To Bind or Not 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 97357, Mexico; kmtzm@unam.mx
2 Facultad de Química, Universidad Nacional Autónoma de México, Avenida Universidad 3000, Mexico City 04510, Mexico
3 Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Unidad Mérida, Universidad Nacional Autónoma de México, Sierra Papacal, Mérida 97302, Mexico
* Correspondence: ferdpm4@hotmail.com

Abstract: Auxins are chemical compounds of wide interest, mostly due to their role in plant metabolism and development. Synthetic auxins have been used as herbicides for more than 75 years and low toxicity in humans is one of their most advantageous features. Extensive studies of natural and synthetic auxins have been made in an effort to understand their role in plant growth. However, molecular details of the binding and recognition process are still an open question. Herein, we present a comprehensive in silico pipeline for the assessment of TIR1 ligands using several structure-based methods. Our results suggest that subtle dynamics within the binding pocket arise from water–ligand interactions. We also show that this trait distinguishes effective binders. Finally, we construct a database of putative ligands and decoy compounds, which can aid further studies focusing on synthetic auxin design. To the best of our knowledge, this study is the first of its kind focusing on TIR1.

Keywords: herbicides; auxins; molecular dynamics; machine learning

1. Introduction

Agricultural production is fundamental from an economic point of view and to feed the world's population. It has been estimated that from 1961 to 2020, food production from agricultural sources has increased significantly, with current per capita values nearing almost 3200 kcal/day at the end of the supply chain [1]. Yet, the annual State of Food Security and Nutrition in the World (SOFI) report published by the Food and Agriculture Organization (FAO) showed that more than 700 million people faced hunger in 2022 [2]. Thus, there is a pressing need to improve food security, which translates to the assurance of higher crop yields.

Some of the main problems that pose a significant challenge to safe, productive, and sustainable agricultural practices are herbicide resistance and toxicity. There are documented cases of resistance of weeds to several chemical groups worldwide, reported since the discovery of herbicide resistance in common groundsel (*Senecio vulgaris*) to triazine herbicides in 1968 [3,4]. Nowadays, herbicide resistance is spread globally, affecting 96 crops across 72 countries. This phenomenon has been documented in 513 distinct instances, involving 267 weed species that have developed resilience against herbicides [5]. Moreover, despite the generally low mammalian toxicity of many newly developed herbicides, growing experimental evidence indicates that exposure to these chemicals can have detrimental effects on the development and/or reproduction of various mammalian species. Notably, 2,4-D (2,4-Dichlorophenoxyacetic acid) and its combination with 2,4,5-T (2,4,5-Trichlorophenoxyacetic acid) have been associated with reproductive issues and malformations in humans [6].

Auxins can be regarded as phytohormones, metabolically derived from L-Tryptophan, responsible for plant growth and development [7]. Indole-3-acetic acid (IAA) is the most well-known example of such a compound, often deemed as the "master hormone" in plants [8]. Since the 1940s, synthetic derivatives from IAA have been developed and tested, with notable examples being naphthalene acids (1-naphthalene acetic acid); phenoxycarboxylic acids (MCPA, 2,4-D); benzoic acids (dicamba); pyridinecarboxylic acids (e.g., picloram), and quinolinecarboxylic acids (quinmerac, quinclorac), all of which showed strong responses in plants, inducing damage at high concentrations while eliciting higher stability when compared to IAA [9]. In the last two decades, synthetic auxins reached commercial success as herbicides, accounting for roughly 1.3 billion USD, which represents approximately 9% of total sales just in 2000 [10].

The intricated auxin signaling pathway is orchestrated by at least three protein families: transport inhibitor response/auxin signaling F-box (TIR1/AFB) proteins, auxin/indole-3-acetic acid (Aux/IAA) transcriptional repressors, and auxin response factor (ARF) transcription factors [11]. At the core of this signaling cascade lies TIR1, belonging to a small family of F-box proteins, comprising five additional AFB proteins (AFB1-AFB5). These six proteins collectively function as auxin receptors [12]. Auxin transport occurs within and between cells via PIN, ABCB, and AUX/LAX transporters. This then interacts with the SCFTIR1/AFB complex, which, upon creation of the SCF-auxin-Aux/IAA complex, triggers the ubiquitination of Aux/IAA transcriptional repressors, leading to the activation of auxin response genes [13]. Therefore, auxin exerts its regulatory influence by directly binding to TIR1, promoting the interaction between SCFTIR1/AFB and Aux/IAAs. This interaction does not induce any conformational changes; instead, auxin acts as a molecular glue, facilitating the formation of a binding pocket for the Aux/IAA protein within the TIR1 structure [14].

Herein, we present an in silico profiling of TIR1 binding coupled with a pipeline for de novo design based on machine learning. We hope this study will aid synthetic auxin design as these molecules can provide additional tools, not just as herbicides but as molecular guidelines for the engineering of plant growth and development [15].

2. Materials and Methods

2.1. De Novo Design of Putative Auxins and Molecular Decoys

The crystal structure of the TIR1/SCP complex of Arabidopsis thaliana was obtained from the Protein Data Bank. The structure with PDBID 2P1P was selected as its values for resolution, R-free, and R-work are very good, also it is cocrystallized with IAA. The protein was then preprocessed with PDBFixer to add hydrogen atoms, IAA, and the cofactors were removed prior to modelling.

OpenGrowth [16] was used for the de novo generation of molecules within the binding site for IAA. As OpenGrowth uses Openbabel as the backend for compound minimization and atom typing, version 2.4.1 was manually compiled. The parameters for OpenGrowth were implemented as described in Table S1 in the Supplementary Information. Minimization was carried out using the MMFF94 force field and the SMoG2016 scoring function was used for ligand ranking. As a result, more than 1200 ligands were generated with these settings. A heuristic threshold of -8.0 kcal/mol was applied to select candidates for virtual screening.

To establish a set of nonbinders, molecular decoys were obtained with DeepCoy [17]; these were generated to match DUD-E descriptors in a 150:1 decoy-to-reference ratio, following the selection of top decoys using the provided scripts within DeepCoy.

As proof of concept, two sets of molecular descriptors were calculated. The first one explored Lipinski's space, composed of molecular weight (MW), topological polar surface area (TPSA), cLogP, rotable bonds, number of H-bond acceptors, and donors. The second set was selected from the molecular descriptors available in Mordred [18]. As an initial filter, features with zero variance or with less than three different values (semiconstant variables) were discarded. Highly correlated descriptors were also excluded, i.e., those with

a Kendall correlation coefficient higher than 0.9; the final set contained 395 of the original 1613 variables. Then, a subset of features was selected using the Boruta algorithm [19]; these are listed in Table S2 in the Supplementary Information. Finally, these two sets of descriptors were visualized via principal component analysis (PCA) and t-SNE using DataWarrior [20].

2.2. Machine Learning

To evaluate the overall pertinence of the generated compounds, we trained different classification models as proof of concept. These included logistic regression (LR), supporting vector machines (SVC), random forest (RF), and extreme gradient boosting classifiers (XGBoost classifier; XGBC). The comparison was performed with Scikit-learn (v.1.2.2) [21] and XGBoost [22] Python modules. Each model was trained and validated by ten-fold cross-validation, while grid search optimization was used to fine-tune their respective hyperparameters [23]. Each compound was represented as a concatenated vector of molecular descriptors selected with Boruta, which was then used for training. In addition, a second iteration was made to train classifiers with topological information; i.e., extended connectivity fingerprints (ECFP4), WHIM, and GETAWAY descriptors, calculated with RDKit (v.2022.9.5) [24]. Given the numerical feature differences, each representation was standardized, subtracting the mean and scaling to unit variance using Scikit-learn.

Classification performance was evaluated with several measures, including validation curves, learning curves, ROC curves, Matthews correlation coefficient, and detection error trade-off curves.

2.3. Molecular Modelling of Auxins

To better understand molecular recognition by TIR1, we propose a pipeline using mixed-solvent molecular dynamics, conventional molecular dynamics, and enhanced sampling with coarse metadynamics. Detailed information is provided in the following subsections.

