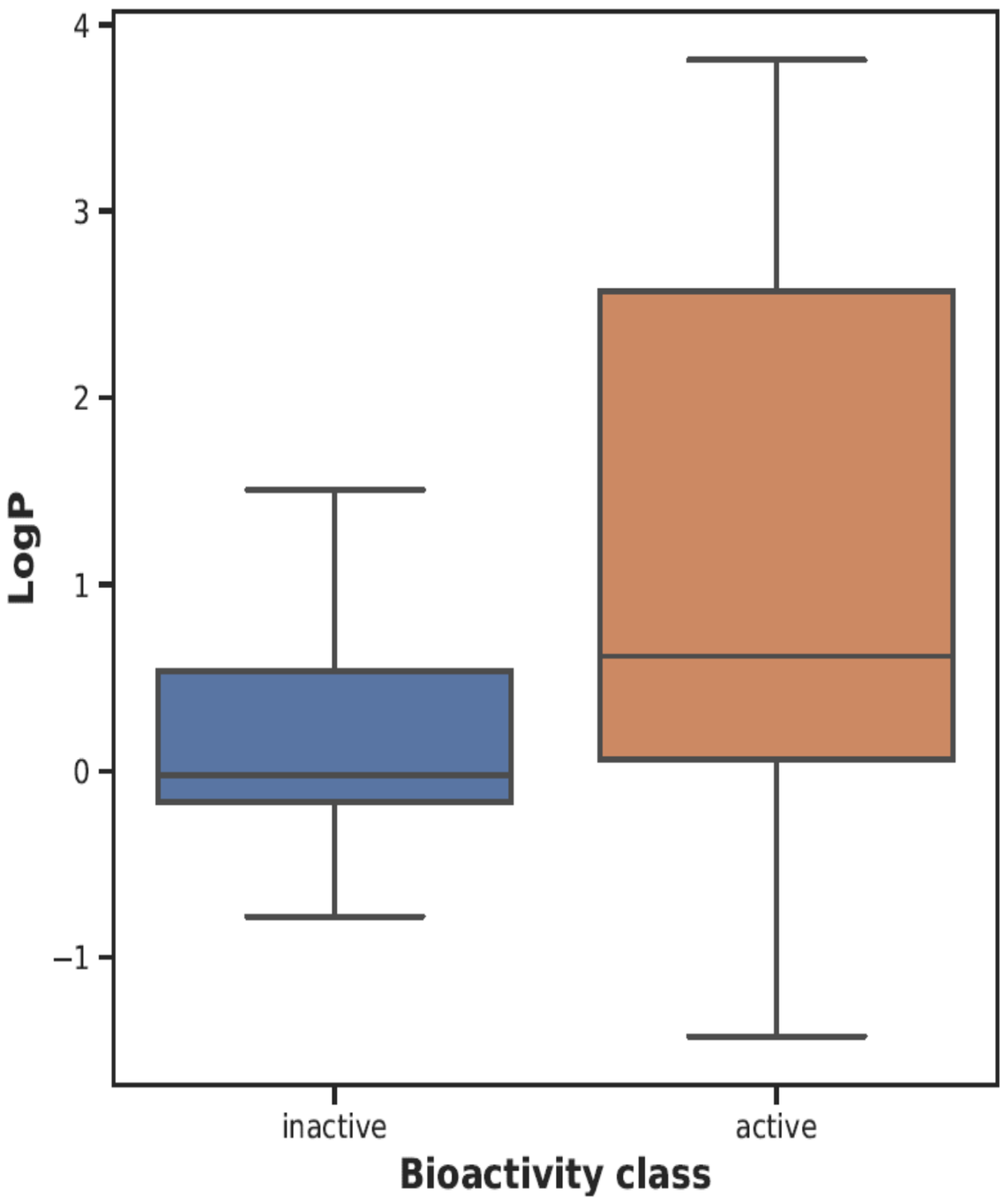

10
8
6
4
2
NumHAcceptors
active
inactive
Bioactivity class

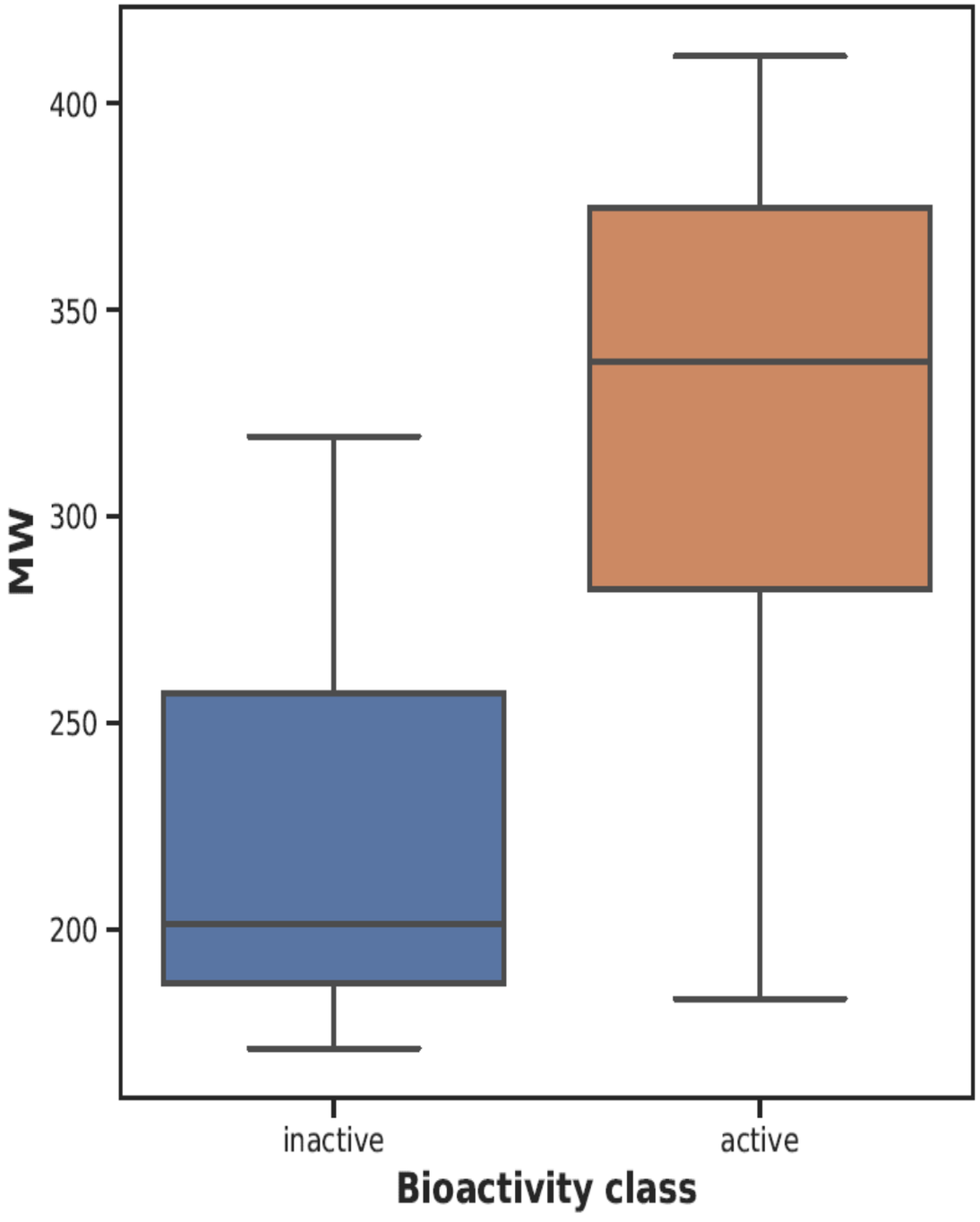

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
NumHDonors
active
inactive
Bioactivity class

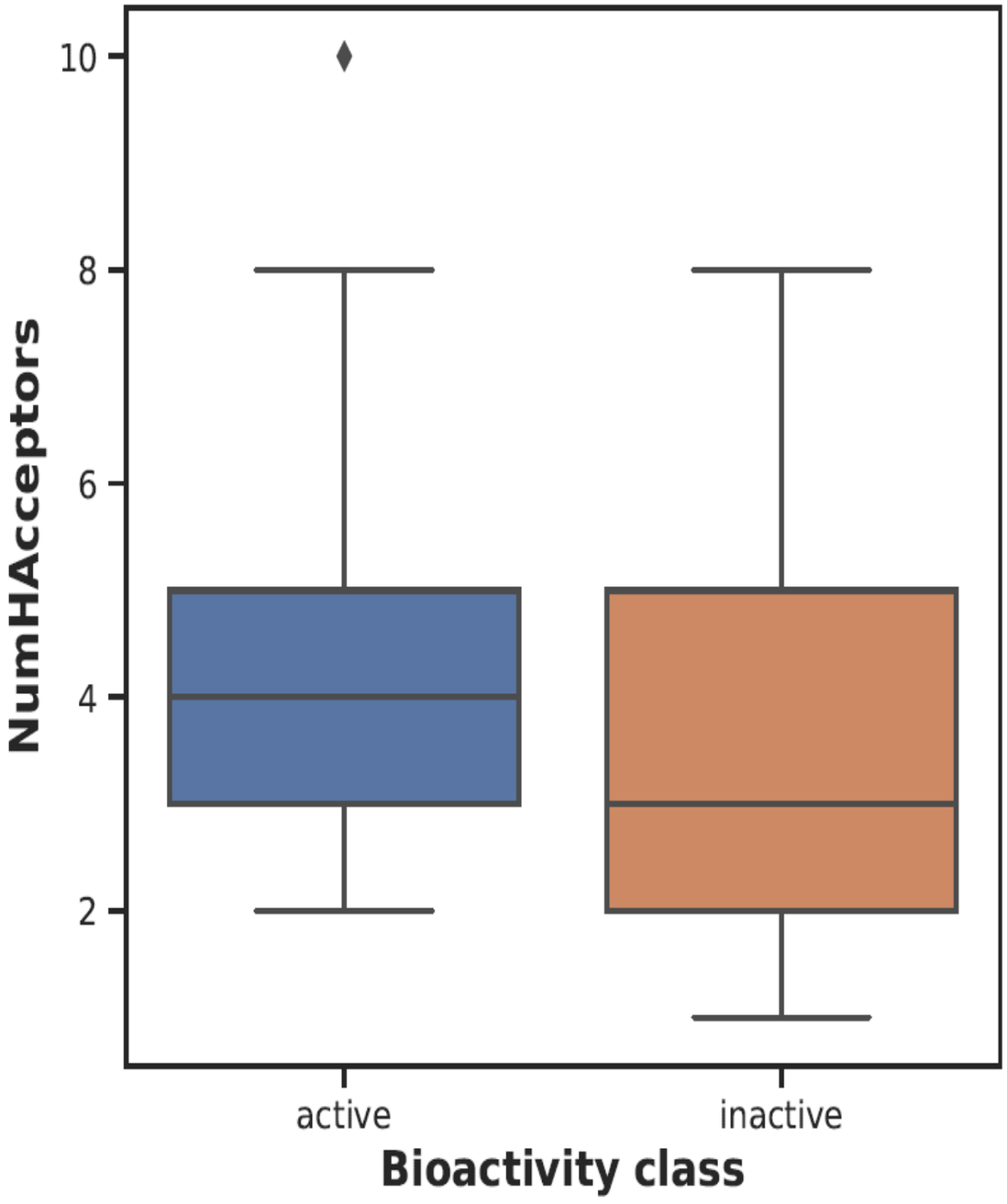
MtbCA2 Exploratory Data Analysis

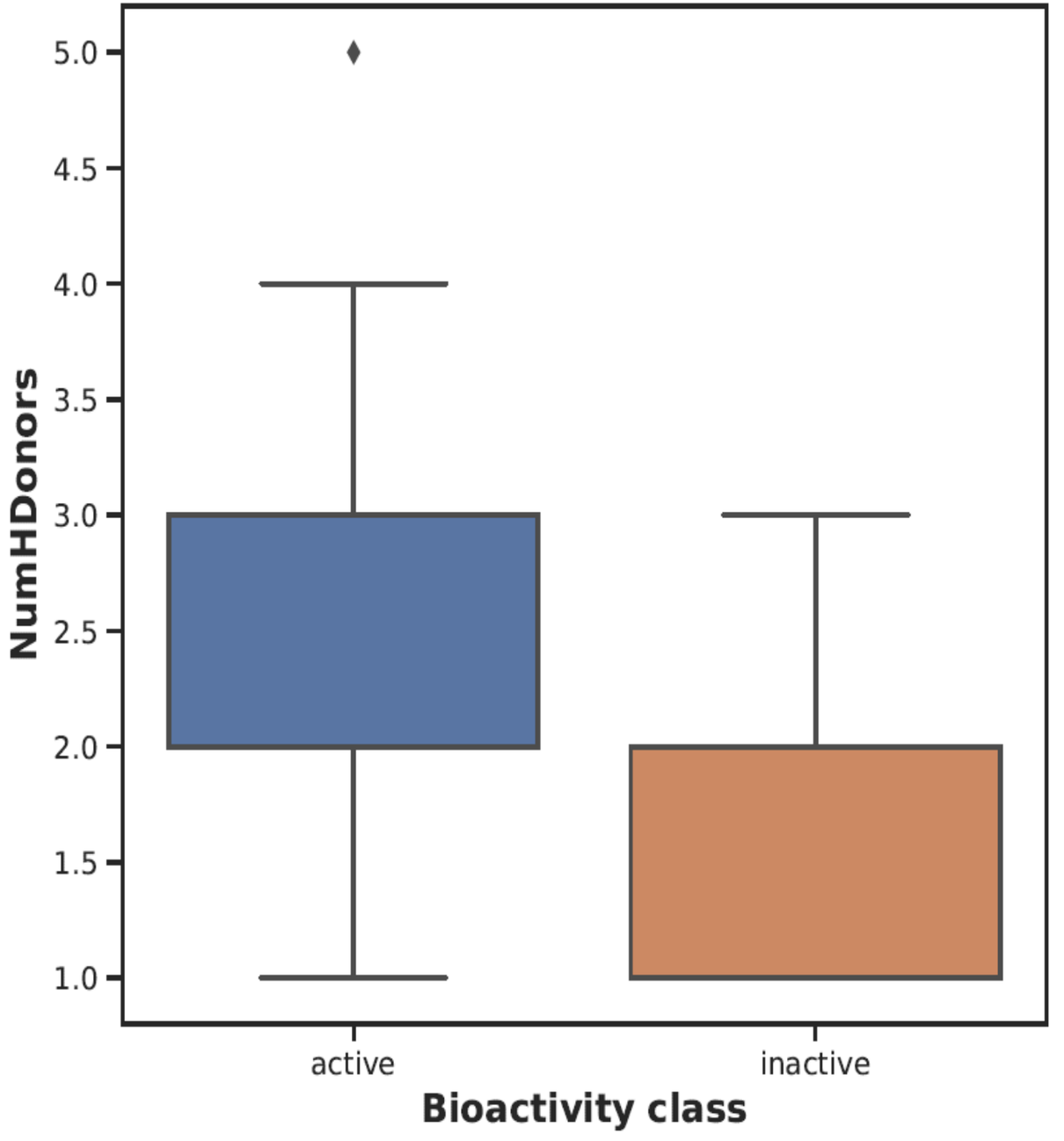

Frequency
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
inactive
active
Bioactivity class


pKi value
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
inactive
active
Bioactivity class


4
3
2
1
0
−1
LogP
inactive
active
Bioactivity class

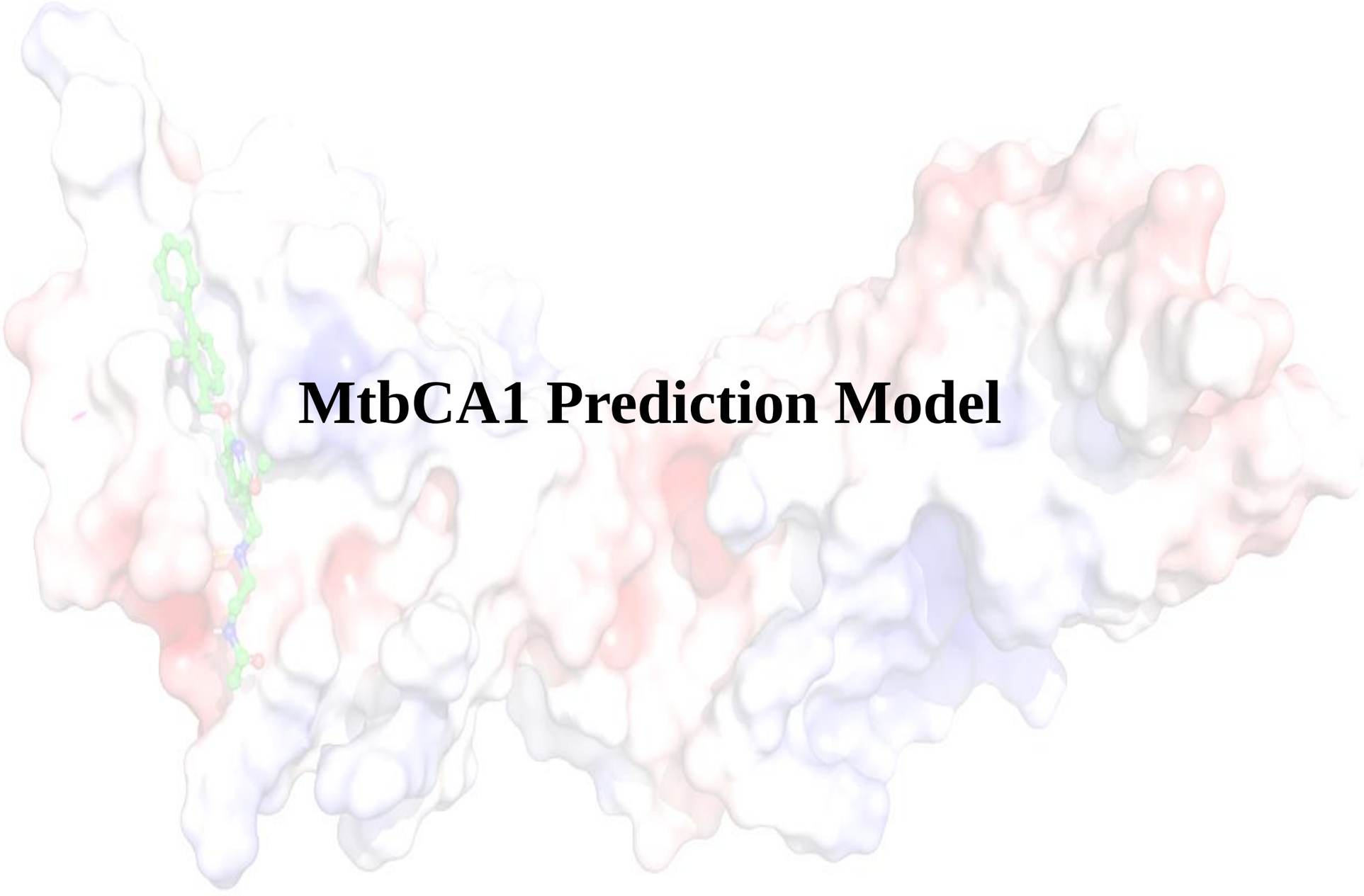

400
350
300
250
200
MW
inactive
active
Bioactivity class


10
8
6
4
2
NumHAcceptors
active
inactive
Bioactivity class


5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
NumHDonors
active
inactive
Bioactivity class

MACHINE LEARNING ASSISTED QUANTITATIVE STRUCTURAL ACTIVITY RELATIONSHIP

MtbCA1 Prediction Model



ML-QSAR model for bioactivity prediction of MtbCA1 inhibitors using substructure fingerprints

ChEMBL

Curation of *MtbCA*1 inhibitors with Ki value

Total inhibitors: 124

Substructure fingerprints: 307

Prediction Parameter: Bioactivity Class ('pKi')

Dataset Splitting into training and test set

Elimination of highly correlated and constant fingerprints

Final number of fingerprints(18), number of molecules in training(99) and test(25)

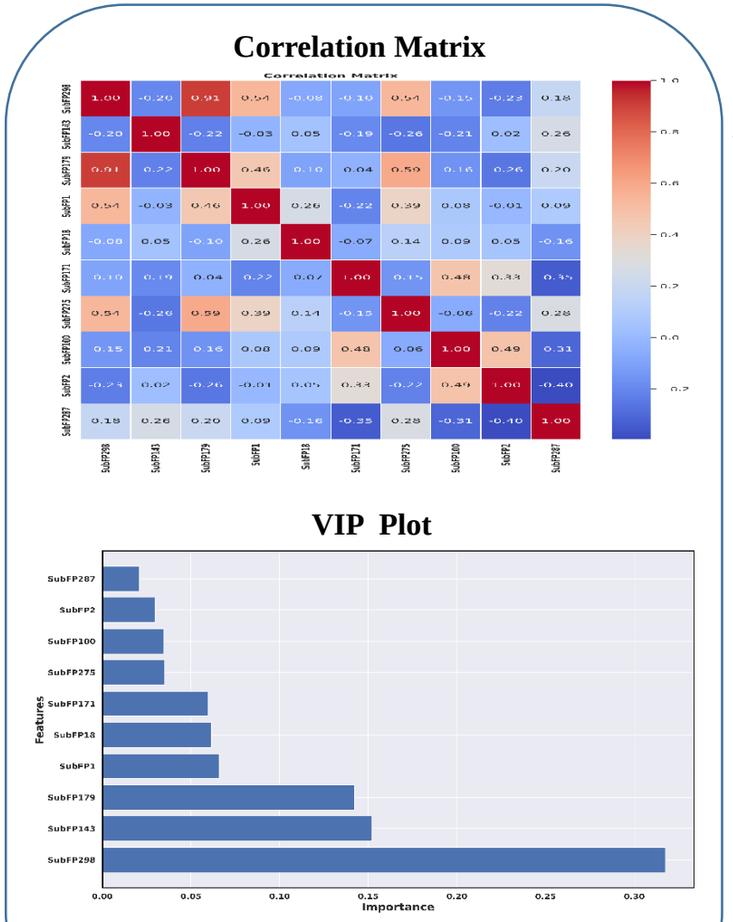
Random Forest Regressor

Initial Model Performance

Model Metrics
R2 (Train: 0.8853, Test: 0.7371)
RMSE (Train: 0.5084, Test: 0.8106)
MAE (Train: 0.3549, Test: 0.7171)

FP and FN Removal

Mention molecules removed (21)



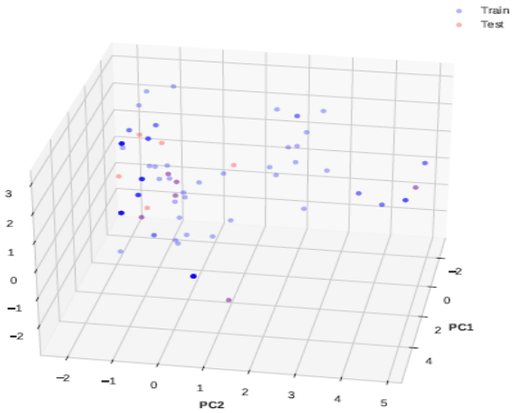
Validation Metrics

Model Metrics
R2 (Train: 0.955, Test: 0.9406)
RMSE (Train: 0.3073, Test: 0.5725)
MAE (Train: 0.2367, Test: 0.4899)

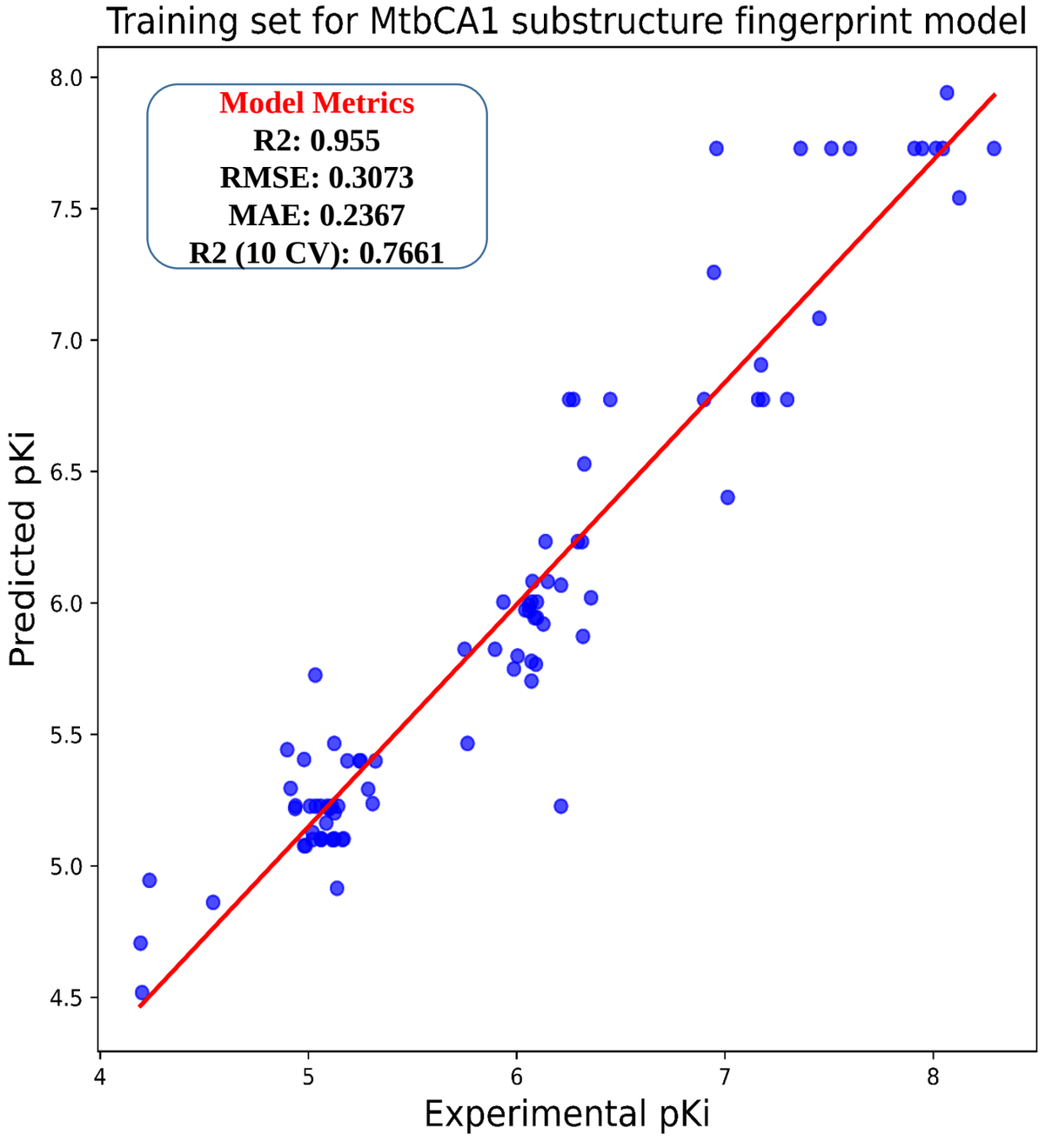
Final Model Performance

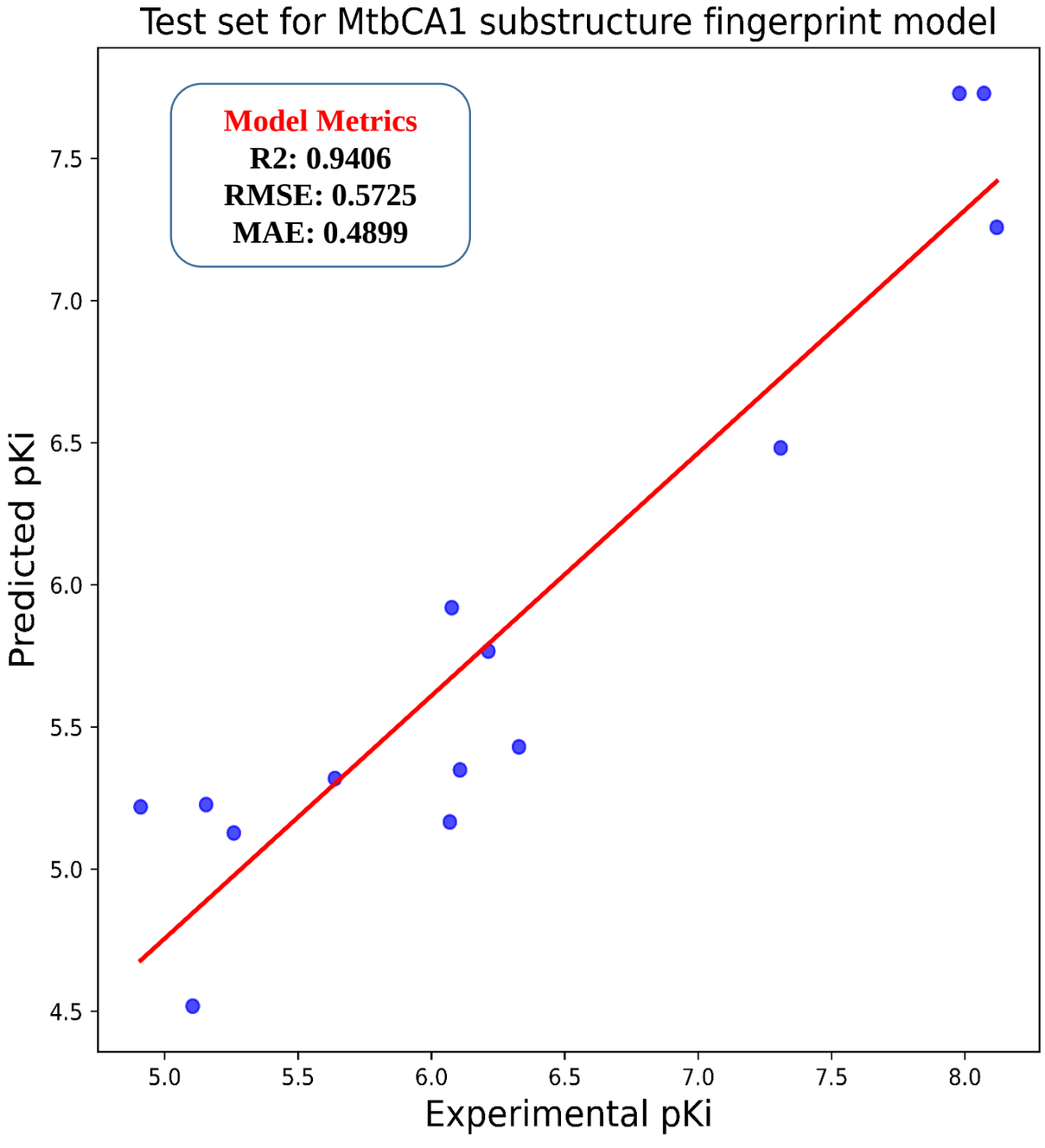
The final machine-learning model

number of molecules in training(89) and test(14)

