To bind or nor to bind? A comprehensive characterization of TIR1 and auxins using consensus *in silico* approaches.

Fernando D. Prieto-Martínez[1], Jennifer Mendoza-Cañas[2] and Karina Martínez-Mayorga[1,3]*

1 Instituto de Química, Unidad Mérida, Universidad Nacional Autónoma de México, Carretera Mérida-Tetiz, Km. 4.5, Ucú, Yucatán, México; ferdpm4@hotmail.com

2 Facultad de Química, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico; kmtzm@unam.mx

3 Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Unidad Mérida, Universidad Nacional Autónoma de México, Sierra Papacal, Mérida, Yucatán, México; kmtzm@unam.mx

**Correspondence: ferdpm4@hotmail.com; (F.D.P.M)*

Table of contents

Figure S9. Performance assessment of extreme gradient boost classifier trained with physicochemical and structural descriptors.

Figure S10. Comparison of IAA root mean square fluctuations (RMSF) from different solvation conditions.

Figure S11. Comparison of NAA root mean square fluctuations (RMSF) from different solvation conditions.

Figure S12. Comparison of 2,4-D root mean square fluctuations (RMSF) from different solvation conditions.

Figure S13. Comparison of TRP root mean square fluctuations (RMSF) from different solvation conditions.

Figure S14. Consensus analysis of MD simulation of indol acetic acid (IAA).

Figure S15. Consensus analysis of MD simulation of tryptophan (TRP).

Figure S16. Consensus analysis of MD simulation of α-naphtyl acetic acid (NAA).

Figure S17. Consensus analysis of MD simulation of β-napthyl acetic acid (2-NAA).

Figure S18. Consensus analysis of MD simulation of fluroxypy (FXY).

Figure S19. Consensus analysis of MD simulation of rinskor (RSK).

Figure S20. Free energy surfaces computed from well-tempered metadynamics for the NAA/TIR1 complexes.

Figure S21. Representative NAA conformations found at basins from metadynamics runs.

Figure S22. Free energy surfaces computed from well-tempered metadynamics for the 2-NAA/TIR1 complexes.

Figure S23. Representative 2-NAA conformations found at basins from metadynamics runs.

Figure S24. Free energy surfaces computed from well-tempered metadynamics for the FXY/TIR1 complexes.

Figure S25. Representative FXY conformations found at basins from metadynamics runs.

Figure S26. Free energy surfaces computed from well-tempered metadynamics for the IAA/TIR1 complexes.

Figure S27. Representative IAA conformations found at basins from metadynamics runs.

Figure S28. Free energy surfaces computed from well-tempered metadynamics for the RSK/TIR1 complexes.

Figure S29. Representative RSK conformations found at basins from metadynamics runs.

Figure S30. Free energy surfaces computed from well-tempered metadynamics for the TRP/TIR1 complexes.

Figure S31. Representative TRP conformations found at basins from metadynamics runs.

Table S1. Parameters used for OpenGrowth in the present study.

Parameter	Value
SCORING_FUNCTION	SMOG2016
BINDING_SITE_X	7.535
BINDING_SITE_Y	-111.72
BINDING_SITE_Z	-25.923
BINDINGBOX_SIZE	0.50
MODE	DENOVO
GROWTH_MODE	FOG
VDW_SCALE_INTER	0.70
VDW_SCALE_INTRA	0.70
ROTATION_PRECISION	24
OPTIMIZATION_MODE	22
OPTIMIZATION_NUMBER	20
OPTIMIZATION_ITERATIONS	10
MAX_FRAGMENTS	20
MAX_ATOMS	60
MAX_MW	550
MAX_ITERATIONS	10
AVERAGE_TYPE	ARITHMETIC

Table S2. Molecular descriptors selected by the Boruta algorithm.

Descriptor	Description
nAromBond	aromatic bonds count
ATS8se	Moreau-Broto autocorrelation of lag 8
AATS1d	averaged Moreau-Broto autocorrelation of lag 1 weighted by sigma electrons
ATSC7i	centered Moreau-Broto autocorrelation of lag 7 weighted by ionization potential
SdssC	Sum of electrotopological state indexes (dssC)
SdO	Sum of dO
ATS6d	Moreau-Broto autocorrelation of lag 6 weighted by sigma electrons

Table S3. Loading values from principal component analyses for the Boruta defined space and RO5 space.

PC#	Boruta space	RO5 space
1	26.67	46.65
2	16.57	25.44
3	14,72	14.82
4	14.24	8.32
5	13.5	3.17
6	8,90	1.6
7	5.4	---

Table S4. Hyperparameter values determined by grid search for the trained classifiers only with physicochemical descriptors.

Classifier	Hyperparameters	Values
Logistic Regression	penalty, solver, C, class_weight	l2, lbfgs, 10000, balanced
Support vector classifier	kernel, C, gamma	rbf, 1, 0.1
Random Forest classifier	criterion, max_features, min_samples_leaf, min_samples_split, n_estimators	gini, auto, 4, 2, 1000
XGboost classifier	n_estimators, booster	50, gblinear

Table S4. Hyperparameter values determined by grid search for the trained classifiers with physicochemical and structural descriptors.

Classifier	Hyperparameters	Values
Logistic Regression	penalty, solver, C, class_weight	l2, newton-cg,10000, balanced
Support vector classifier	kernel, C	linear, 1
Random Forest classifier	criterion, max_features, min_samples_leaf, min_samples_split, n_estimators	entropy, sqrt, 1, 10, 800
XGboost classifier	n_estimators, booster	200, dart

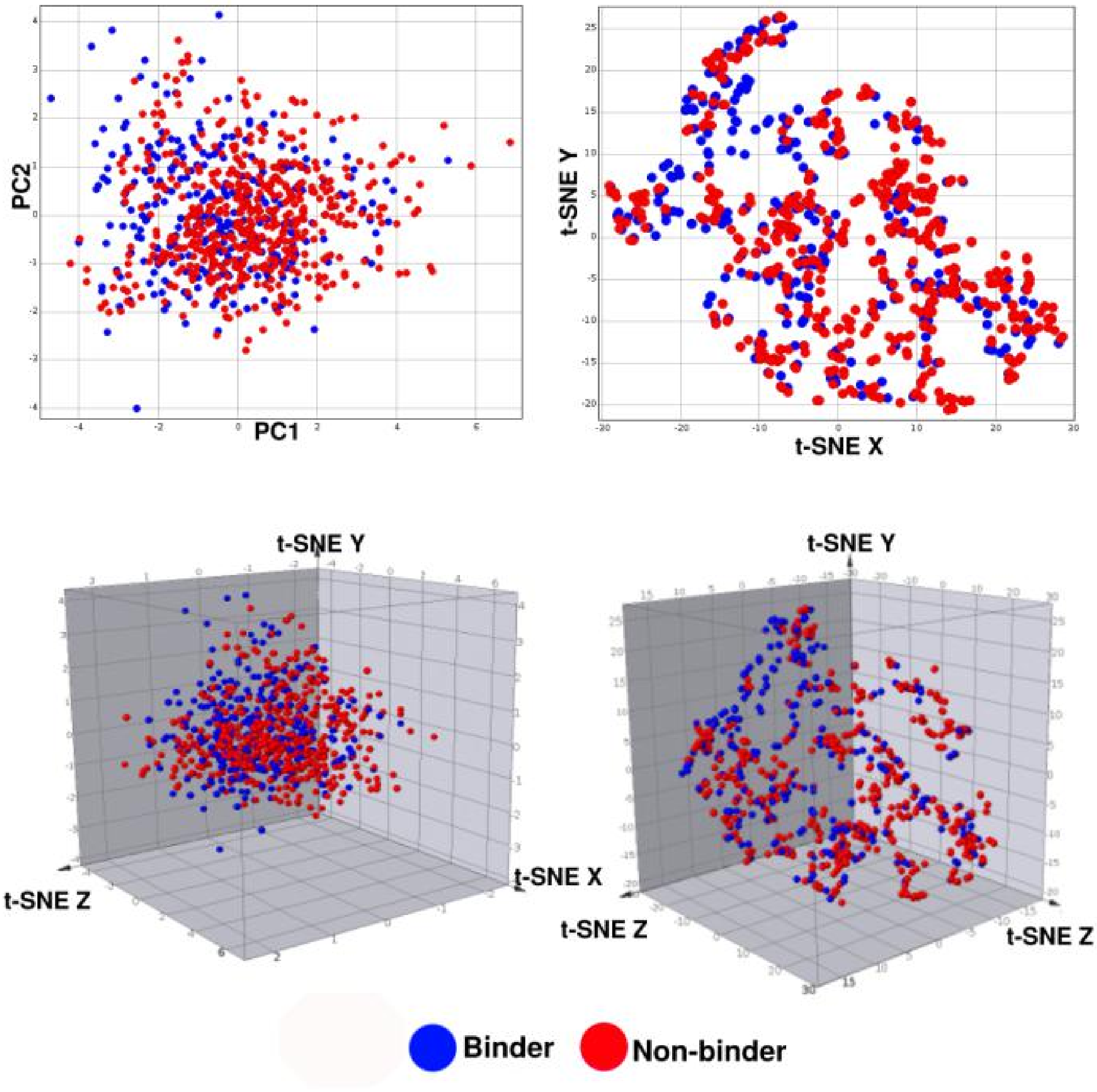

Figure S1. Chemical space of the generated compounds; using descriptors from the Rule of Five. Projection comes from principal component analysis (left) and t-SNE (right).

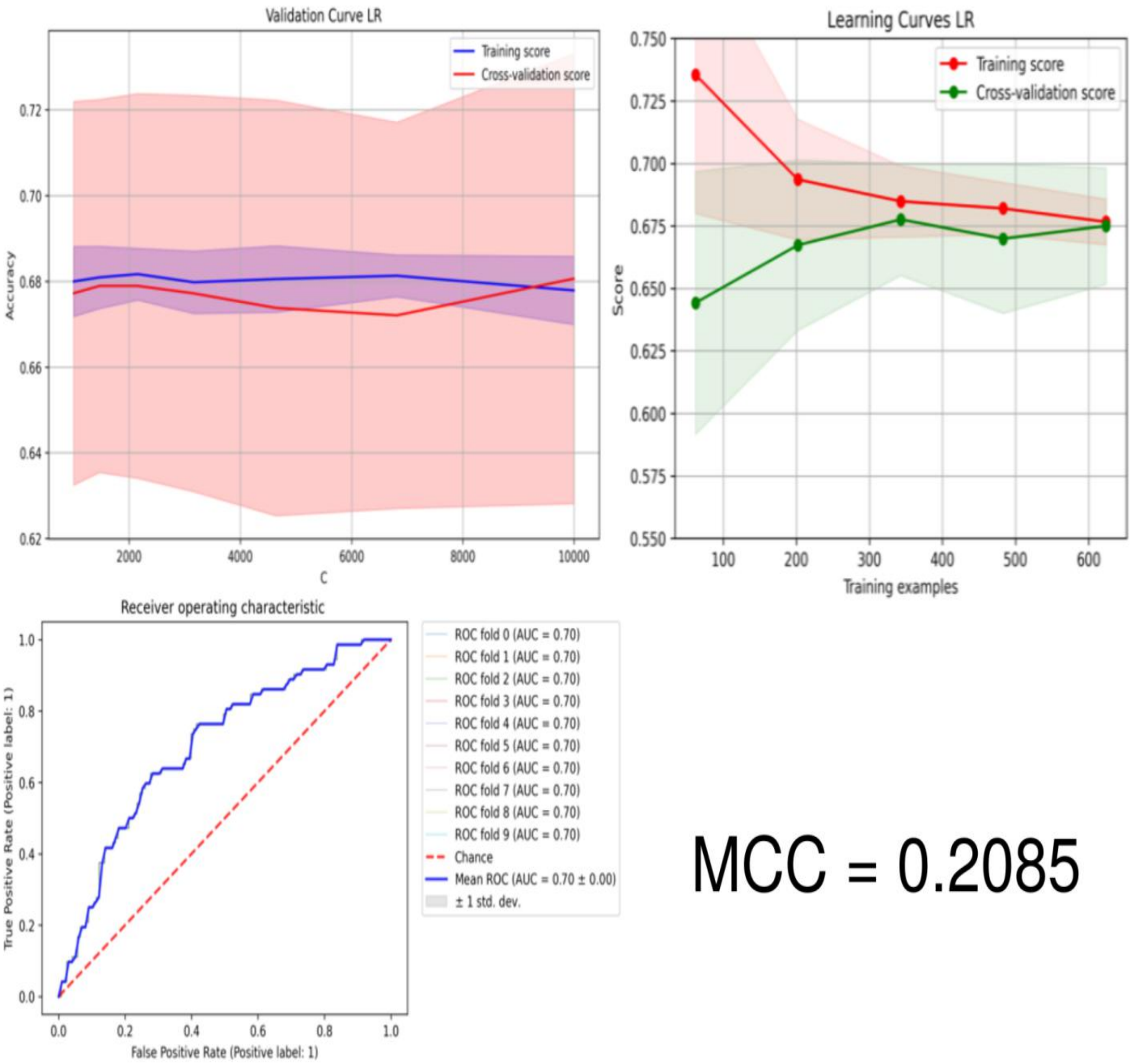

Figure S2. Performance assessment of logistic regression classifier trained with physicochemical descriptors. In order of appereance figure shows model's validation curve, learning curve, k-fold cross validation ROC curve and Matthews correlation coefficient.