2.3.1. Mixed Solvent Molecular Dynamics

To explore the plasticity and overall presence of hotspots within TIR1, molecular dynamics with mixed solvents was used. The MXMD script [25] was modified to include the probes included in Scheme 1.
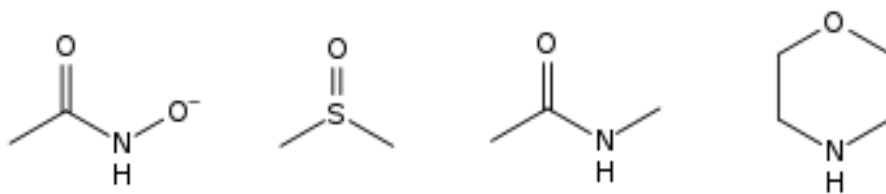
Scheme 1. Chemical structures of the selected probes for hotspot identification.

A modified version of the Ghanakota et al. protocol was used [26]; briefly, the protocol involves the creation of replica systems with a volume/volume concentration of 95/5% (water/probe). MXMD uses pre-equilibrated boxes of the selected cosolvent probe, which is then incorporated in a cubic box of water buffering the whole system by 15 Å. These systems were then submitted to default relaxation protocol and 15 ns of equilibration time. Production times were set to 20 ns, as it provides a good compromise between sampling and computing time [27]. Each MXMD run consisted of 15 simulations; three for each probe, accounting for an overall sampling of 300 ns. MXMD simulations were carried out with PDBIDs 2P1M, 2P1P, and 2P1N, which correspond to TIR1 with IHP cofactor; TIR1 bound to IAA and IHP; and TIR1 bound to 2,4-D and IHP, respectively.

Hotspots were then evaluated using the following score:

$$Score = \sum_{p}^{probes} \sum_{s}^{spots} \sum_{g}^{grids} Z_{xyz}(p,s,g) \tag{1}$$

where Z_{xyz} is the occupancy of functional groups of the probes, converted to Z-scores using the following equation:

$$Z_{xyz} = \frac{X_{xyz} - <X>}{\sigma} \tag{2}$$

2.3.2. Assessment of Pocket Solvation and Its Role in Auxin Recognition

Molecular recognition is often grounded on the "desolvation paradigm", which considers that water molecules need to be displaced by ligands during binding events [28]. Only recently, this notion has been challenged with the discovery of protein families, which possess a tightly bound water network, such as bromodomains. It has been shown that these complex arrangements serve as enthalpy control for ligand binding [29]. Additionally, it has been suggested that a robust description of binding modes is more consistent when ligands are properly solvated [30,31].

Thus, we conducted a series of simulations under different starting conditions using 3D structures with the following PDBIs 2P1N, 2P1O, 2P1P, and 2P1Q. Structures were prepared with pdbfixer to add missing atoms and/or residues followed by hydrogen bond optimization using PropKa and Maestro's protein preparation wizard. From here, two independent systems were built: with and without preserving water molecules in the vicinity of cocrystal ligands. Both systems had the protein buffered in orthorhombic boxes of TIP3P water and NaCl at 0.15 M.

TIR1-ligand complexes were then submitted to molecular dynamics simulations using Desmond (v.6.1) [32]. Systems were parametrized with the OPLS_2005 force field simulation and integration was carried out with the RESPA algorithm using 4 and 8 fs values for timesteps. A hydrogen mass repartition scheme was implemented to allow the timesteps to increase [33]. The prepared systems were then relaxed using Desmond's default protocol. This includes Brownian dynamics and an equilibration period to attain NVT and NPT conditions. The target values were 300 K for temperature and 1 atm for pressure. Following system relaxation, the complex was equilibrated for 50 ns under the NPT ensemble using MTTK barostat and Nosé–Hoover chain thermostat. Production runs of 500 ns for both systems of each PDBID were obtained and analyzed for RMSD and RMSF with MDTraj (v.1.98) [34]. In addition, interaction fractions were obtained with Maestro. TIR1-auxin contacts were also used to define microstates to construct transition networks to assess binding mode metastability.

Finally, we tested pocket solvation using grand-canonical Monte Carlo (GCMC) as it is a well-documented method with notable results [35–38]. As proof of concept, we selected the IAA/Tryptophan pair. As positive and negative controls, respectively, IAA was used as is from PDBID 2P1P and tryptophan was docked to the structure using PLANTS (more details in the following section). Each complex was prepared as described above, without preserving the water molecules of the binding site. Pocket solvation was achieved using the GCMC plugin in Desmond. This required defining the MC region in a buffer of 6 Å around ligands; solvent parameters were default values for TIP3P water, similar to those described elsewhere [39].

Simulations were carried out in five stages, beginning with GCMC/MD with restraints on heavy atoms, consisting of 10,000 MC moves per cycle and a total of 500 ps of MD. This stage was followed by GCMC/MD with 100,000 moves per cycle, 1 ns of MD, and no restraints. Systems were then relaxed with short runs of Brownian dynamics and Langevin dynamics for equilibration of volume and pressure. For these simulations, hydrogen mass repartitioning was not used; thus, timestep values were kept at default. An additional equilibration of 10 ns using the MTTK barostat and the Nosé–Hoover thermostat was carried out. Hydrogen mass repartition was implemented during this stage.

Finally, production runs of 500 ns were made for each ligand, keeping conditions from the previous step.

2.3.3. Molecular Dynamics

We compared a series of pairs, consisting of known ligands and chemically related "decoys". Decoys were intended to be low affinity/nonbinders for TIR1 based on experimental data [40,41] while also having structural similarity to known ligands. Selected pairs were IAA/Tryptophan; α-naphthylacetic acid/β-naphthylacetic acid; Diclophenac/Rinskor and Fluroxypyr/Quinclorac.

As a first step, compounds were docked to PDBID 2P1P using PLANTS software (v.1.2) as it is a well-established method with high accuracy and consistency [42–45]. Ligands were prepared using SPORES [46] to obtain proper input for PLANTS. The binding site was defined in a sphere with a radius of 15 Å centered with IAA's crystallographic coordinates. Additional parameters were kept at default values.

The top-ranked pose was selected for GCMC/MD as detailed previously and production times were kept at 500 ns for each ligand. In addition to the previously mentioned analyses, concurrence maps were calculated using ContactMap explorer (v.0.7.0) [47].

2.3.4. Coarse Metadynamics

To evaluate the metastability and get a rough estimate of the affinity differences between binders and decoys, well-tempered metadynamics was used. This enhanced sampling method has been successfully applied to study binding events [48–52]. Metadynamics uses specific degrees of freedom, which are related to the desired reaction coordinate (collective variables; CVs). In this case, we applied two CVs; the first was defined using distances between centers of mass (protein and ligand). For the protein, residues 435–437 were selected and the ligand's heavy atoms delimited the second center of mass. The remaining CV was defined from WHIM's second principal moment using the ligand's heavy atoms. The bias factor was set to 5.0 while the Gaussian height was set to 0.08 kcal/mol with a width of 0.25 Å. The Gaussians were deposited every 250 steps; additional parameters were similar to those detailed in Section 2.3.2.

Additionally, we examined the role of the so-called engagement niche, a region composed of residues K410, S440, G441, A464, and P465 [53]. For this case, the compounds were docked at a centroid of these residues in PDBID 2P1M. The same definition of CVs was used. Both metadynamics simulations had production times of 50 ns.