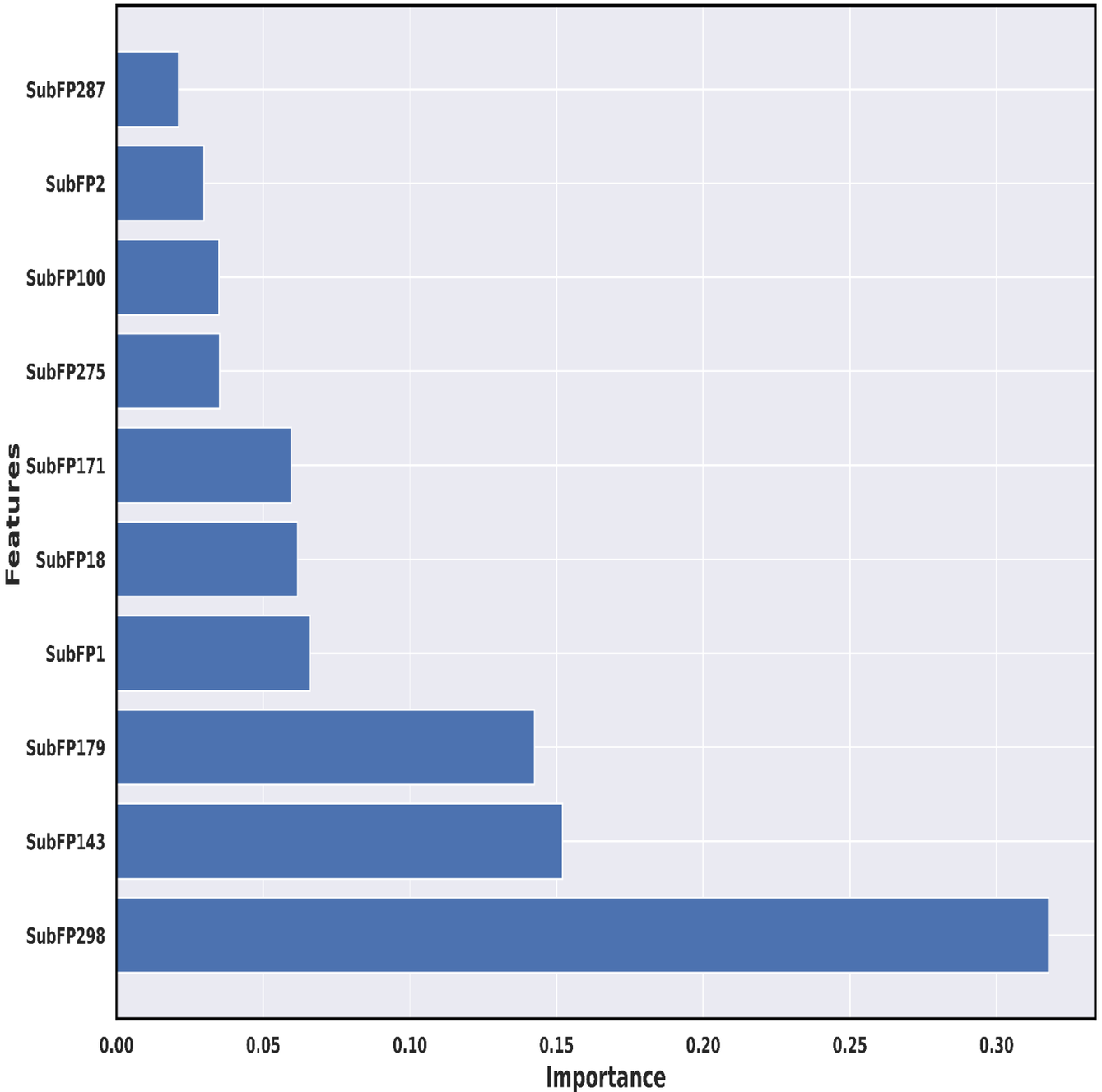


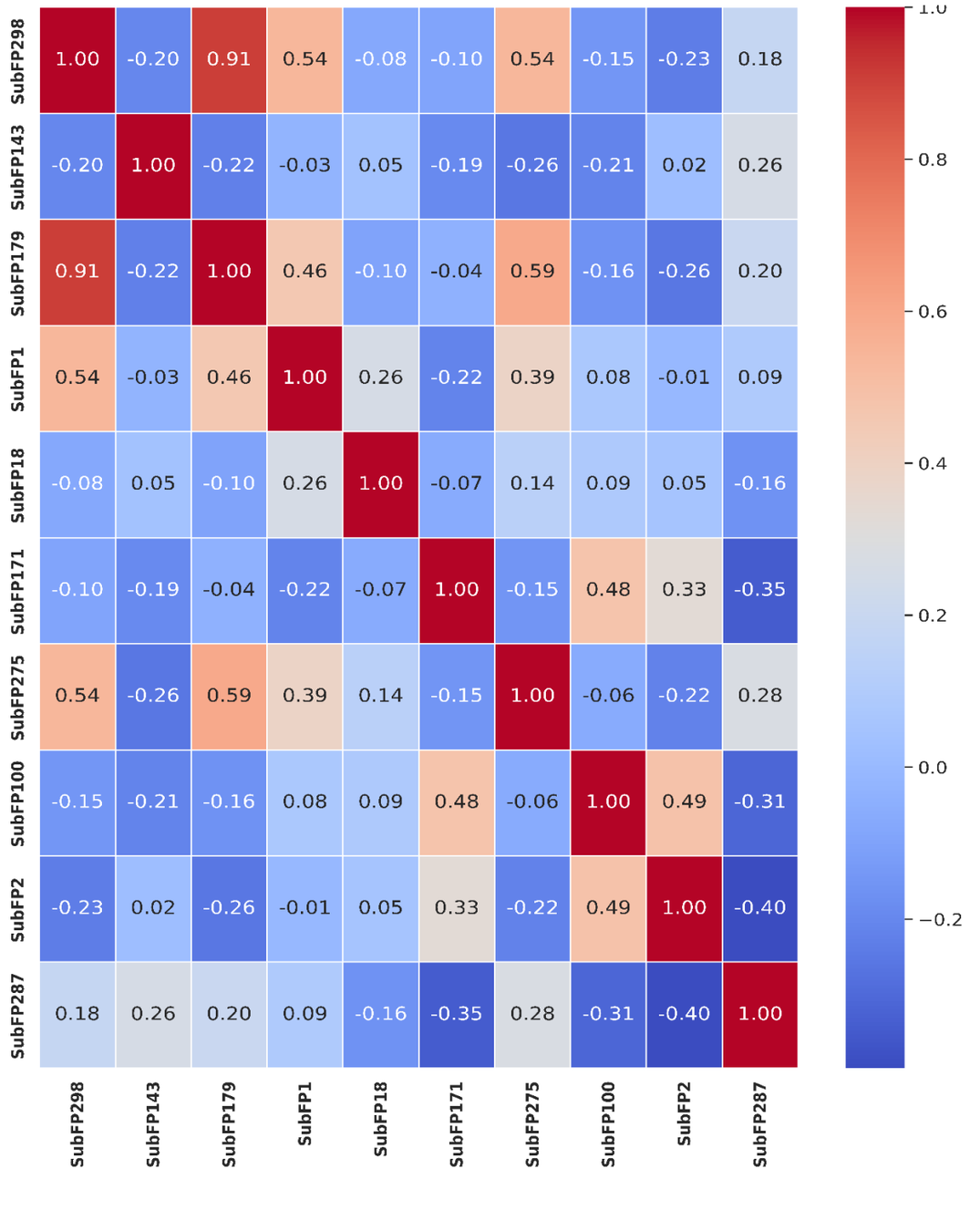
Applicability Domain Analysis

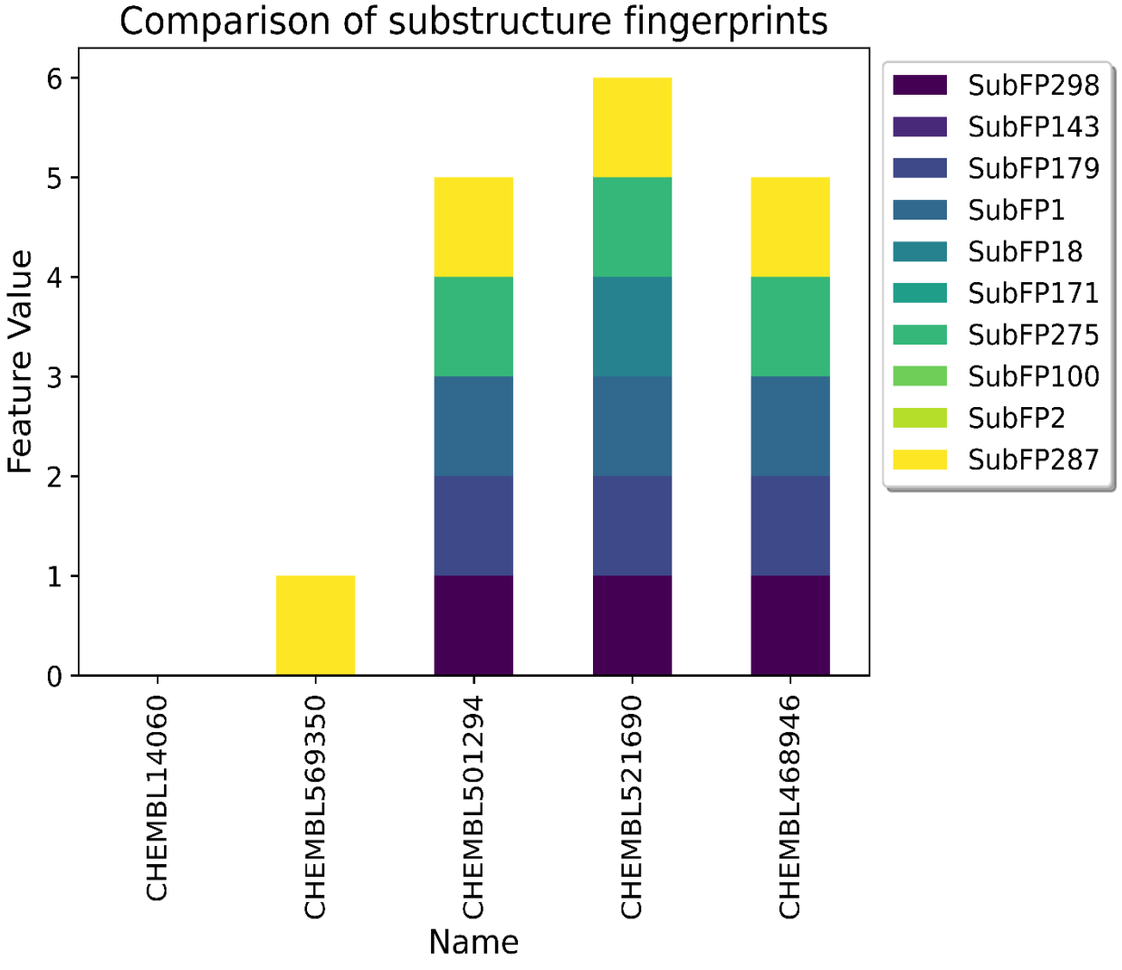

Training set for MtbCA1 substructure fingerprint model
Model Metrics
R2: 0.955
RMSE: 0.3073
MAE: 0.2367
R2 (10 CV): 0.7661
Predicted pKi
Experimental pKi
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4
5
6
7
8


Test set for MtbCA1 substructure fingerprint model
Model Metrics
R2: 0.9406
RMSE: 0.5725
MAE: 0.4899
Predicted pKi
Experimental pKi
7.5
7.0
6.5
6.0
5.5
5.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0

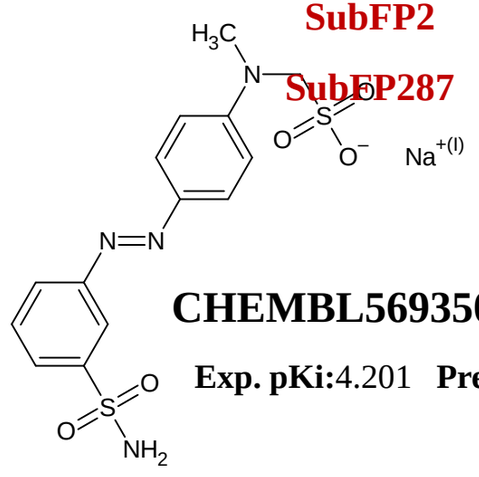
VIP plot and correlation matrix analysis for *MtbCA1* substructure prediction model





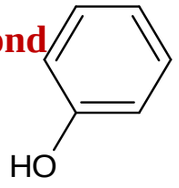


SubFP298	**Cation**
SubFP143	**Carbonic acid derivatives**
SubFP179	**Hetero N basic H**
SubFP1	**Primary carbon**
SubFP18	**Alkylarylether**
SubFP171	**Arylchloride**
SubFP275	**Heterocyclic**
SubFP100	**Secondary amide**
SubFP2	**Secondary carbon**
SubFP287	**Conjugated double bond**



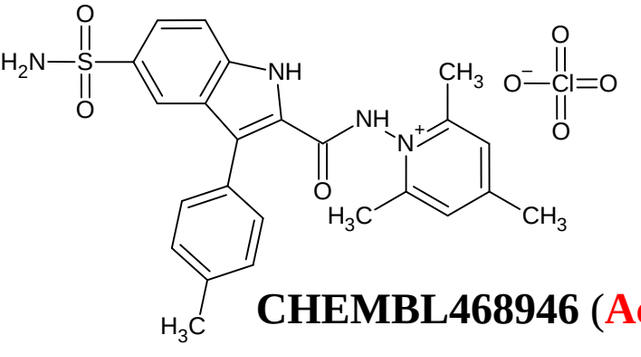
CHEMBL569350 (Inactive)

Exp. pKi:4.201 **Pred. pKi:**4.518



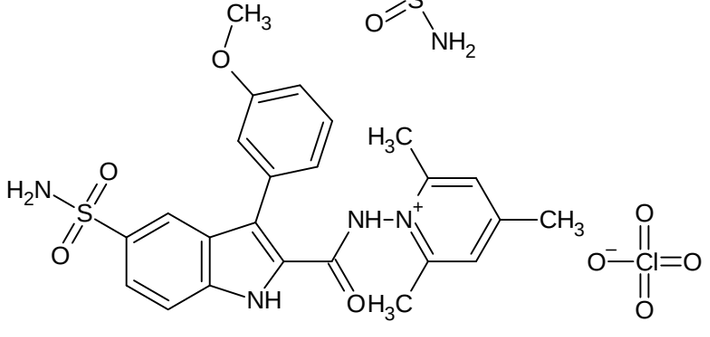
CHEMBL14060 (Inactive)

Exp. pKi:4.194 **Pred. pKi:**4.706



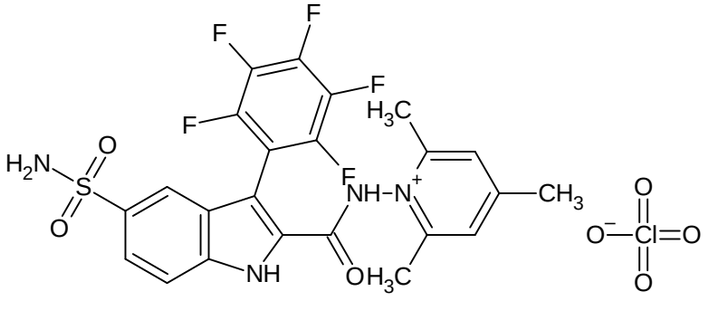
CHEMBL468946 (Active)

Exp. pKi:8.292 **Pred. pKi:**7.729



CHEMBL521690 (Active)

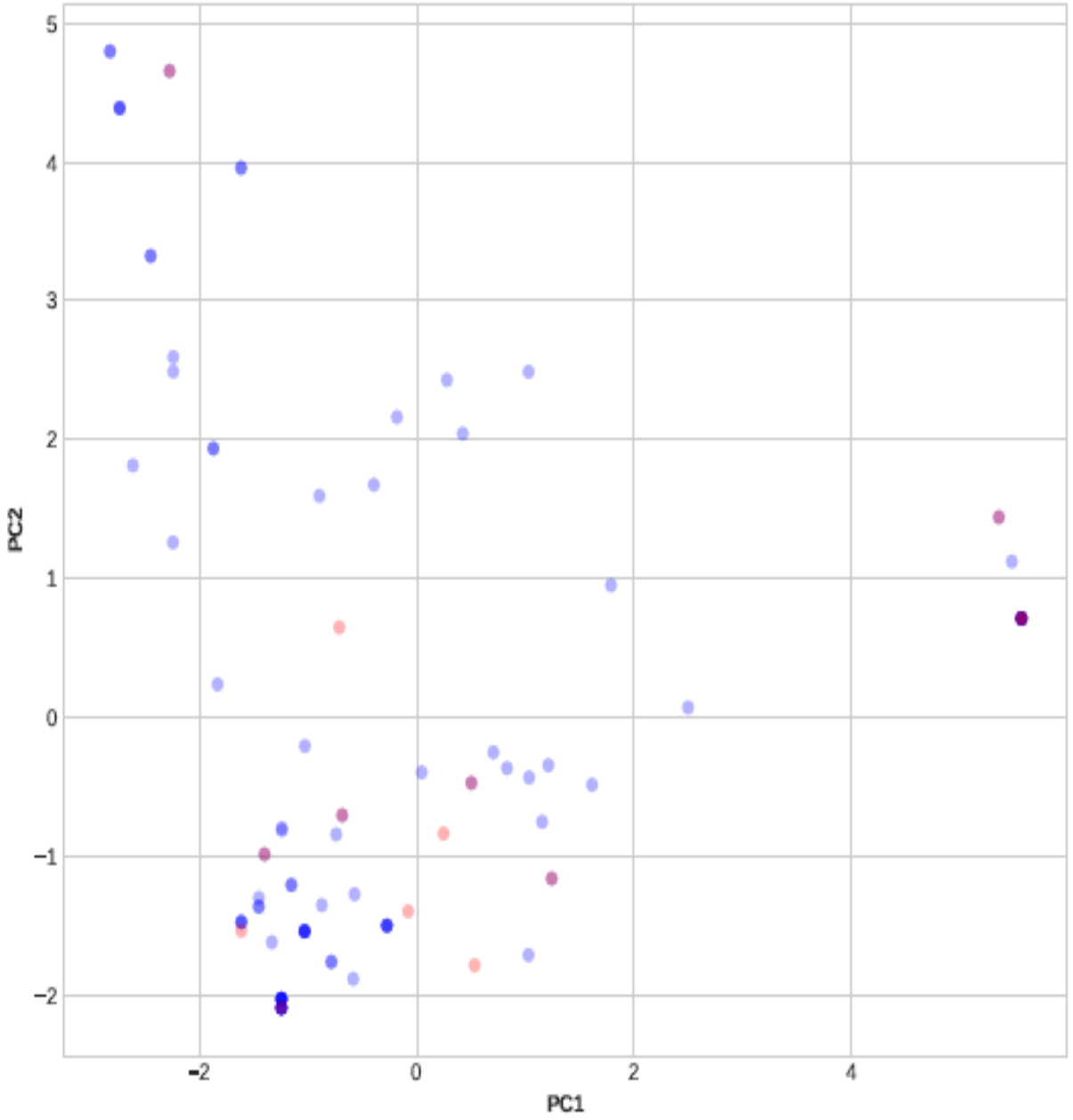
Exp. pKi:8.066 **Pred. pKi:**7.941

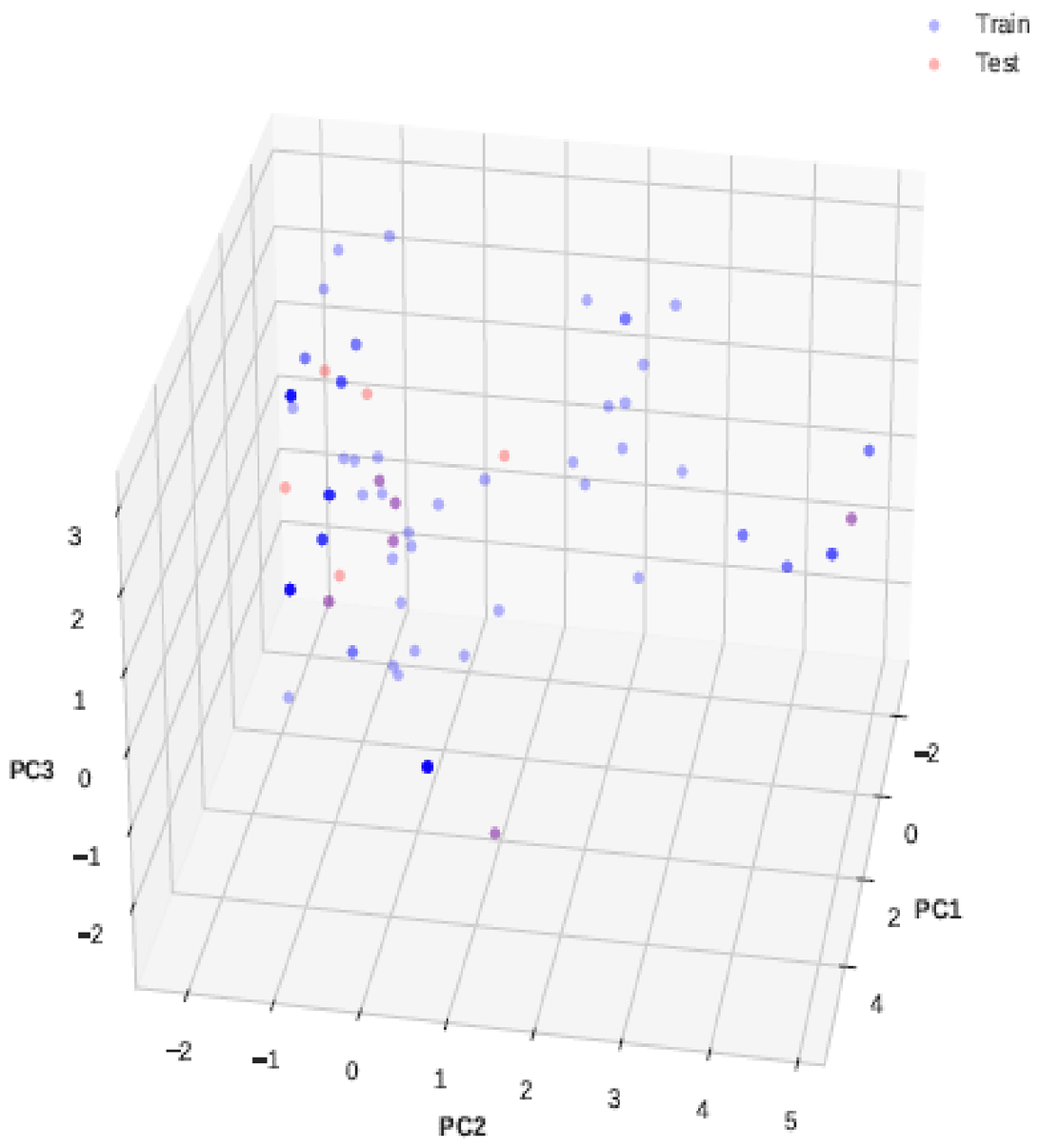


CHEMBL501294 (Active)

Exp. pKi:8.013 **Pred. pKi:**7.729

Applicability domain analysis through 2D and 3D PCA plots for *MtbCA1* substructure prediction model





ML-QSAR model for bioactivity prediction of MtbCA1 inhibitors using 1D and 2D molecular descriptors

ChEMBL

Curation of *MtbCA*1 inhibitors with Ki value

Total inhibitors: 124

1D and 2D Molecular Descriptors: 1444

Prediction Parameter: Bioactivity Class ('pKi')

Dataset Splitting into training and test set

Elimination of highly correlated and constant fingerprints

Final number of descriptors(763), number of molecules in training(99) and test(25)

Random Forest Regressor

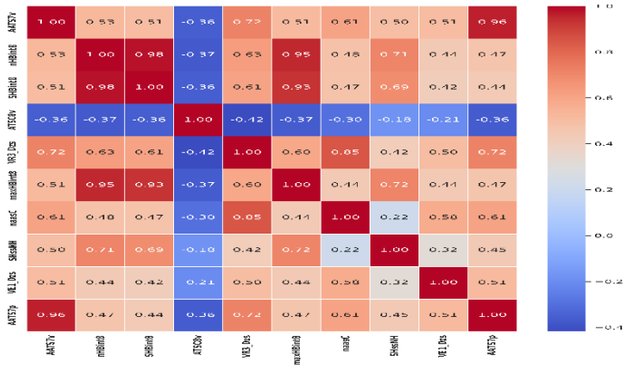
Initial Model Performance

Model Metrics
R2 (Train: 0.9769, Test: 0.6047)
RMSE (Train: 0.2793, Test: 0.9688)
MAE (Train: 0.2134, Test: 0.7608)

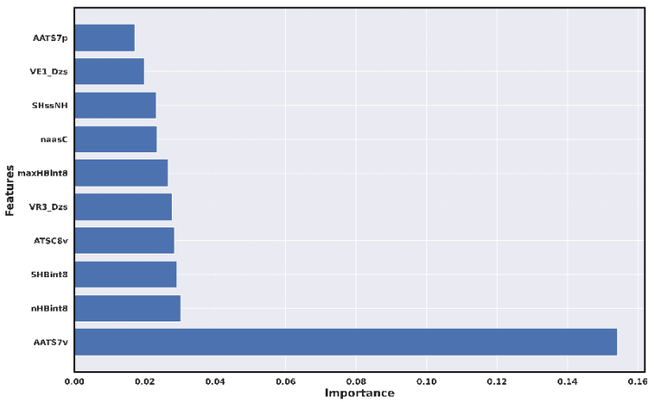
FP and FN Removal

Mention molecules removed (9)

Correlation Matrix



VIP Plot



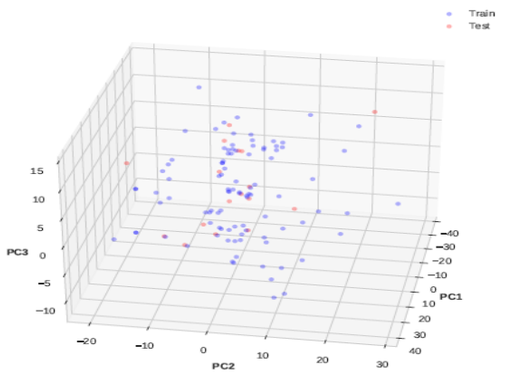
Validation Metrics

Model Metrics
R2 (Train: 0.9787, Test: 0.9333)
RMSE (Train: 0.2551, Test: 0.4083)
MAE (Train: 0.2042, Test: 0.4899)