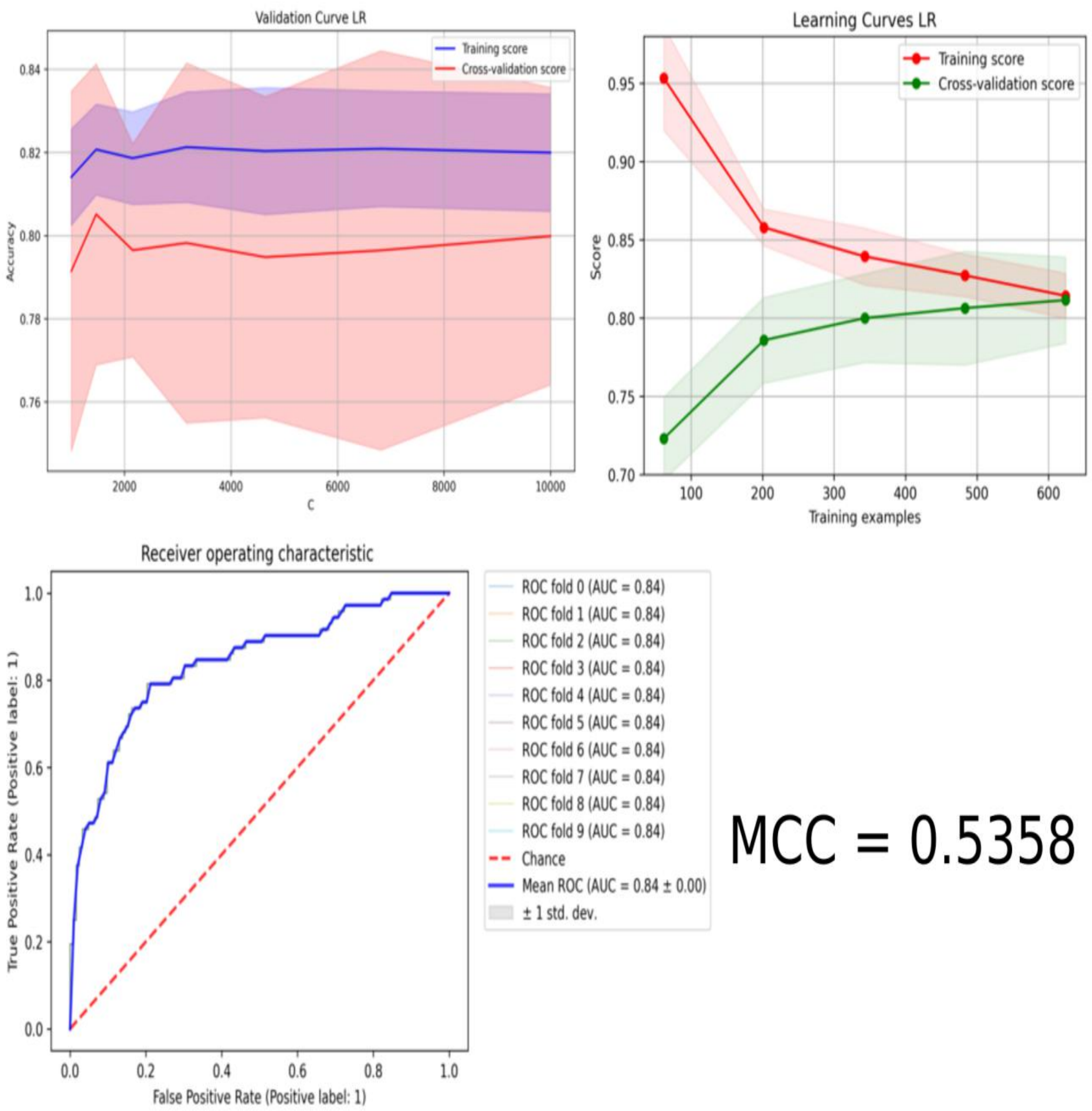

Figure S3. Performance assessment of logistic regression classifier trained with physicochemical and structural descriptors. In order of appereance figure shows model's validation curve, learning curve, k-fold cross validation ROC curve and Matthews correlation coefficient.

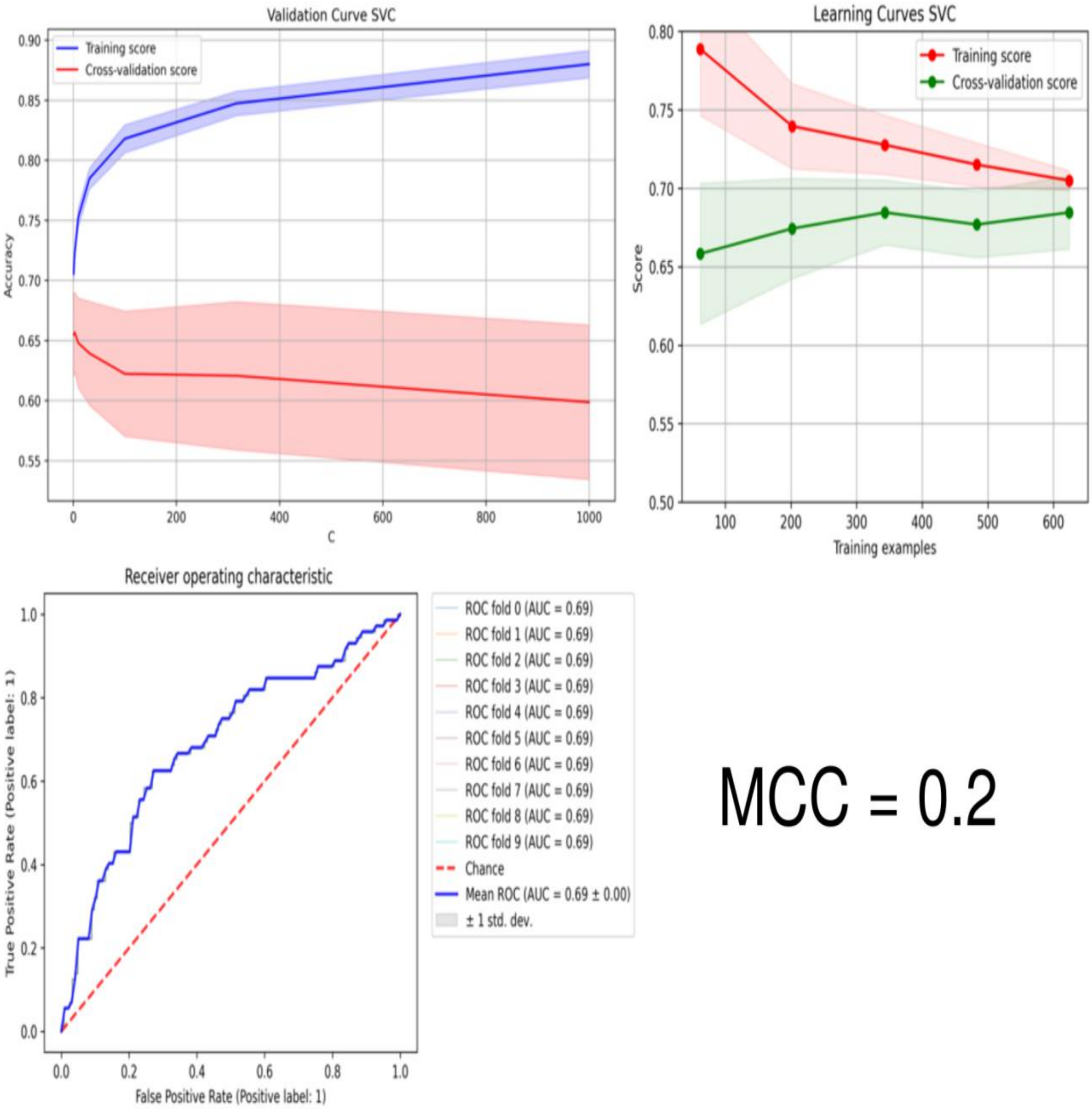

Figure S4. Performance assessment of supporting vectors classifier trained with physicochemical descriptors. In order of appereance figure shows model's validation curve, learning curve, k-fold cross validation ROC curve and Matthews correlation coefficient.

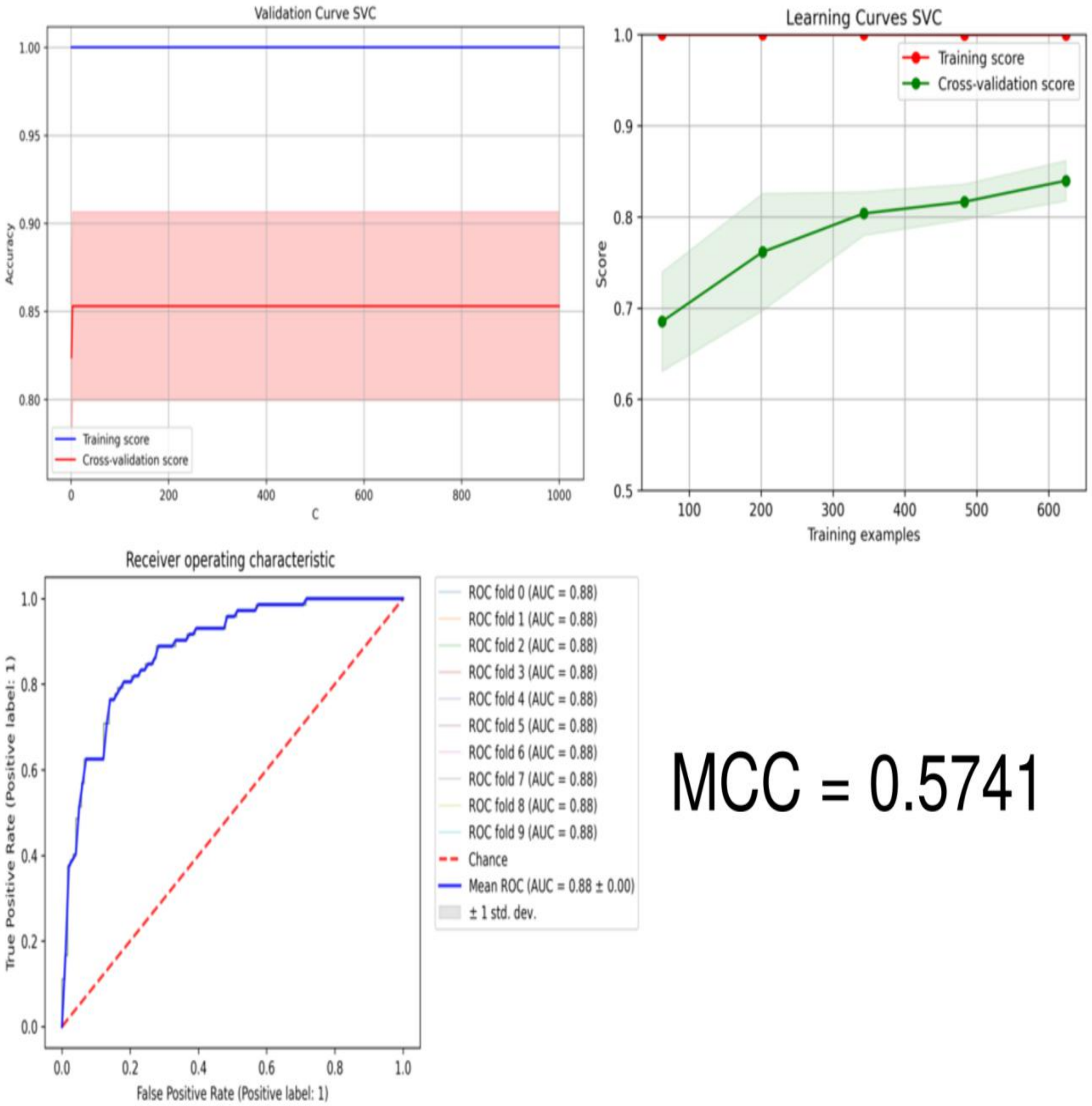

Figure S5. Performance assessment of supporting vectors classifier trained with physicochemical descriptors. In order of appereance figure shows model's validation curve, learning curve, k-fold cross validation ROC curve and Matthews correlation coefficient.

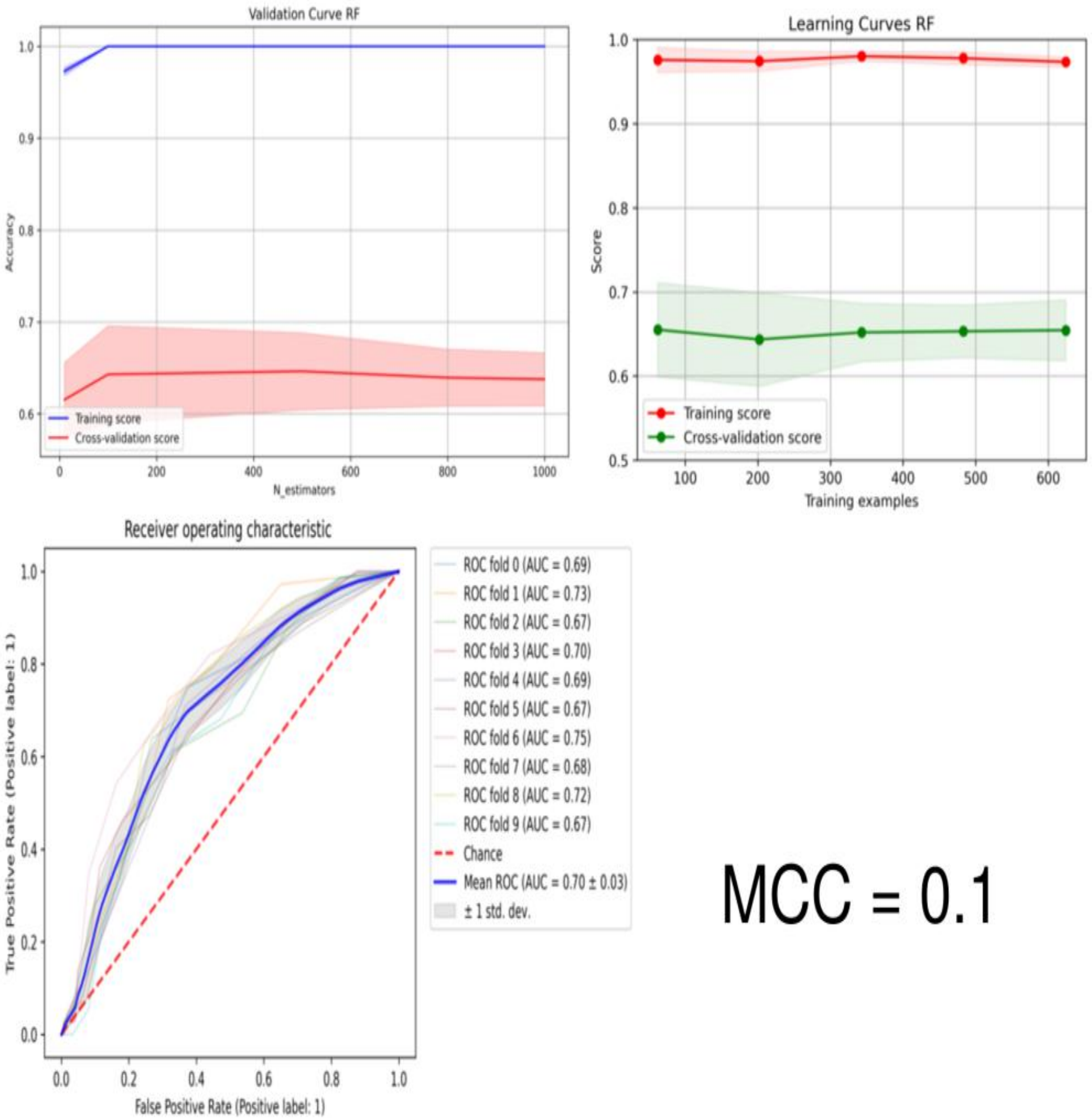

Figure S6. Performance assessment of random forest classifier trained with physicochemical descriptors. In order of appereance figure shows model's validation curve, learning curve, k-fold cross validation ROC curve and Matthews correlation coefficient.