3. Results and Discussion

3.1. De Novo Design of Putative Auxins and Decoys

Figure 1 shows the chemical space delimited by the selected descriptors with Boruta. Loading values can be found in Table S3 in the Supplementary Information. A good superposition between generated binders and decoys was observed in both principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE). This suggests that classes may not be linearly separable; a similar behavior was found using the Rule of Five descriptors (Figure S1). Such a trait is positive, as it shows that decoys are indeed robust due to sharing similar physicochemical spaces, which in turn may lead to effective discrimination of its latent space with a good value as screening tools. In addition, this lack of clear separation underscores the challenge of effective classification using only physicochemical descriptors. Overreliance on molecular properties leading to negative effects, such as molecular inflation, has been reported [54,55]. Having established a good baseline, we trained different classifiers using machine learning algorithms.
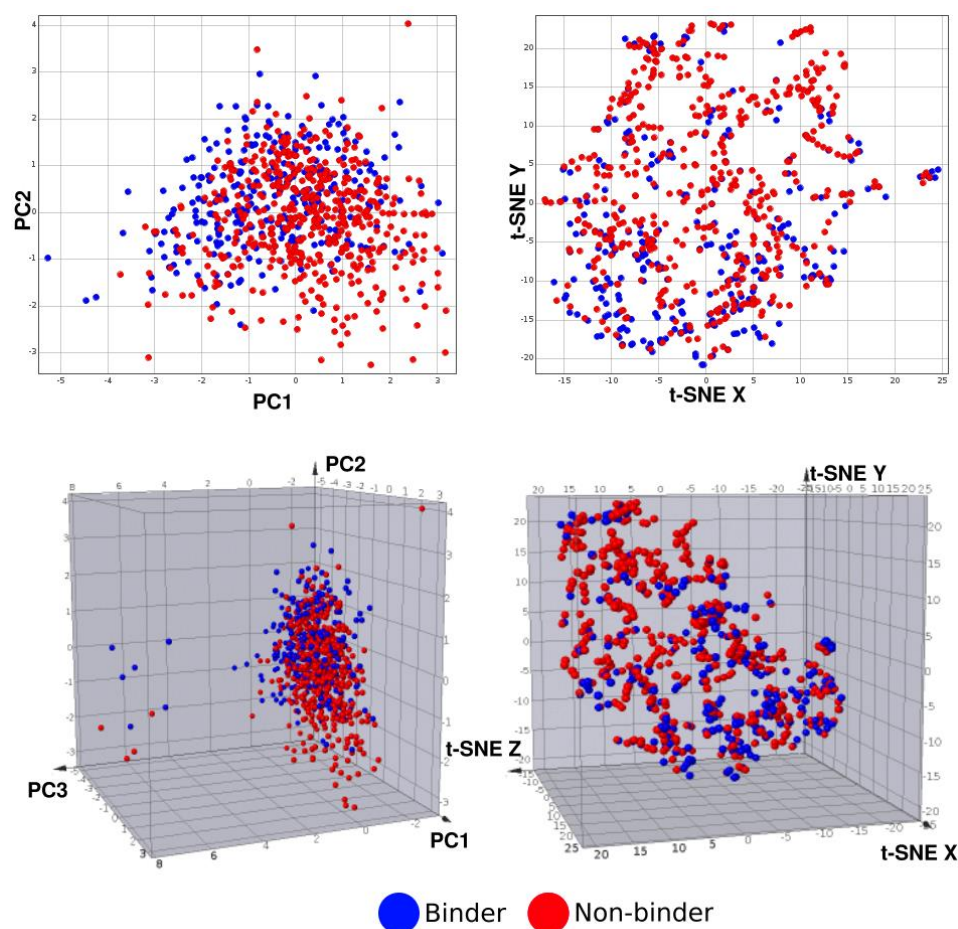
Figure 1. Chemical spaces of the generated compounds using descriptors selected with Boruta.

3.2. Machine Learning

Classification and prediction of biological activity based on latent spaces have been attempted since early quantitative structure activity relationship (QSAR) studies. Still, after several decades and unexpected results, it became clear that the QSAR is hardly linear in nature [56], with the added complexity of highly similar compounds with highly variable biological activity, an issue now defined as activity cliffs [57]. Such a concept has evolved to activity ridges due to the consistent presence of such behavior in certain scaffolds and not just compounds [58].

The emergence of empirical rules, such as Lipinski's Rule of Five [59], suggested the existence of latent spaces for drug discovery. The notion of a drug-likeness has been challenged and perhaps clearly debunked [60]. Herein, we did not intend to find a set of empirical rules to characterize putative auxins but rather a starting point for chemical space exploration and expansion. We aim to aid the rational design of auxins and related agents. To this end, we used the generated compounds to train commonly reported classifying algorithms in order to verify chemical space coverage and the overall significance of structural information within our data set. For classification tasks, several methods are used in conjunction to evaluate the accuracy and consistency of predictions. Hereunder, we selected a combination of four measurements that convey a general picture of both the data and the model's performance:

1. Validations curves: To visualize the trend of variance and suggest the presence of bias; either underfitting or overfitting.
2. Learning curves: Convey information on the learning rate as a function of *n* in data; i.e., if the provided examples are enough for effective classification or if the addition of data may benefit scoring.

3. ROC curves and Matthews correlation coefficient (MCC): These provide a general description of the predictive power due to their relation with the confusion matrix. The inclusion of both is complementary. As MCC can take values between -1 and 1, it is possible to assess if a model is indeed providing a significant difference when applied or if its success relies more on chance, even when the area under the curve of the ROC plot would suggest otherwise.

4. Detection error trade-off: Similar to the ROC curve, this metric gives information on the success of classification tasks. During classification, there is always a trade-off between the rate of falsely classified values. Therefore, this plot conveys the trend of said trade-off; an ideal classifier would keep a low rate of both false positives and false negatives, resulting in a curve downward and to the left, an inverse of sorts to an expected ROC curve. All these plots can be found in the Supplementary Information (Figures S2–S9). ROC curves and error trade-off for the models' best models are shown in Figure 2.
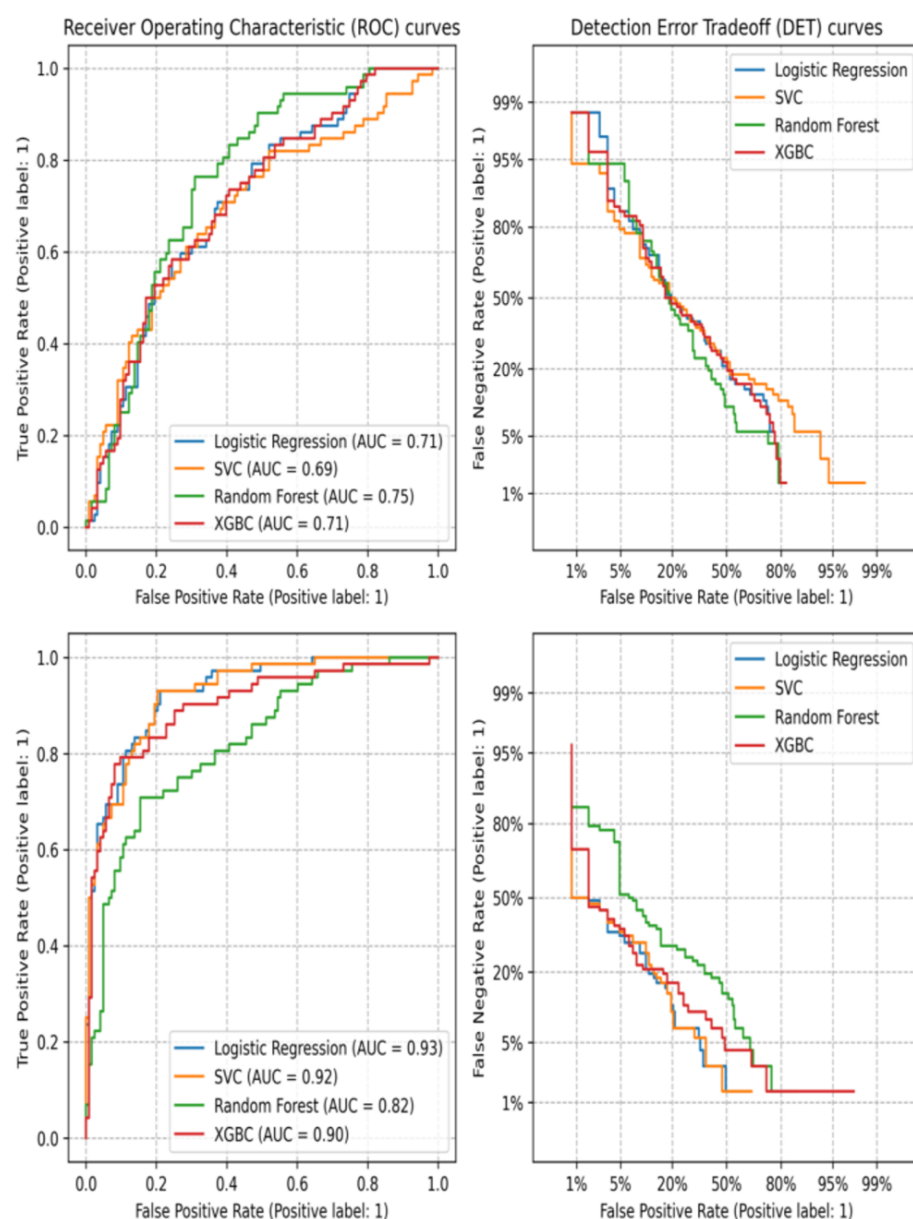
Figure 2. Performance assessment of the trained classifiers: using only physicochemical descriptors (**top**); using physicochemical and structural descriptors (**bottom**).