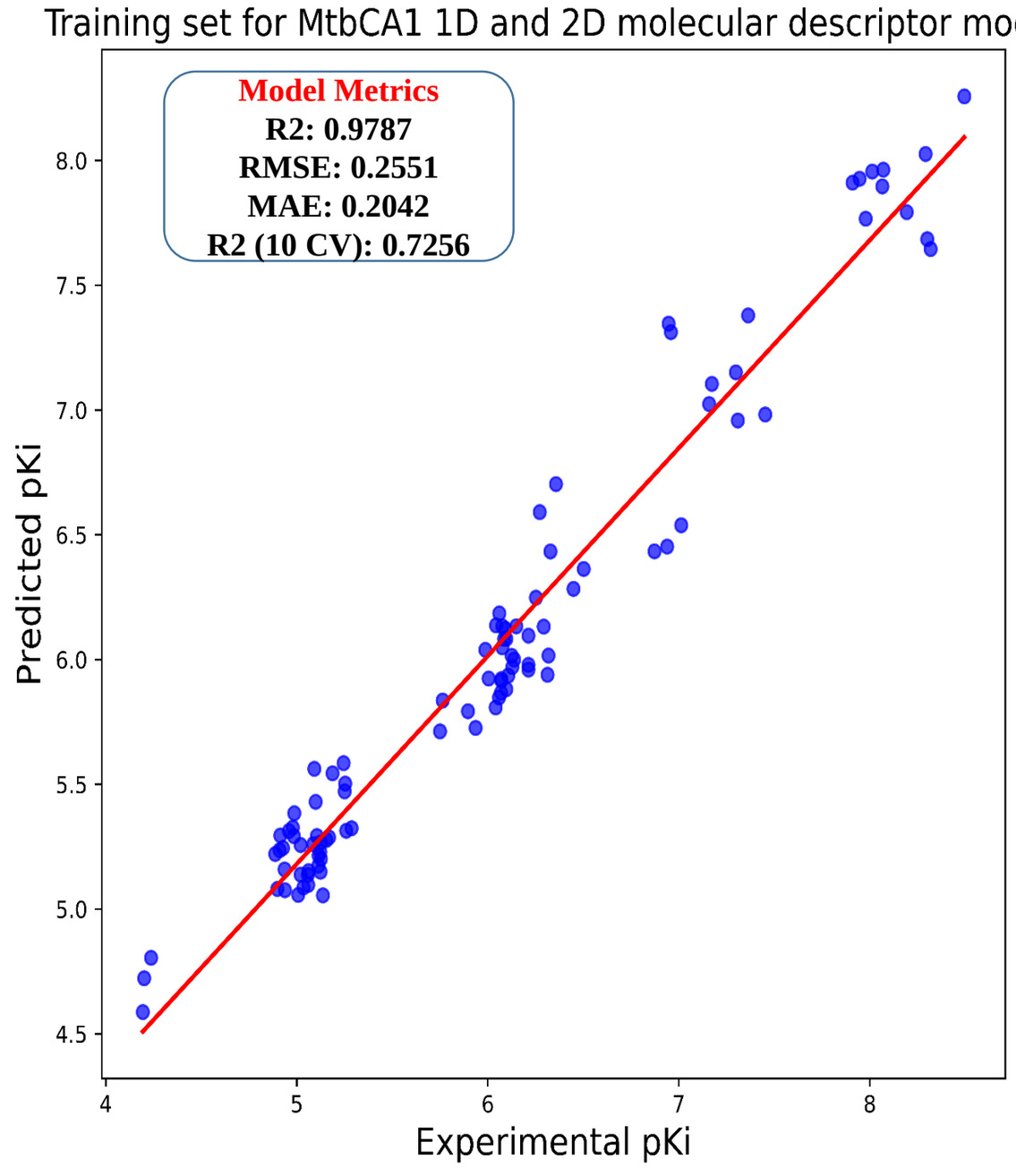
Final Model Performance

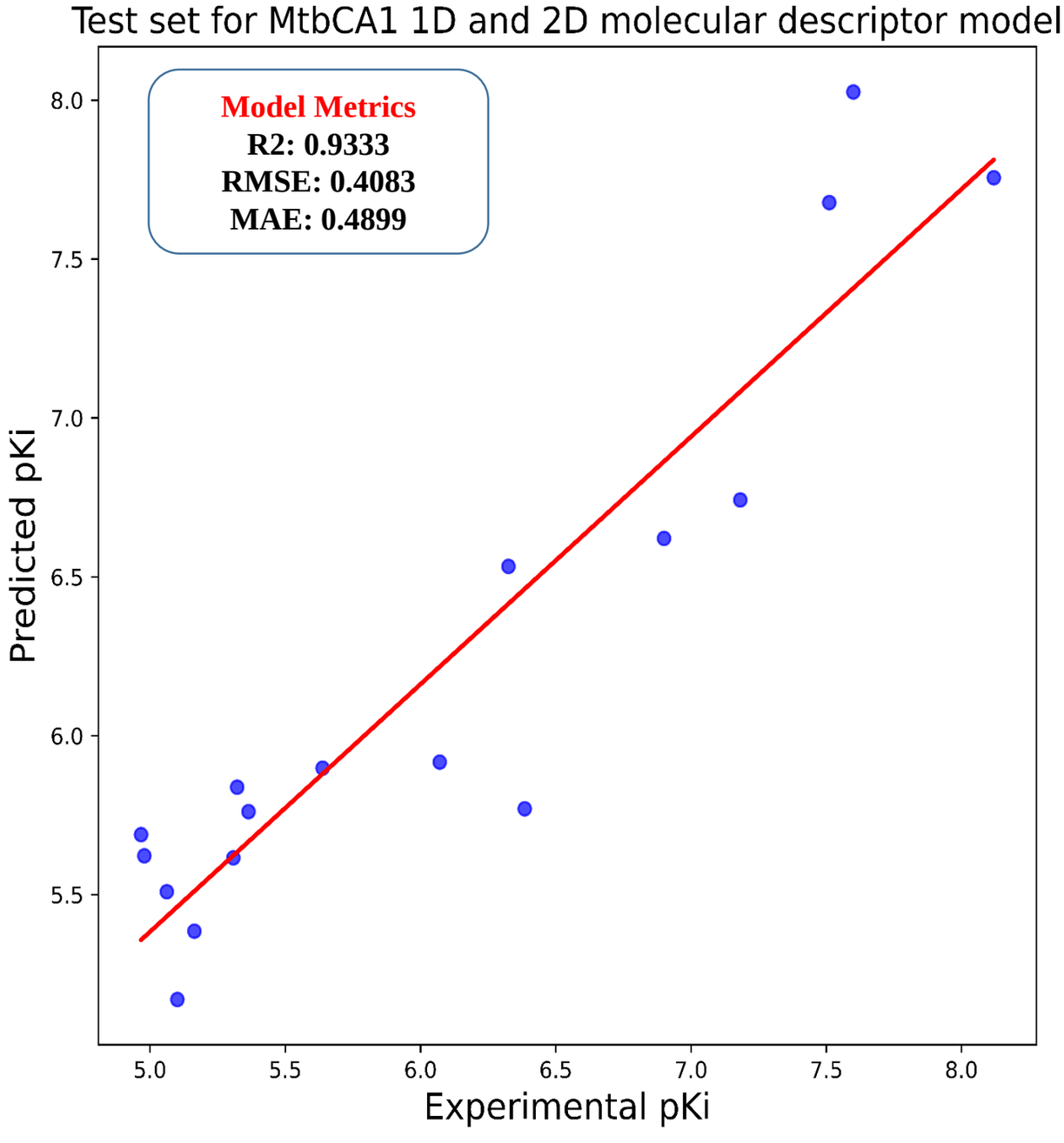
The final machine-learning model

number of molecules in training(98) and test(17)

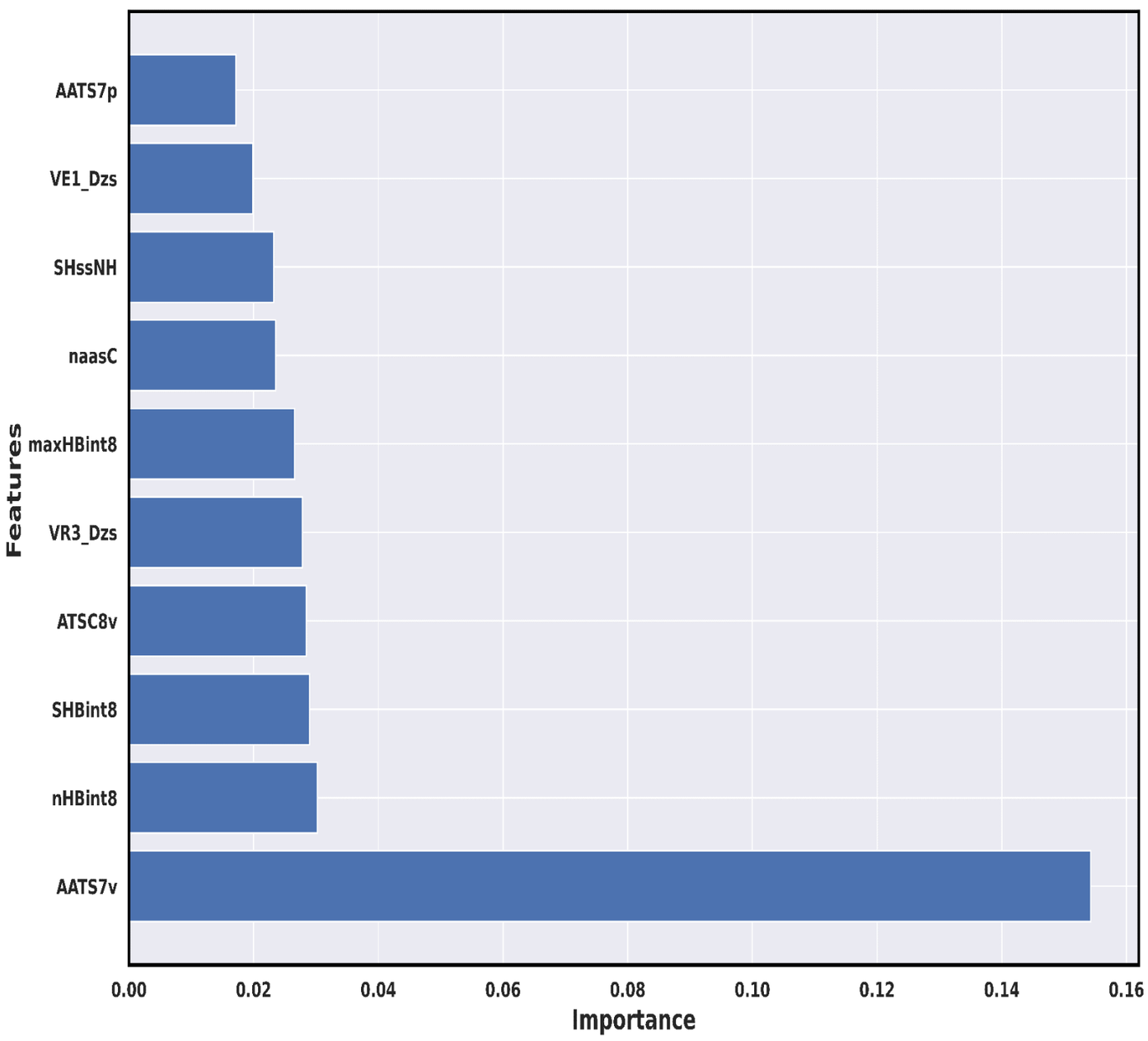


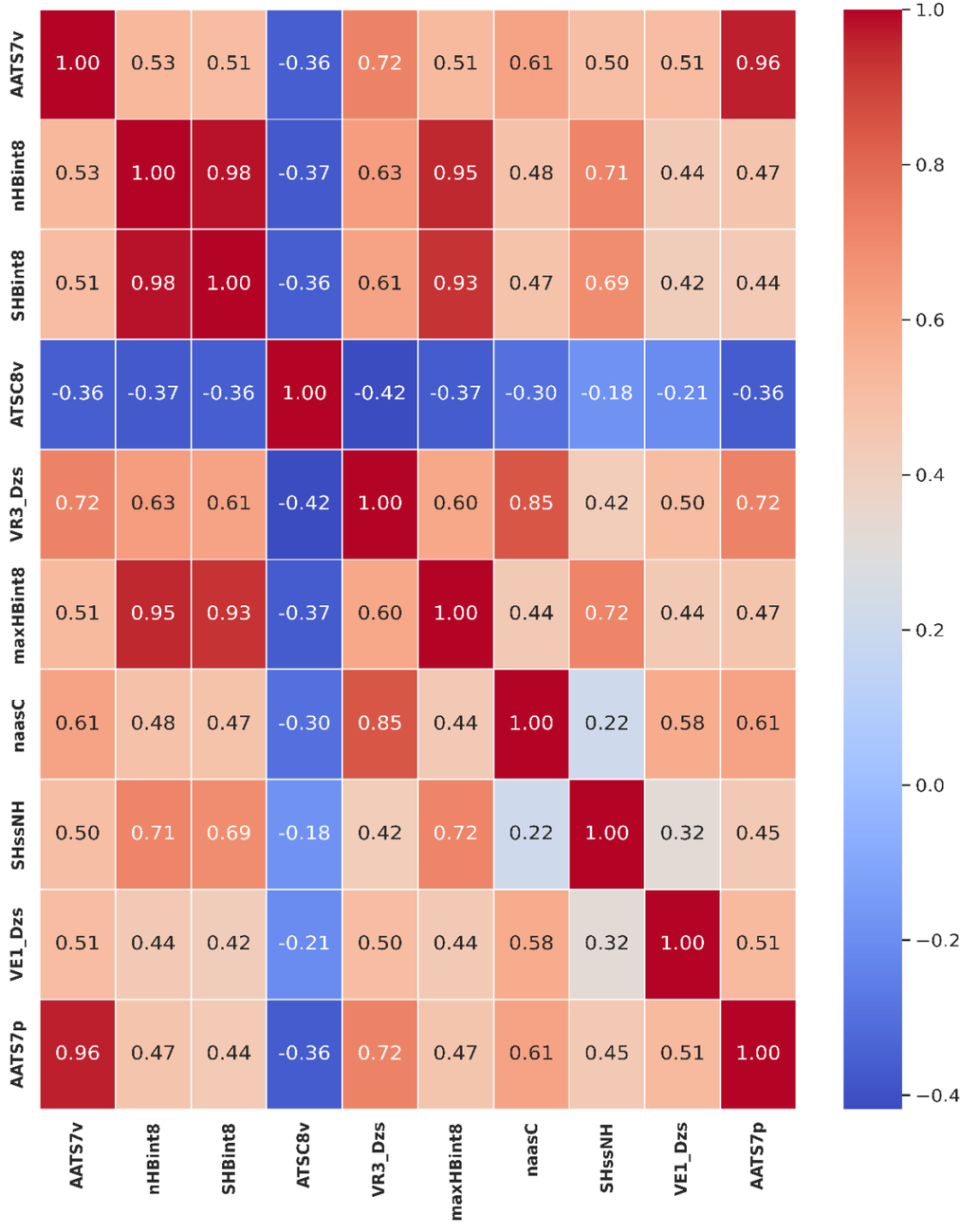
Applicability Domain Analysis

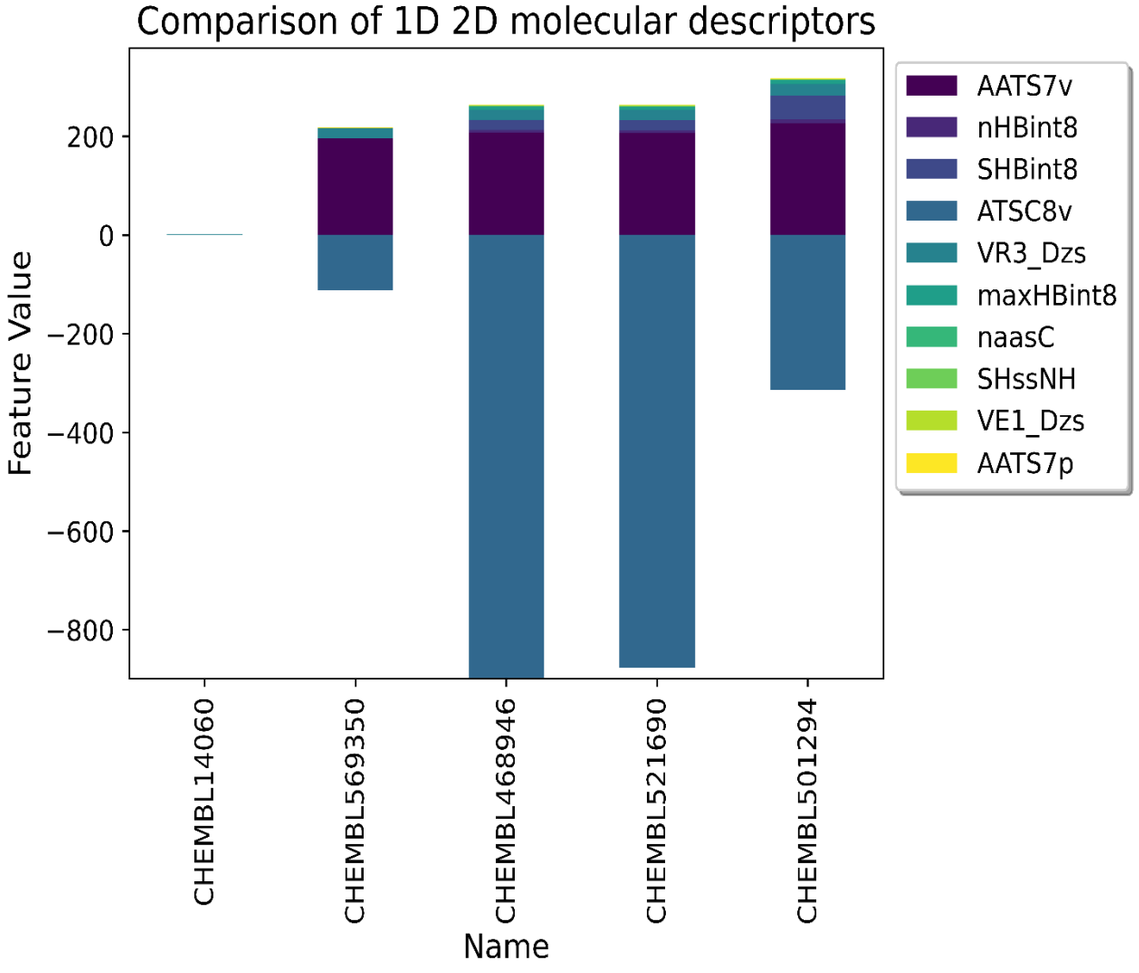

Training set for MtbCA1 1D and 2D molecular descriptor model
Model Metrics
R2: 0.9787
RMSE: 0.2551
MAE: 0.2042
R2 (10 CV): 0.7256
Predicted pKi
Experimental pKi
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
4
5
6
7
8


Test set for MtbCA1 1D and 2D molecular descriptor model
Model Metrics
R2: 0.9333
RMSE: 0.4083
MAE: 0.4899
Predicted pKi
Experimental pKi
5.5
6.0
6.5
7.0
7.5
8.0
5.0
5.5
6.0
6.5
7.0
7.5
8.0

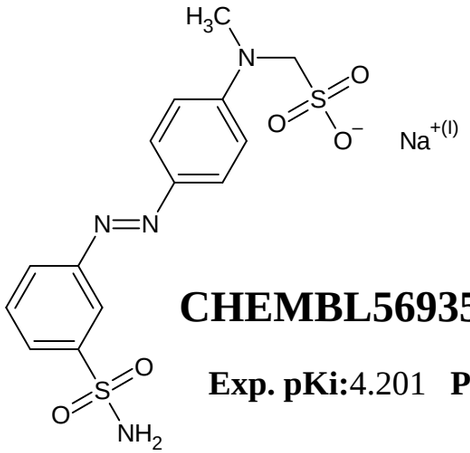
VIP plot and correlation matrix analysis for *MtbCA1* 1D 2D molecular descriptor prediction model





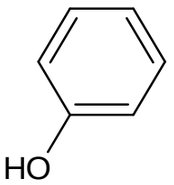


AATS7v	**Average Broto-Moreau autocorrelation - lag 7 / weighted by van der Waals volumes**
nHBint8	**Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 8**
SHBint8	**Sum of E-State descriptors of strength for potential hydrogen bonds of path length 8**
ATSC8v	**Centered Broto-Moreau autocorrelation - lag 8 / weighted by van der Waals volumes**
VR3_Dzs	**Logarithmic Randic-like eigenvector-based index from Barysz matrix / weighted by I-state**
maxHBint8	**Maximum E-State descriptors of strength for potential Hydrogen Bonds of path length 8**
naasC	**Count of atom-type E-State: :C:-**
SHssNH	**Sum of atom-type H E-State: -NH-**
VE1_Dzs	**Coefficient sum of the last eigenvector from Barysz matrix / weighted by I-state**
AATS7p	**Average Broto-Moreau autocorrelation - lag 7 / weighted by polarizabilities**



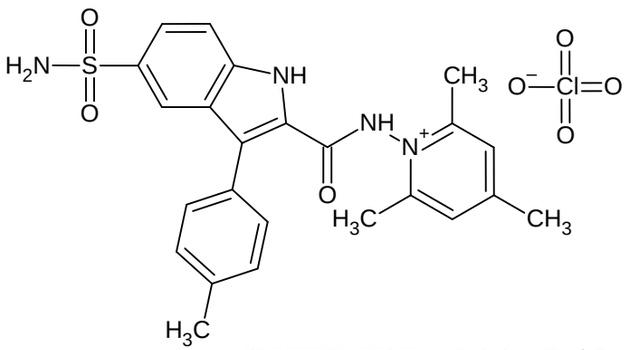
CHEMBL569350 (Inactive)

Exp. pKi:4.201 **Pred. pKi:**4.723



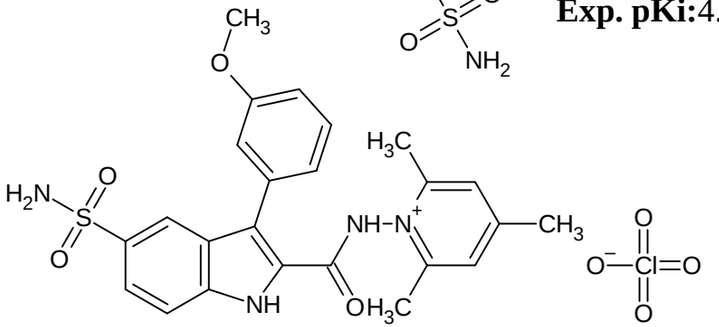
CHEMBL14060 (Inactive)

Exp. pKi:4.194 **Pred. pKi:**4.587



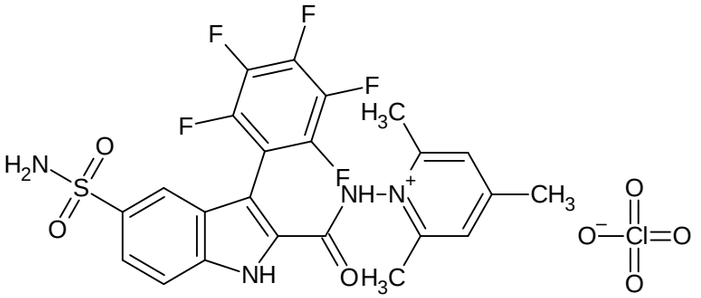
CHEMBL468946 (Active)

Exp. pKi:8.292 **Pred. pKi:**8.026



CHEMBL521690 (Active)

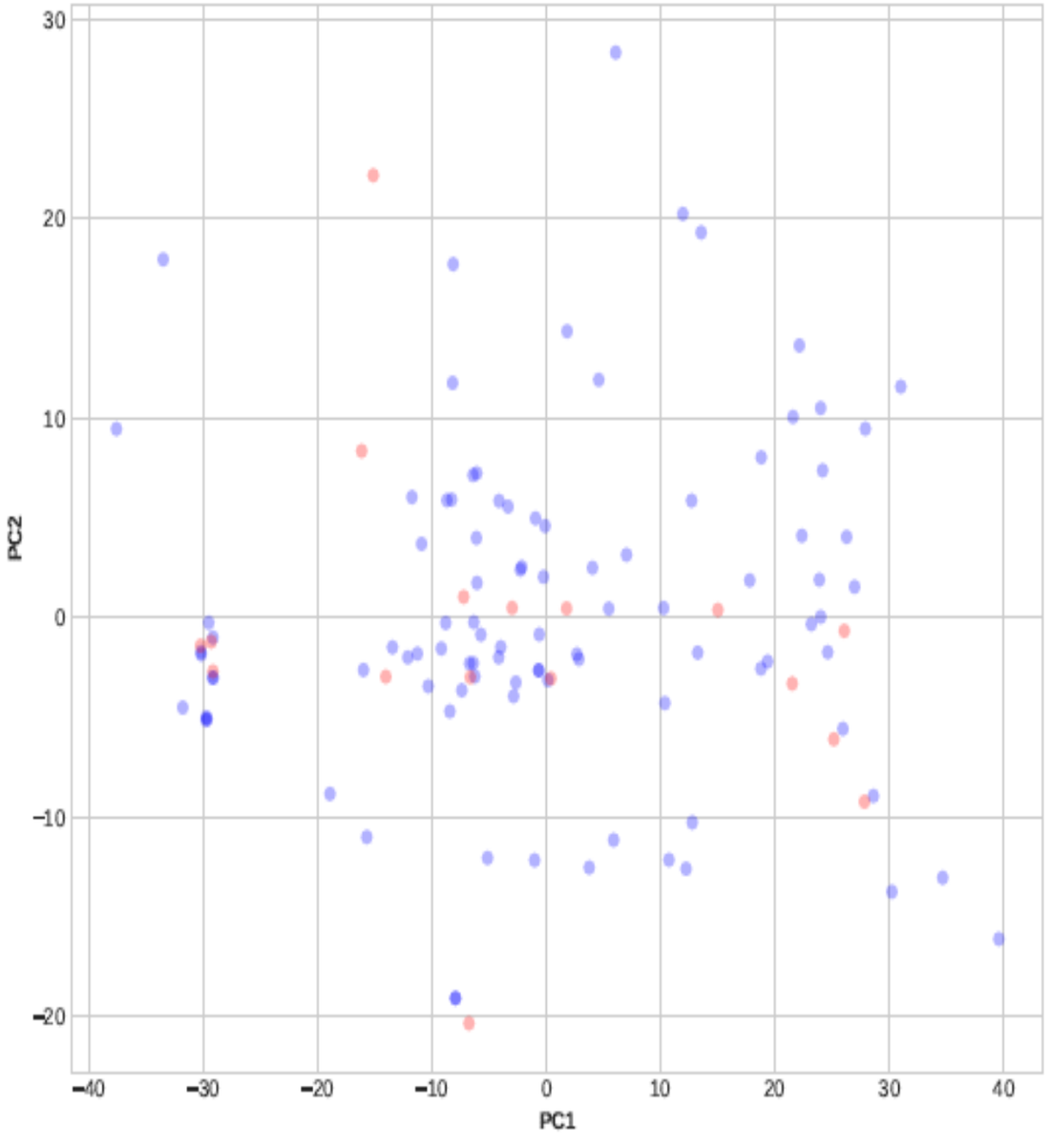
Exp. pKi:8.066 **Pred. pKi:**7.896

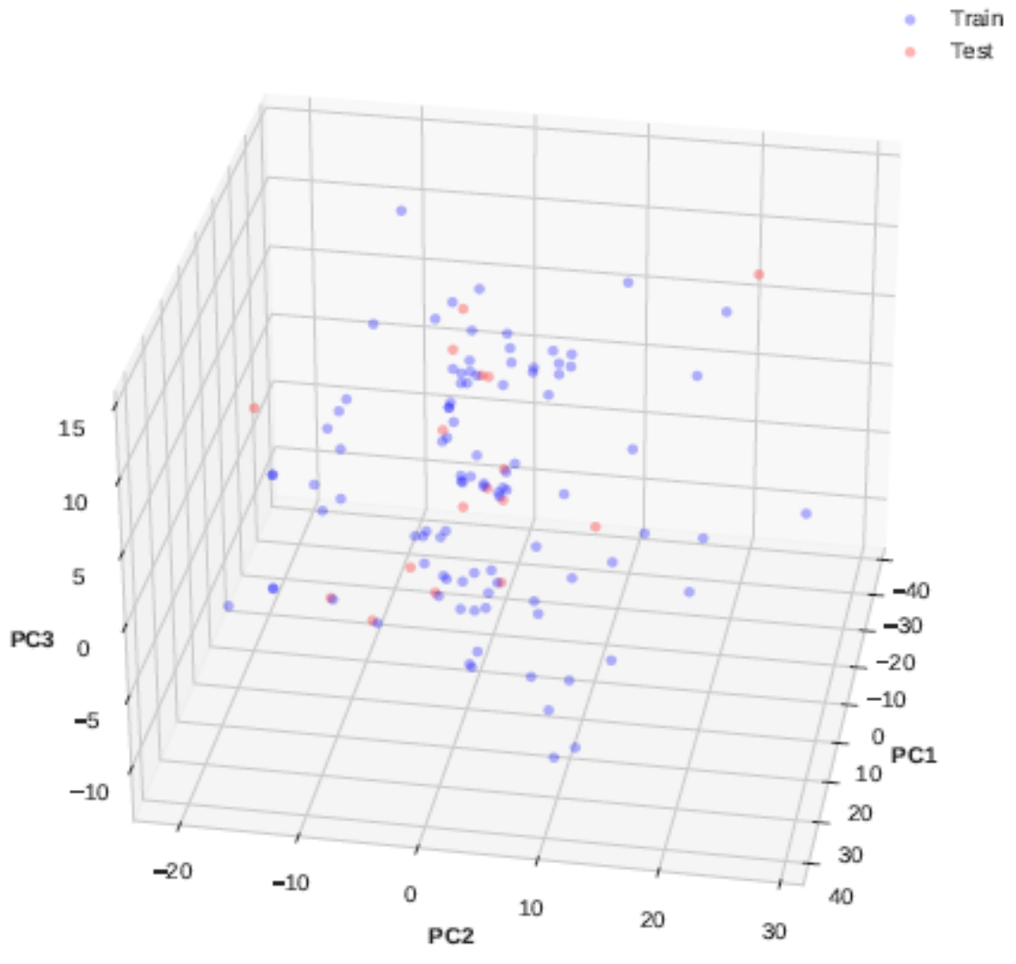


CHEMBL501294 (Active)

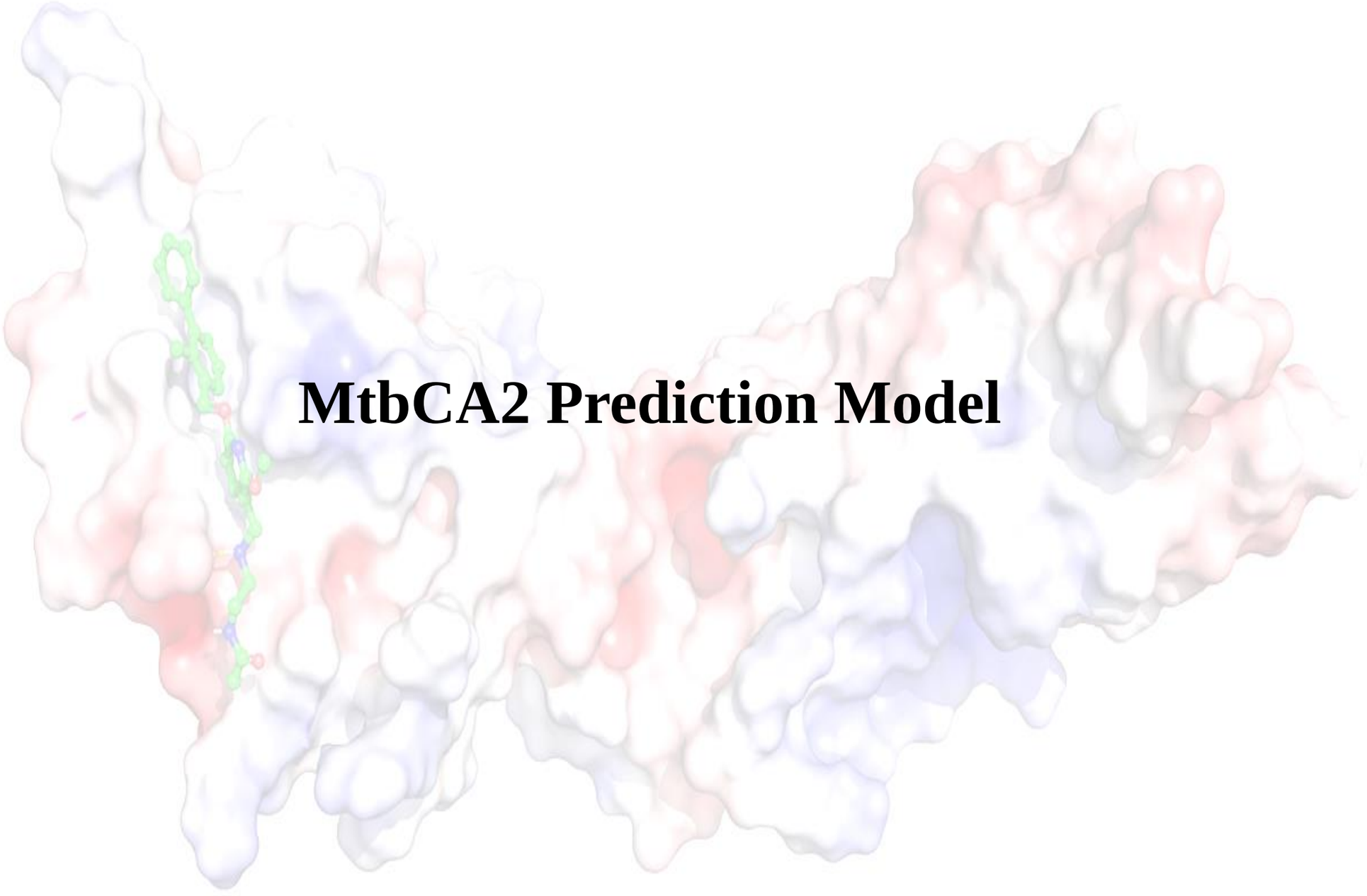
Exp. pKi:8.013 **Pred. pKi:**7.956

Applicability domain analysis through 2D and 3D PCA plots for *MtbCA1* 1D 2D molecular descriptor prediction model