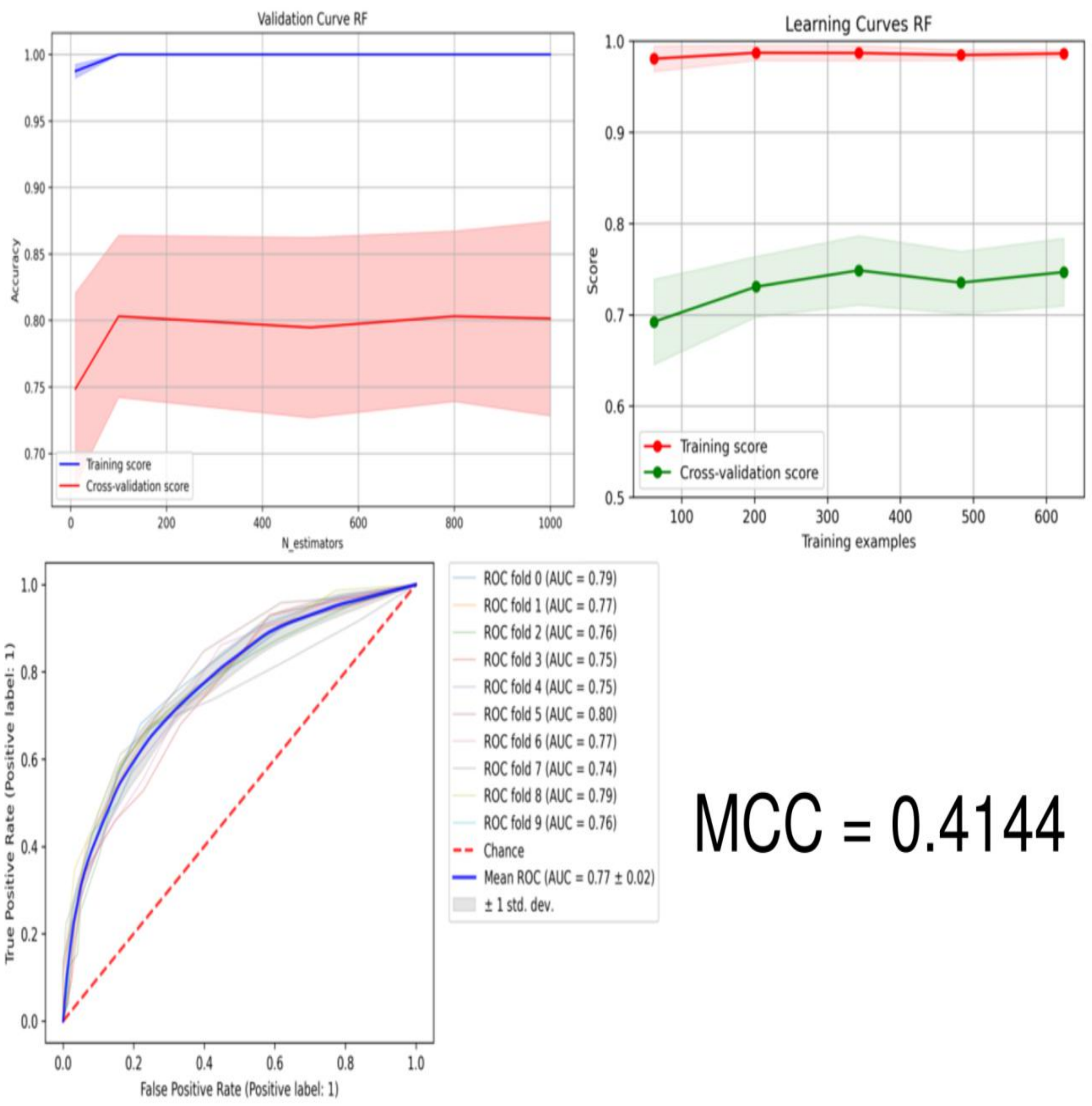

Figure S7. Performance assessment of random forest classifier trained with physicochemical descriptors. In order of appereance figure shows model's validation curve, learning curve, k-fold cross validation ROC curve and Matthews correlation coefficient.

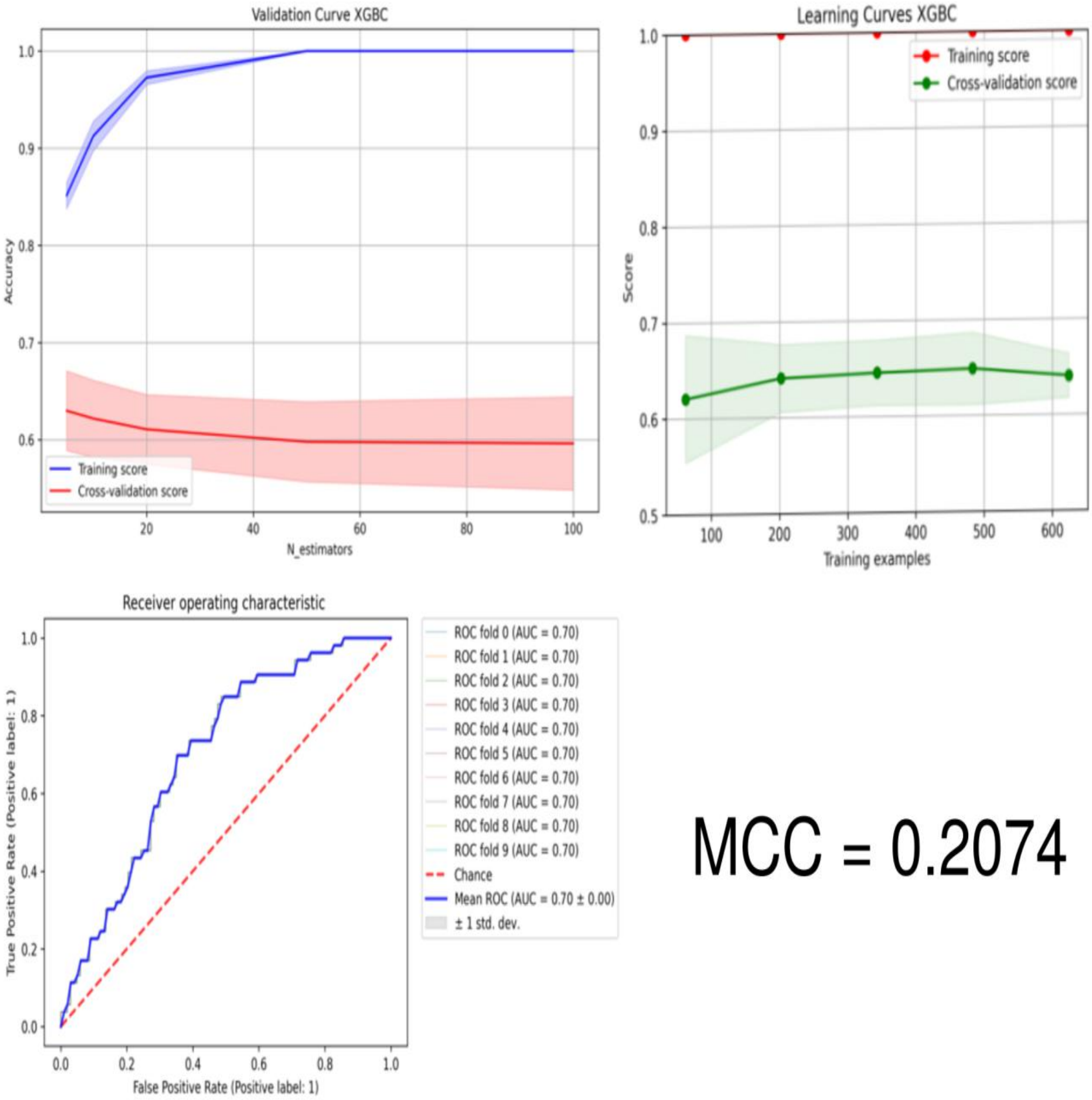

Figure S8. Performance assessment of extreme gradient boost classifier trained with physicochemical descriptors. In order of appereance figure shows model’s validation curve, learning curve, k-fold cross validation ROC curve and Matthews correlation coefficient.

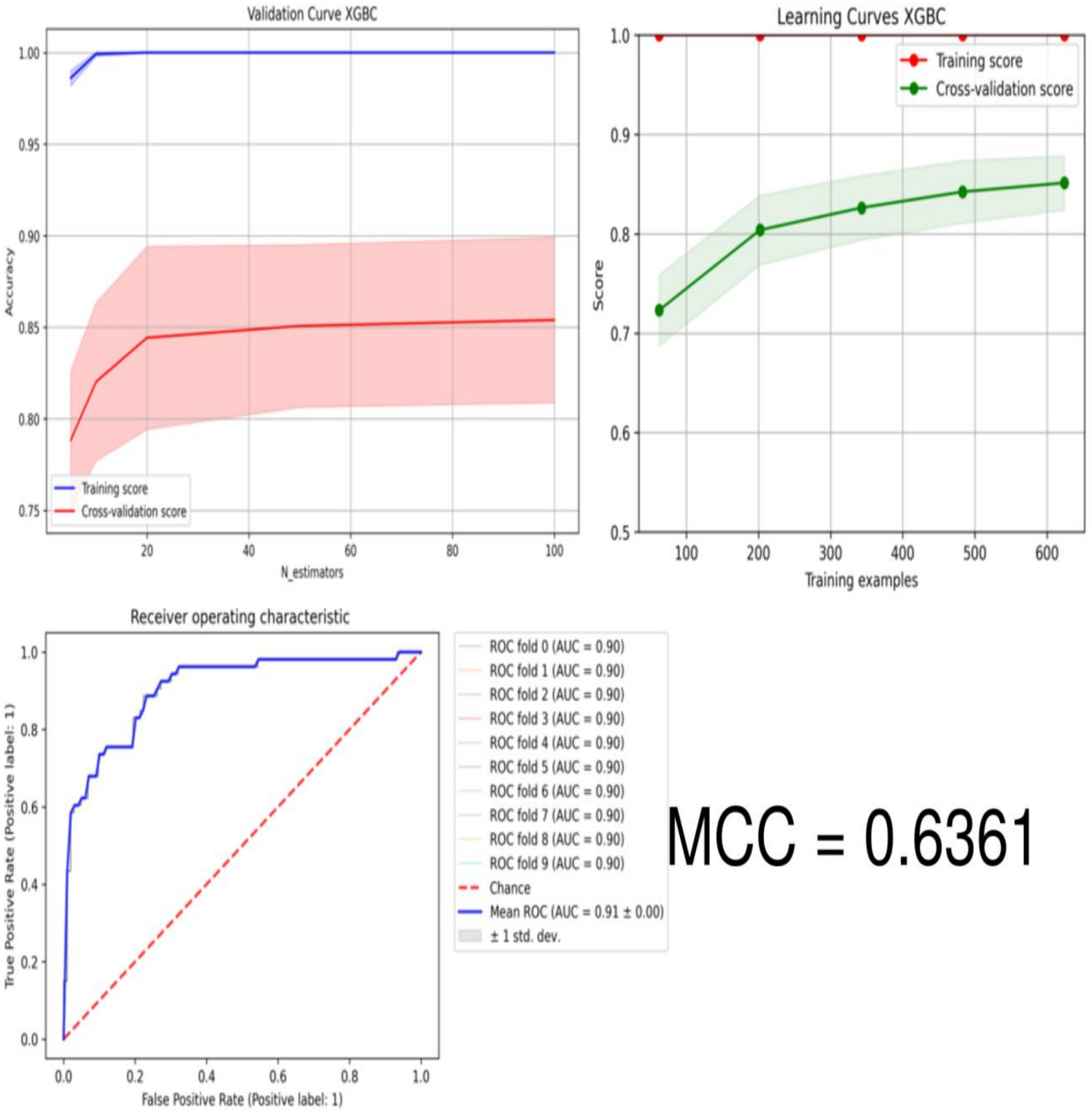

Figure S9. Performance assessment of extreme gradient boost classifier trained with physicochemical and structural descriptors. In order of appereance figure shows model's validation curve, learning curve, k-fold cross validation ROC curve and Matthews correlation coefficient.

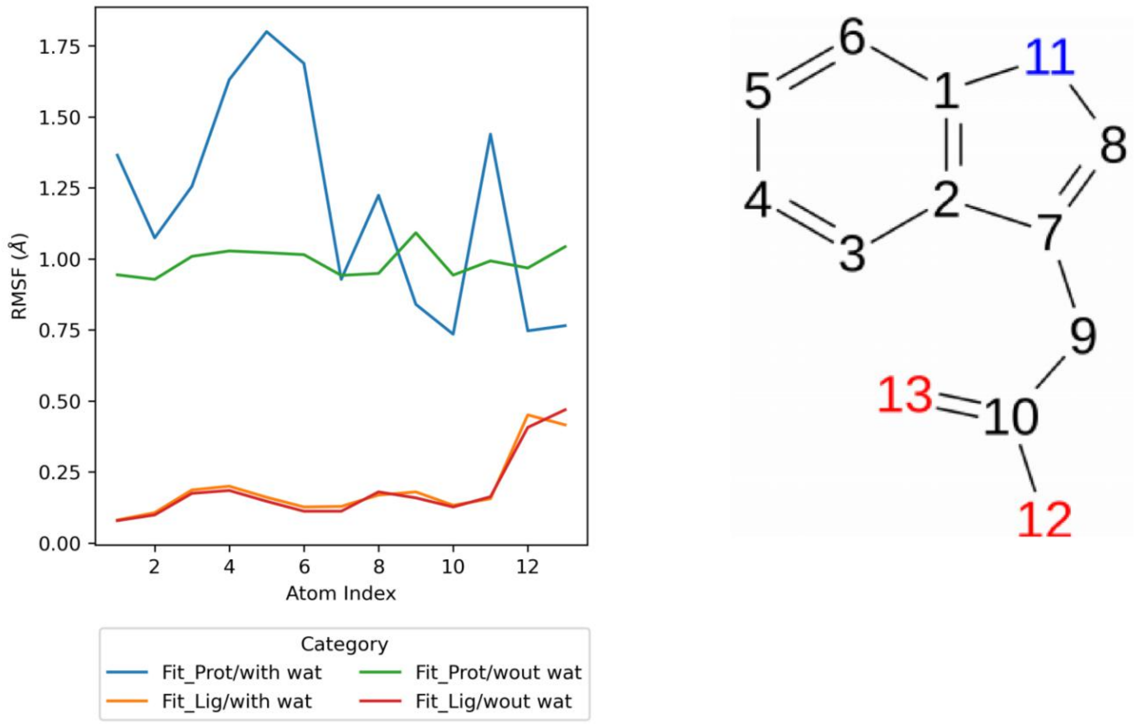

Figure S10. Comparison of IAA root mean square fluctuations (RMSF) from different solvation conditions (left); with preserved crystal waters around ligand (blue/orange) and without (green/red). RMSF values are presented with respect to TIR1 and ligand.