As expected, models trained solely with physicochemical data showed low performance. Based on the area under the ROC curves, algorithms based on linear methods performed worst. Moreover, it is quite significant that the classifiers were barely superior to random chance. When analyzing the detection error trade-off, it also gives an almost linear trend, suggesting that any improvement in the discrimination of either false positives or negatives inherently results in an increase in misclassification of the other. This underscores that proper discrimination between auxins and nonbinders is no easy task. Even random forest and supporting vector classifiers, which are often regarded as very robust [61–65], showed poor performance. In addition, a strong tendency to overfitting was found for these models (see Figures S4 and S6 in the Supplementary Information).

Such results verify the convenience of generating molecular decoys. A common use of molecular decoys is for structure-based design. Thus, decoys ought to be molecules that share chemical space with actives, with little to no pharmacophoric similarity. The quest for such compounds resulted in the construction of datasets such as DEKOIS [66] or, most notably, DUDE-Z [67]. Only recently have use cases been extended to ligand-based methods and scoring function construction [68,69]. Hence, we have confirmed that decoys do share the chemical space using two sets of very different descriptors.

To confirm the second half of the provided definition, topological and structural features shall be more conclusive in this classification test. Indeed, consideration of these features proved very significant, as the area under the ROC curves increased roughly by 0.2 in all cases. A random forest classifier showed the lowest AUC value, consistent with its MCC and error trade-off. Following this logic, it would seem that logistic regression was the top performer. Yet, such is not the case, as its MCC was 0.53. For this test, nonlinear methods were more consistent and showed significantly better performance. At first glance, SVC would seem the superior choice; then again, comparing MCC values suggests that XGBC shall be preferred.

Our goal was achieved, as the generated compounds do serve as a reasonable sample for further analysis while also conveying enough information on true binders; still, there is no denying that the rate of false negatives remains quite high. To some extent, this result was to be expected, as no computational method is completely accurate and reliable on its own. Thus, we used structure-based methods to gain further information on what characterizes a true binder of TIR1.

3.3. Molecular Modelling of Auxins

3.3.1. Mixed Solvents Molecular Dynamics

Protein plasticity is paramount for their function, still this results in an outstanding challenge to tackle, computationally speaking. A quick comparison between crystallographic structures of TIR1 (Table 1) shows little conformational variations. Nonetheless, in a dynamic context, substantial differences may arise, with critical implications for binding events [70,71]. Hence, to better understand the overall flexibility and features of the auxin binding pocket, we conducted molecular dynamics with mixed solvents.
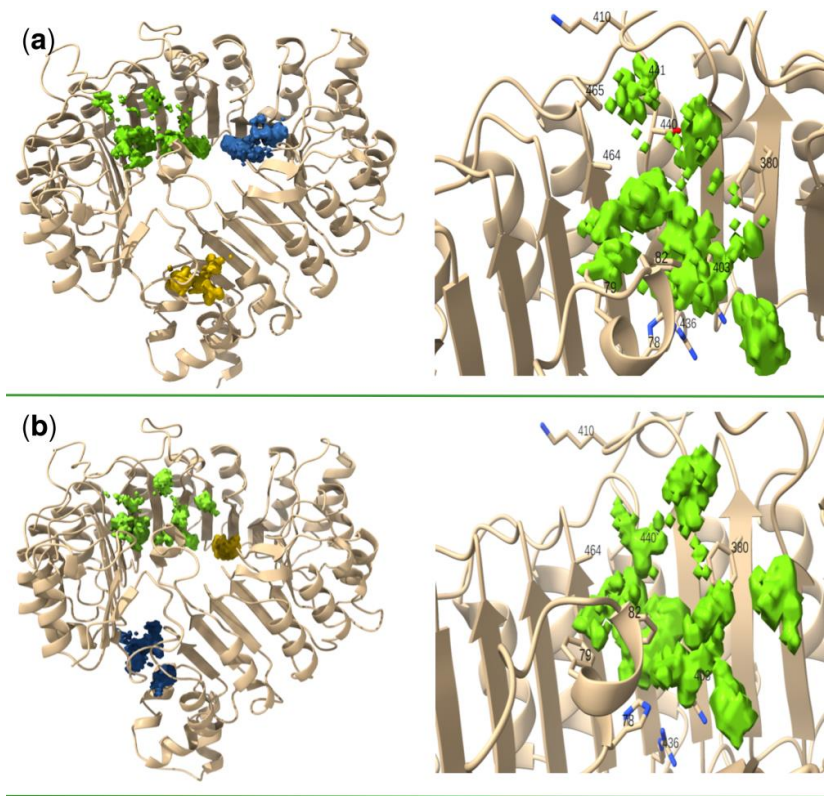
During drug discovery campaigns, molecular design often follows a fragment-based approach where protein structures are crystallized iteratively in the presence of probe scaffolds or leads, thus creating an empirical mapping for binding phenomena [72]. A computational equivalent of this can be found in molecular dynamics with mixed solvents, where a rather discrete amount of polar solvent molecules is introduced to the simulation cell. Following an equilibration period, the probe molecules will interact with the protein. Therefore, the occupancy of these can be evaluated and further compared. A notable advantage of such protocol is that the calculation of occupancy can be made from different probes as a means of attaining consensus, which, in turn, increases descriptive power and may be used to delimit allosteric or even cryptic pockets within proteins [73].

For this work, 45 independent simulations of TIR1 structures were produced, accounting for 900 ns. Figure 3 shows the computed surfaces from this sampling. Notably, the auxin binding pocket had the highest scoring.
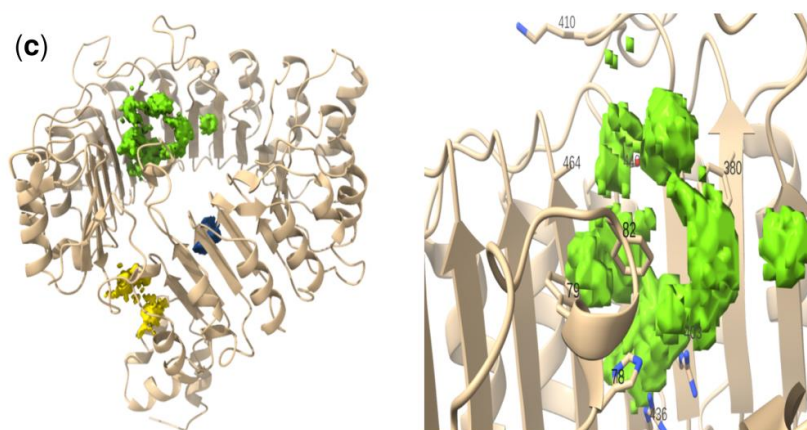
Table 1. Root mean squared deviation values from the structural alignment of TIR1 structures. Binding site residues with above-average deviations are shown.