MtbCA2 Prediction Model



ML-QSAR model for bioactivity prediction of MtbCA2 inhibitors using substructure fingerprints

ChEMBL

Curation of *MtbCA2* inhibitors with Ki value

Total inhibitors: 42

Substructure fingerprints: 307

Prediction Parameter: Bioactivity Class ('pKi')

Dataset Splitting into training and test set

Elimination of highly correlated and constant molecular descriptors

Final number of fingerprints(13), number of molecules in training(33) and test(9)

Random Forest Regressor

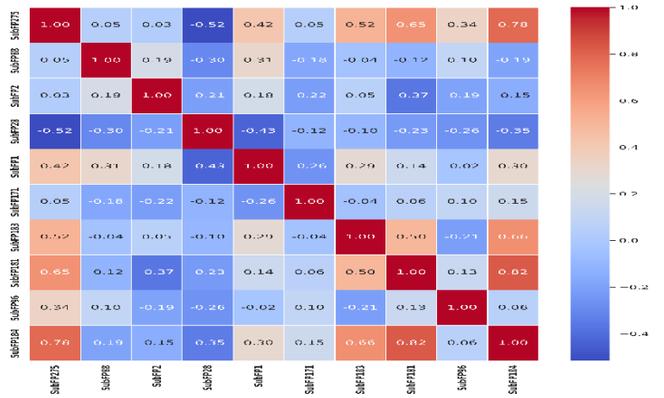
Initial Model Performance

Model Metrics
R2 (Train: 0.9645, Test: 0.8764)
RMSE (Train: 0.2785, Test: 0.5463)
MAE (Train: 0.1948, Test: 0.4097)

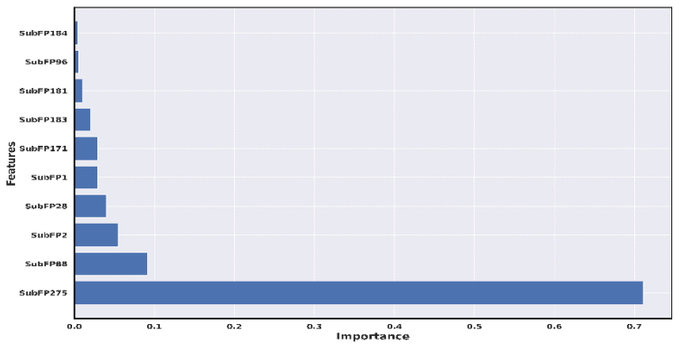
FP and FN Removal

Mention molecules removed (0)

Correlation Matrix



VIP Plot



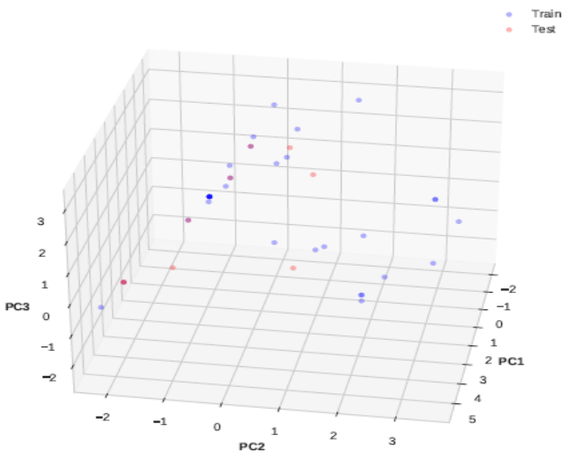
Validation Metrics

Final Model Performance

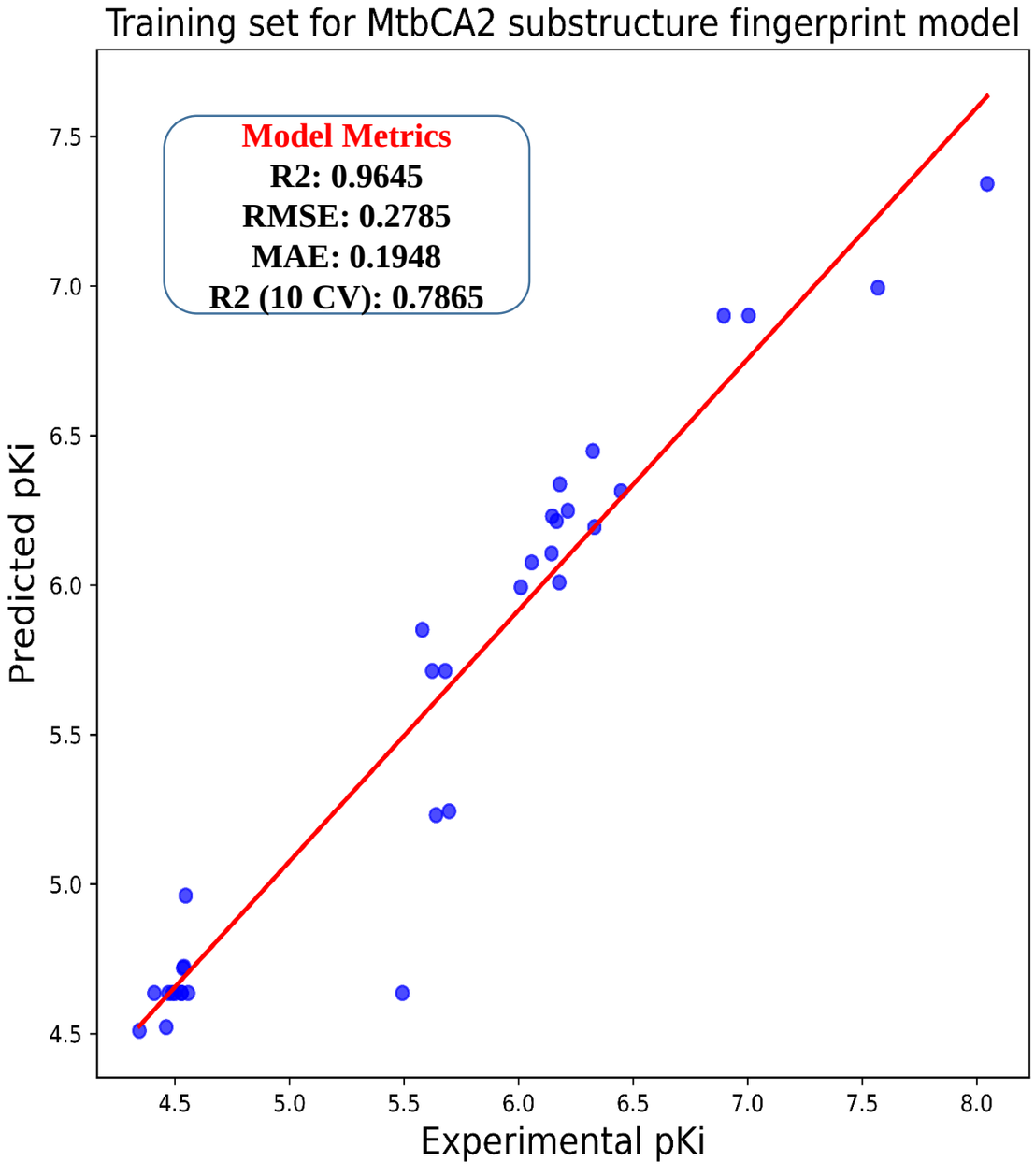
Model Metrics
R2 (Train: 0.9645, Test: 0.8764)
RMSE (Train: 0.2785, Test: 0.5463)
MAE (Train: 0.1948, Test: 0.4097)

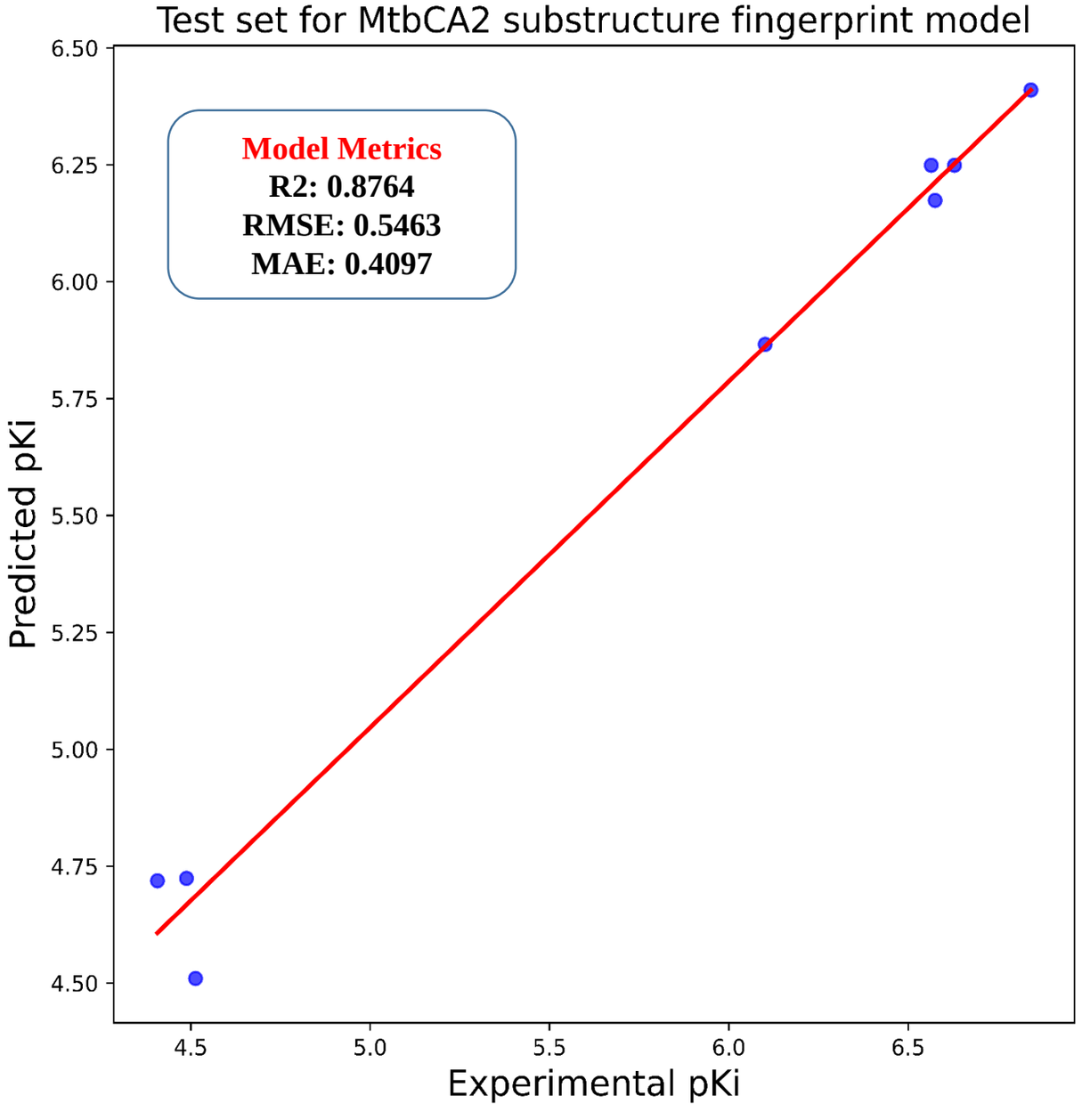
The final machine-learning model

number of molecules in training(33) and test(9)

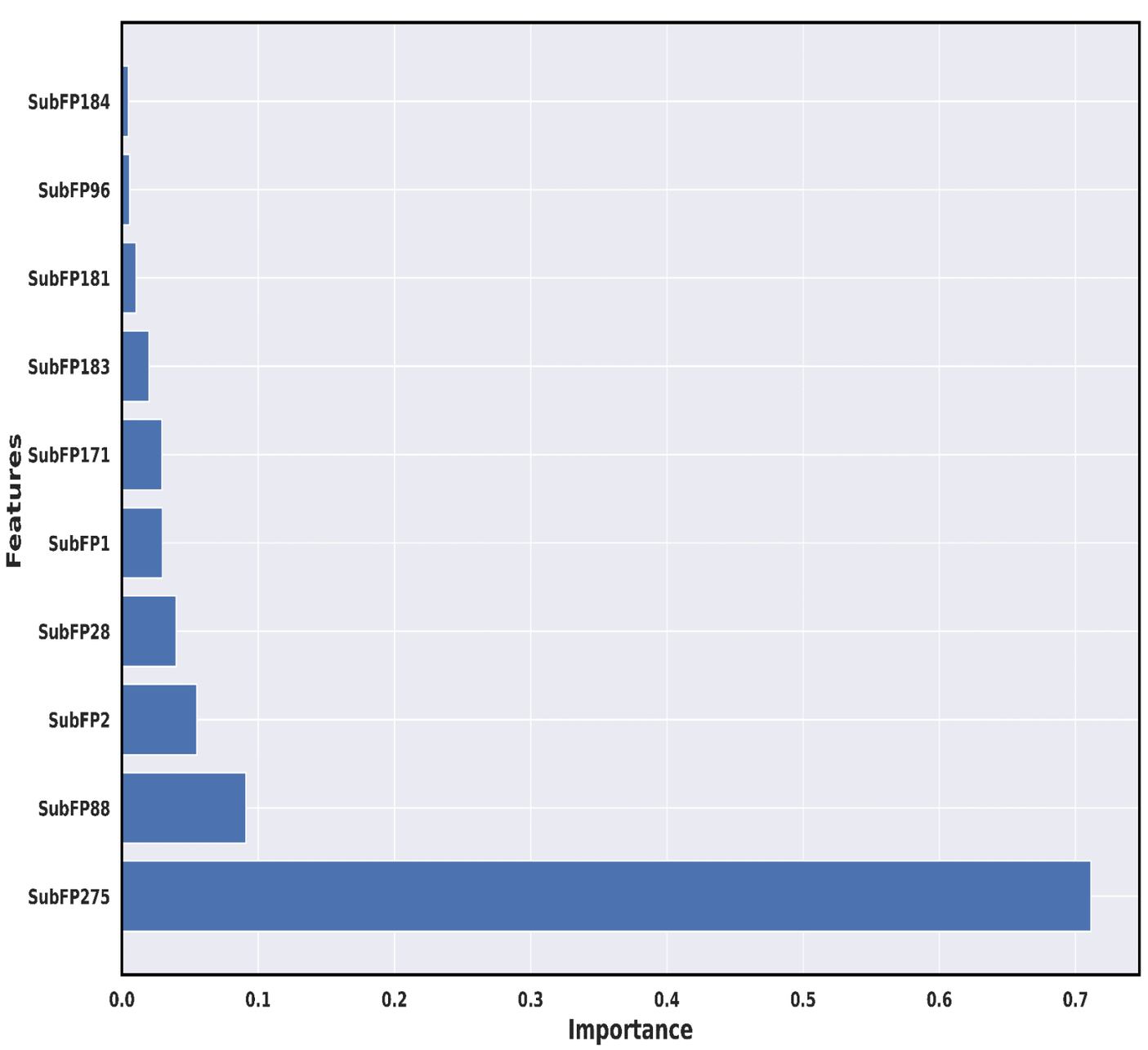


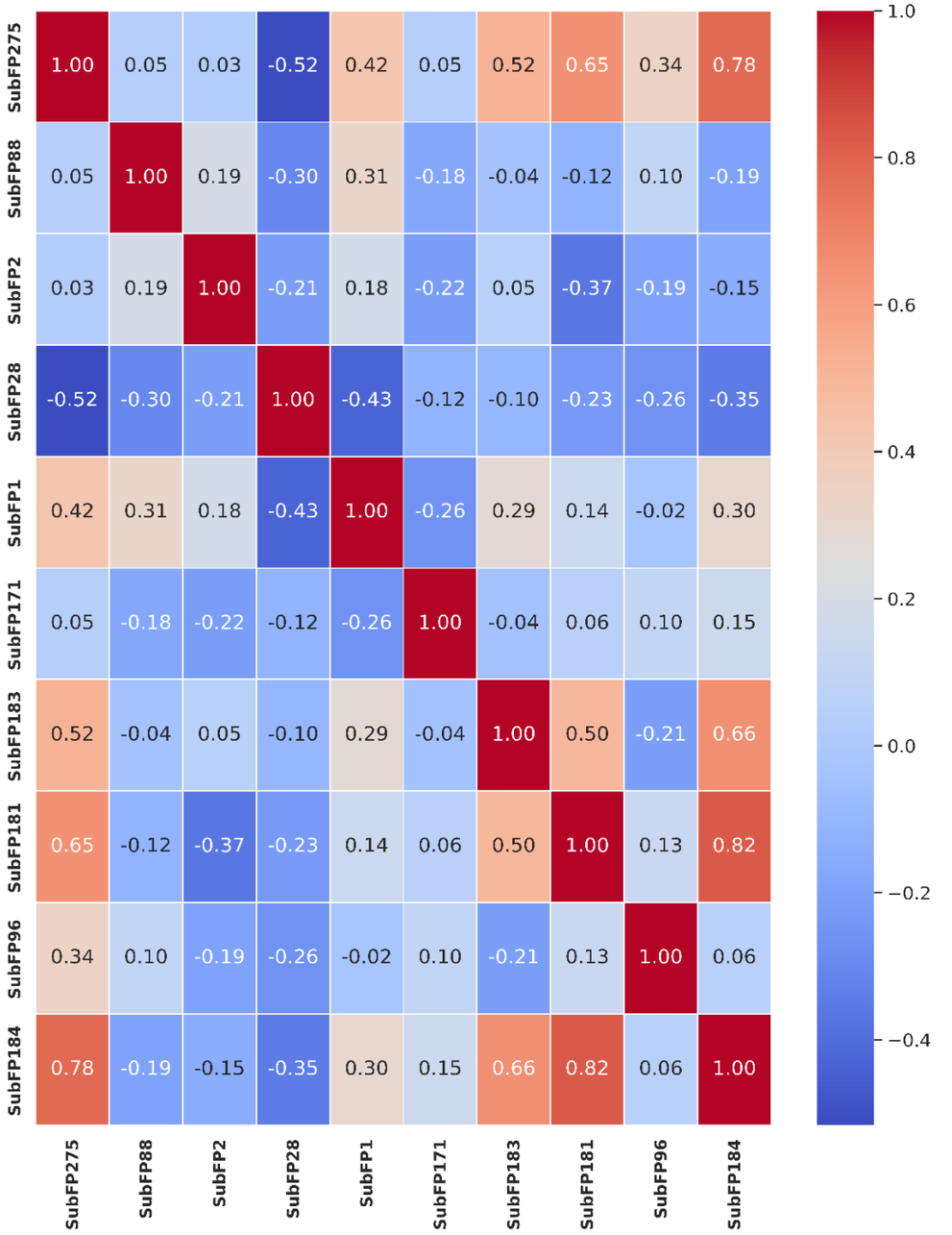
Applicability Domain Analysis

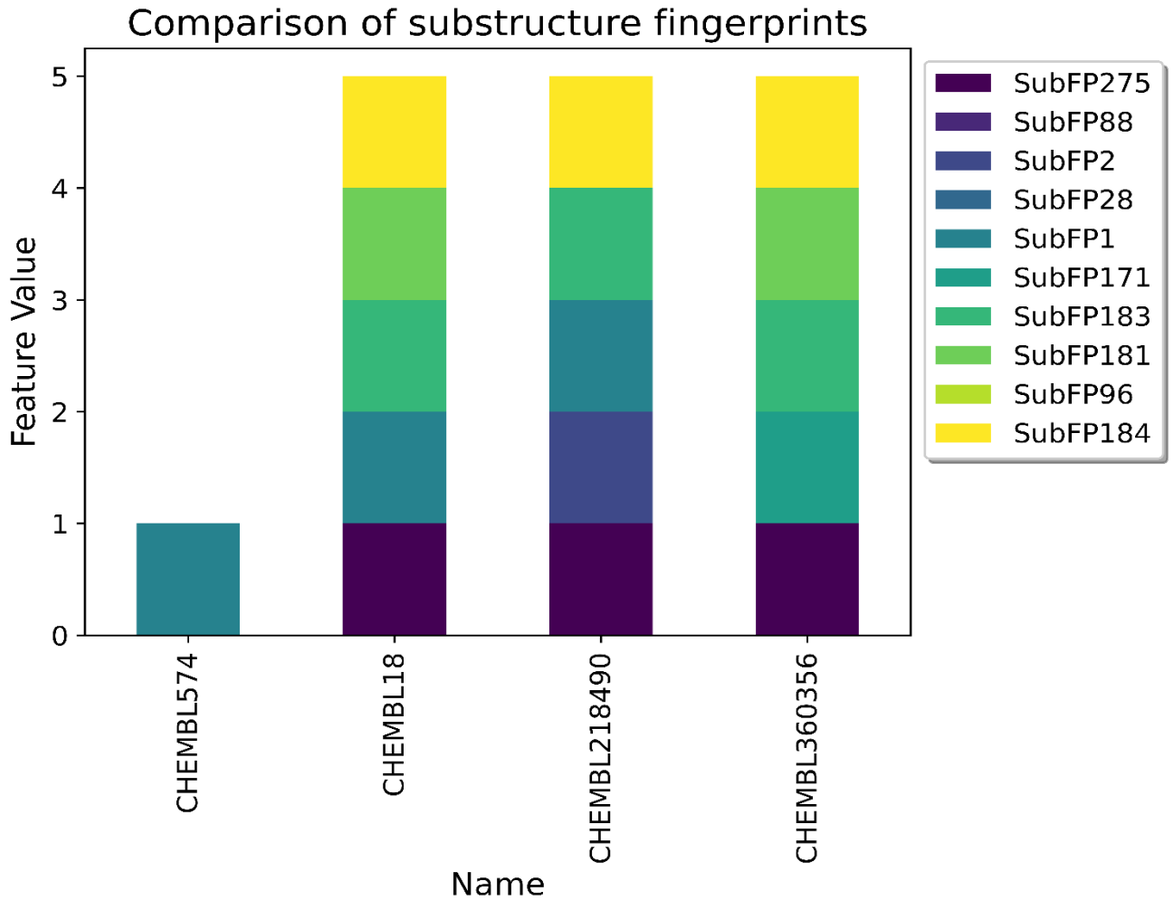

Training set for MtbCA2 substructure fingerprint model
Model Metrics
R2: 0.9645
RMSE: 0.2785
MAE: 0.1948
R2 (10 CV): 0.7865
Predicted pKi
Experimental pKi
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0


Test set for MtbCA2 substructure fingerprint model
Model Metrics
R2: 0.8764
RMSE: 0.5463
MAE: 0.4097
Predicted pKi
Experimental pKi
4.50
4.75
5.00
5.25
5.50
5.75
6.00
6.25
6.50
4.5
5.0
5.5
6.0
6.5

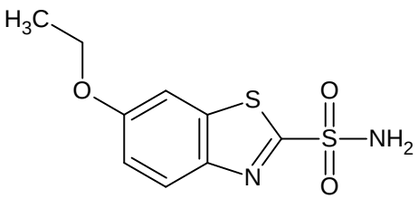
VIP plot and correlation matrix analysis for *MtbCA2* substructure prediction model





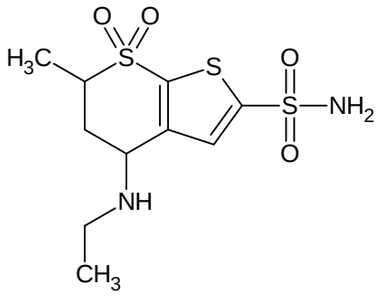


SubFP275	**Heterocyclic**
SubFP88	**Carboxylic acid derivative**
SubFP2	**Secondary carbon**
SubFP28	**Primary aromatic amine**
SubFP1	**Primary carbon**
SubFP171	**Arylchloride**
SubFP183	**Hetero S**
SubFP181	**Hetero N nonbasic**
SubFP96	**Carbodithioic ester**
SubFP184	**Heteroaromatic**



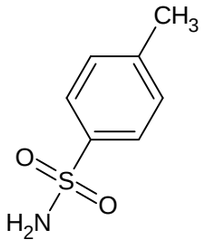
CHEMBL18/Ethoxzolamide
(Active)

Exp. pKi:7.569 **Pred. pKi:**6.994



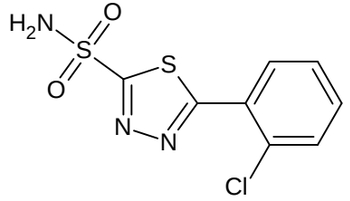
CHEMBL218490/Dorzolamide
(Active)

Exp. pKi:7.004 **Pred. pKi:**6.901



CHEMBL574/P-Toluenesulfonamide
(Inactive)

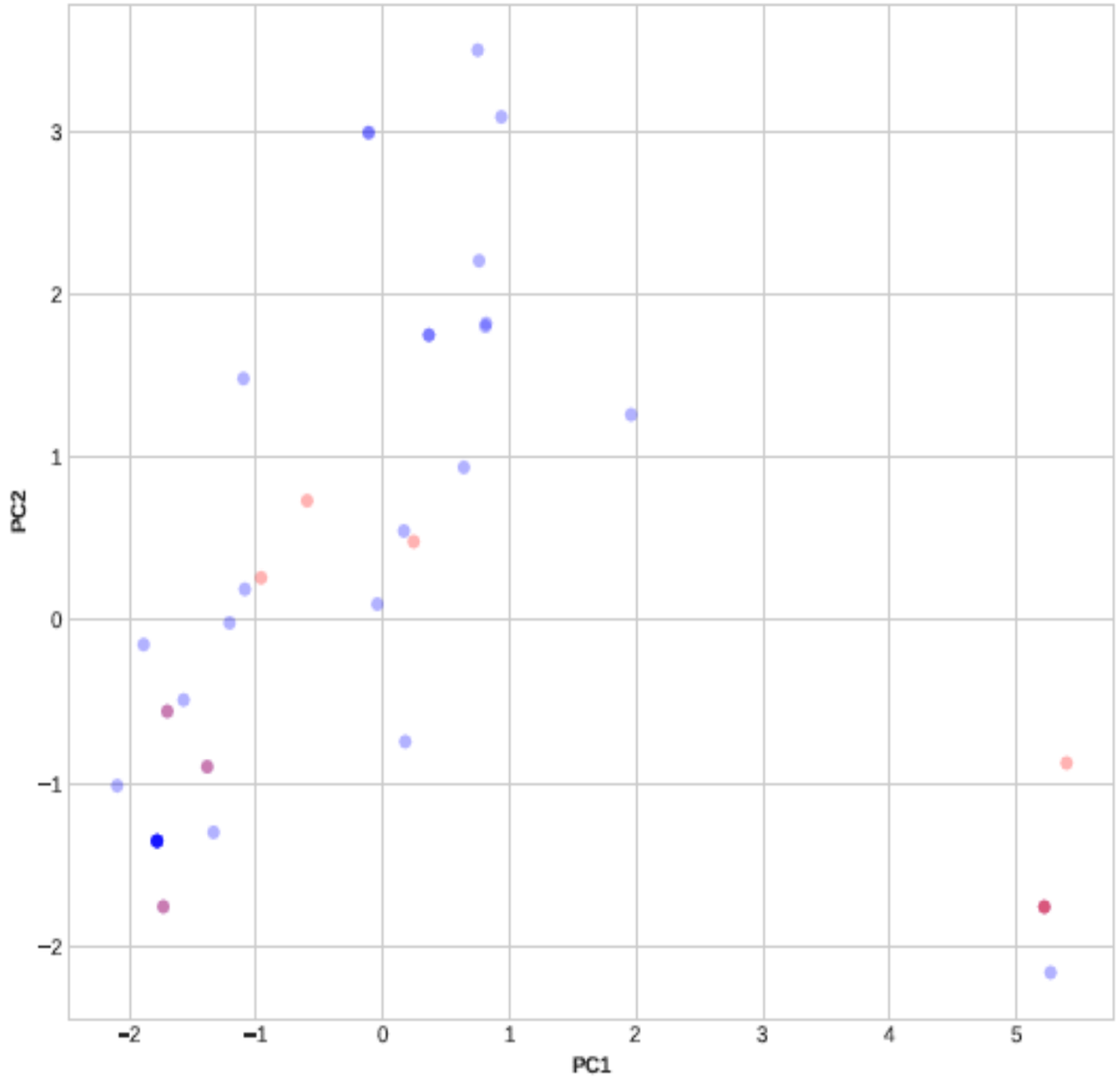
Exp. pKi:4.41 **Pred. pKi:**4.636

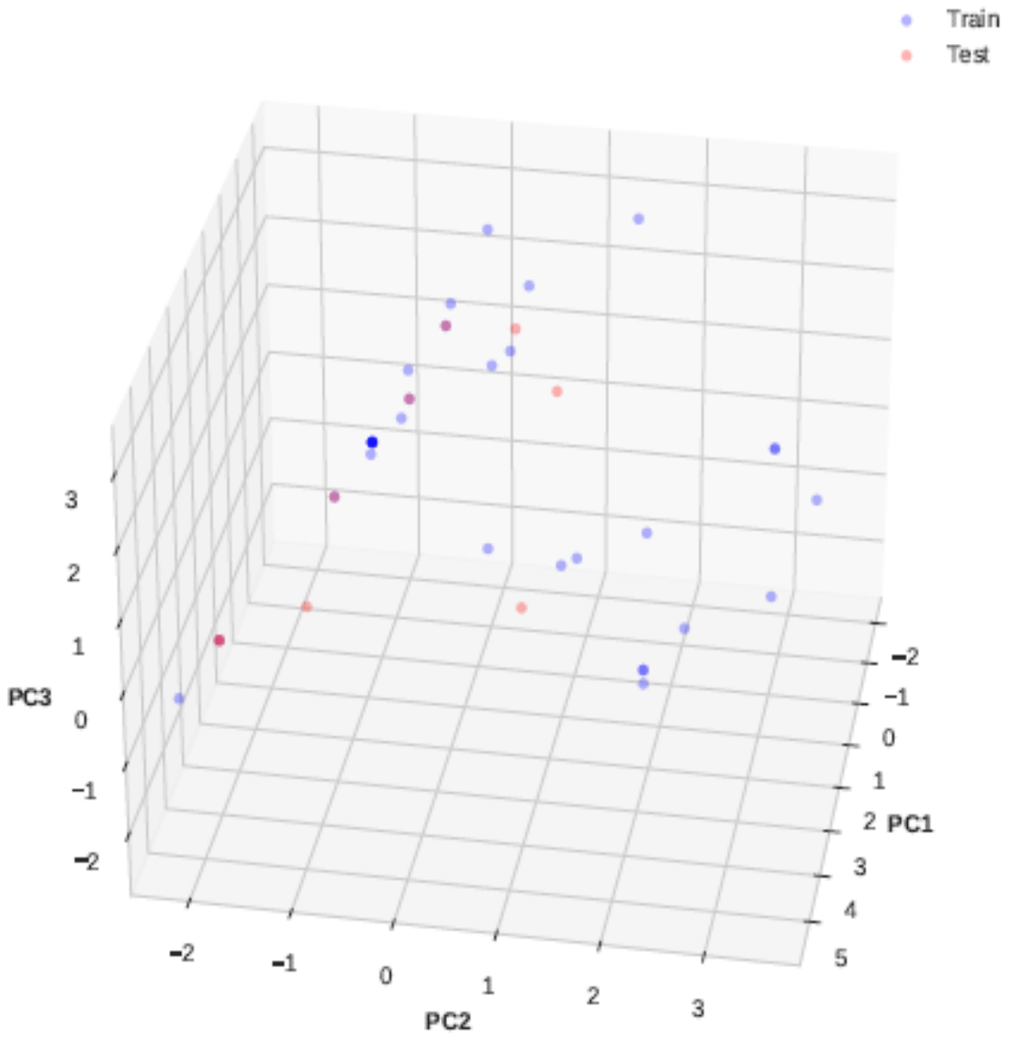


CHEMBL360356/Chlorzolamide
(Inactive)