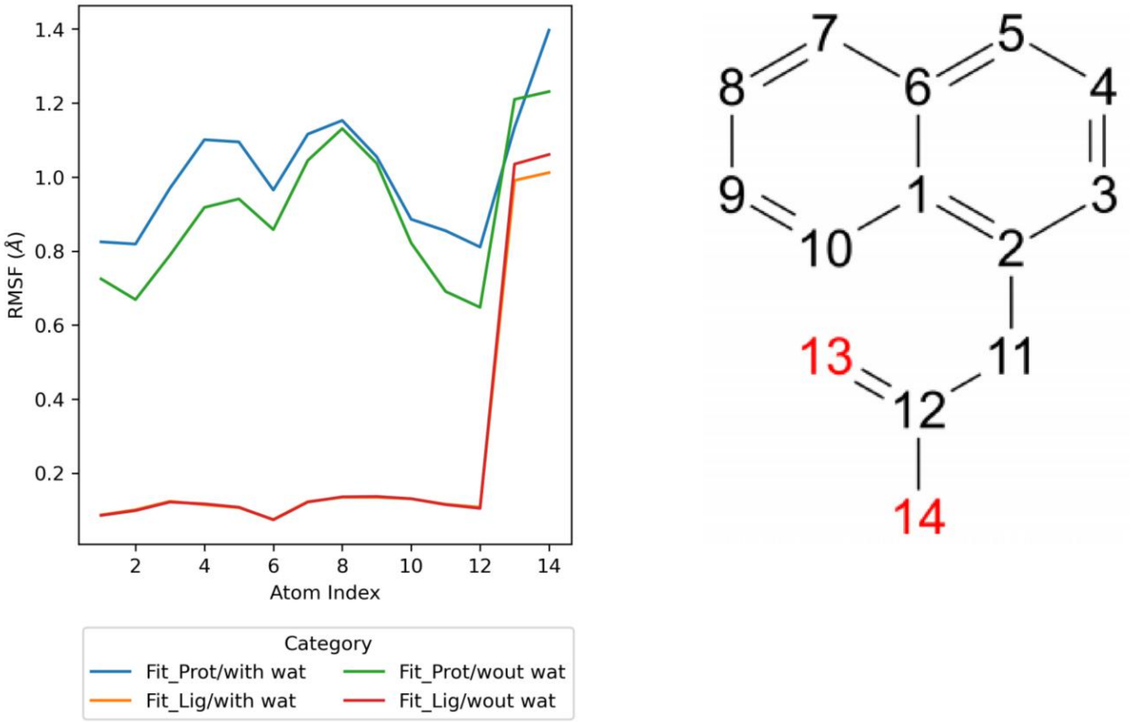

Figure S11. Comparison of NAA root mean square fluctuations (RMSF) from different solvation conditions (left); with preserved crystal waters around ligand (blue/orange) and without (green/red). RMSF values are presented with respect to TIR1 and ligand.

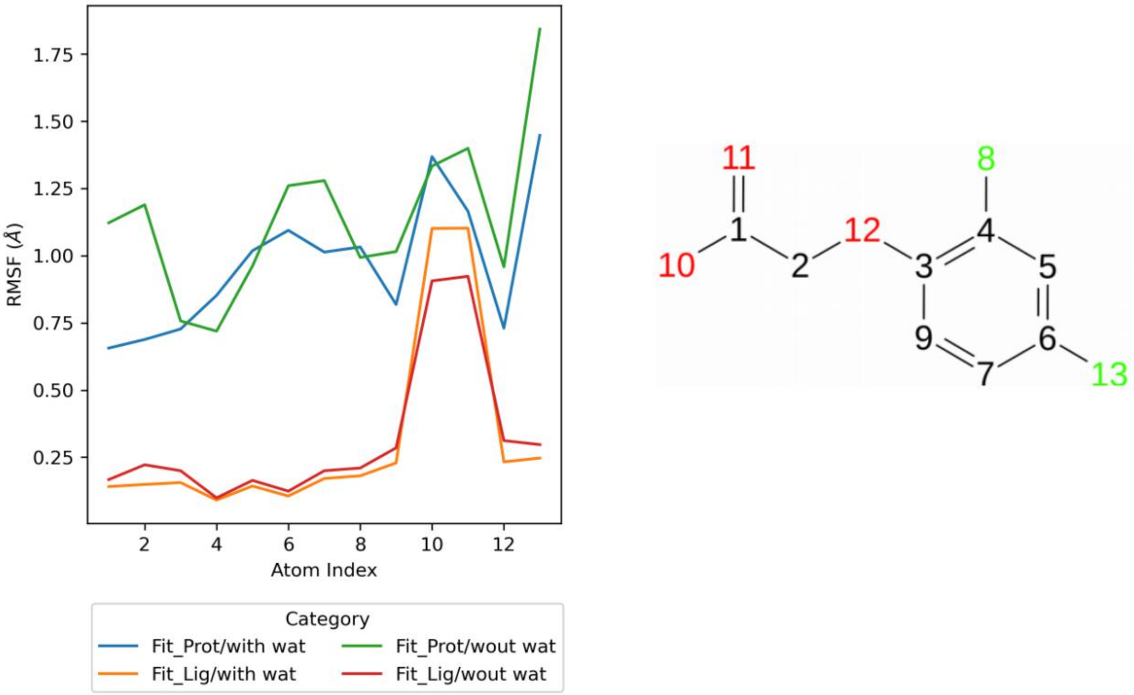

Figure S12. Comparison of 2,4-D root mean square fluctuations (RMSF) from different solvation conditions (left); with preserved crystal waters around ligand (blue/orange) and without (green/red). RMSF values are presented with respect to TIR1 and ligand.

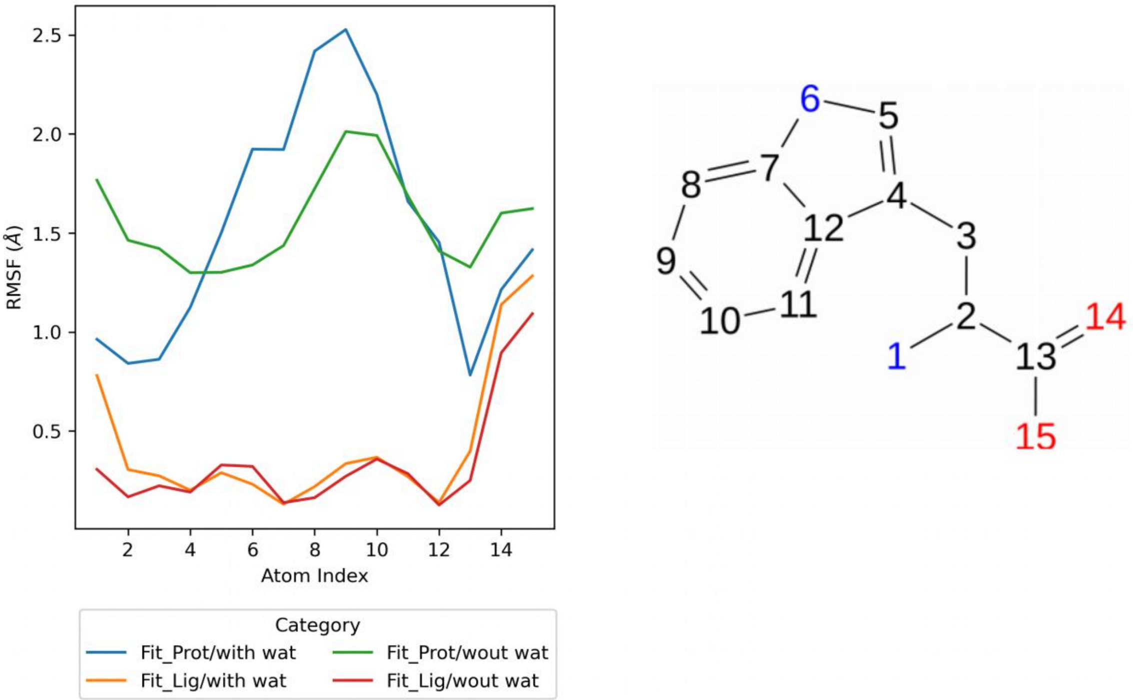

Figure S13. Comparison of TRP root mean square fluctuations (RMSF) from different solvation conditions (left); with preserved crystal waters around ligand (blue/orange) and without (green/red). RMSF values are presented with respect to TIR1 and ligand.

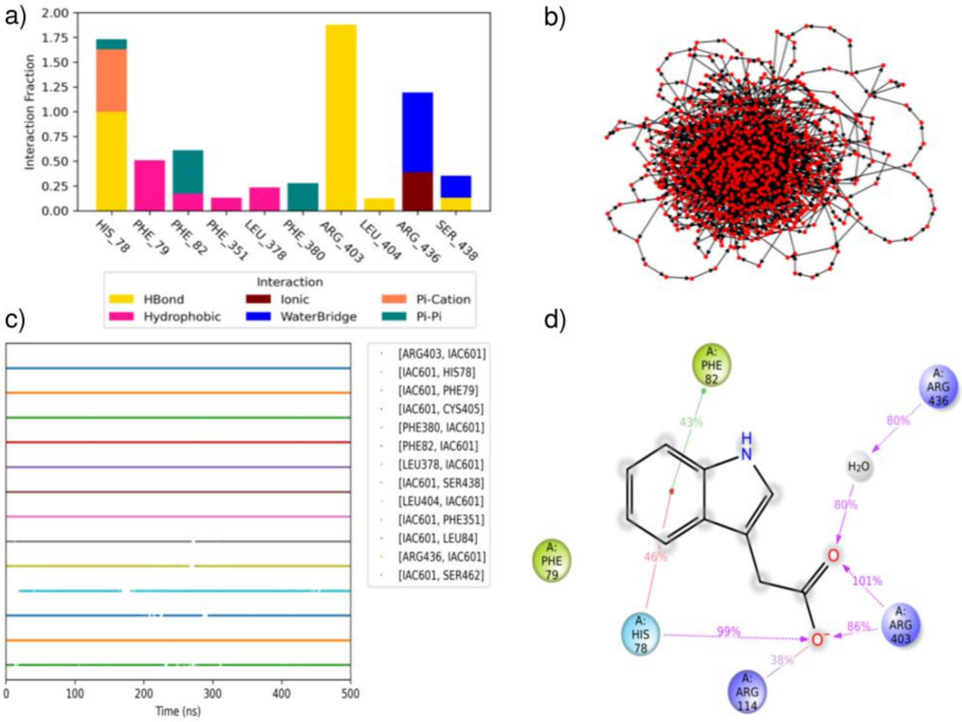

Figure S14. Consensus analysis of MD simulation of indol acetic acid (IAA); a) Protein-ligand interaction fraction; b) Concurrence of contacts; c) Transition network for protein-ligand interactions; d) Ligand-protein interaction.

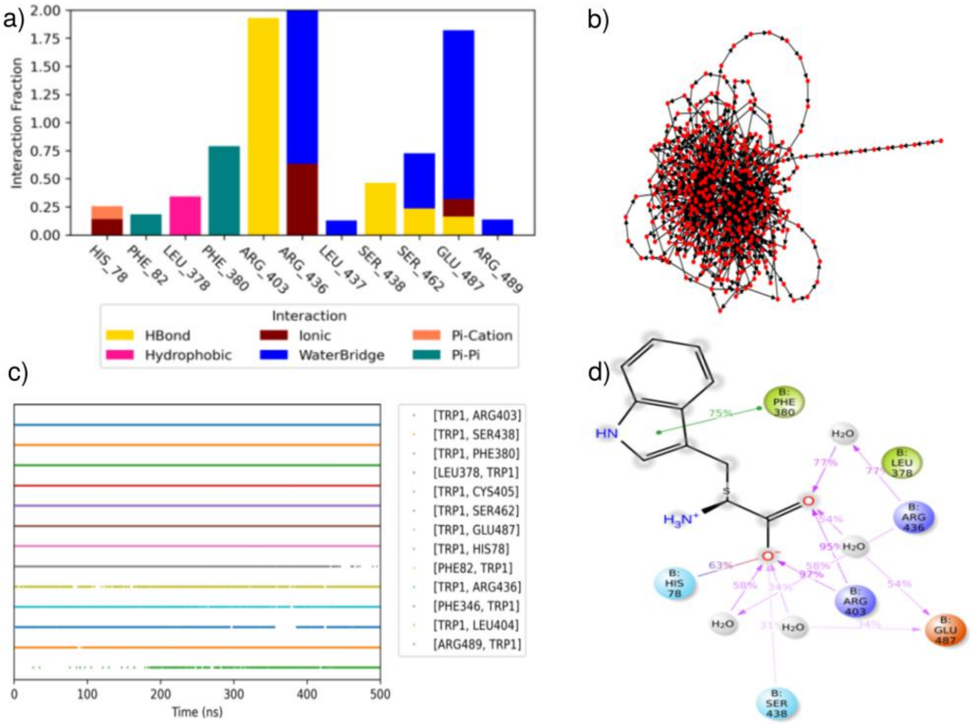

Figure S15. Consensus analysis of MD simulation of tryptophan (TRP); a) Protein-ligand interaction fraction; b) Concurrence of contacts; c) Transition network for protein-ligand interactions; d) Ligand-protein interaction.