PDBID	2P1M	2P1P	2P1Q	2P1O	2P1N
2P1M		(RMSD 0.16 Å) H78 − 0.32 F79 − 0.18 D81 − 0.31 F82 − 0.19 C405 − 0.27 S438 − 0.19 L439 − 0.22 S440 − 0.19 S462 − 0.18	(RMSD 0.22 Å) H78 − 0.3 F79 − 0.3 D81 − 0.58 F82 − 0.45 C405 − 0.27 R435 − 0.23 S438 − 0.23 L439 − 0.3 R489 − 0.22	(RMSD 0.27 Å) H78 − 0.36 D81 − 0.59 F82 − 0.52 C405 − 0.29 L439 − 0.87	(RMSD 0.33 Å) H78 − 0.41 F79 − 0.38 D81 − 0.55 F82 − 0.35 C405 − 0.54 L439 − 0.85 V463 − 0.49 A464 − 0.58 R489 − 0.33
2P1P			(RMSD 0.16 Å) F79 − 0.18 D81 − 0.28 F82 − 0.52 L439 − 0.2 A464 − 0.2 R484 − 0.22	(RMSD 0.22 Å) D81 − 0.29 F82 − 0.6 A464 − 0.22	(RMSD 0.26 Å) F79 − 0.26 D81 − 0.29 F82 − 0.45 C405 − 0.37 R435 − 0.27 L439 − 0.66 V463 − 0.37 A464 − 0.54
2P1Q				(RMSD 0.14 Å) F49 − 0.15 L439 − 0.62 S440 − 0.16	(RMSD 0.22 Å) H78 − 0.28 C405 − 0.29 R435 − 0.35 L439 − 0.62 V463 − 0.42 A464 − 0.4
2P1O					(RMSD 0.19 Å) H78 − 0.3 F82 − 0.22 C405 − 0.31 R435 − 0.28 V463 − 0.43 A464 − 0.42

Figure 3. *Cont*.

Figure 3. Probe occupancy surfaces recovered by molecular dynamics with mixed solvents using PBDIDs 2P1M (**a**), 2P1P (**b**), and 2P1N (**c**). Pockets with top-three scores are shown in green, blue, and yellow, respectively.

In general, the conformational variations in TIR1 did result in noticeable changes in the shape of the druggable regions. Overall, the observed ring shape is suggestive of the known mechanism of auxins, as these stabilize the protein–protein interface. Considering that said shape is more prominent in 2P1N, a link between ligand polarity and pocket dynamics is suggested.

As for individual probes, in 2P1M major occupancy was found for N-methylacetamide and acetohydroxamic acid. In contrast, 2P1P and 2P1N showed a major occupancy of drug-like probes, particularly imidazole.

Additionally, significant changes in the neighboring region to the engagement niche were found. This region comprises a phenylalanine triad (F79, 82, and 380), of which F82 has been reported as significant for the molecular perception of auxins [74]. Based on the alignments of Table 1, F79 and 82 are consistently flexible between structures. A higher deviation can be found precisely between 2P1P and 2P1Q, suggesting a transition between hydrophobic and stacking interactions. This, together with the observed changes in molecular dynamics with mixed solvents, reinforces the importance of stacking interactions within the binding pocket.

3.3.2. Assessment of Pocket Solvation and Its Role in Auxin Recognition

As stated in the Methods Section, only recently has the role of pocket solvation been critically assessed within the context of small molecule binding. Based on crystallographic structures, it does seem that TIR1 has several water molecules that ought to participate in auxin binding [75]. To determine if such contribution is relevant for auxin perception, we compared protein–ligand interactions under different starting conditions of solvation (Figure 4).

Major differences between the interaction profiles include the contact with R436. This interaction seems to be mostly driven by water bridges, suggesting a change in ligand orientation and directionality. From the previous section, residues showing changes between TIR1 conformations are H78, L439, and A464, with the latter being more displaced when 2,4-D is bound to TIR1. This trait is noteworthy, as A464 belongs to the engagement niche. Perhaps the observed variation for L439 can also be linked to the presence of water bridges. Unfortunately, this was only observed for IAA, so a clear reason remains elusive. Nevertheless, this may be due to the presence of a nitrogen atom in the indole scaffold, a feature absent in other binders under study. Still, previous studies have suggested that the presence of this heteroatom is paramount for the potency and overall affinity of IAA [76]. Hence, it does seem that water interactions are significant within the engagement niche.

H78 also showed subtle but significant differences; unfortunately, it is difficult to assess if these are the result of solvation or merely the nature of numerical simulation. Another subtle effect is the change in hydrophobic contacts to stacking interactions. It can be argued

that, due to the lipophilicity of auxins, these are prone to the latter. However, considering the cases of NAA and 2,4-D (2P1O and 2P1N, respectively), interaction distribution and persistence did change due to the solvation environment. Figures S11 and S12 show some examples of this (vide infra).
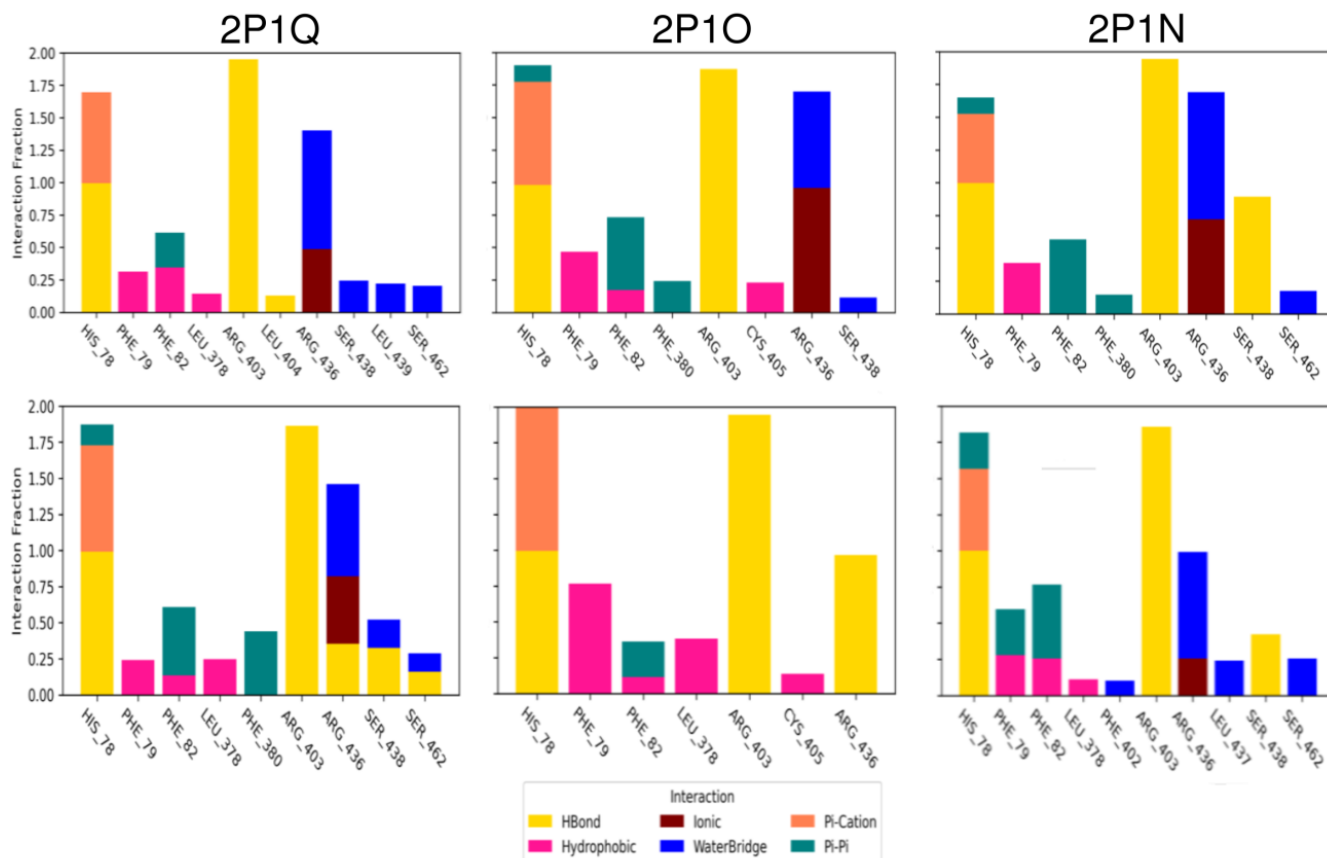
Figure 4. Comparison of protein–ligand interactions for the auxin/TIR1 complexes (PDBIDs 2P1Q, 2P1O, and 2P1N). Whenever the interaction fraction goes beyond 1, it is indicative of more than one contact with said residue. Top row shows results from simulations preserving crystal water molecules, bottom row shows results discarding said waters.