Exp. pKi:4.345 **Pred. pKi:**4.51

Applicability domain analysis through 2D and 3D PCA plots for *MtbCA2* substructure prediction model





ML-QSAR model for bioactivity prediction of MtbCA2 inhibitors using 1D and 2D molecular descriptors

ChEMBL

Curation of *MtbCA2* inhibitors with Ki value

Total inhibitors: 42

1D and 2D Molecular Descriptors: 1444

Prediction Parameter: Bioactivity Class ('pKi')

Dataset Splitting into training and test set

Elimination of highly correlated and constant molecular descriptors

Final number of descriptors(736), number of molecules in training(33) and test(9)

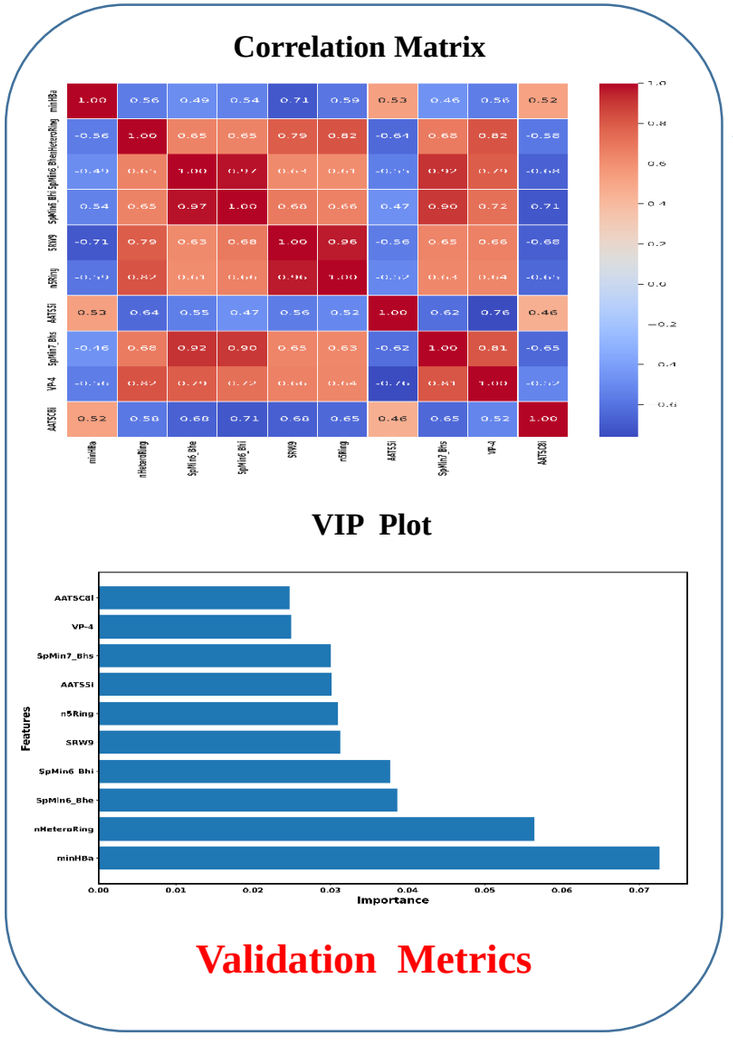
Random Forest Regressor

Initial Model Performance

Model Metrics
R2 (Train: 0.9651, Test: 0.8572)
RMSE (Train: 0.326, Test: 0.4517)
MAE (Train: 0.2082, Test: 0.3368)

FP and FN Removal

Mention molecules removed (0)



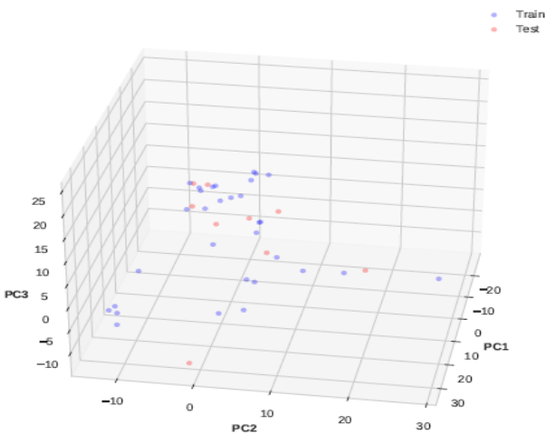
Validation Metrics

Model Metrics
R2 (Train: 0.9864, Test: 0.8874)
RMSE (Train: 0.1971, Test: 0.4122)
MAE (Train: 0.1553, Test: 0.3101)

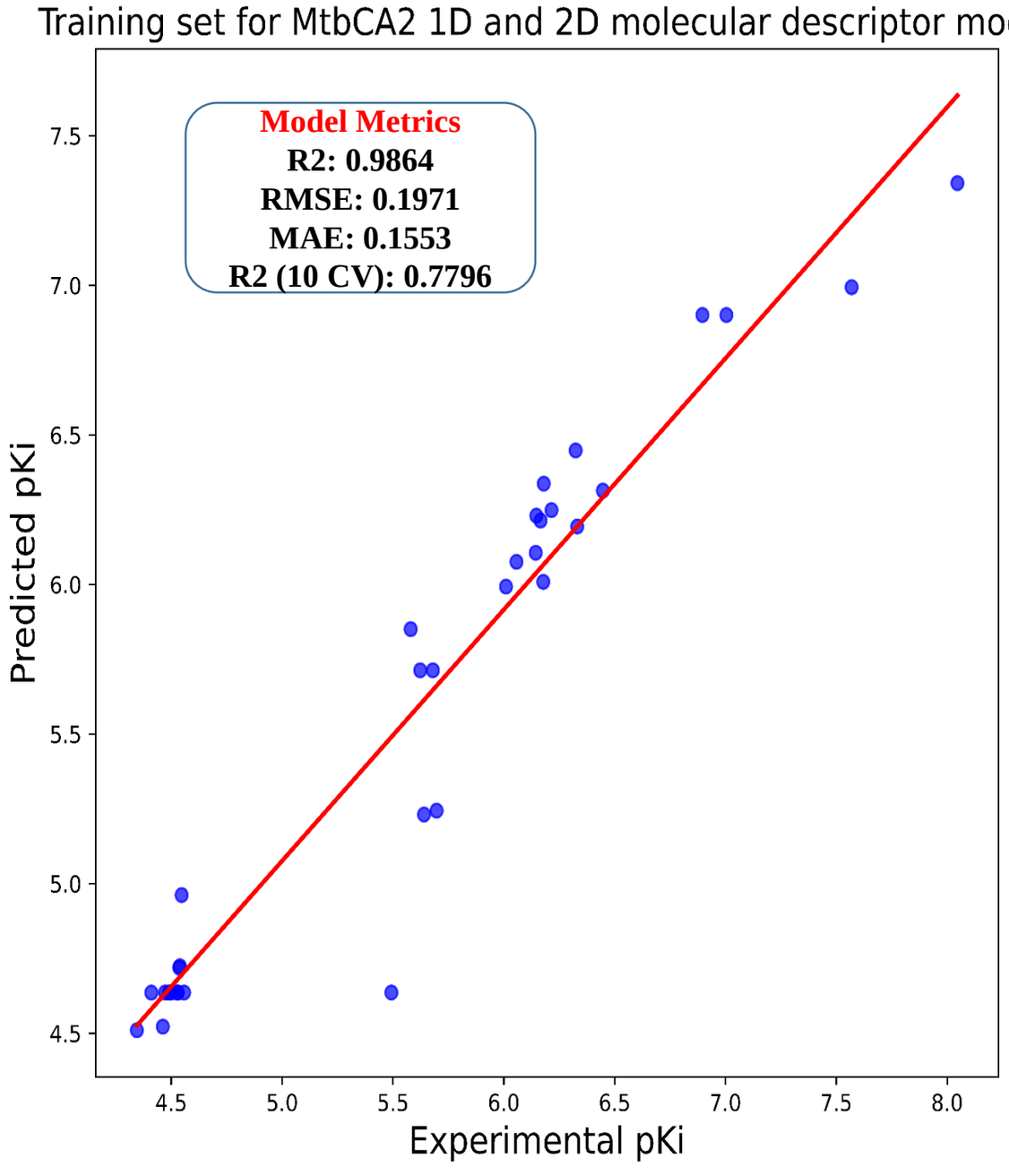
Final Model Performance

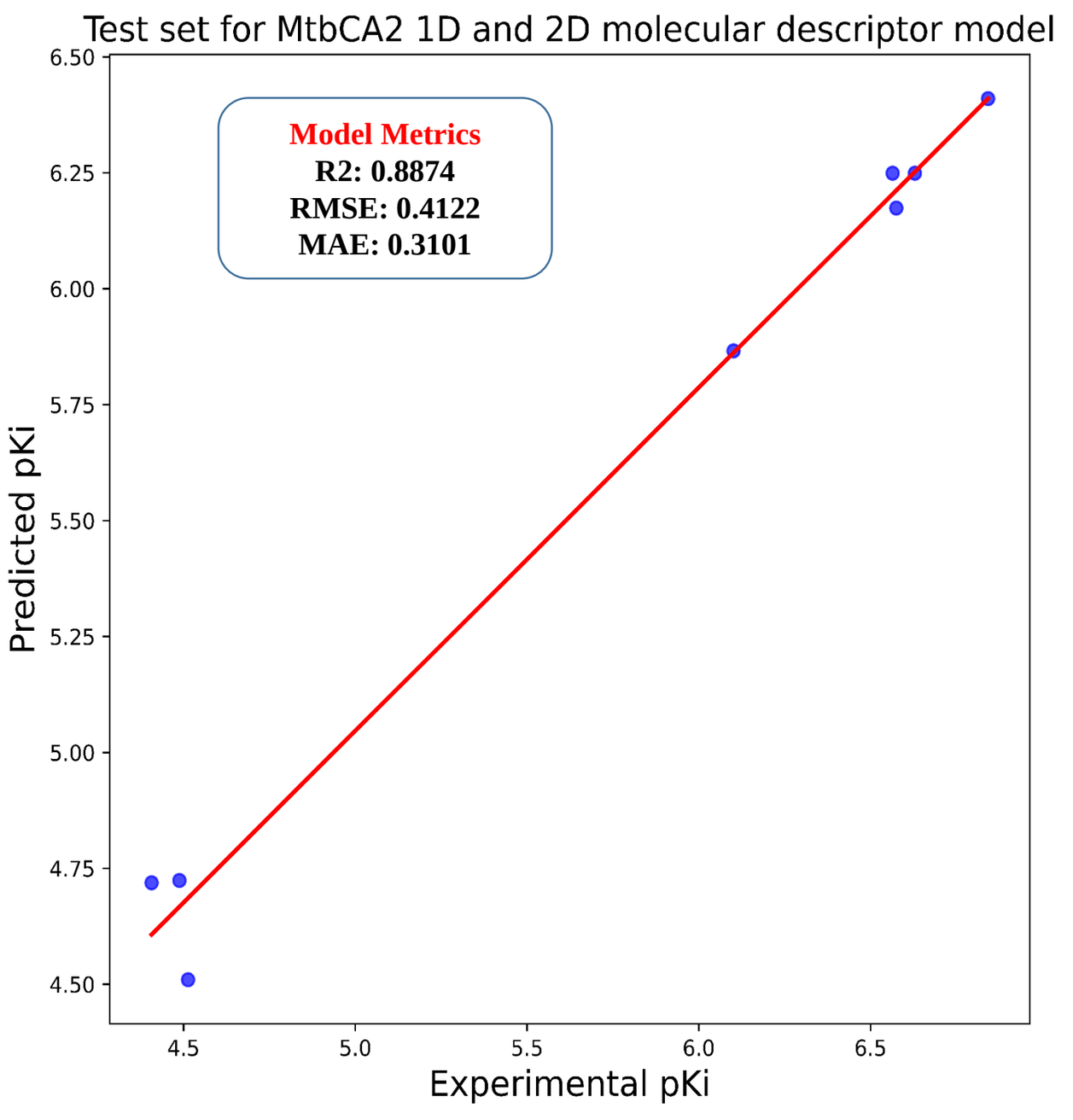
The final machine-learning model

number of molecules in training(31) and test(9)

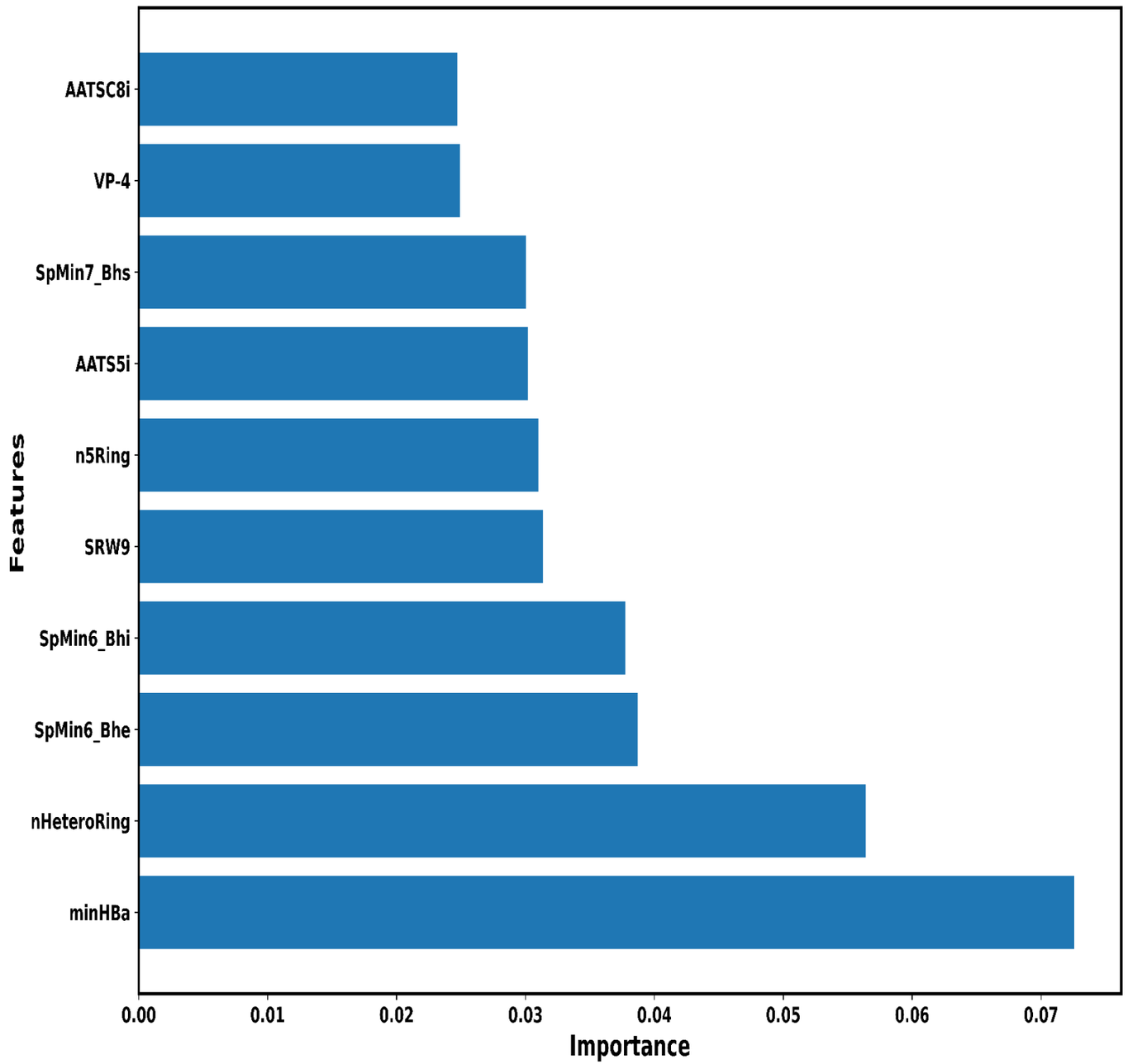


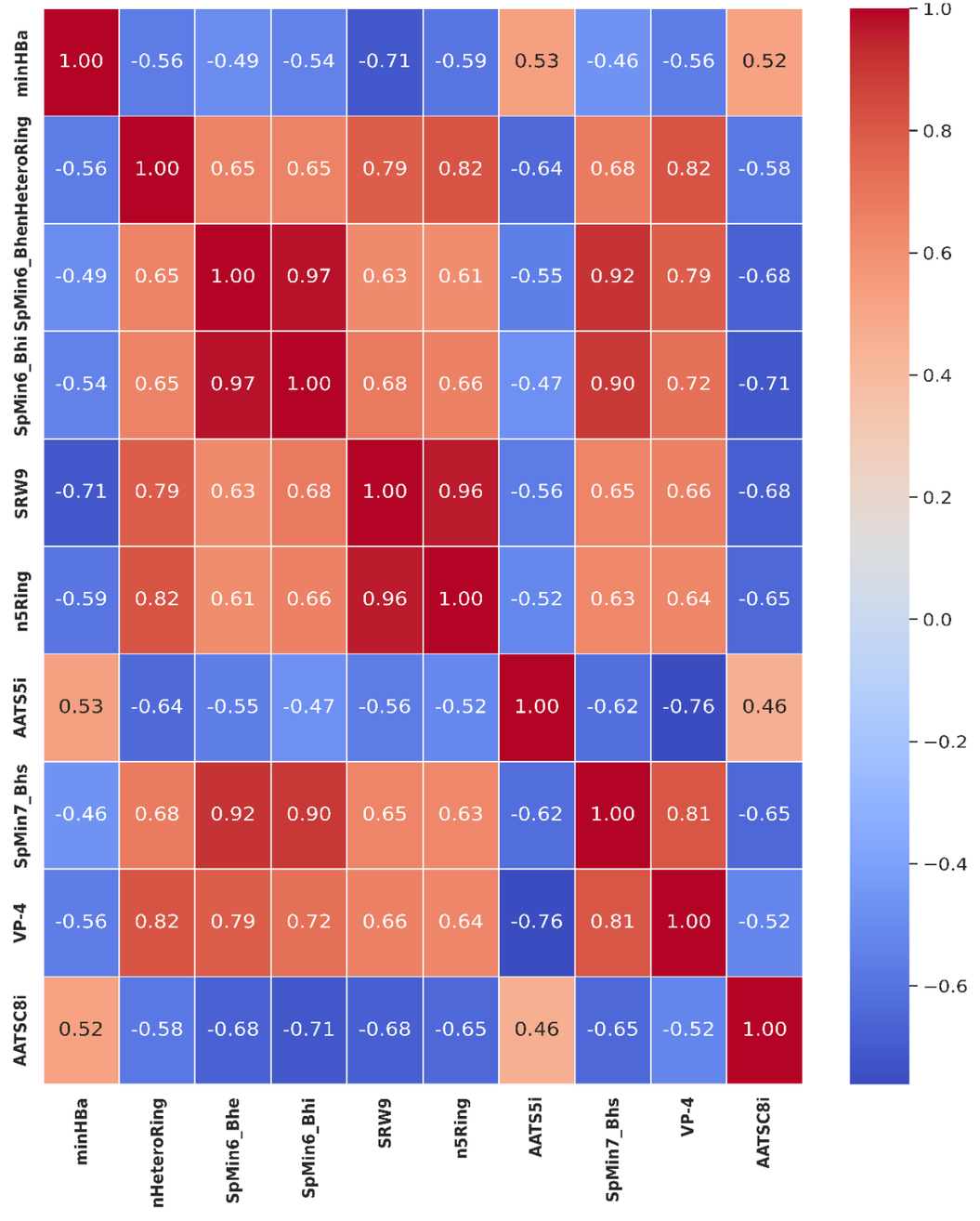
Applicability Domain Analysis

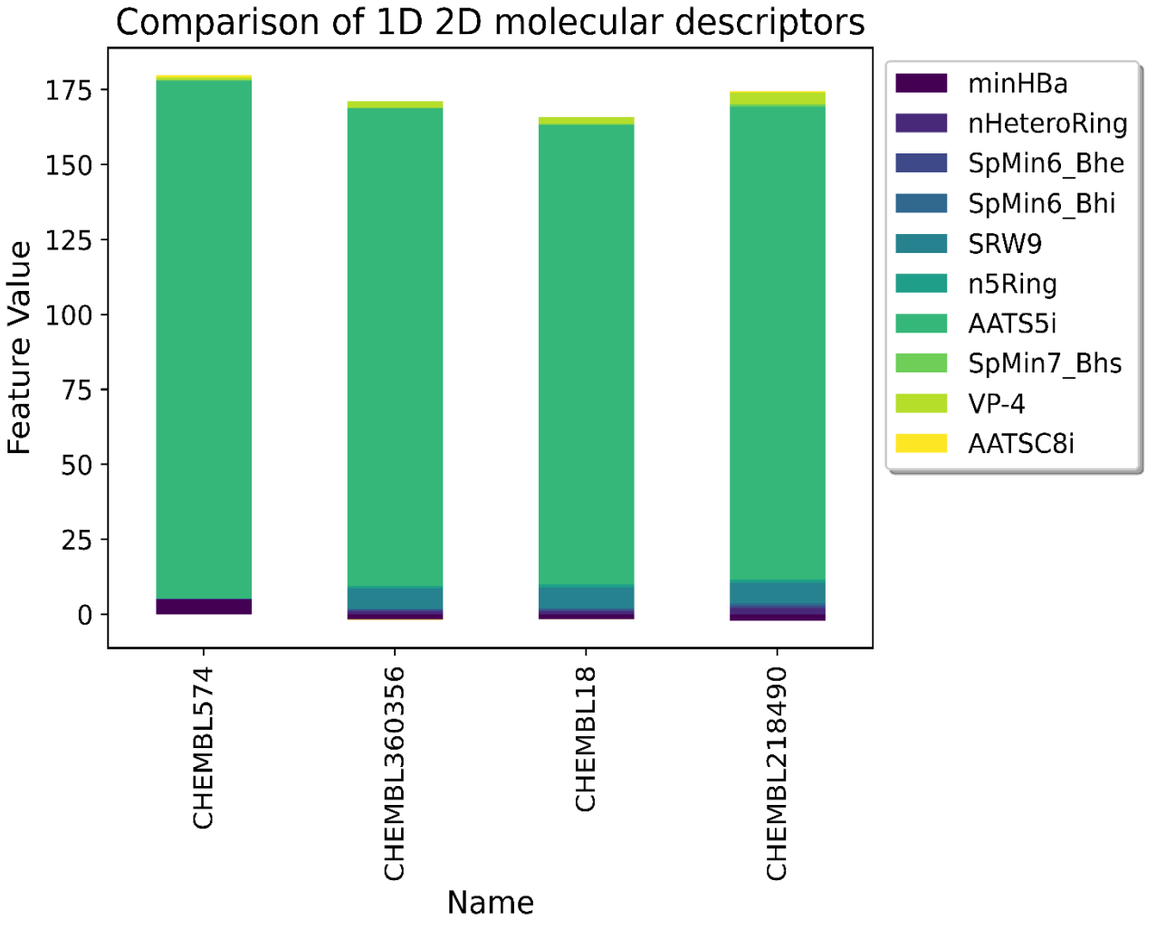

Training set for MtbCA2 1D and 2D molecular descriptor model
Model Metrics
R2: 0.9864
RMSE: 0.1971
MAE: 0.1553
R2 (10 CV): 0.7796
Predicted pKi
Experimental pKi


Test set for MtbCA2 1D and 2D molecular descriptor model
Model Metrics
R2: 0.8874
RMSE: 0.4122
MAE: 0.3101
Predicted pKi
Experimental pKi

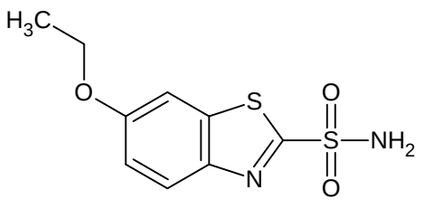
VIP plot and correlation matrix analysis for *MtbCA2* 1D 2D molecular descriptor prediction model





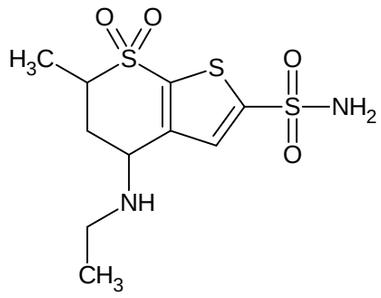


minHBa	**Minimum E-States for (strong) Hydrogen Bond acceptors**
nHeteroRing	**Number of rings containing heteroatoms (N, O, P, S, or halogens)**
SpMin6_Bhe	**Smallest absolute eigenvalue of Burden modified matrix - n 6 / weighted by relative Sanderson electronegativities**
SpMin6_Bhi	**Smallest absolute eigenvalue of Burden modified matrix - n 6 / weighted by relative first ionization potential**
SRW9	**Self-returning walk count of order 9 (ln(1+x)**
n5Ring	**Number of 5-membered rings**
AATS5i	**Average Broto-Moreau autocorrelation - lag 5 / weighted by first ionization potential**
SpMin7_Bhs	**Smallest absolute eigenvalue of Burden modified matrix - n 7 / weighted by relative I-state**
VP-4	**Valence path, order 4**
AATSC8i	**Average centered Broto-Moreau autocorrelation - lag 8 / weighted by first ionization potential**



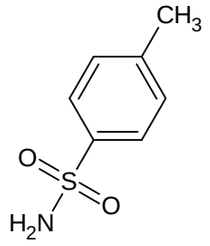
CHEMBL18/Ethoxzolamide (Active)

Exp. pKi:7.569 **Pred. pKi:**6.823



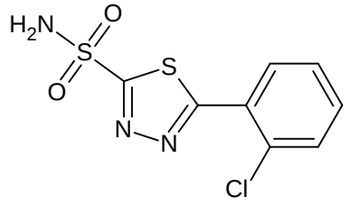
CHEMBL218490/Dorzolamide (Active)

Exp. pKi:7.004 **Pred. pKi:**6.784



CHEMBL574/P-Toluenesulfonamide (Inactive)