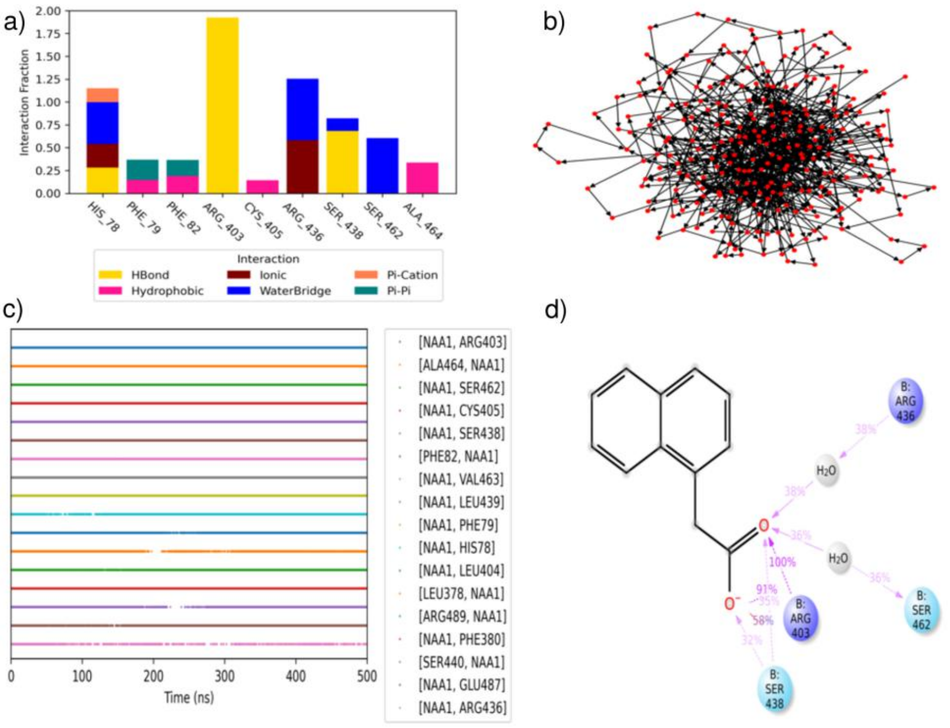

Figure S16. Consensus analysis of MD simulation of 1-naphtyl acetic acid (NAA); a) Protein-ligand interaction fraction; b) Concurrence of contacts; c) Transition network for protein-ligand interactions; d) Ligand-protein interaction.

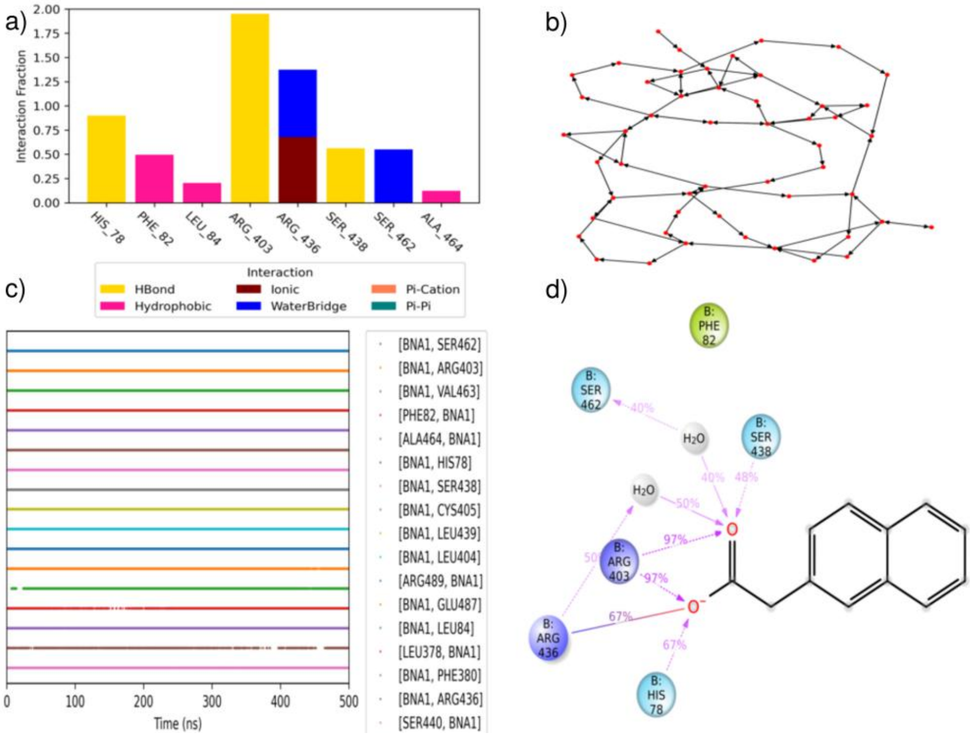

Figure S17. Consensus analysis of MD simulation of 2-naphtyl acetic acid (2-NAA); a) Protein-ligand interaction fraction; b) Concurrence of contacts; c) Transition network for protein-ligand interactions; d) Ligand-protein interaction.

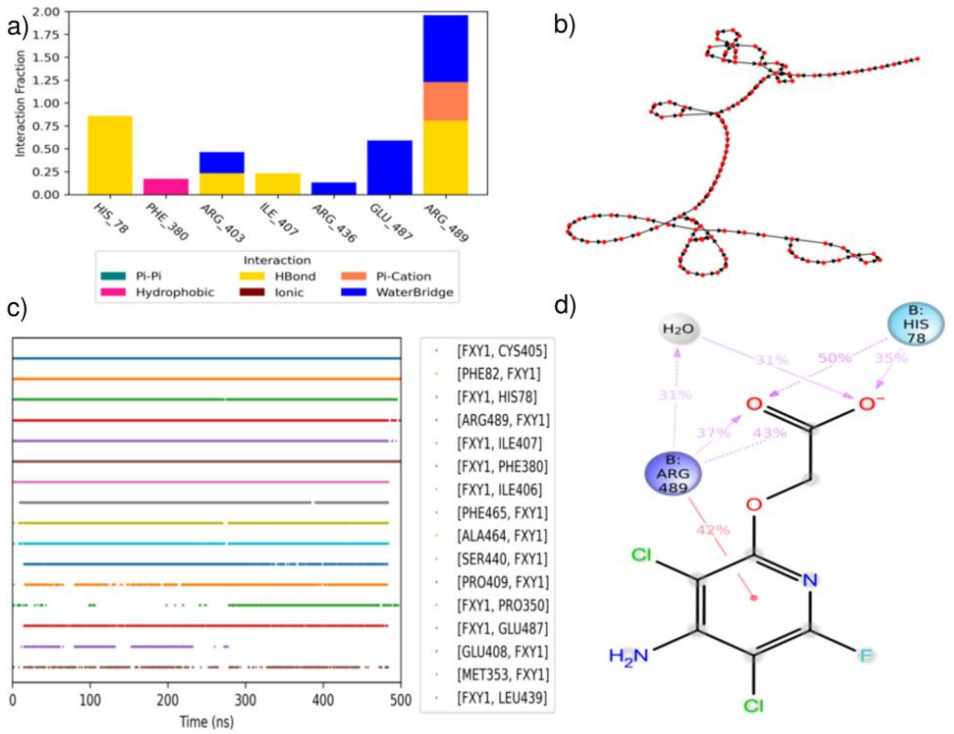

Figure S18. Consensus analysis of MD simulation of fluroxypyr (FXY); a) Protein-ligand interaction fraction; b) Concurrence of contacts; c) Transition network for protein-ligand interactions; d) Ligand-protein interaction.

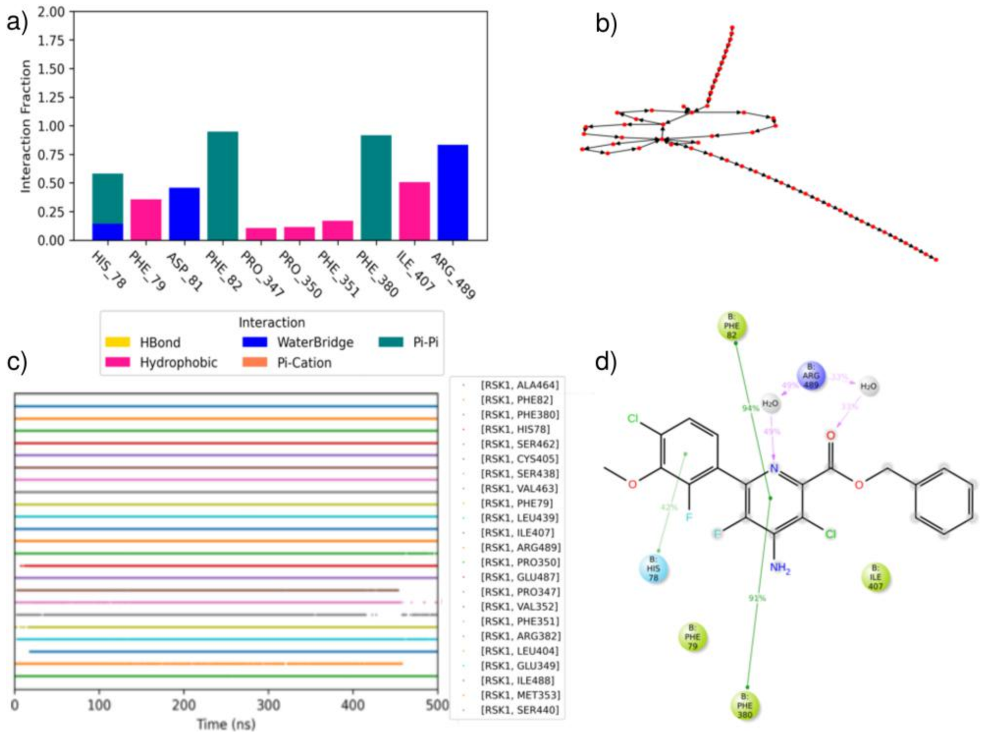

Figure S19. Consensus analysis of MD simulation of rinskor (RSK); a) Protein-ligand interaction fraction; b) Concurrence of contacts; c) Transition network for protein-ligand interactions; d) Ligand-protein interaction.

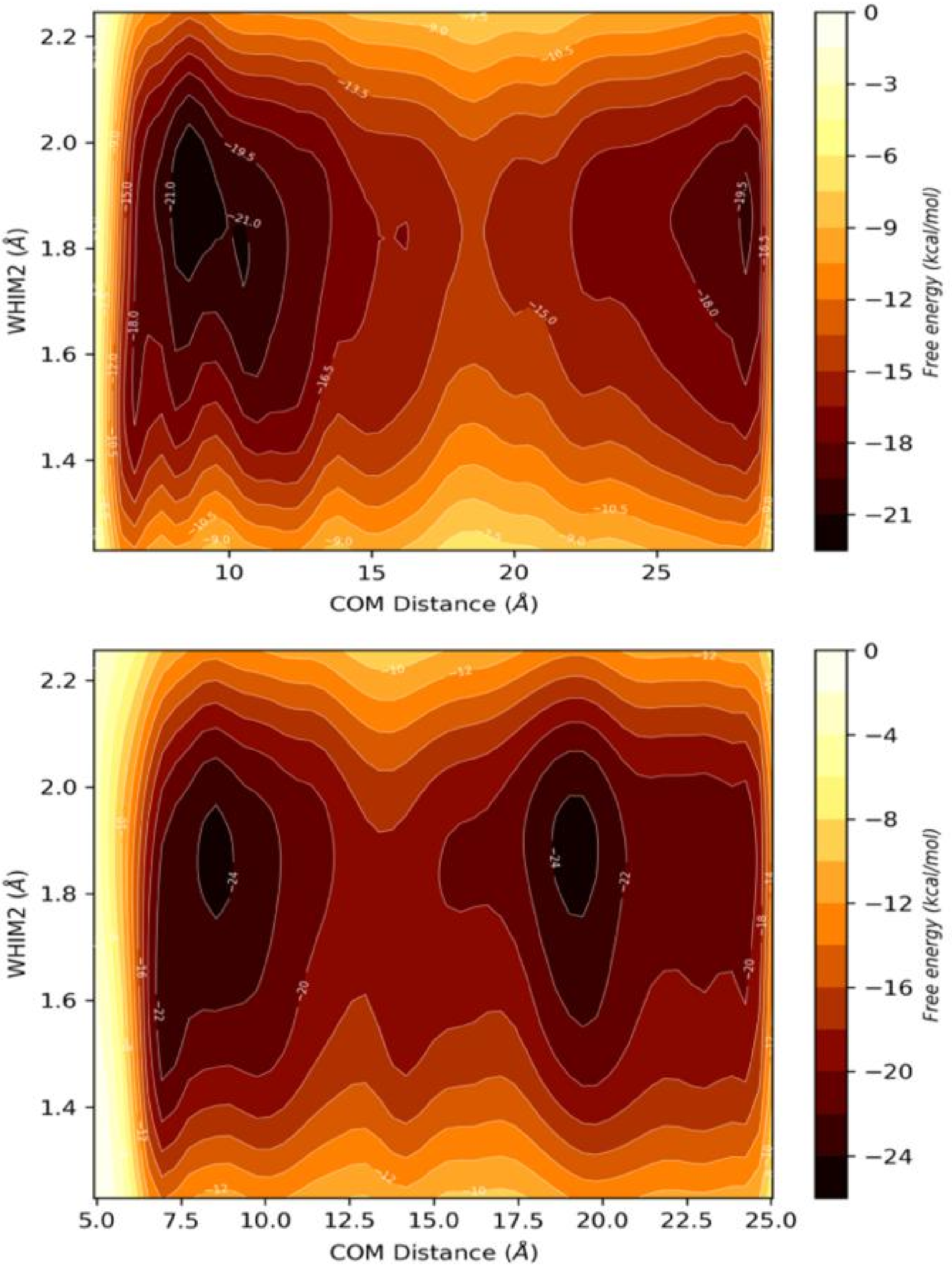

Figure S20. Free energy surfaces computed from well-tempered metadynamics for the NAA-TIR1 complexes. With ligand positioned within the binding site (top) and positioned within the engagement niche (bottom).