For further proof, we constructed transition networks based on protein–ligand interactions. Briefly, these networks encode microstate information. Hence, neighboring microstates share similar information. If these clusters are densely populated, this can be related to metastability, delimiting a local minimum [77]. Plus, given enough data, these can even confer kinetic information [78]. Thus, it is very important to delimit which information is used to describe said microstates. Due to this, an additional criterion for comparison was introduced. Transition networks were then constructed with and without the inclusion of water bridges in their definition. These involve a hydrogen bond with the following criteria: a distance of 2.8 Å between the donor and acceptor atoms; an angle $\geq 110°$ between donor-hydrogen-acceptor atoms; an acceptor angle $\geq 90°$ between hydrogen-acceptor-bonded atoms. Figure 5 shows the obtained transition networks.

As suggested previously, water–ligand interactions do have a significant impact on auxin binding. Metastability of binding modes was present in all three simulations with crystal waters; plus, this trait is further accentuated with the introduction of water bridges to microstate definition. Such behavior is more noticeable for the synthetic auxins, reinforcing the established link between ligand polarity and pocket dynamics.
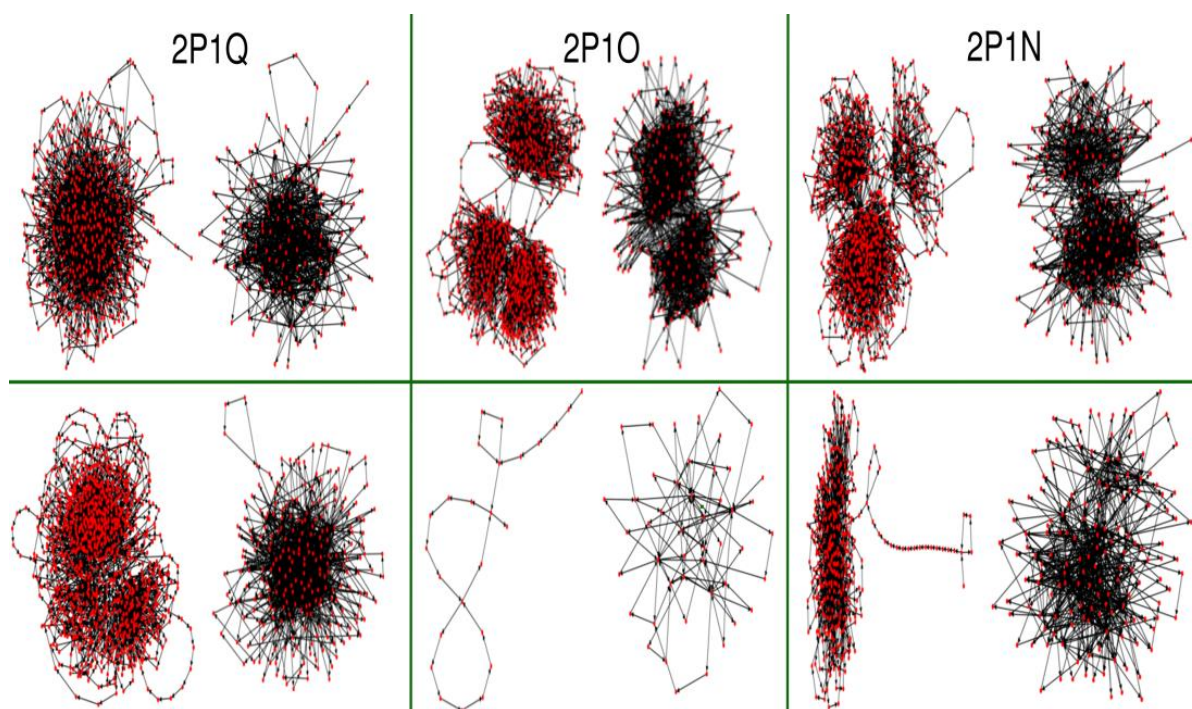
Figure 5. Transition networks built from protein–ligand interactions. Top row shows results from simulations preserving crystal water molecules; bottom row shows results discarding these waters.

In stark contrast, transition networks from conventional protocols were highly variable. For IAA's case (2P1Q), an inverse pattern is observed as it shows some degree of metastability based on the three clusters and a densely populated network. NAA, on the other hand, showed highly transient interactions. While some states are indeed revisited, no apparent clustering arises under these conditions. The omission of water bridges does result in a more complex network; then again, the transient nature of contacts persists. Finally, for 2,4-D, metastability is clearly shown by three well-defined clusters, suggesting a local minima or basin.

The inclusion of water-mediated interactions did improve the representation and gave a more robust description of basins and metastability. It can be argued that such observations are to be expected due to the complex nature of binding and the balance of enthalpic and entropic contributions. Nonetheless, our aim is not to strictly evaluate the role of water interactions but their general contribution to discern between binders and nonbinders. To this end, we conducted GCMC/MD simulations with IAA and TRP. For these simulations, PDBID 2P1P was used. The comparison was between a common MD protocol and GCMC/MD using similar analyses described above. Figure 6 shows protein–ligand interactions retrieved from IAA and TRP. For the most part, the GCMC/MD protocol was successful in establishing a water network within the pocket. The prominent water bridge between the ligand and R436 was also recovered. This is a positive result, as it highlights the robustness of the method and its placement of solvent molecules.

On the other hand, the interaction profile remains consistent and similar to those in previous simulations, providing further proof of the significance of water bridges in microstate definition. For IAA, perhaps a notable change is the dominant stacking interaction with F79 instead of a hydrophobic contact.

In contrast, one of the most consistent residues was R436, which, in turn, suggests that this water bridge can be regarded as "canonical" and not anecdotal behavior. Moreover, the complete absence of said bridge in conventional protocols is very interesting. A similar case was found in 2P1O; this interaction was not able to form under standard conditions. Of note, simulations of both 2P1Q and 2P1N were able to form this interaction. The fact