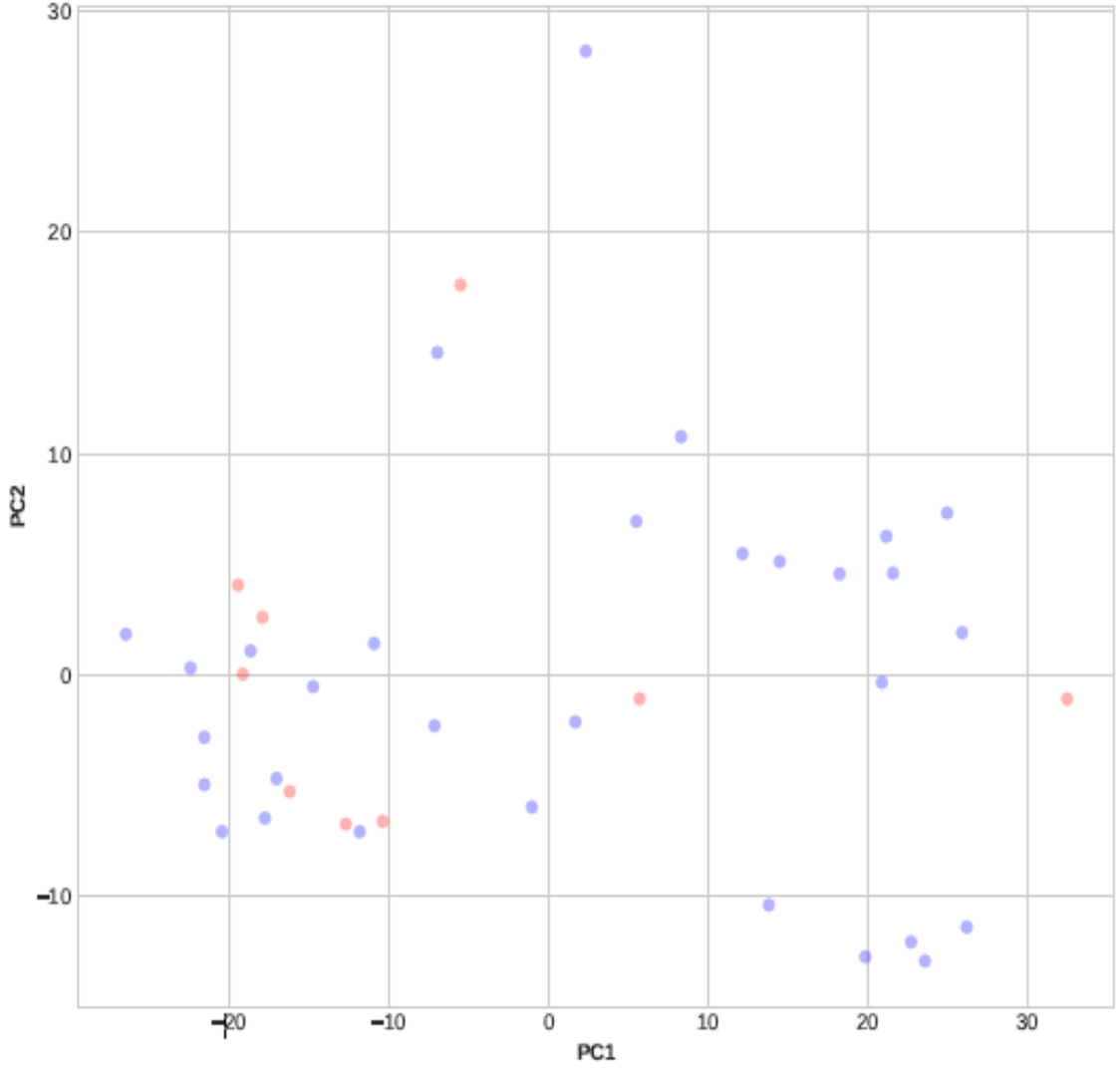
Exp. pKi:4.41 **Pred. pKi:**4.907

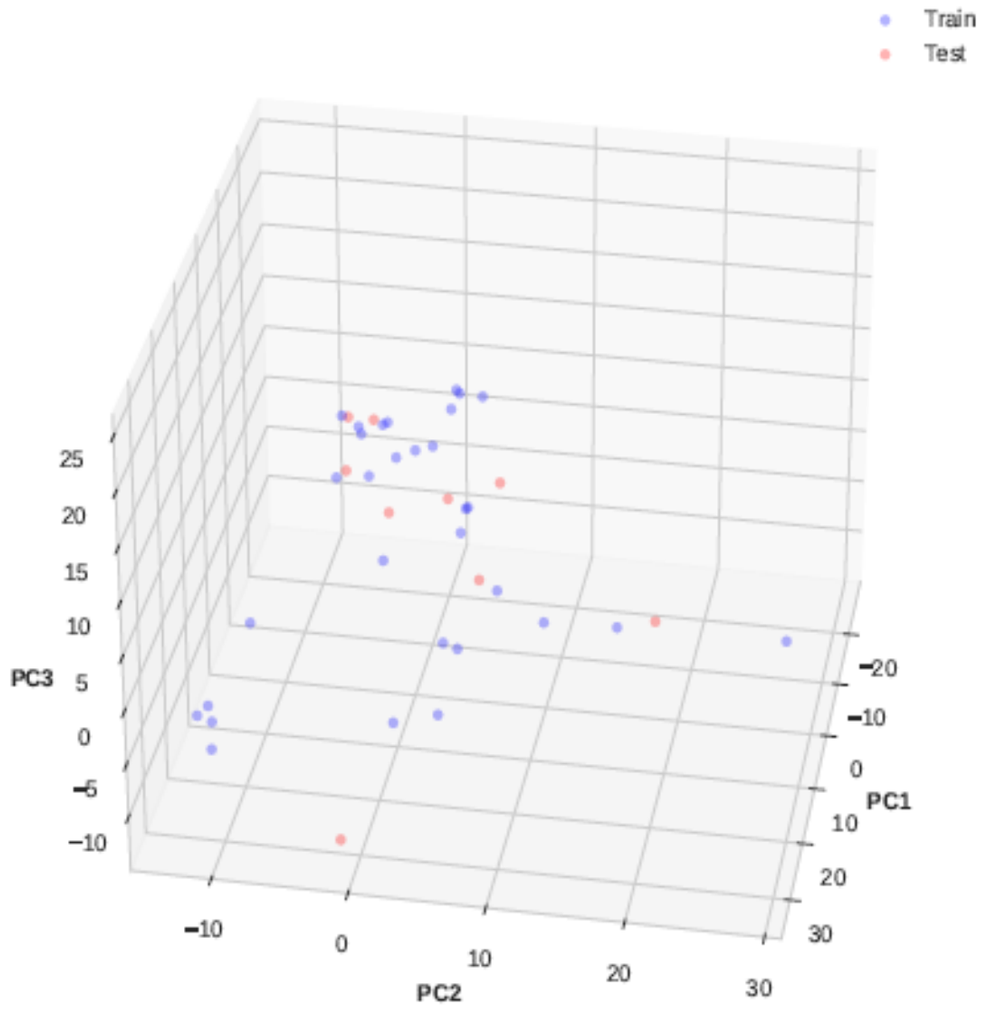


CHEMBL360356/Chlorzolamide (Inactive)

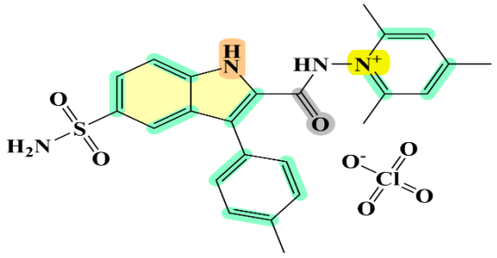
Exp. pKi:4.345 **Pred. pKi:**4.491

Applicability domain analysis through 2D and 3D PCA plots for *MtbCA2* 1D 2D molecular descriptor prediction model

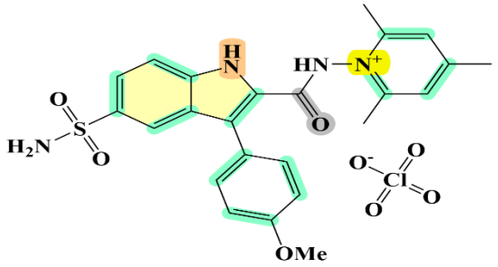




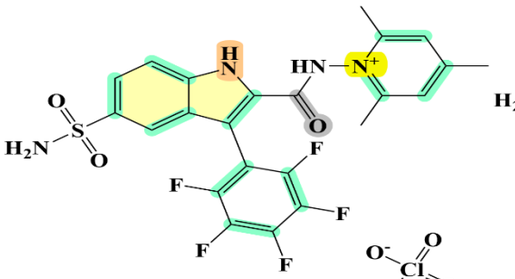
ML-QSAR model for bioactivity prediction of MtbCA1 inhibitors using substructure fingerprints



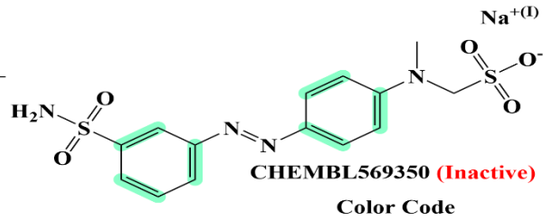
CHEMBL468946 (Active)



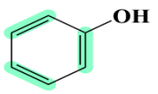
CHEMBL521690 (Active)



CHEMBL501294 (Active)



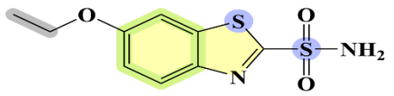
CHEMBL569350 (Inactive)



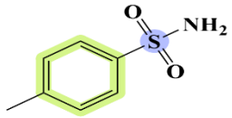
CHEMBL14060 (Inactive)

Color Code
SubFP179
SubFP298
SubFP287, SubFP2
SubFP143
SubFP275

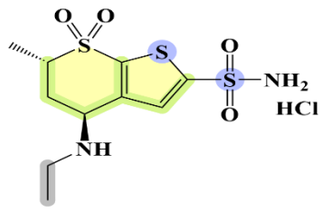
ML-QSAR model for bioactivity prediction of MtbCA2 inhibitors using substructure fingerprints



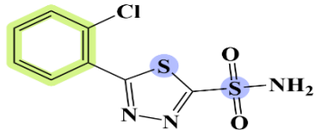
CHEMBL18/Ethoxzolamide
(Active)



CHEMBL574/P-Toluenesulfonamide
(Inactive)



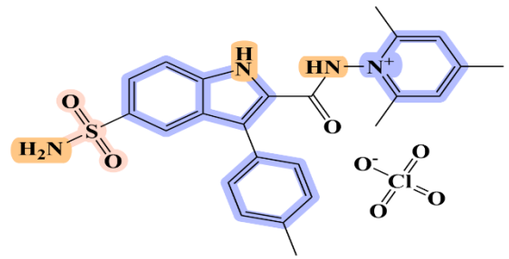
CHEMBL218490/Dorzolamide
(Active)



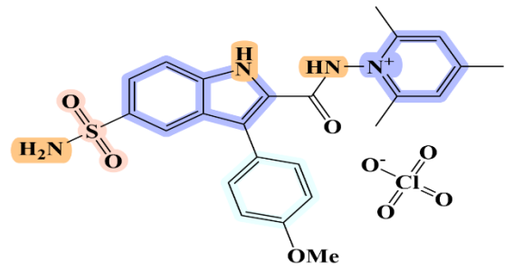
CHEMBL360356/Chlorzolamide
(Inactive)

Color Code
SubFP2
SubFP143
SubFP275, SubFP184
SubFP1

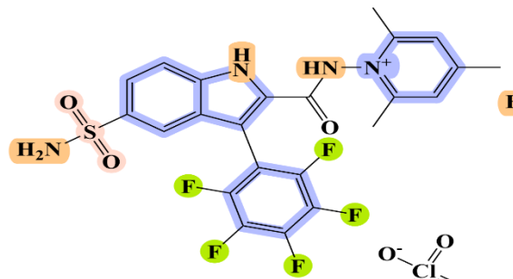
ML-QSAR model for bioactivity prediction of MtbCA1 inhibitors using 1D and 2D molecular descriptors



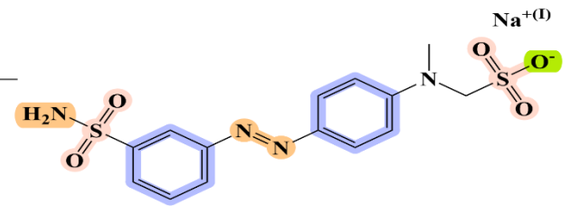
CHEMBL468946 (Active)



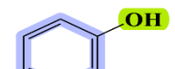
CHEMBL521690 (Active)



CHEMBL501294 (Active)



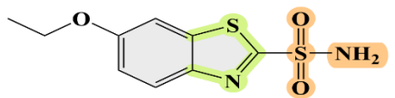
CHEMBL569350 (Inactive)



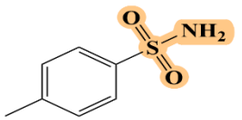
CHEMBL14060 (Inactive)

Color Code
nHBint8
SHssNH$HBint8
naasC
VR3_Dzs /VE1_Dzs

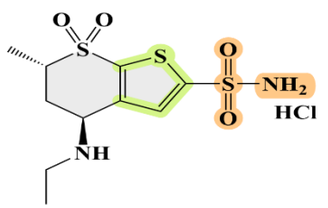
ML-QSAR model for bioactivity prediction of MtbCA2 inhibitors using 1D and 2D molecular descriptors



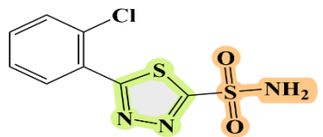
CHEMBL18/Ethoxzolamide
(Active)



CHEMBL574/P-Toluenesulfonamide
(Inactive)



CHEMBL218490/Dorzolamide
(Active)



CHEMBL360356/Chlorzolamide
(Inactive)

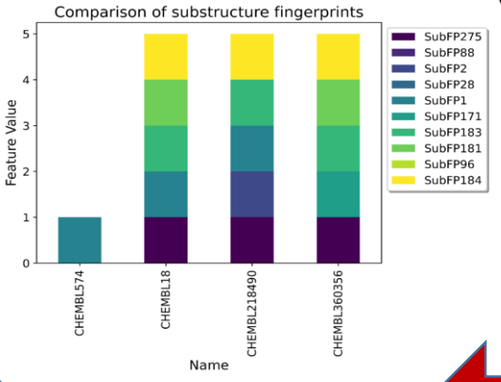
Color Code
minHBa
nHeteroRing
n5Ring

Initial ML-QSAR models deployment using two different molecular signatures with respect to MtbCA1 inhibitors with their respective pKi values)
(S: 307 ; D: 1444)

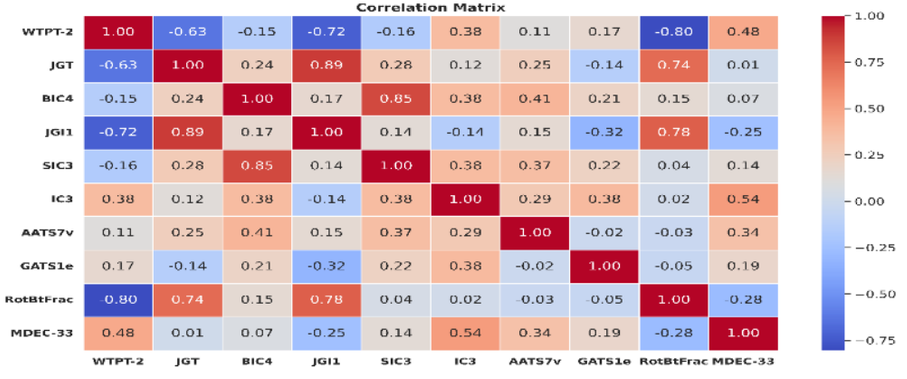
ML-QSAR models generation and statistical analysis
(S:18; D:763)

Machine Learning models generation with significant molecular signatures

Web app development



ML-QSAAR
(Dual inhibitors)



MtbCA-Selec-Pred

https://mtbca-selec-pred.streamlit.app/

Initial ML-QSAR models deployment using two different molecular signatures with respect to MtbCA2 inhibitors with their respective pKi values)
(S: 307 ; D: 1444)

ML-QSAR models generation and statistical analysis
(S:13; D:736)

Machine Learning models generation with significant molecular signatures

Web app development

Machine Learning-assisted web application for selectivity bioactivity prediction for MtbCAIs, against MtbCA1 and MtbCA2

S: Substructure Fingerprints D: 1D & 2D Descriptors

Fork this app

Choose a prediction model

MtbCA1 prediction model using ...

1. Upload your CSV data

Upload your input file

Drag and drop file here

Limit 200MB per file • TXT

Browse files

Example input file

Predict

MtbCA-Selec-Pred app

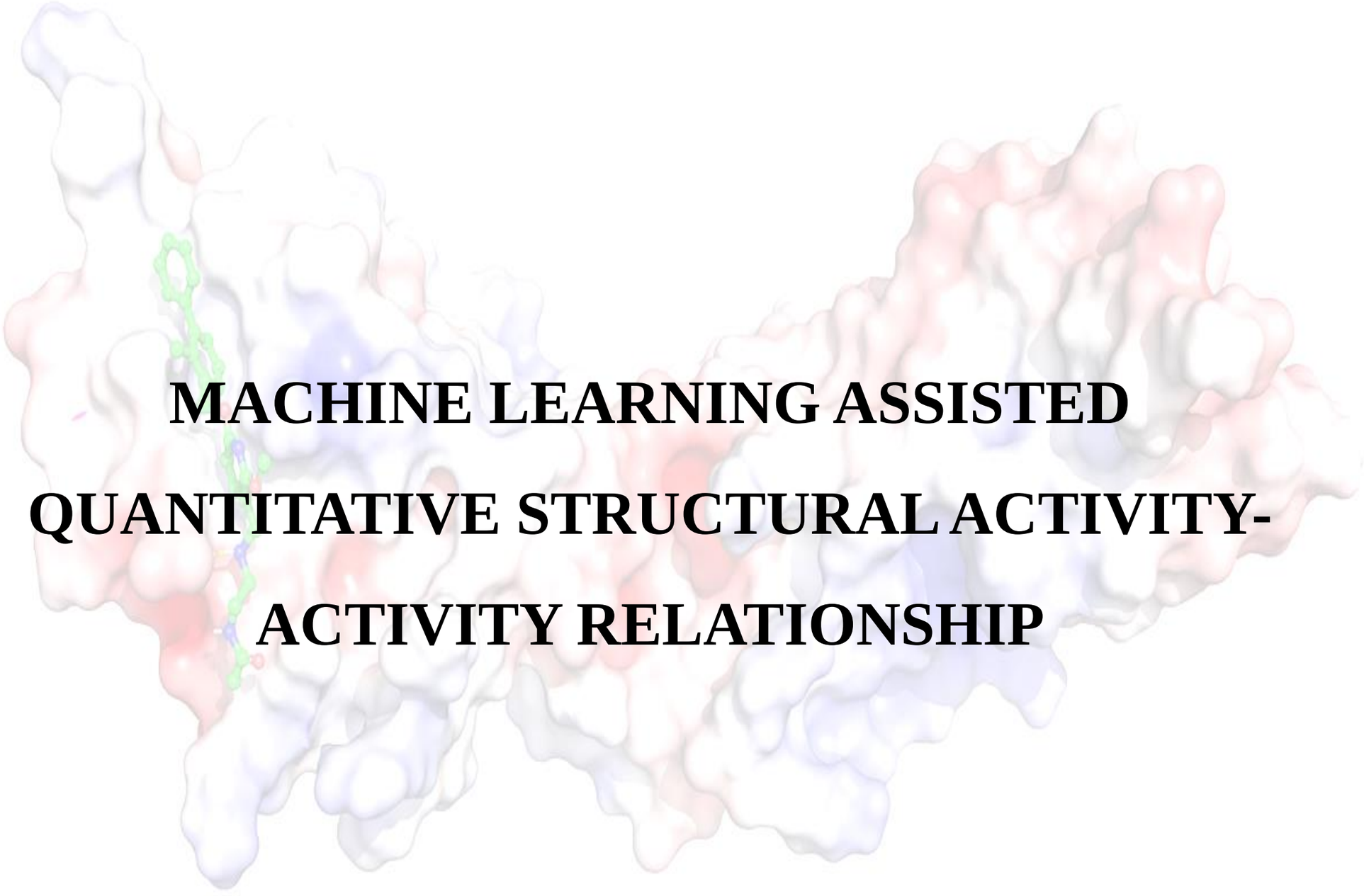
MtbCA-Selec-Pred allows users to predict bioactivity and selectivity of a query molecule separately against the Mycobacterium tuberculosis carbonic anhydrase target protein isoforms, MtbCA1 and MtbCA2

Main | About | What is Mycobacterium tuberculosis carbonic anhydrase (MtbCA)? | Dataset | Model performance | Pyt

Application Description

This module of **MtbCA-Selec-Pred** has been built to predict bioactivity and identify potent inhibitors against Mycobacterium tuberculosis carbonic anhydrases, MtbCA1 and MtbCA2 using robust machine learning algorithms.

Predict bioactivity of molecules against MtbCA1 using substructure fingerprints



MACHINE LEARNING ASSISTED QUANTITATIVE STRUCTURAL ACTIVITY-ACTIVITY RELATIONSHIP

ChEMBL

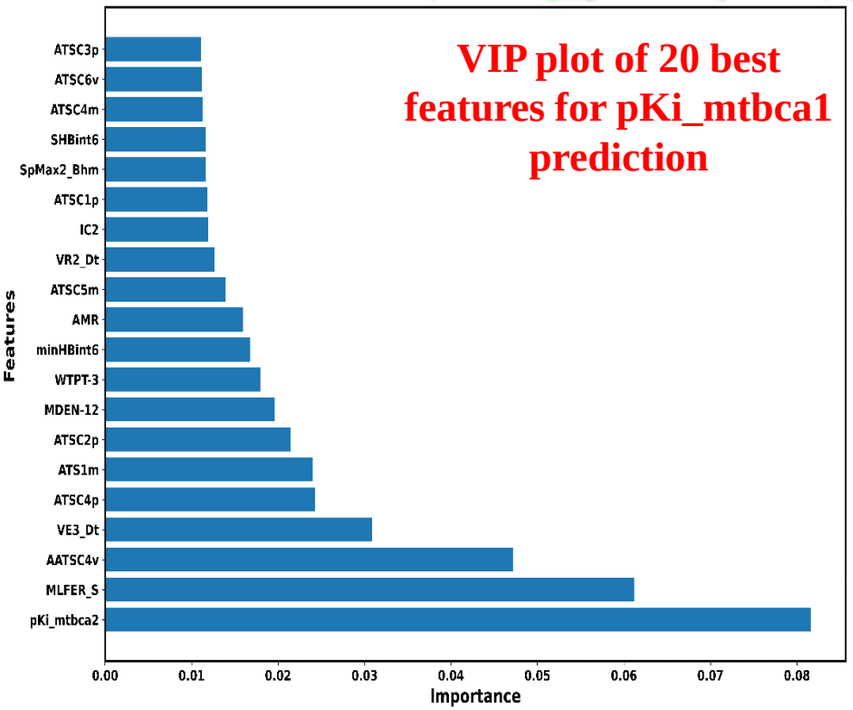
Curation of 42 molecules with activity (Ki) against both *MtbCA1* and *MtbCA2*

Calculation of 1D and 2D Molecular Descriptors (1444)

Elimination of highly correlated and constant molecular descriptors combined with VIP selection

ML-QSAAR models

VIP plot of 20 best features for pKi_mtbca1 prediction



Feature importance score >0.02

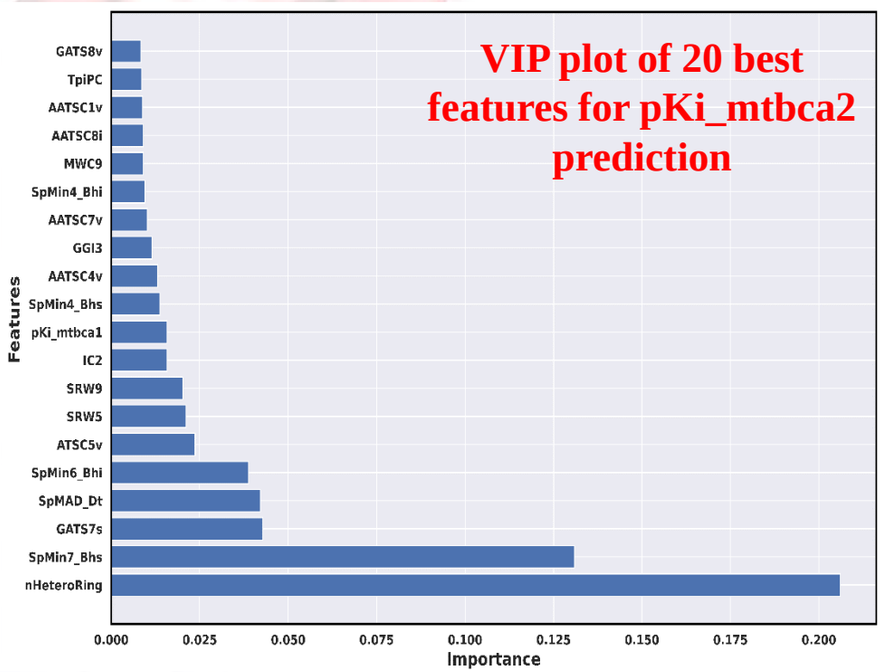
Selection of significant descriptors

Significant 1D 2D molecular descriptors + pKi_mtbca2

pKi_mtbca1

pKi_mtbca2

VIP plot of 20 best features for pKi_mtbca2 prediction



Selection of significant descriptors

Feature importance score >0.02

Significant 1D 2D molecular descriptors + pKi_mtbca1

ML-QSAAR model for bioactivity prediction of MtbCA1 inhibitors using 1D and 2D molecular descriptors

ChEMBL

Curation of 42 (MtbCA1 + *MtbCA2)* inhibitors with Ki value

Total inhibitors: 42

1D and 2D Molecular Descriptors: 1444

Prediction Parameter: Bioactivity Class ('pKi')

Dataset Splitting into training and test set

Elimination of highly correlated and constant molecular descriptors

Final number of descriptors(7), number of molecules in training(33) and test(9)

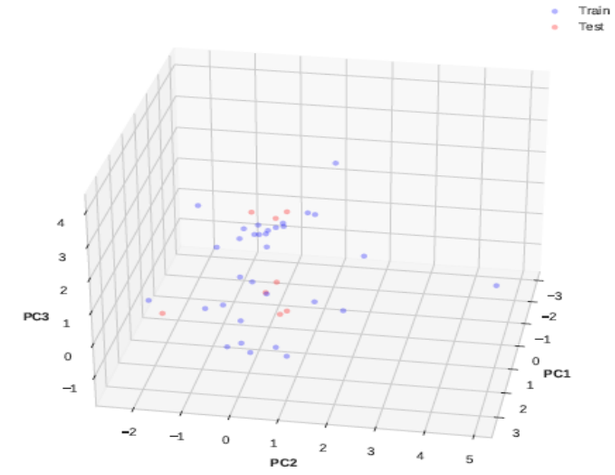
MLP Regressor

Initial Model Performance

Model Metrics
R2 (Train: 0.9187, Test: 0.4816)
RMSE (Train: 0.2439, Test: 0.6908)
MAE (Train: 0.1691, Test: 0.5028)

FP and FN Removal

Mention molecules removed (2)



Applicability Domain Analysis

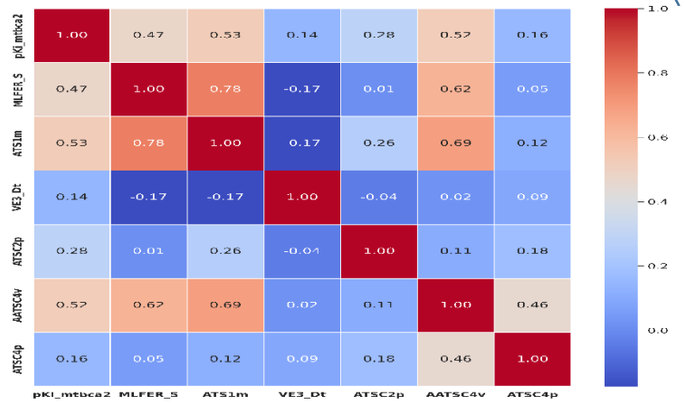
number of molecules in training(32) and test(8)

The final machine-learning model

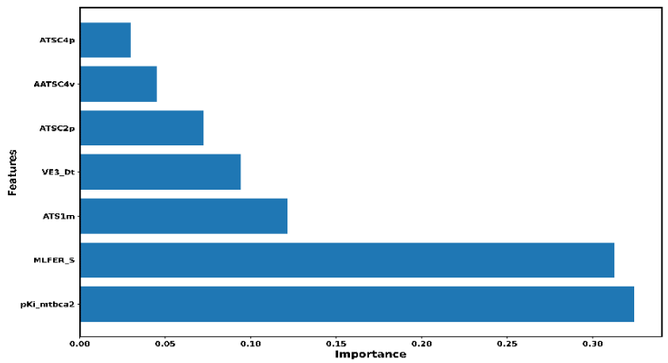
Final Model Performance

Model Metrics
R2 (Train: 0.9789, Test: 0.794)
RMSE (Train: 0.122, Test: 0.4878)
MAE (Train: 0.0904, Test: 0.3647)