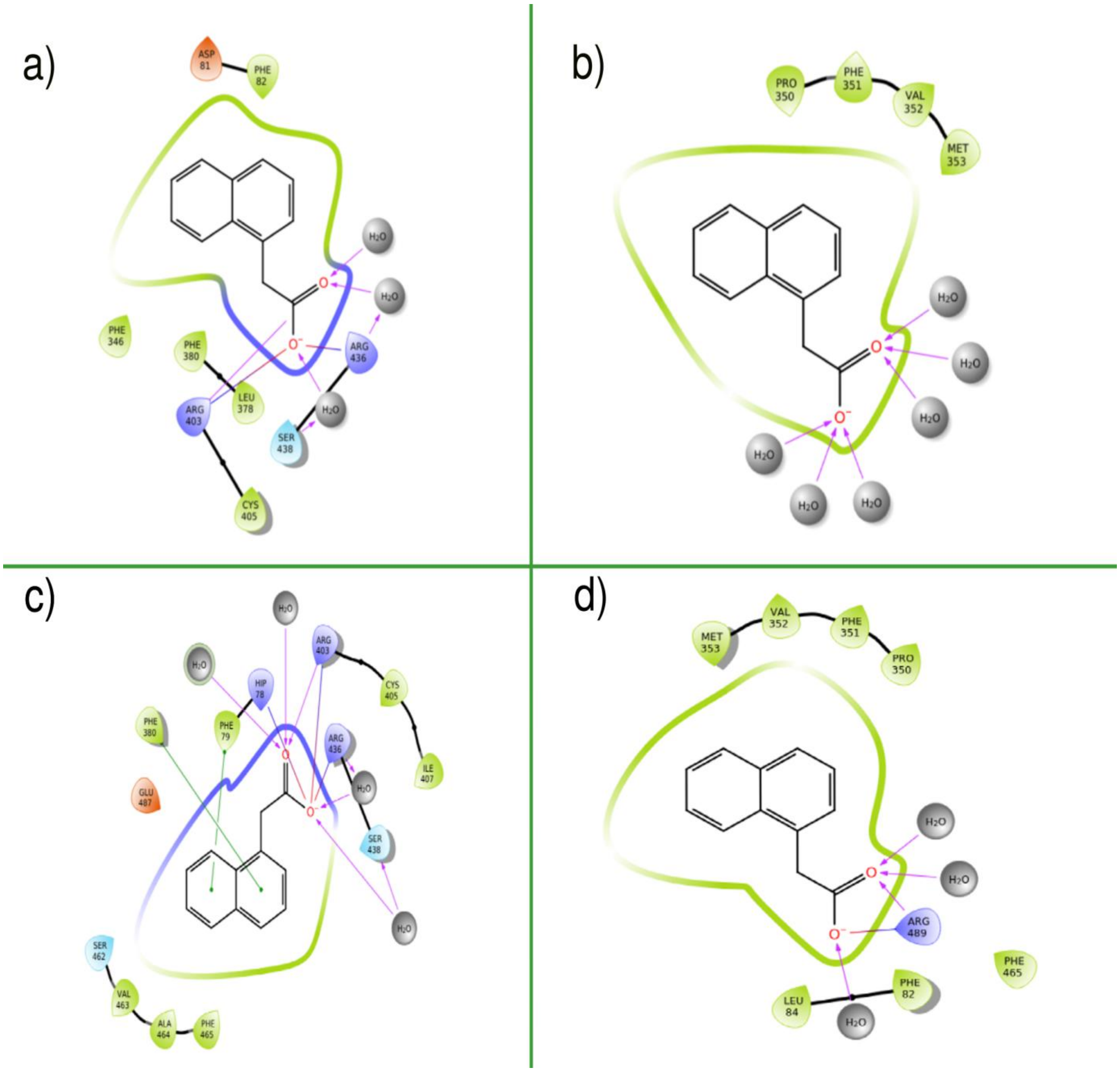

Figure 21. Representative NAA conformations found at basins from metadynamics runs. With ligand positioned within the binding site (a & b) and positioned within the engagement niche (c & d).

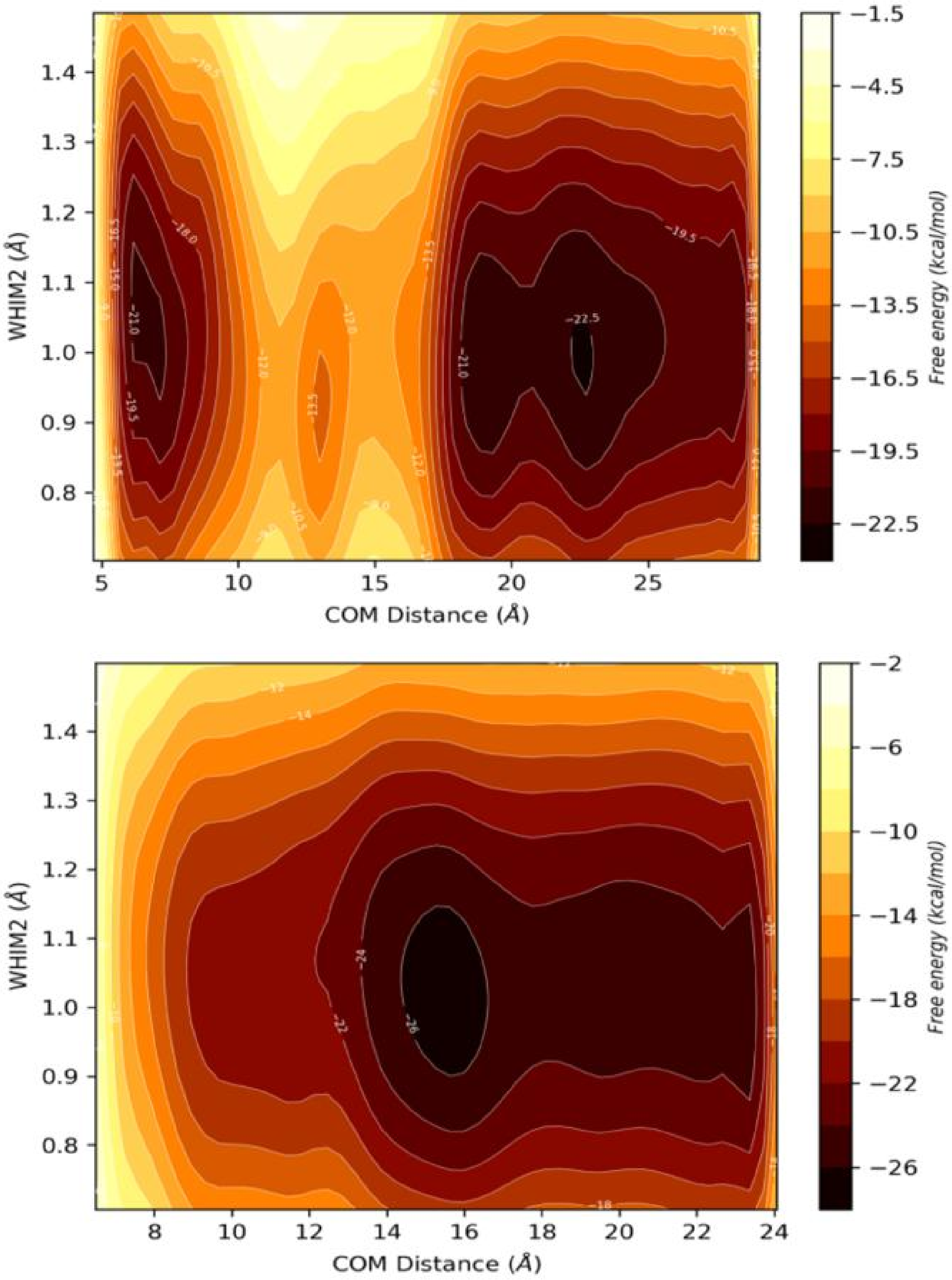

Figure 22. Free energy surfaces computed from well-tempered metadynamics for the 2-NAA-TIR1 complexes. With ligand positioned within the binding site (top) and positioned within the engagement niche (bottom).

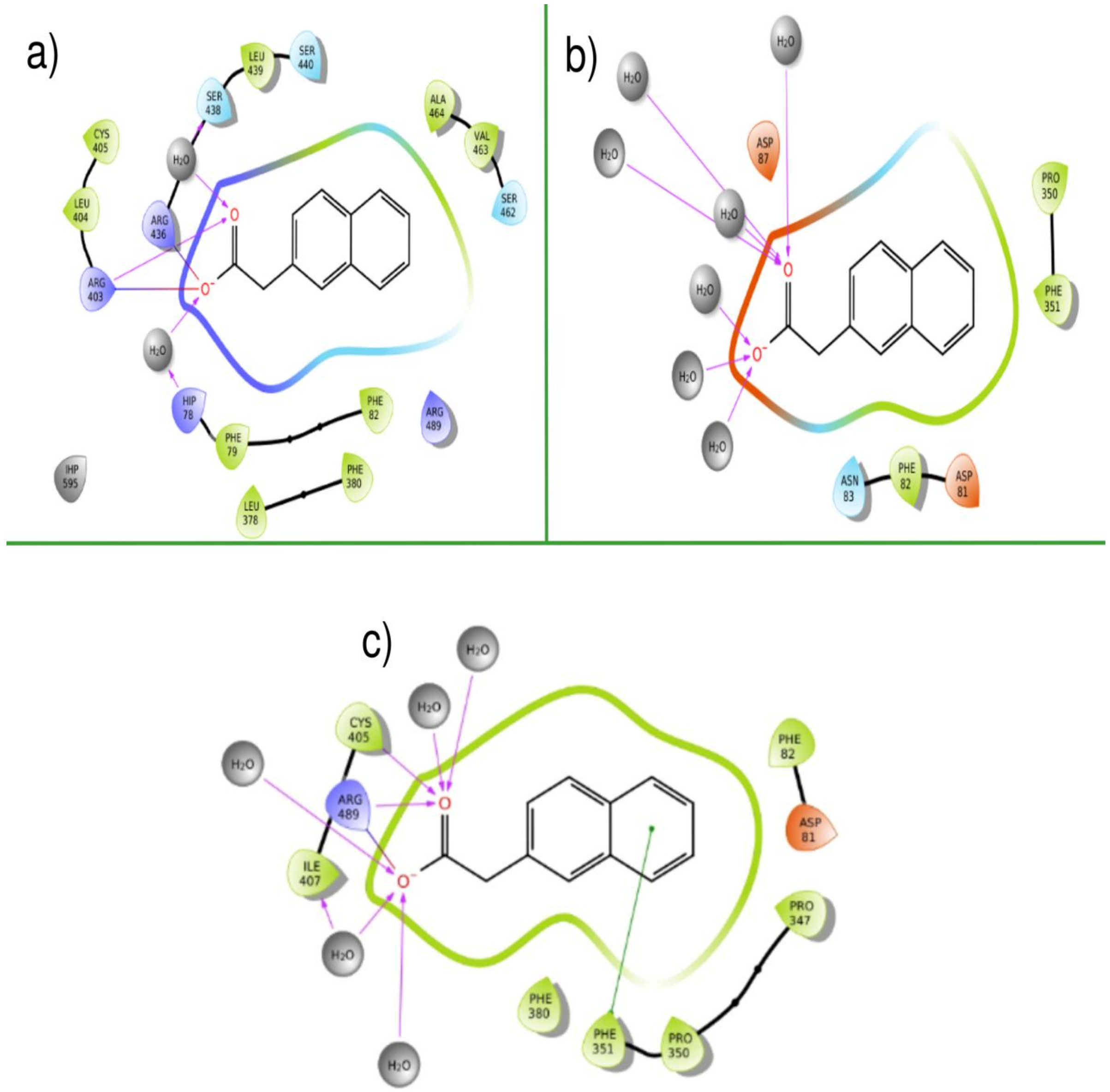

Figure 23. Representative 2-NAA conformations found at basins from metadynamics runs. With ligand positioned within the binding site (a & b) and positioned within the engagement niche (c).

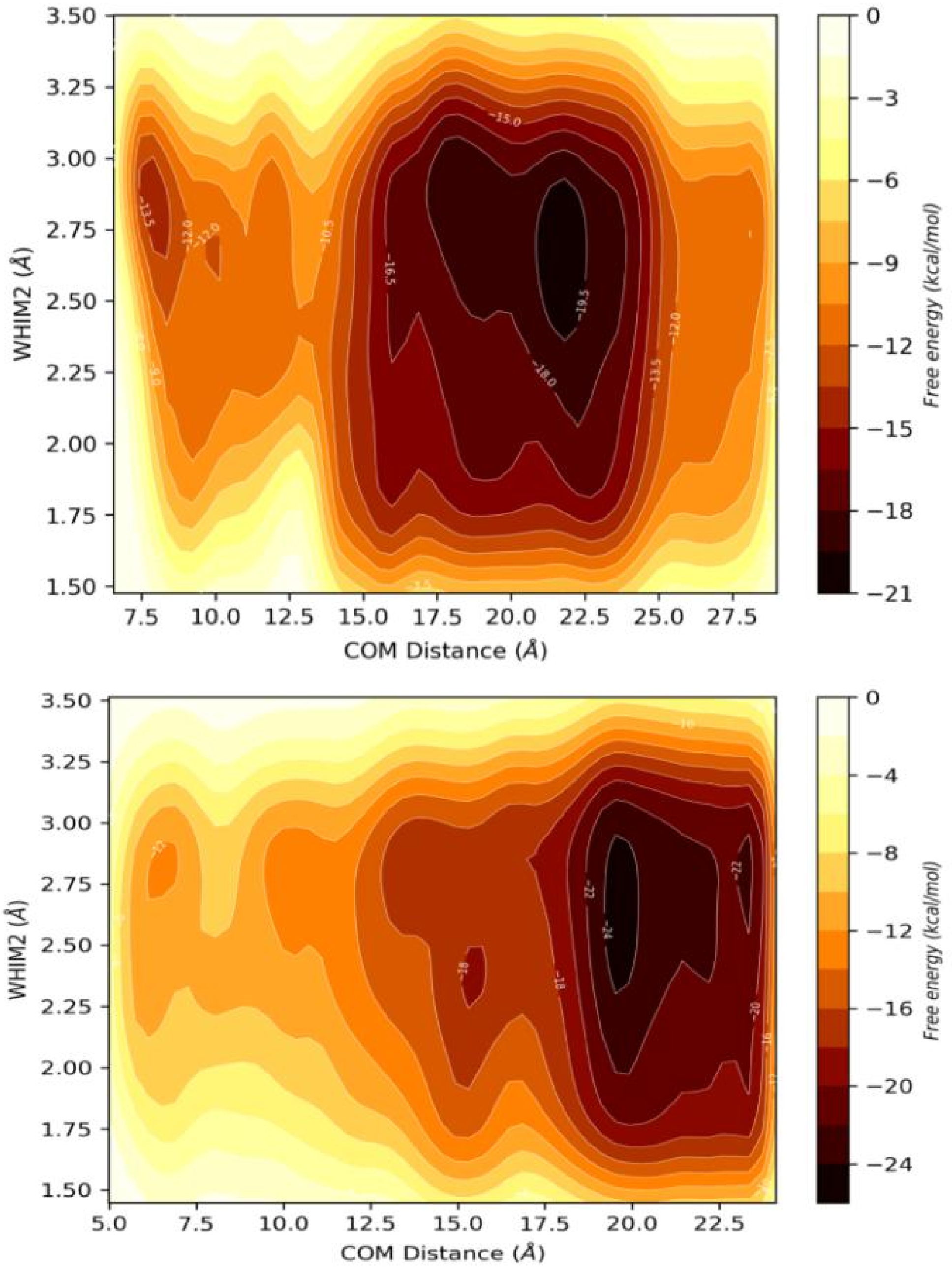

Figure 24. Free energy surfaces computed from well-tempered metadynamics for the FXY-TIR1 complexes. With ligand positioned within the binding site (top) and positioned within the engagement niche (bottom).