these subtle variations within the binding site of TIR1 introduce such divergent results is remarkable. This stands as a cautionary tale for the well-known limitations of structure base methods, while also highlighting that structure selection is no trivial matter [79,80]. Herein, it was found that 2P1P is prone to the loss of ligand–solvent interactions, whereas 2P1Q and 2P1N may yield more reliable and consistent results. For further assessment, we analyzed the root mean square fluctuation of auxins in these sets of simulations (Figures S10–S13).
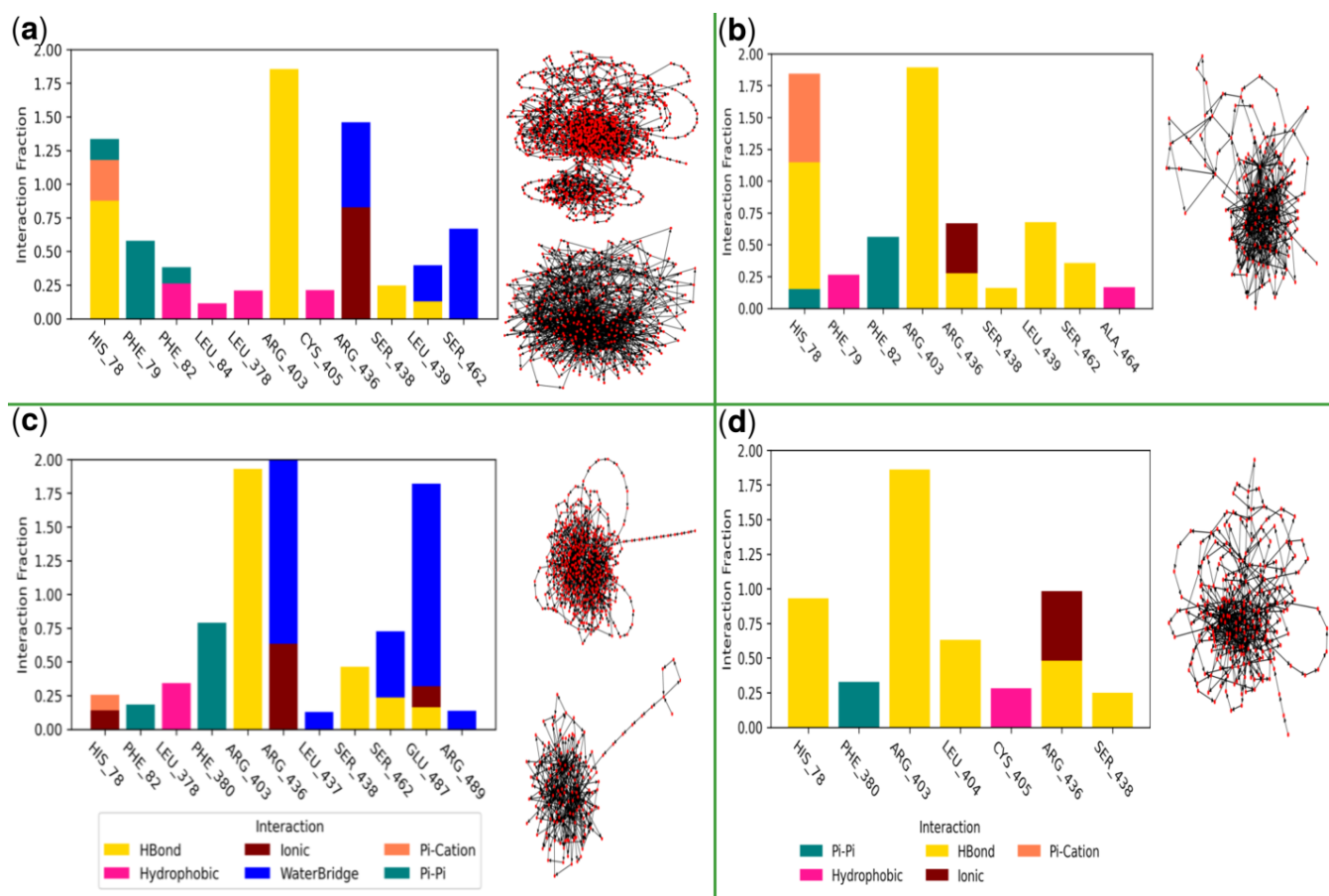
Figure 6. Influence of pocket solvation in TIR1. Interaction fractions and transition networks are compared between IAA (**a**,**b**) and TRP (**c**,**d**). Simulation protocols under conventional MD (**left**) and GCMC/MD (**right**) are also compared.

Considering IAA and TRP in 2P1P, RMSF plots indicate higher fluctuation when water molecules are correctly placed in the pocket. Under a standard protocol, the fluctuations are significantly suppressed, more so for IAA's case. For both ligands, higher fluctuation was observed on the nitrogen side of the scaffold. This is worth mentioning, as this part of the molecule faces in the vicinity of F79.

When comparing NAA and 2,4-D, interesting patterns were found. For the former, the fluctuation profile was very persistent with respect to its fit within the binding site, with the main difference being found in the carboxylic acid moiety, which may be explained by the loss of the water bridge with R436. For 2,4-D, the fluctuation profile changed but in contrast to IAA, TRP, and NAA, higher values were observed under standard conditions. Taking into account that atoms showing this behavior were both chlorines and the carbon atoms of this side of the phenyl ring, this is yet another case where the solvation environment ought to be the main driving force.

As for the ligand's fit, the fluctuation was very consistent and almost unaffected. In all cases, the highest value was found in the charged oxygen. For TRP, this was more prominent and even expected due to the amino group. Beyond that, only 2,4-D showed differences in fluctuation for this atom between protocols.

3.3.3. Molecular Dynamics

In view of prior results, we used an integrative approach for the analyses of known binders and nonbinders. In an attempt to facilitate the correct assessment of binding, we propose the use of protein–ligand interaction fractions, concurrence maps, transition networks, and ligand–protein interactions simultaneously. We present two cases as proof of concept: diclofenac (DCF) and quinclorac (QCL). Both are arylic carboxilic acids, with quinclorac being an auxin herbicide. However, the molecular target of quinclorac is not TIR1 but rather F-box protein ABP5 [40].