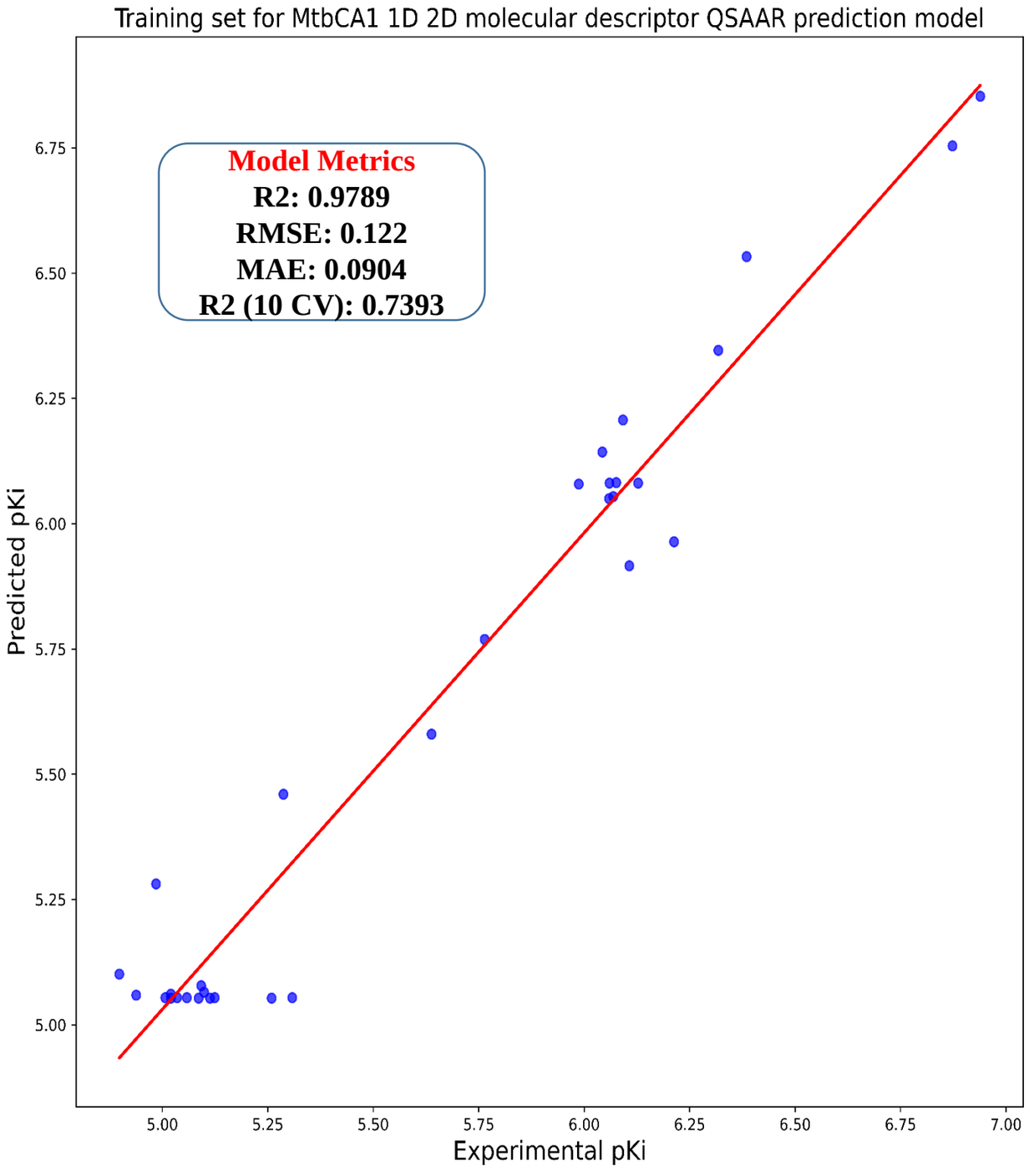
Correlation Matrix

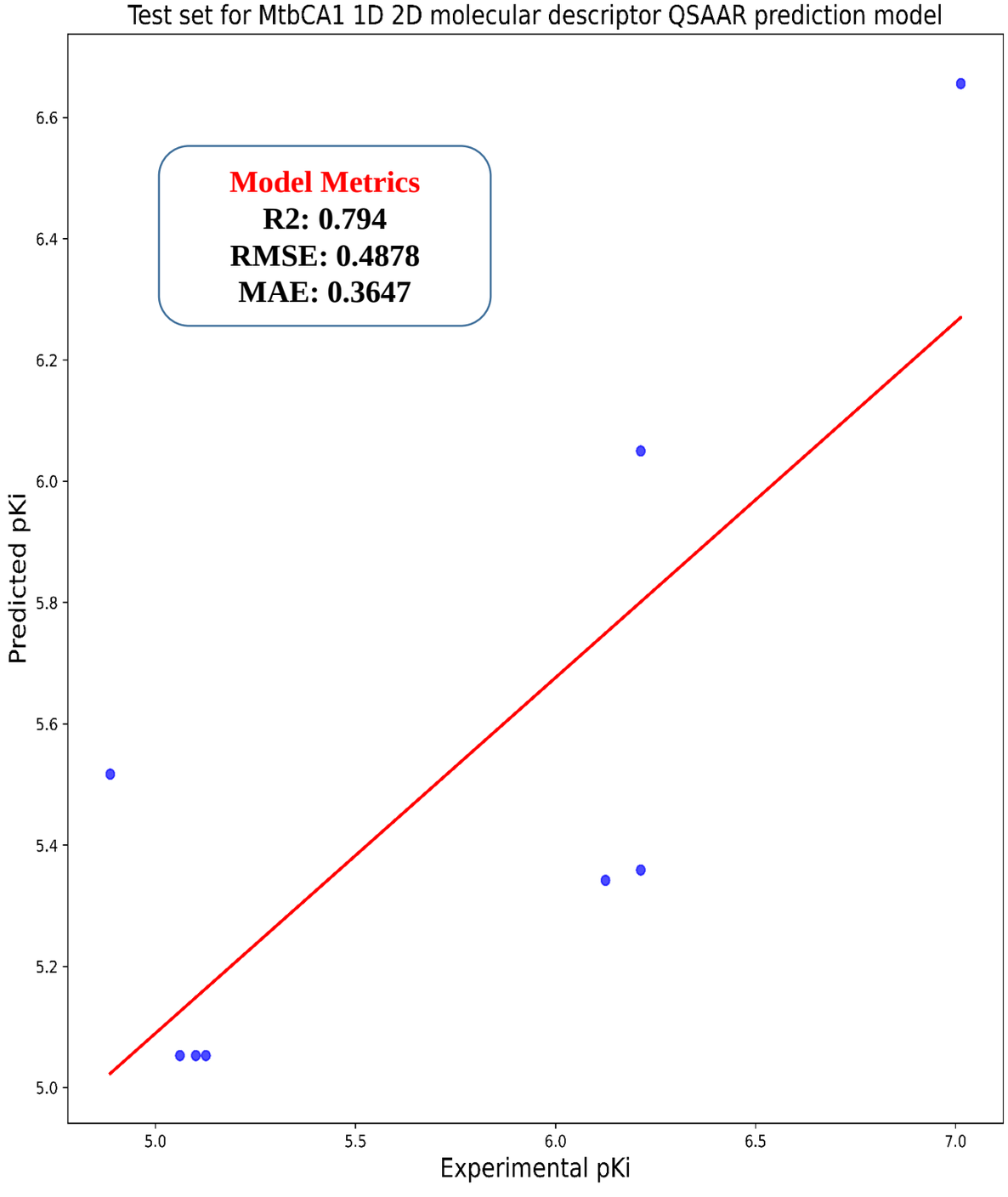


VIP Plot

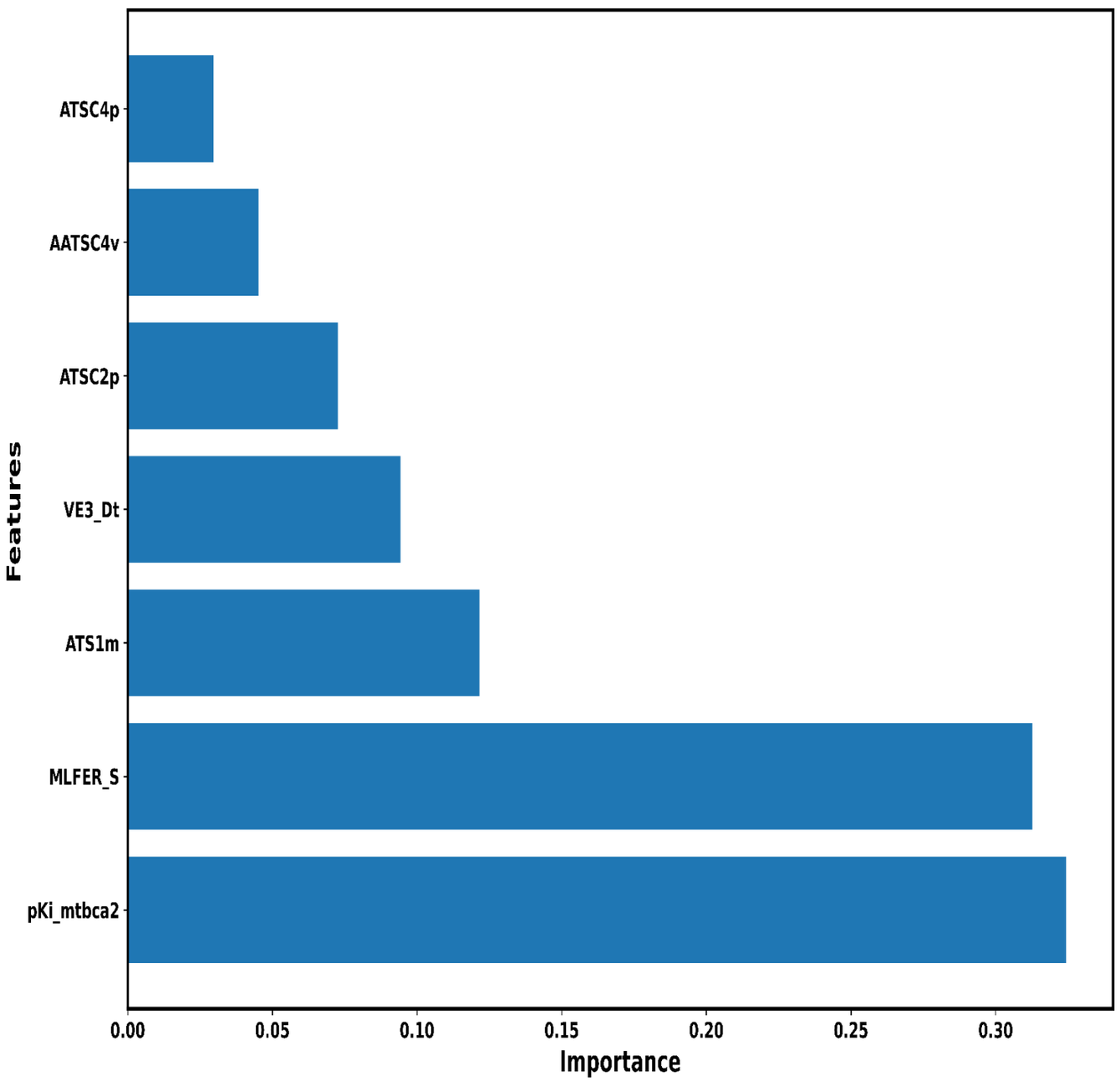


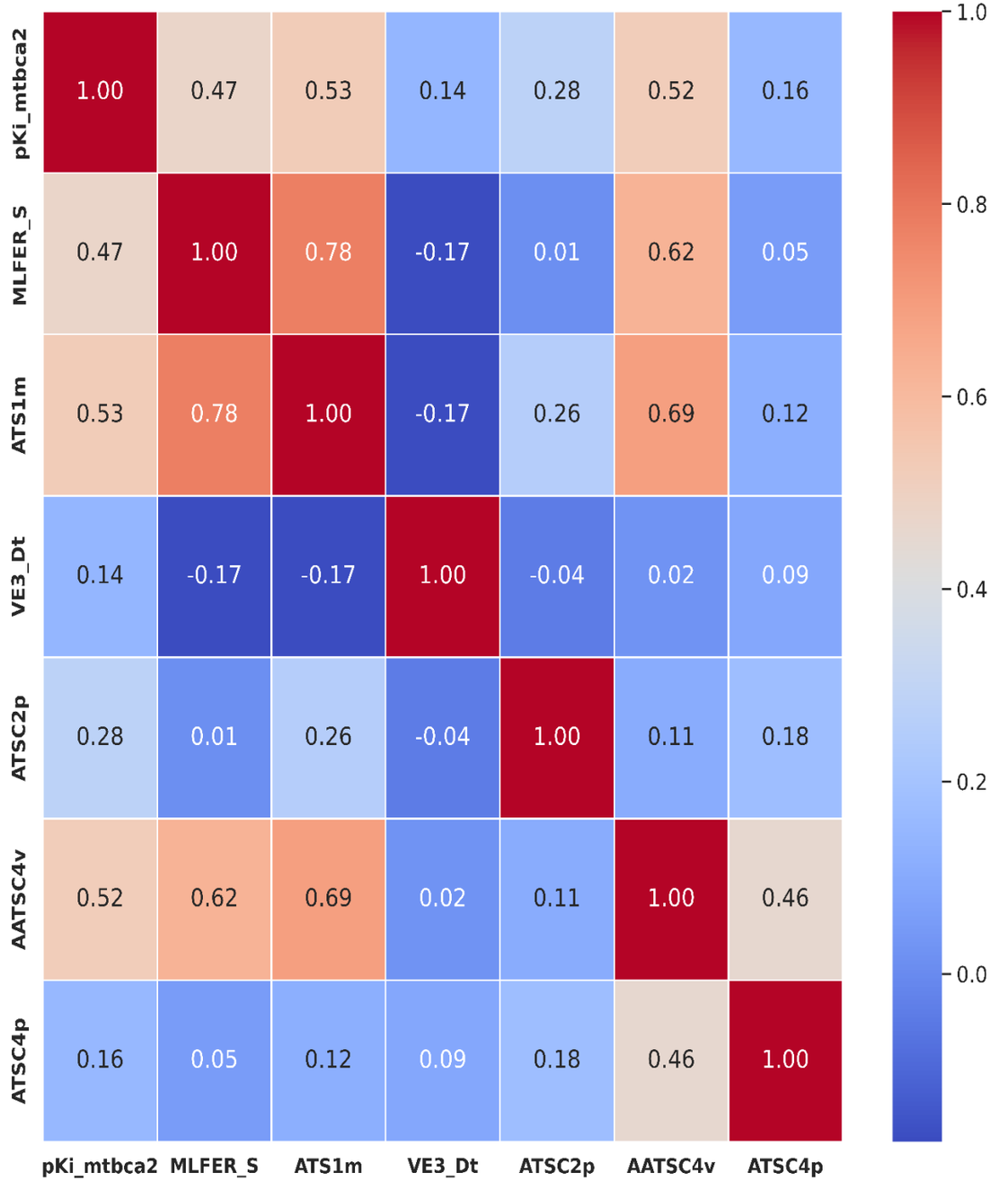
Validation Metrics


Training set for MtbCA1 1D 2D molecular descriptor QSAAR prediction model
Model Metrics
R2: 0.9789
RMSE: 0.122
MAE: 0.0904
R2 (10 CV): 0.7393
Predicted pKi
Experimental pKi
5.00
5.25
5.50
5.75
6.00
6.25
6.50
6.75
7.00


Test set for MtbCA1 1D 2D molecular descriptor QSAAR prediction model
Model Metrics
R2: 0.794
RMSE: 0.4878
MAE: 0.3647
Predicted pKi
Experimental pKi
5.0
5.2
5.4
5.6
5.8
6.0
6.2
6.4
6.6
5.0
5.5
6.0
6.5
7.0

VIP plot and correlation matrix analysis for *MtbCA1* 1D 2D molecular descriptor QSAAR prediction model





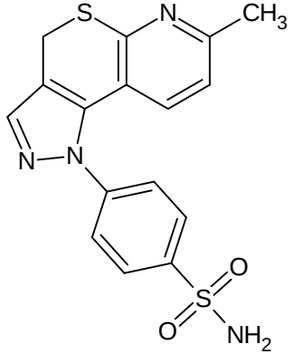
MtbCA1 inhibitors with *MtbCA2 inhibition property*

MLFER_S

pKi (*MtbCA2*)

ATS1m

Positive values



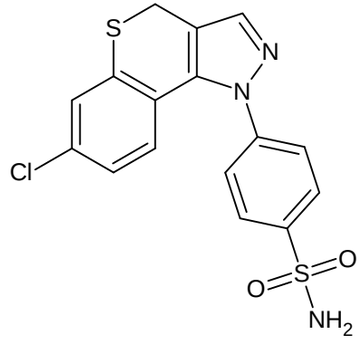
CHEMBL2179304

pKi (MtbCA1): 6.939 (Active)

pKi (MtbCA2): 6.842 (Active)

***MtbCA1* inhibition signatures**

Presence of SubFP1, SuFP179, SubFP287, SubFP100, nHBint8, SHBint8, naasC, and SHssNH



CHEMBL2179307

pKi (MtbCA1): 6.385 (Active)

pKi (MtbCA2): 6.628 (Active)

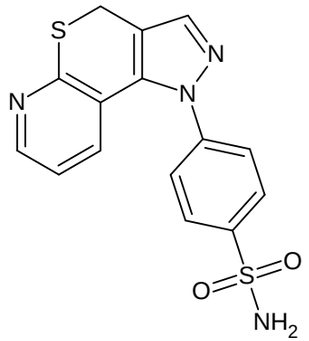
ATSC2p

ATSC4p

AATSC4v

VE3_Dt

Negative values



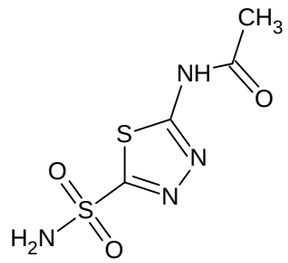
CHEMBL2179305

pKi (MtbCA1): 6.873 (Active)

pKi (MtbCA2): 6.318 (Active)

***MtbCA2* inhibition signatures**

Presence of SubFP275, SubFP28, SubFP1, SubFP183, SubFP184, minHBa, nHeteroRing, n5Ring



CHEMBL20 (Acetazolamide)

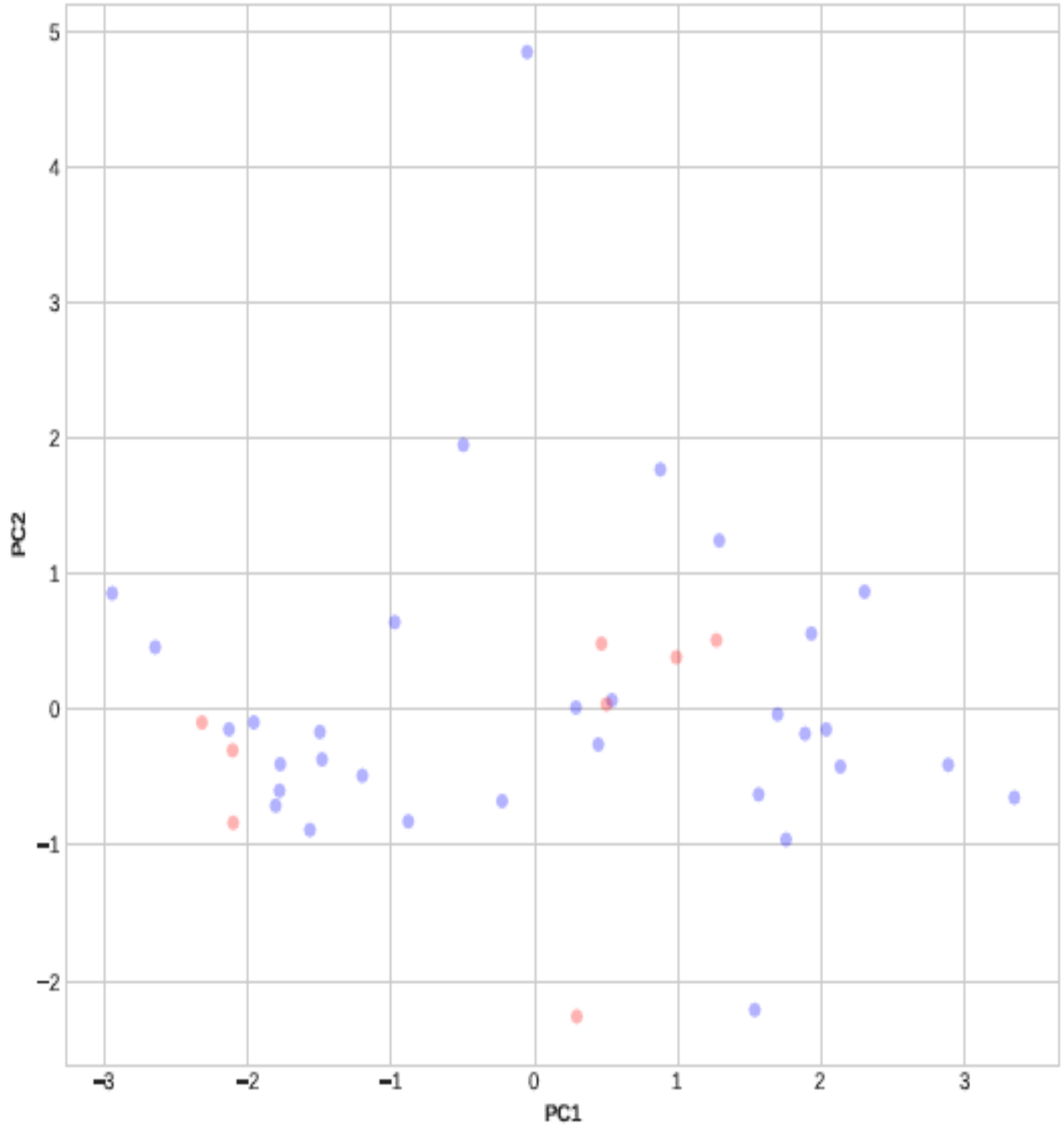
pKi (MtbCA1): 6.318 (Active)

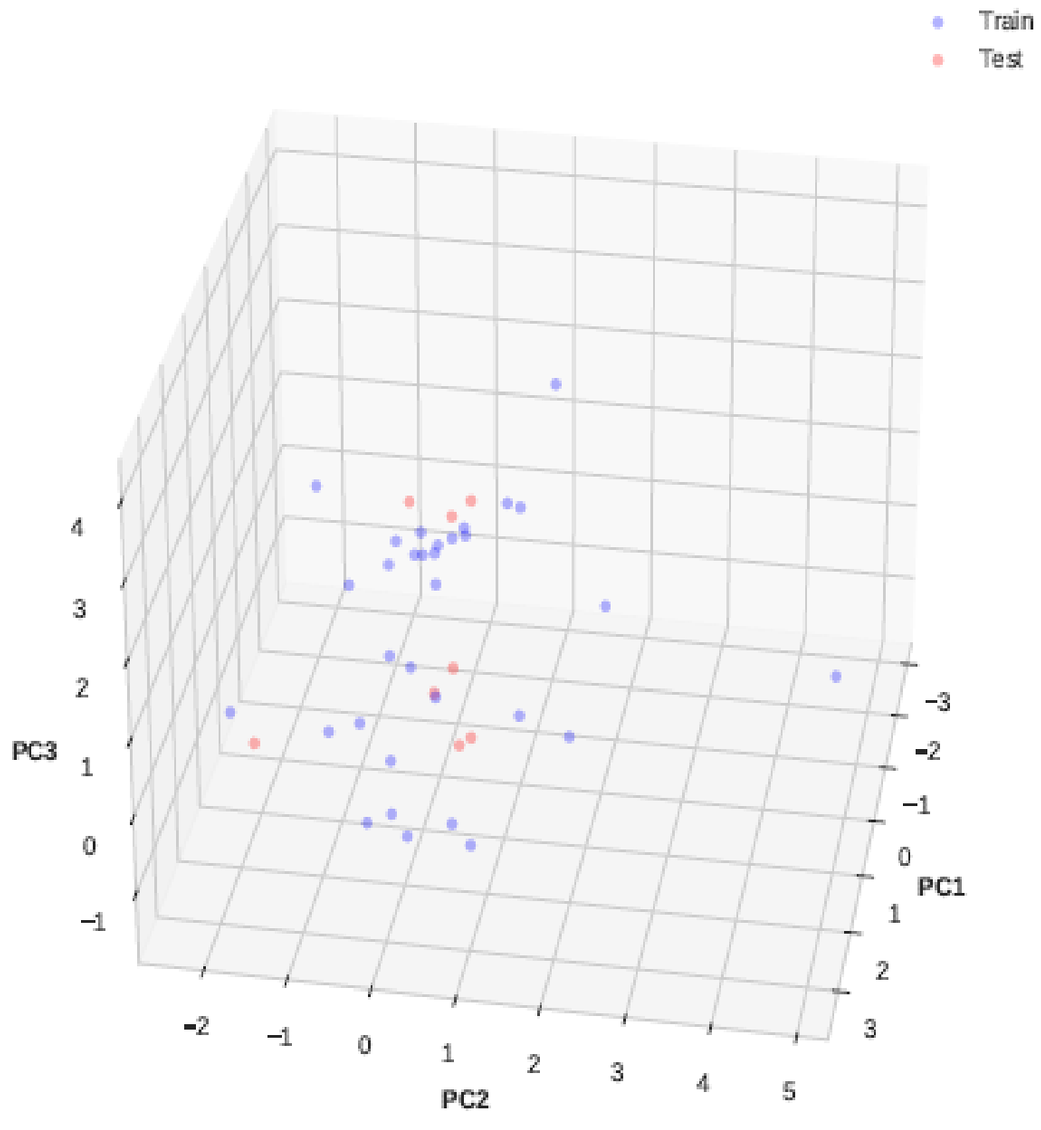
pKi (MtbCA2): 8.046 (Active)

pKi (*MtbCA2*)

pKi (*MtbCA1*)

Applicability domain analysis through 2D and 3D PCA plots for *MtbCA1* 1D 2D molecular descriptor QSAAR prediction model





ML-QSAAR model for bioactivity prediction of MtbCA2 inhibitors using 1D and 2D molecular descriptors

ChEMBL

Curation of 42 (MtbCA1 + *MtbCA2)* inhibitors with Ki value

Total inhibitors: 42

1D and 2D Molecular Descriptors: 1444

Prediction Parameter: Bioactivity Class ('pKi')

Dataset Splitting into training and test set

Elimination of highly correlated and constant molecular descriptors

Final number of descriptors(7), number of molecules in training(33) and test(9)

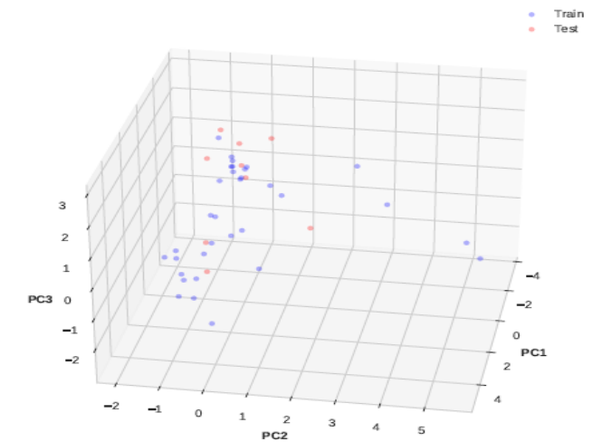
MLP Regressor

Initial Model Performance

Model Metrics
R2 (Train: 0.9916, Test: 0.8612)
RMSE (Train: 0.1585, Test: 0.7041)
MAE (Train: 0.1222, Test: 0.5742)

FP and FN Removal

Mention molecules removed (2)



Applicability Domain Analysis

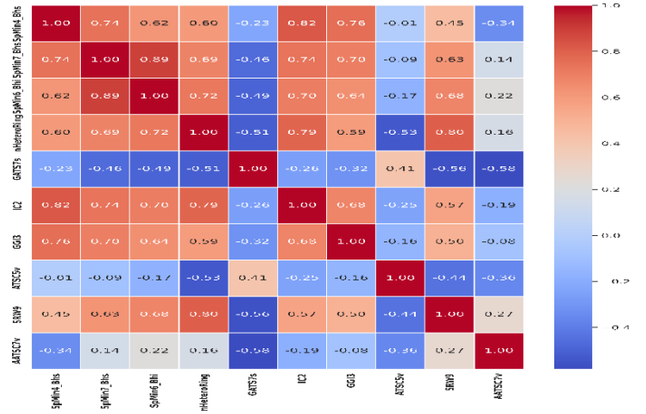
number of molecules in training(33) and test(6)

The final machine-learning model

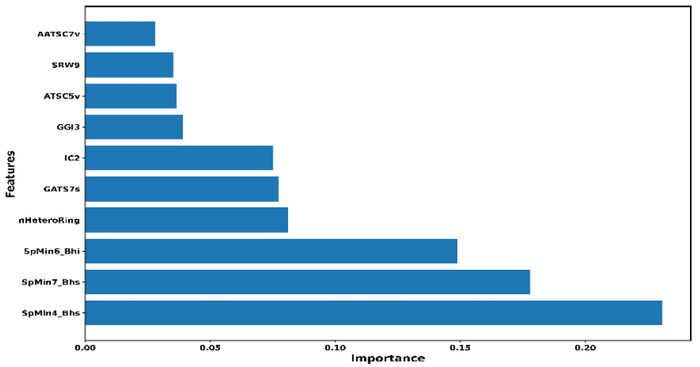
Final Model Performance

Model Metrics
R2 (Train: 0.9916, Test: 0.9713)
RMSE (Train: 0.1585, Test: 0.3641)
MAE (Train: 0.1222, Test: 0.2697)