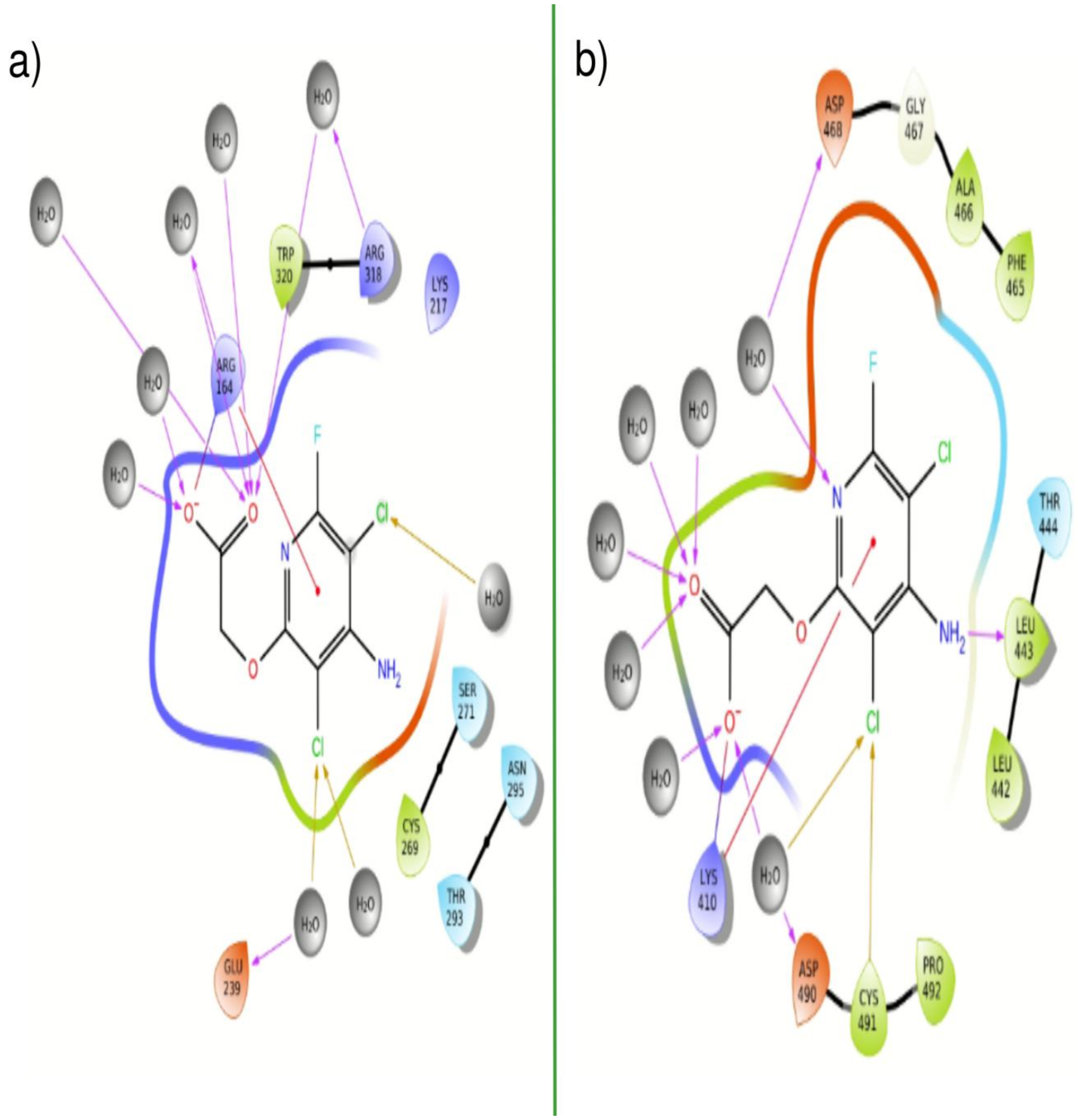

Figure 25. Representative FXY conformations found at basins from metadynamics runs. With ligand positioned within the binding site (a) and positioned within the engagement niche (b).

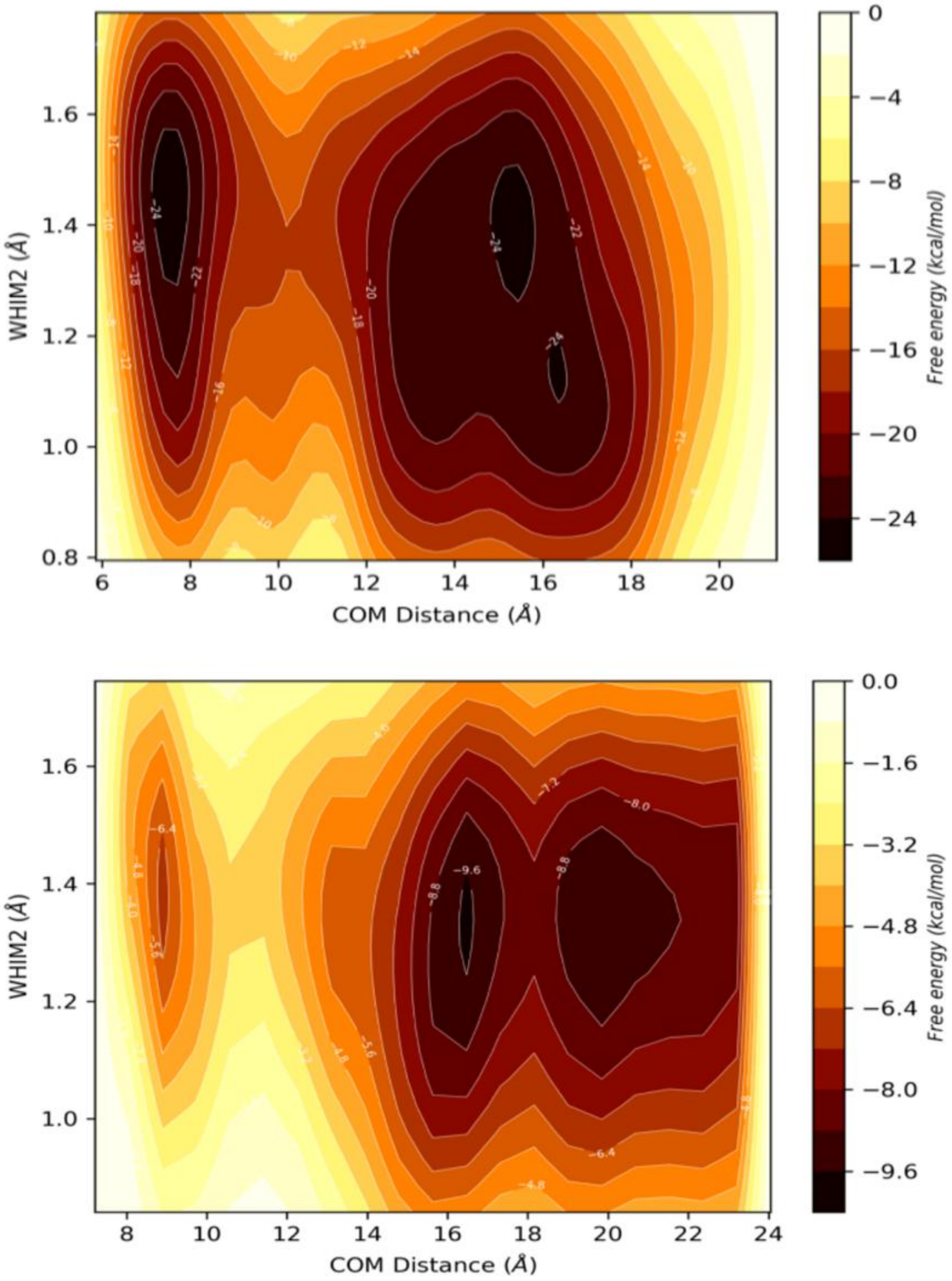

Figure 26. Free energy surfaces computed from well-tempered metadynamics for the IAA-TIR1 complexes. With ligand positioned within the binding site (top) and positioned within the engagement niche (bottom).

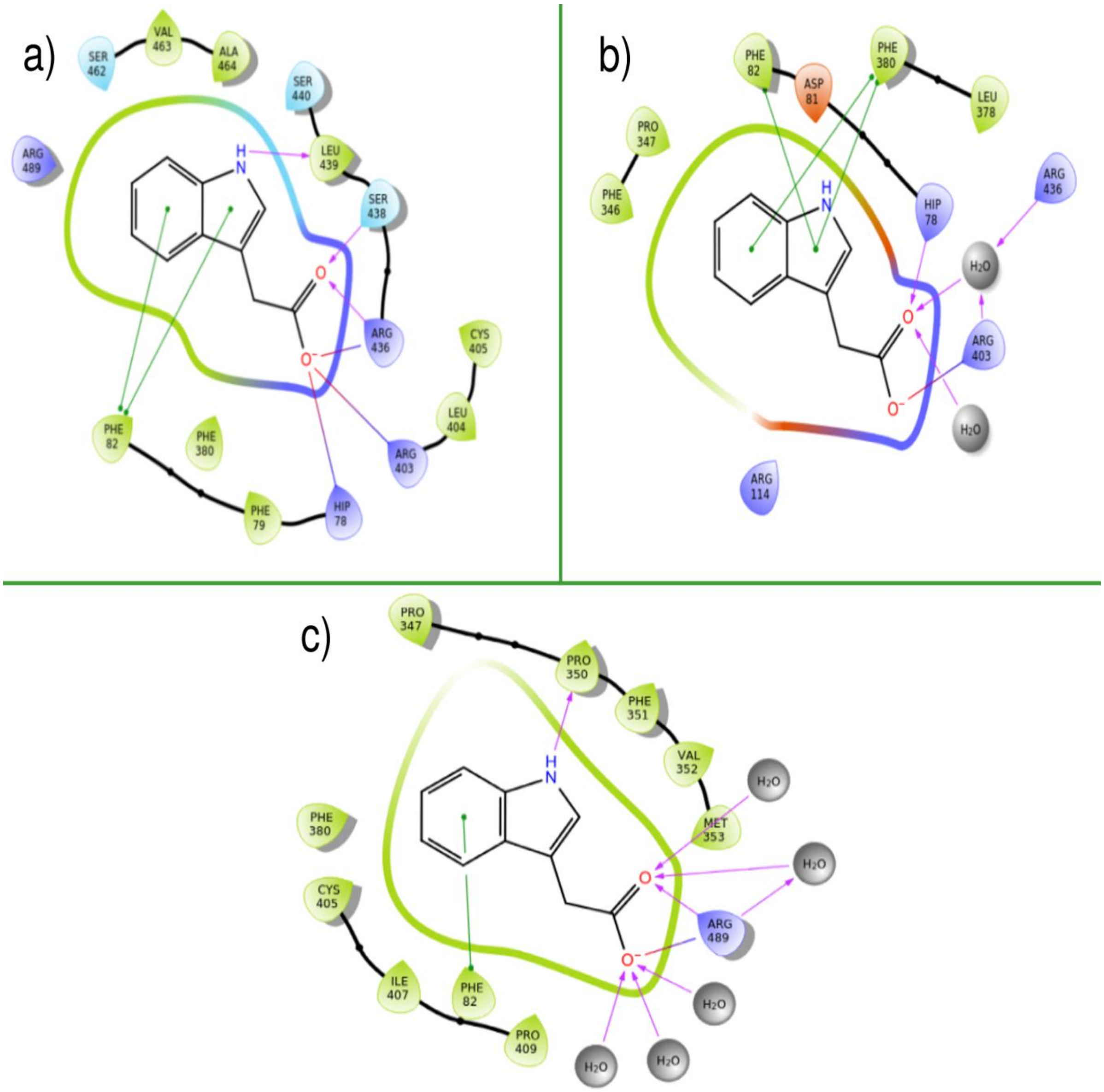

Figure 27. Representative IAA conformations found at basins from metadynamics runs. With ligand positioned within the binding site (a & b) and positioned within the engagement niche (c).