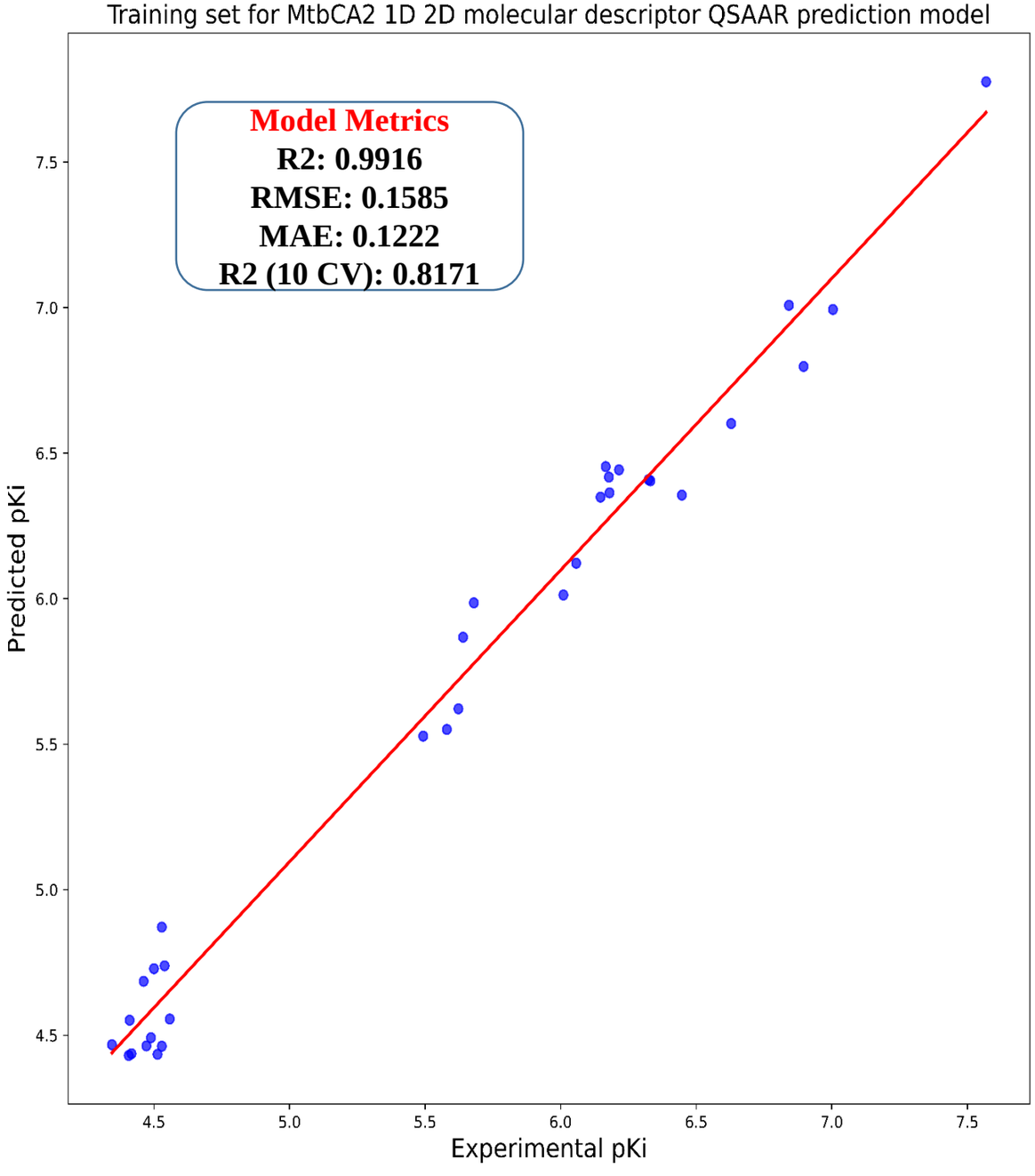
Correlation Matrix

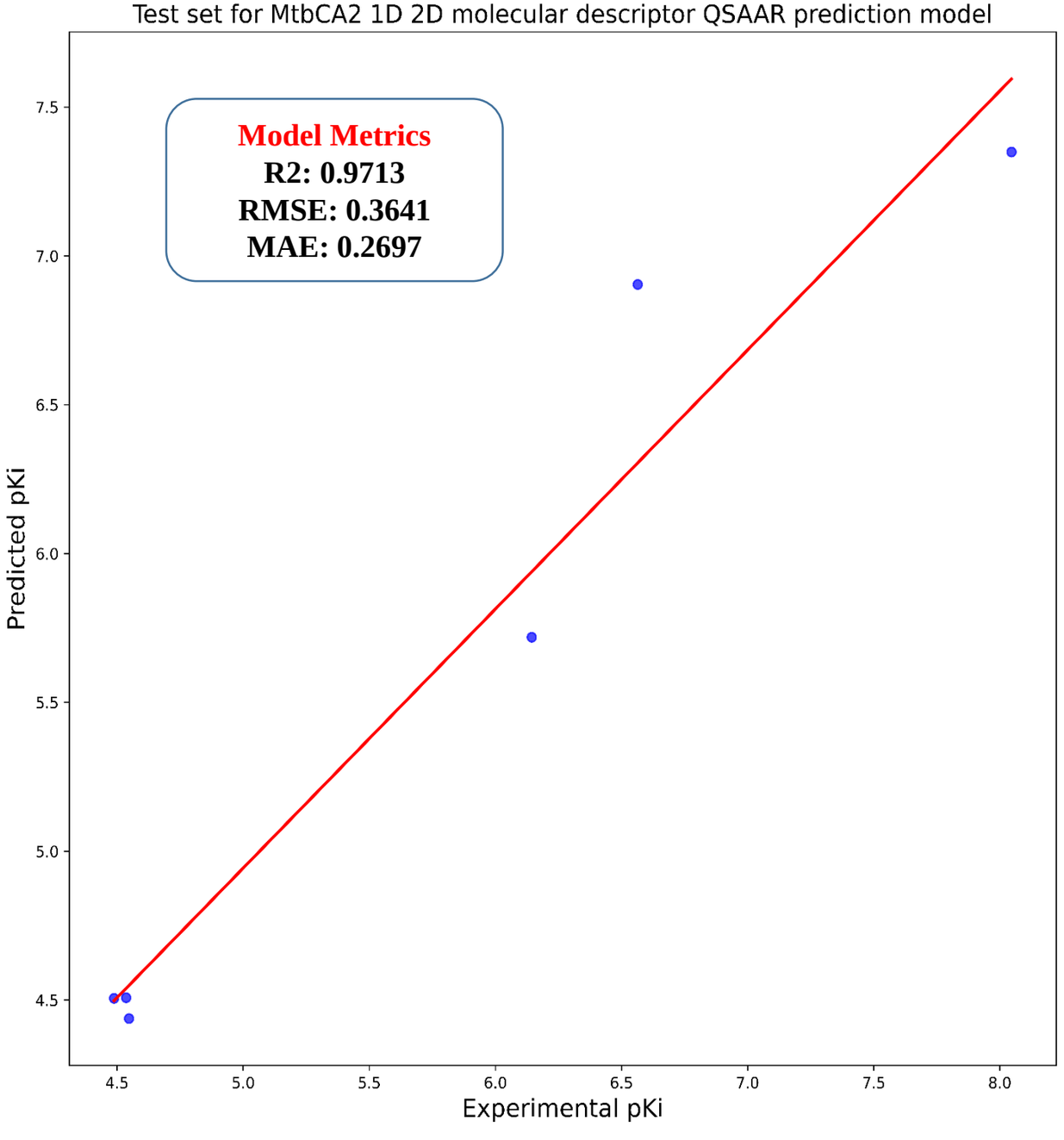


VIP Plot

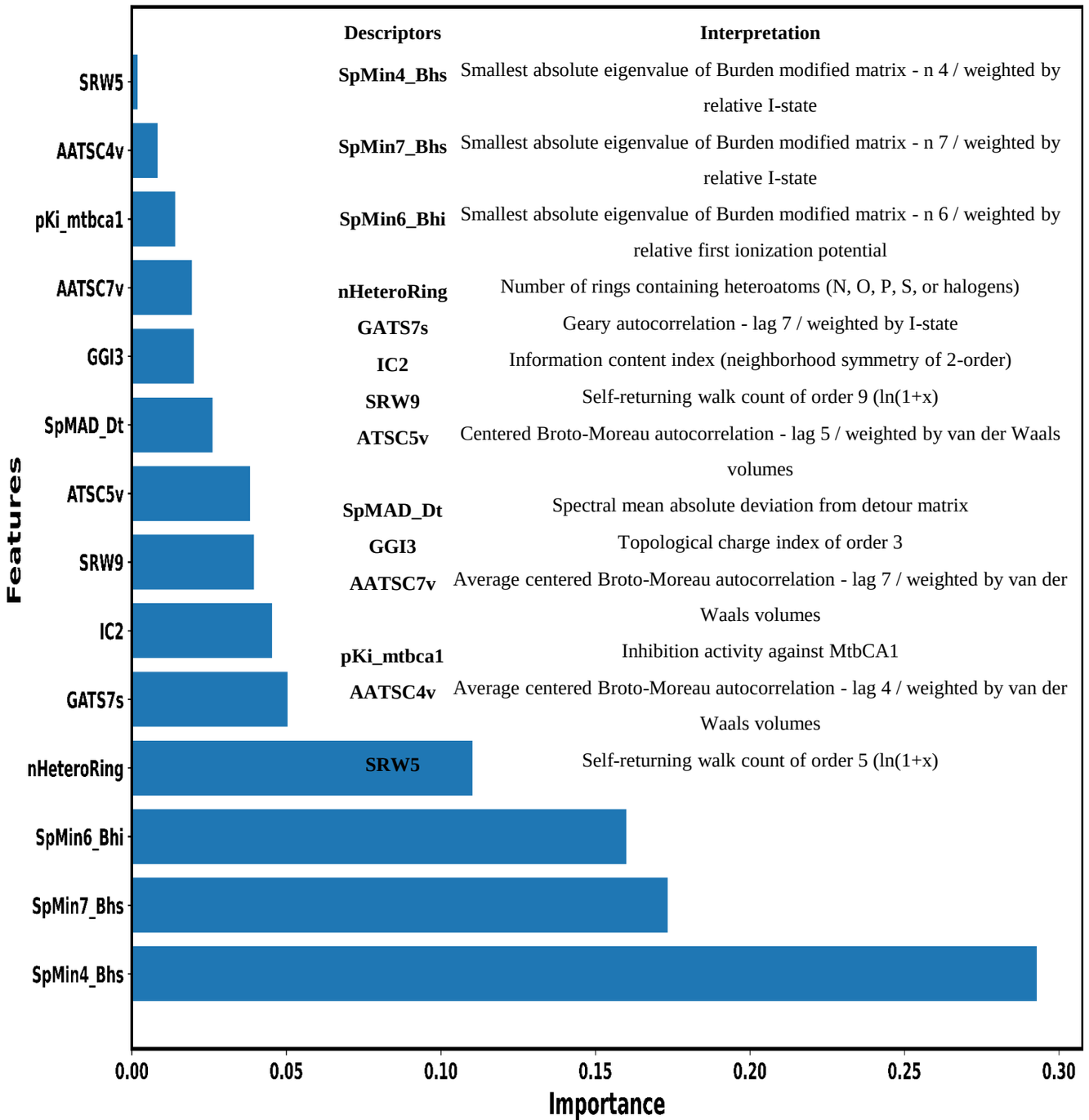


Validation Metrics

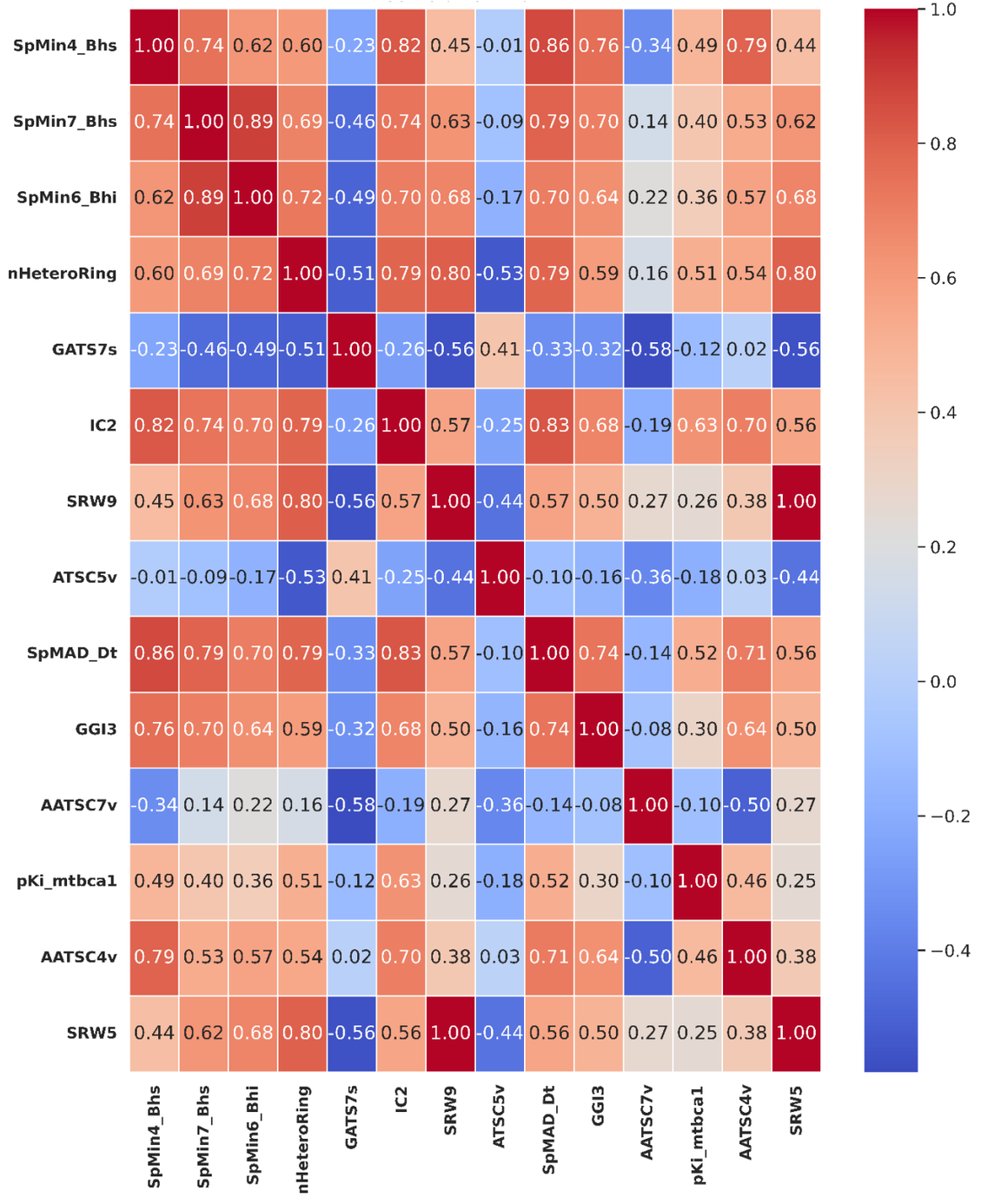

Training set for MtbCA2 1D 2D molecular descriptor QSAAR prediction model
Model Metrics
R2: 0.9916
RMSE: 0.1585
MAE: 0.1222
R2 (10 CV): 0.8171
Predicted pKi
Experimental pKi
4.5
5.0
5.5
6.0
6.5
7.0
7.5


Test set for MtbCA2 1D 2D molecular descriptor QSAAR prediction model
Model Metrics
R2: 0.9713
RMSE: 0.3641
MAE: 0.2697
Predicted pKi
Experimental pKi
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0

VIP plot and correlation matrix analysis for *MtbCA2* 1D 2D molecular descriptor QSAAR prediction model



Descriptors	Interpretation
SpMin4_Bhs	Smallest absolute eigenvalue of Burden modified matrix - n 4 / weighted by relative I-state
SpMin7_Bhs	Smallest absolute eigenvalue of Burden modified matrix - n 7 / weighted by relative I-state
SpMin6_Bhi	Smallest absolute eigenvalue of Burden modified matrix - n 6 / weighted by relative first ionization potential
nHeteroRing	Number of rings containing heteroatoms (N, O, P, S, or halogens)
GATS7s	Geary autocorrelation - lag 7 / weighted by I-state
IC2	Information content index (neighborhood symmetry of 2-order)
SRW9	Self-returning walk count of order 9 (ln(1+x)
ATSC5v	Centered Broto-Moreau autocorrelation - lag 5 / weighted by van der Waals volumes
SpMAD_Dt	Spectral mean absolute deviation from detour matrix
GGI3	Topological charge index of order 3
AATSC7v	Average centered Broto-Moreau autocorrelation - lag 7 / weighted by van der Waals volumes
pKi_mtbca1	Inhibition activity against MtbCA1
AATSC4v	Average centered Broto-Moreau autocorrelation - lag 4 / weighted by van der Waals volumes
SRW5	Self-returning walk count of order 5 (ln(1+x)



MtbCA2 inhibitors with *MtbCA1* inhibition property

GATS7s

pKi (*MtbCA1*)

SpMin6_Bhi

SpMin4_Bhs

SpMin7_Bhs

SpMAD_Dt

IC2

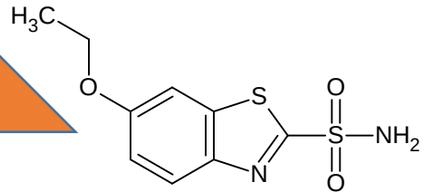
nHeteroRing

GGI3

SRW5

SRW9

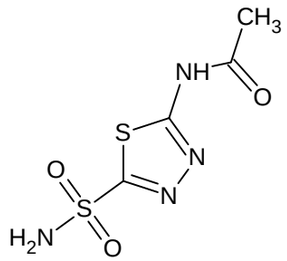
Positive values



CHEMBL18 (Ethoxzolamide)
pKi (MtbCA2): 7.569 (Active)
pKi (MtbCA1): 5.99 (Moderate)

MtbCA1 inhibition signatures

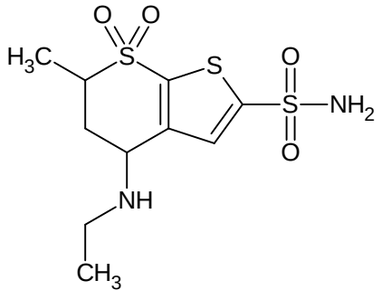
Presence of SubFP1, SuFP179, SubFP287, SubFP143, SubFP100, nHBint8, SHBint8, naasC, and SHssNH



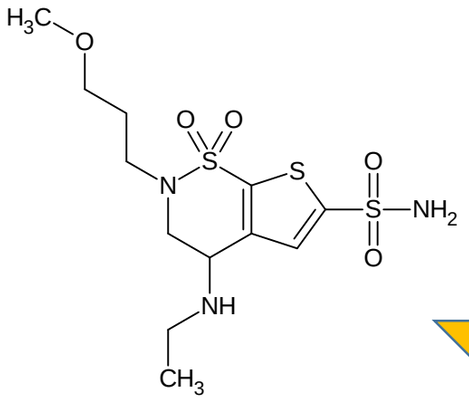
CHEMBL20 (Acetazolamide)
pKi (MtbCA2): 8.046 (Active)
pKi (MtbCA1): 6.318 (Active)

MtbCA2 inhibition signatures

Presence of SubFP275, SubFP28, SubFP1, SubFP183, SubFP184, minHBa, nHeteroRing, n5Ring



CHEMBL218490 (Dorzolamide)
pKi (MtbCA2): 7.004 (Active)
pKi (MtbCA1): 6.13 (Active)



CHEMBL220491 (Brinzolamide)
pKi (MtbCA2): 6.896 (Active)
pKi (MtbCA1): 6.076 (Active)

ATSC5v

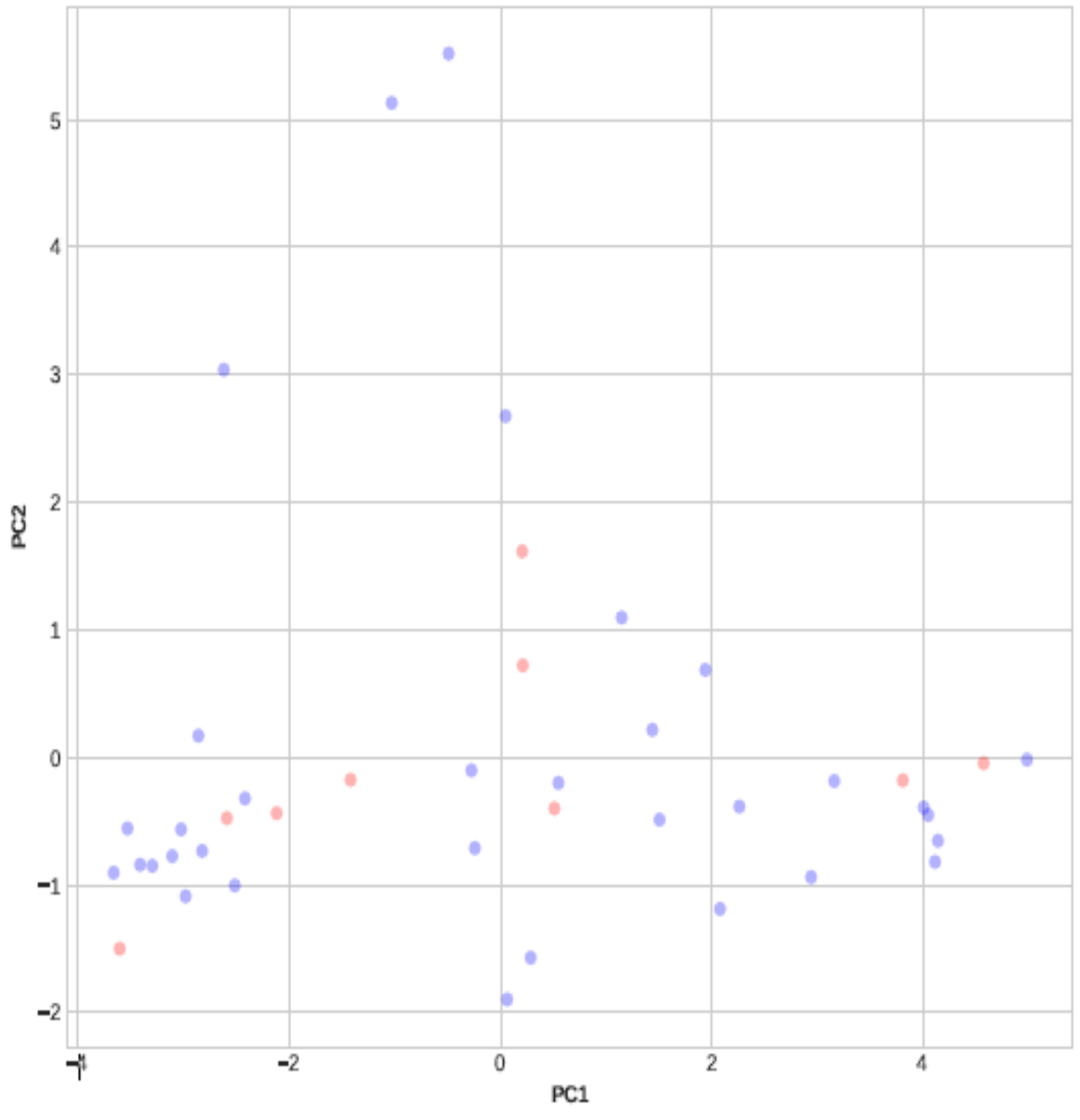
AATSC4v

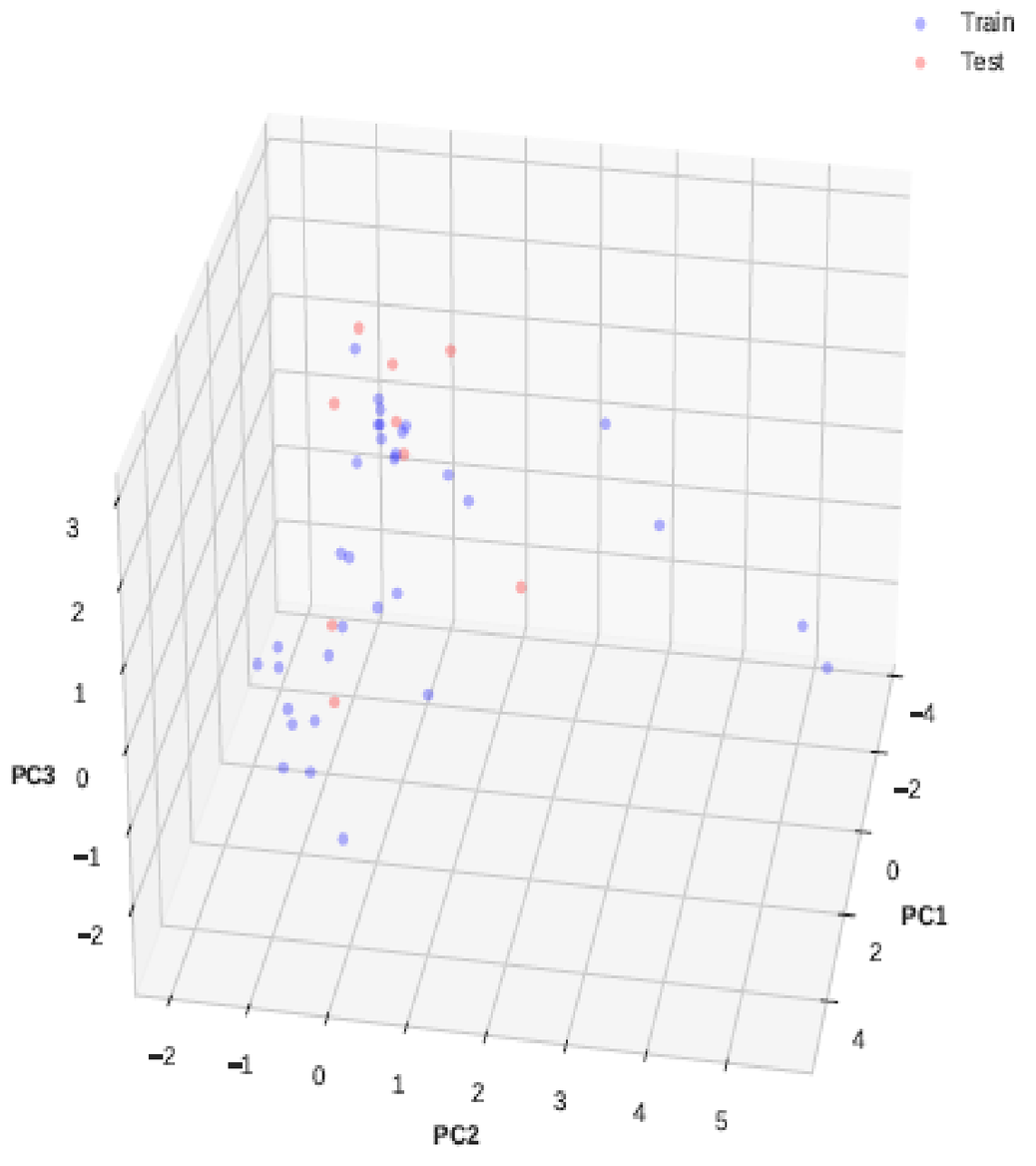
AATSC7v

Negative values

pKi (*MtbCA2*) does not always imply pKi (*MtbCA1*)

Applicability domain analysis through 2D and 3D PCA plots for *MtbCA2* 1D 2D molecular descriptor QSAAR prediction model



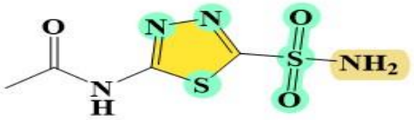


ML-QSAAR model for bioactivity prediction of MtbCA1 inhibitors using 1D and 2D molecular descriptors

CHEMBL2179304
pKi (MtbCA1): 6.939 (Active)
pKi (MtbCA2): 6.842 (Active)

CHEMBL2179305
pKi (MtbCA1): 6.873 (Active)
pKi (MtbCA2): 6.318 (Active)

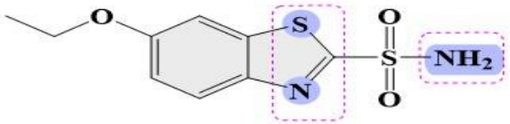
CHEMBL2179307
pKi (MtbCA1): 6.385 (Active)
pKi (MtbCA2): 6.628 (Active)



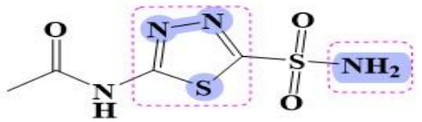
CHEMBL20 (Acetazolamide)
pKi (MtbCA1): 6.318 (Active)
pKi (MtbCA2): 8.046 (Active)

Color Code
MLFER_S
VE3_Dt
ATS1m

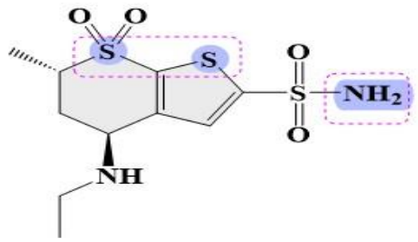
ML-QSAAR model for bioactivity prediction of MtbCA2 inhibitors using 1D and 2D molecular descriptors



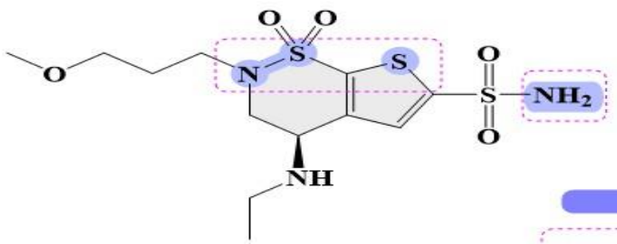
CHEMBL18 (Ethoxzolamide)
pKi (MtbCA2): 7.569 (Active)
pKi (MtbCA1): 5.99 (Moderate)



CHEMBL20 (Acetazolamide)
pKi (MtbCA1): 6.318 (Active)
pKi (MtbCA2): 8.046 (Active)

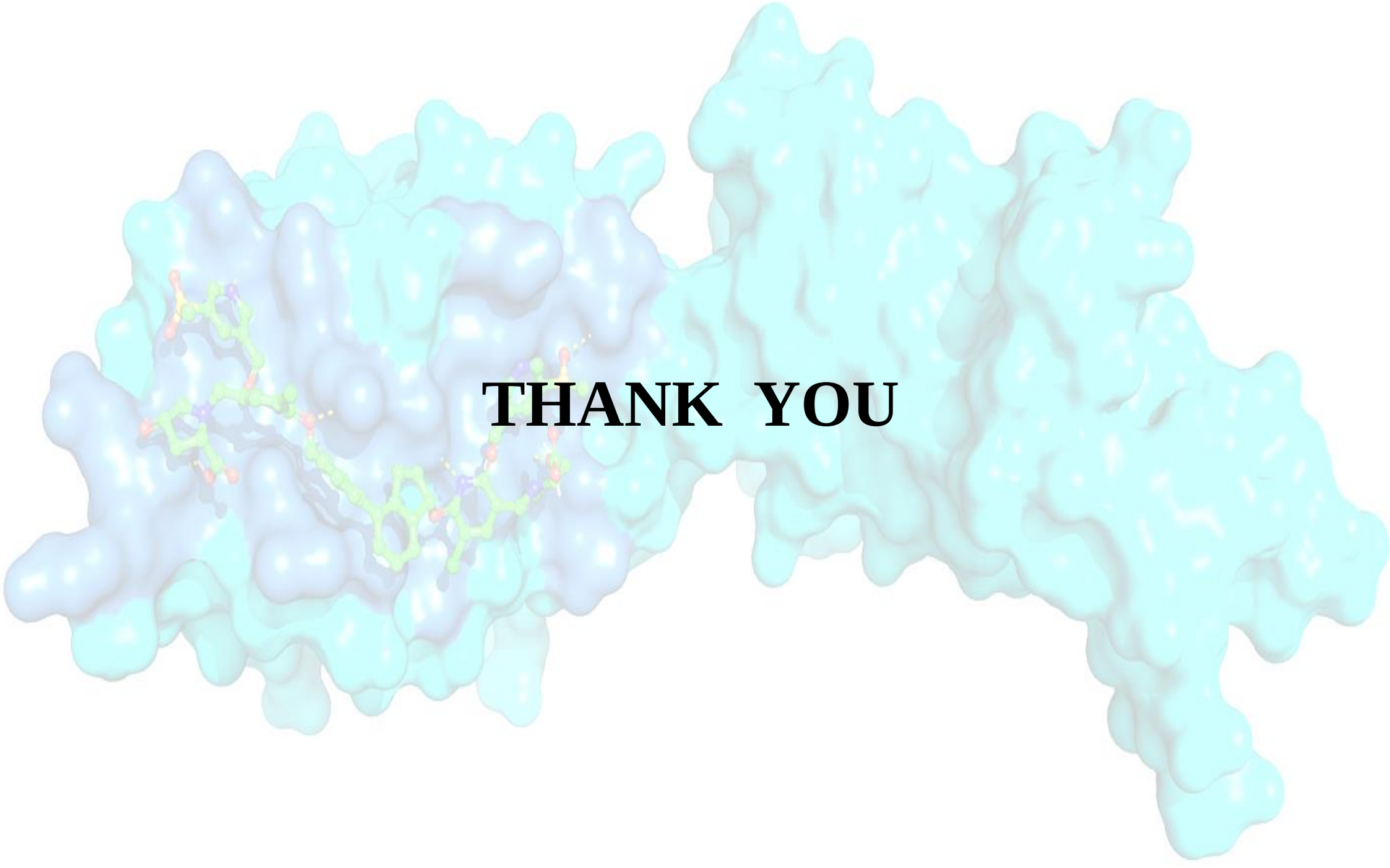


CHEMBL218490 (Dorzolamide)
pKi (MtbCA2): 7.004 (Active)
pKi (MtbCA1): 6.13 (Active)



CHEMBL220491 (Brinzolamide)
pKi (MtbCA2): 6.896 (Active)
pKi (MtbCA1): 6.076 (Active)

Color Code
GGI3
SpMin4_Bhs, SpMin7_Bhs, and SpMin6_Bhs
nHeteroRing



THANK YOU