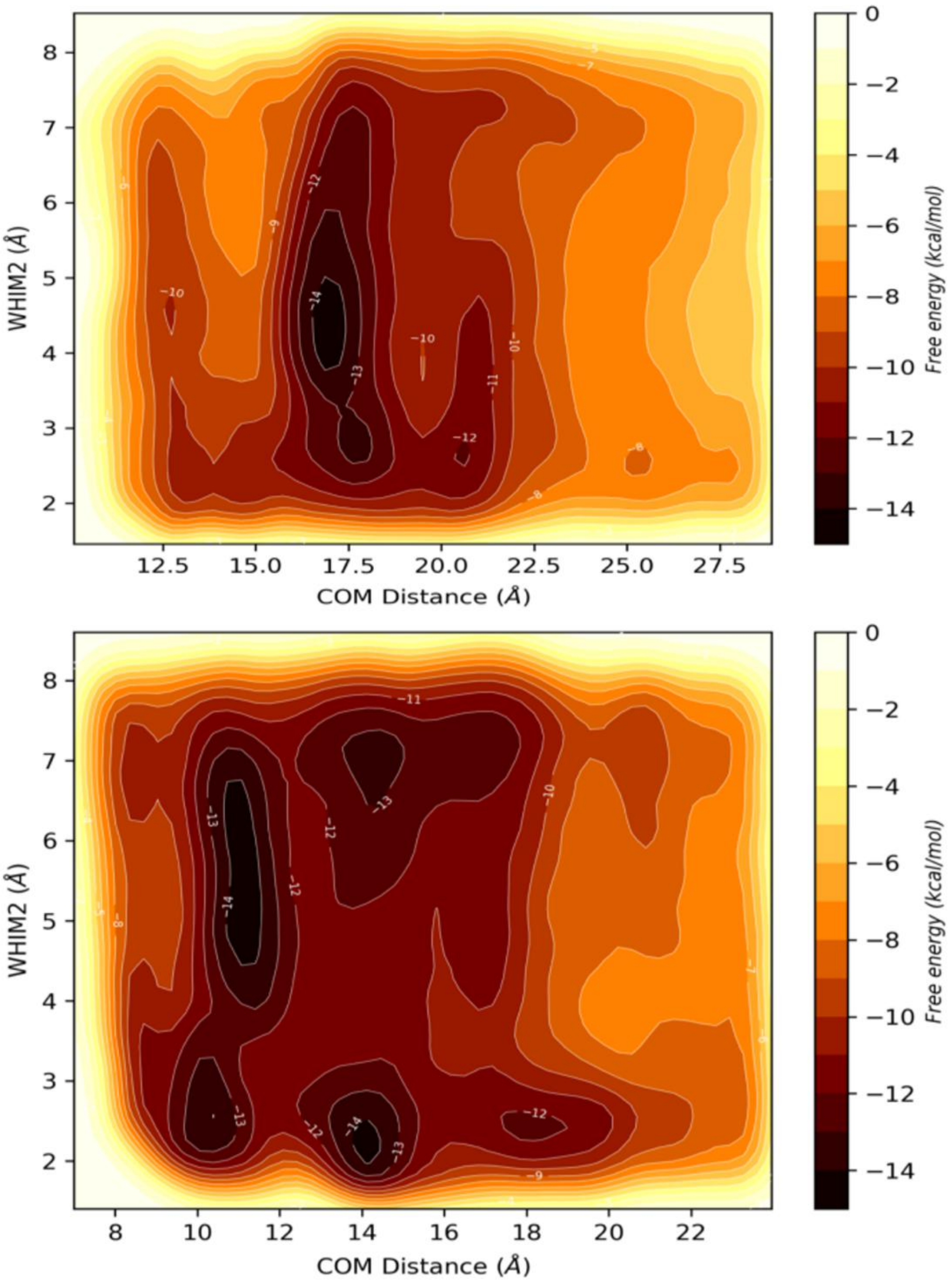

Figure 28. Free energy surfaces computed from well-tempered metadynamics for the RSK-TIR1 complexes. With ligand positioned within the binding site (top) and positioned within the engagement niche (bottom).

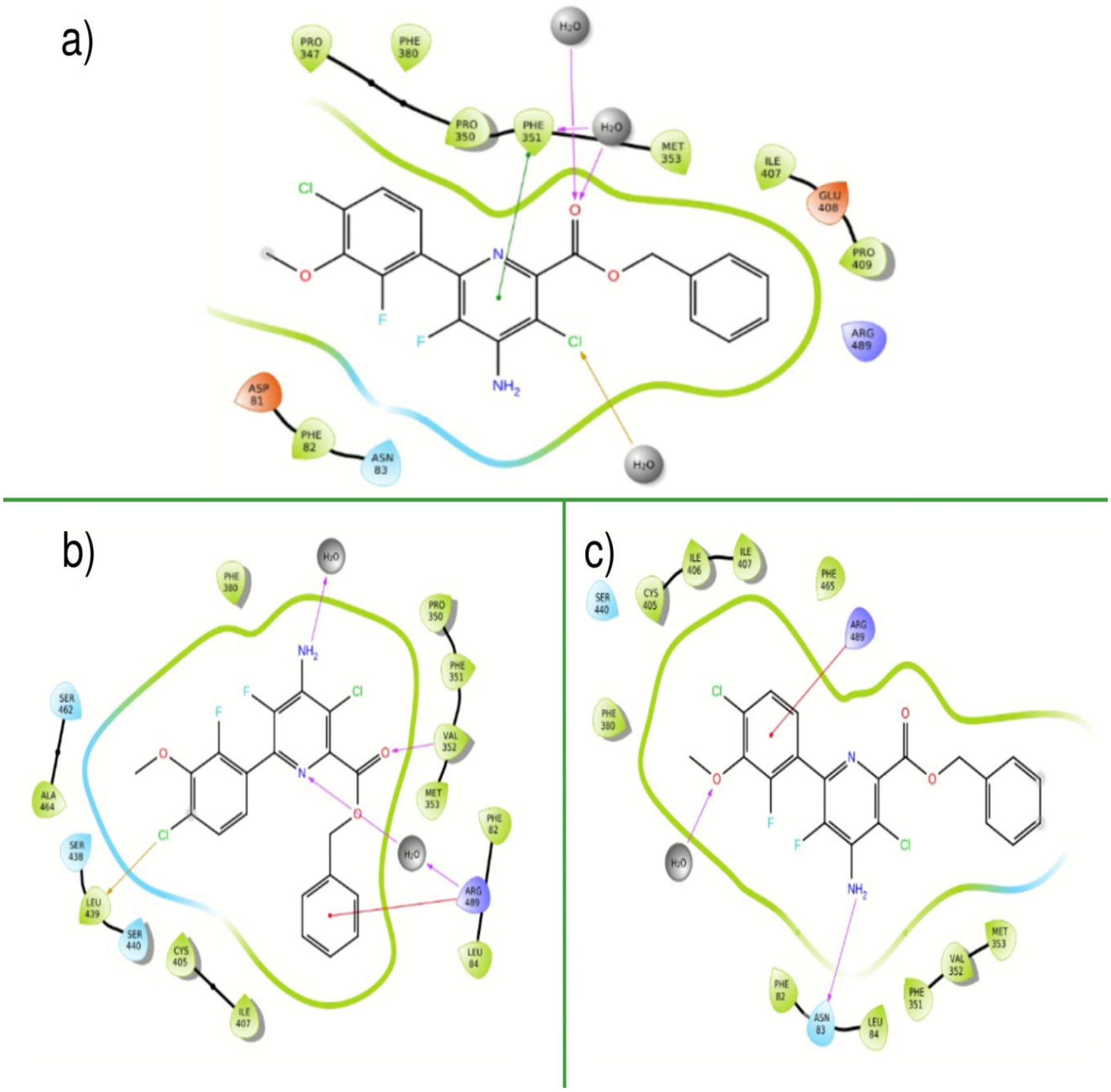

Figure 29. Representative RSK conformations found at basins from metadynamics runs. With ligand positioned within the binding site (a) and positioned within the engagement niche (b).

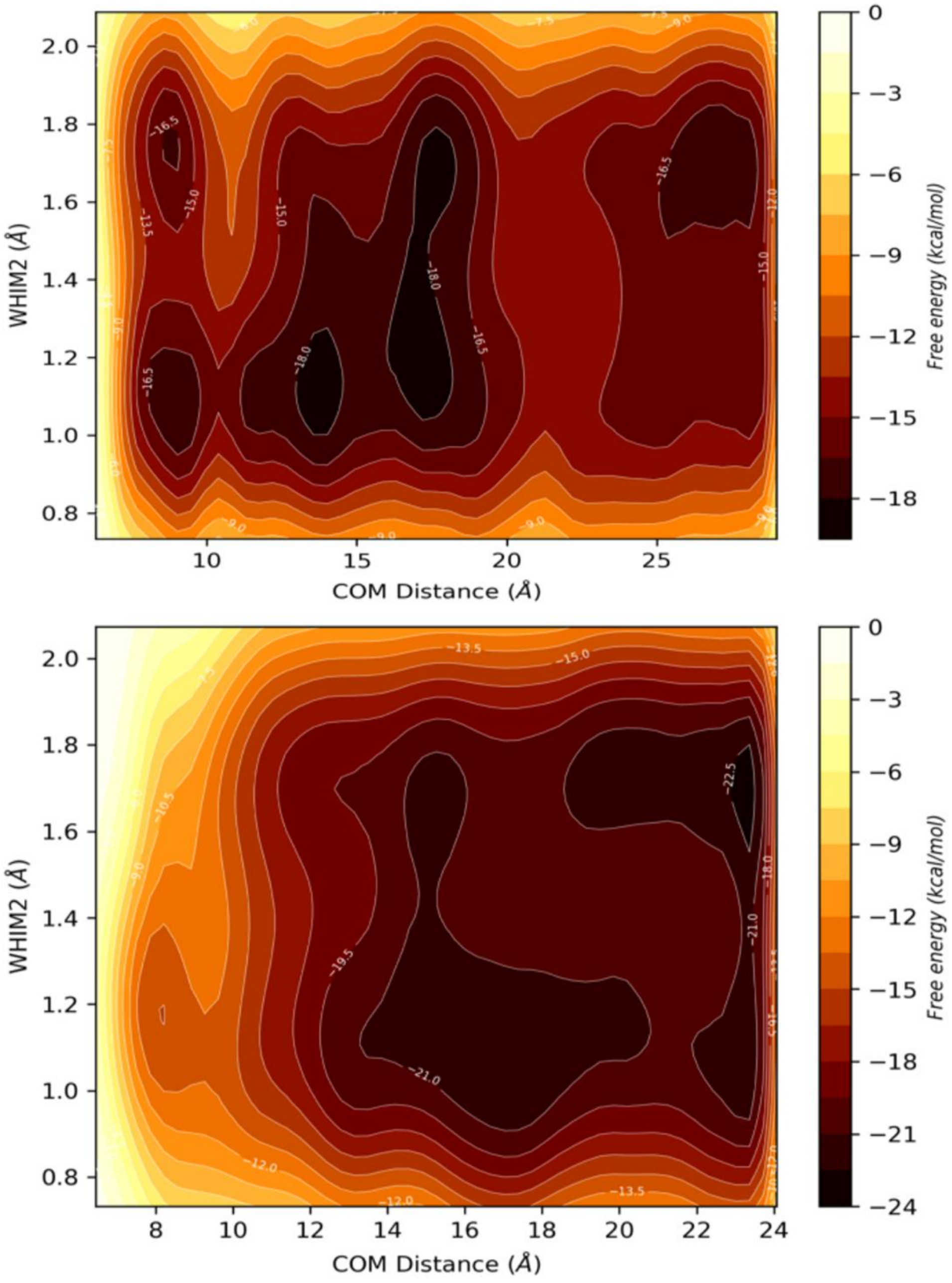

Figure 30. Free energy surfaces computed from well-tempered metadynamics for the TRP-TIR1 complexes. With ligand positioned within the binding site (top) and positioned within the engagement niche (bottom).

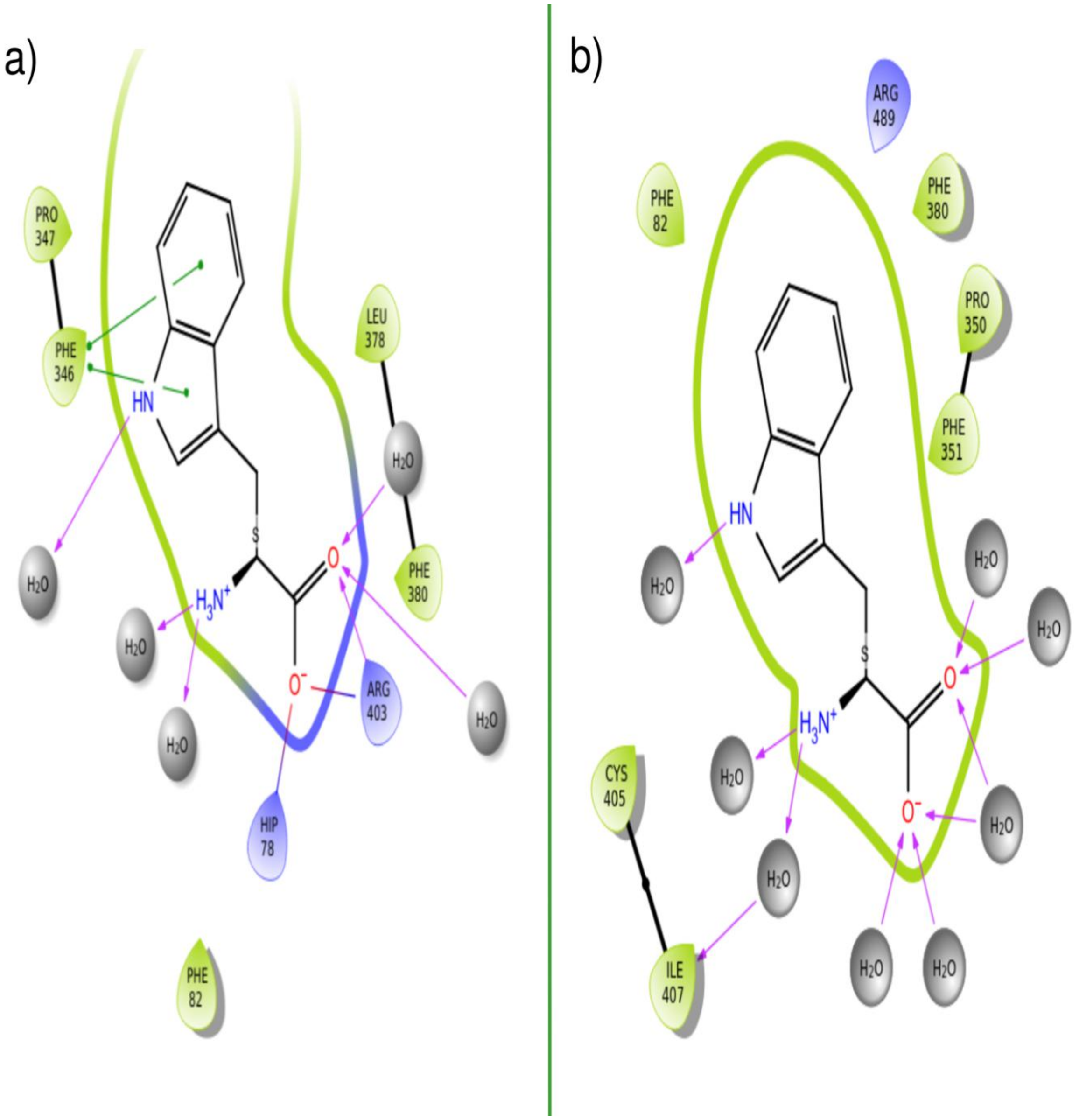

Figure 31. Representative TRP conformations found at basins from metadynamics runs. With ligand positioned within the binding site (a) and positioned within the engagement niche (b).