DCF is a nonsteroidal anti-inflammatory agent (NSAID) and, thus, not an auxin per se but it has been shown that exposure to this drug can cause effects similar to those of IAA [41]. Figure 7 presents the consensus analyses of the molecular dynamics simulation of DCF after docking in TIR1. Indeed, it can be observed that its interaction profile is similar to that of synthetic auxins. Nonetheless, there are no ionic or stacking interactions with H78. This may be attributed to the ring's orientation due to the hindrance of the chlorine atoms. Another significant observation is the diminished presence of the water bridge with R436. Both traits are consistent with our previous results and hypothesis on the role of auxin polarity, water bridges, and pocket dynamics.
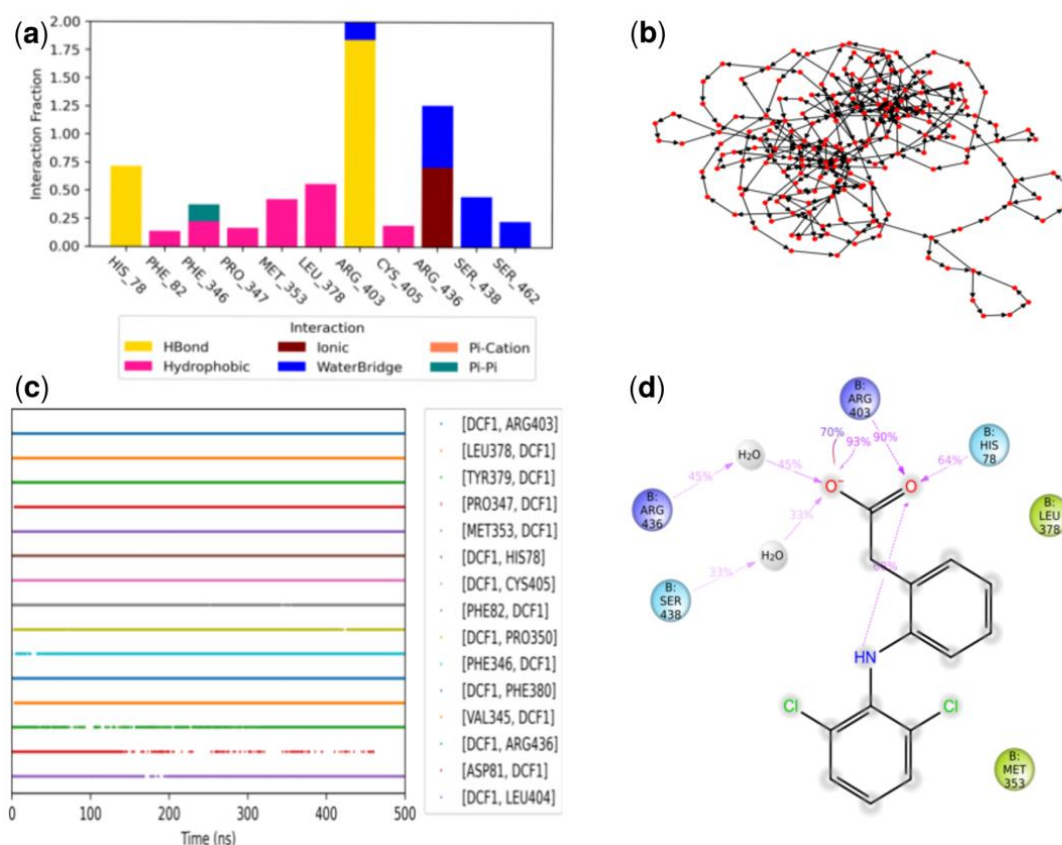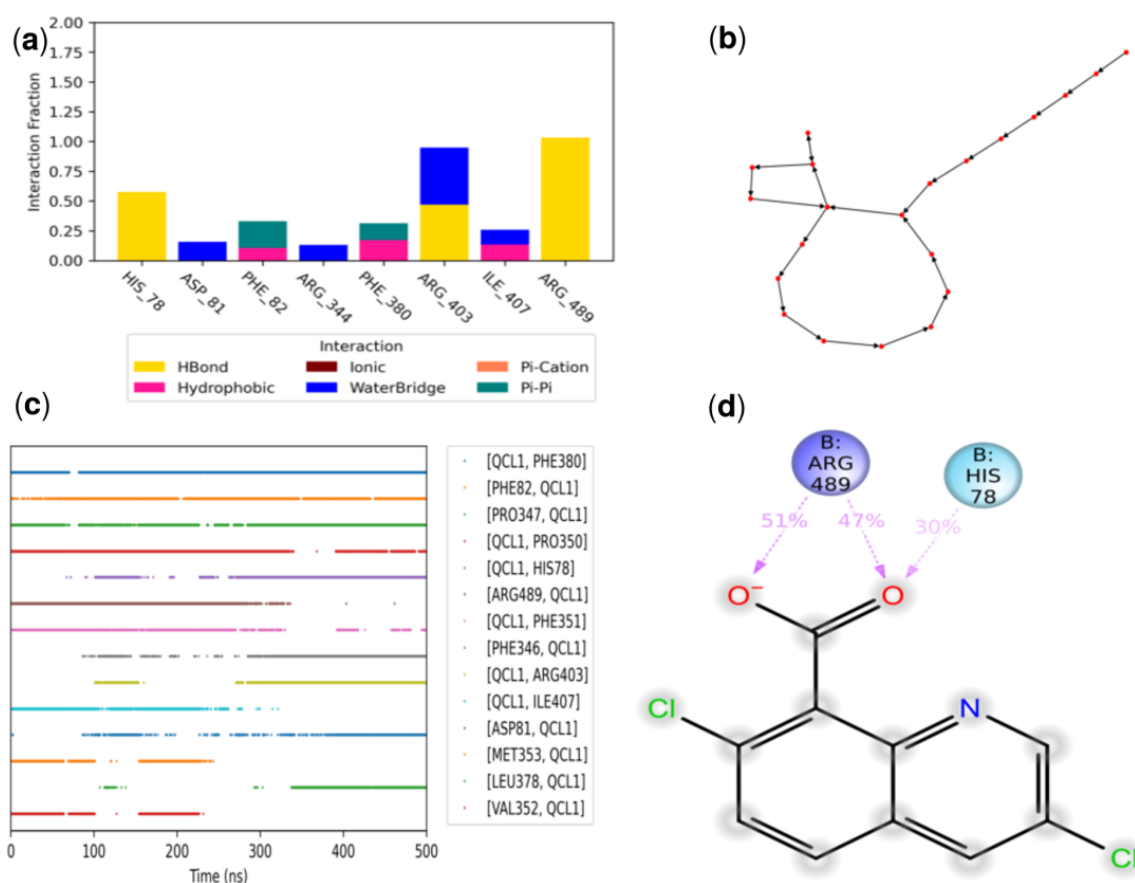

Figure 7. Consensus analysis of MD simulation of diclofenac (DCF); (**a**) Protein–ligand interaction fraction; (**b**) Concurrence of contacts; (**c**) Transition network for protein–ligand interactions; (**d**) Ligand–protein interaction.

When analyzing contact concurrence and interaction network, these suggest that DCF remains in the general vicinity of the main pocket. This, in addition to the observed basins and metastability, can be taken as an indication of true binding. These observations provide a strong case in favor of the proposed protocol; still, a criterion related to binding affinity is needed to assess the relative potency of a given ligand.

On the other hand, QCL has no experimental affinity for TIR1. Figure 8 shows the results of the consensus analyses from MD. Based on the interaction profile, all contacts are

significantly weaker for quinclorac. Additionally, the water bridge with R436 is nonexistent; perhaps it is due to this that the interaction network shows a highly transient behavior. No basins were observed and this is consistent with the concurrence plot as it suggests that quinclorac remains near hydrophobic residues; i.e., F82 and 380. Due to its planarity, the presence of stacking interactions with these hydrophobic residues may be enough to stabilize QCL within the pocket. Beyond that, any significant contact is barely present; the main anchor was R489, as confirmed by ligand–protein interactions. Yet, even this interaction is lost in favor of hydrogen bonding with H78.

Figure 8. Consensus analysis of MD simulation of quinclorac (QCL); (**a**) Protein–ligand interaction fraction; (**b**) Concurrence of contacts; (**c**) Transition network for protein–ligand interactions; (**d**) Ligand–protein interaction.

Results for the other pairs under study can be found in the Supplementary Information, Figures S14–S19. Akin to results from Section 3.3.2, IAA showed a highly stable binding. Stacking interactions with F82 and 380 were observed. The transition network showed only one highly populated basin with significant metastability. Also, the persistence of the water bridge with R436 is noteworthy.

Interestingly, as discussed in the previous section, a densely populated basin was found for TRP. Such a trait by itself would suggest that TRP is a putative ligand of TIR1. Nevertheless, inspection of the interaction profile is divergent when compared to that of known binders, just as QCL contact with H78 is diminished and mostly ionic in nature. Moreover, it must be stated that, based on concurrences and ligand protein interactions, the recovered binding mode for TRP is notably stable. Even after 500 ns, this nonbinder was able to remain within the pocket. This may be attributed to the indole scaffold due to the stacking interactions with both F82 and 380, the strong presence of the water bridge with R436, and additional water bridges with the engagement niche. May this be a notable warning that even within "long" simulations, nonbinders can be falsely identified as hits

due to pharmacophore similarity. As seen with DCF, this is yet another case where affinity would provide a decisive criterion for true binding assessment and pruning.

As for NAA, both isomers had similar concurrence maps and ligand–protein interactions. Despite this, the interaction profile of the α-napthyl acetic acid showed stacking interactions with H78, F79, and 82; traits absent for the β isomer. The main difference, however, was found in transition networks. With one of the NAA showed a clear basin with significant metastability, whereas 2-NAA showed a network more in line with the one observed for QCL. Experimental evidence confirms that, indeed, NAA has a higher potency.

Finally, we discuss the results for fluroxypyr (FXY) and rinskor (RSK); both auxin herbicides belonging to the picolinate subclass [81]. Similarly to QCL, these have a preference for ABF proteins. Still, fluroxypyr can bind significantly to TIR1. Based on the interaction profile, FXY seems to be a binder whose main anchors are H78 and R489. Interestingly, this arginine exhibits an analogous behavior to that of R436.

This added to its general positioning, makes us hypothesize that auxin orientation initiates with K410 just outside the engagement niche, with R489 serving as the pulling mechanism toward the auxin site, a premise supported by the transition network, where some minor basins seemed to be found with shared microstates. Based on the interaction profile and the presence of additional water bridges, this could be the origin of said transitions. The concurrence plot also seems to favor this, since FXY stays in the general vicinity of H78 during the simulation, making close contact with D487 only during the last segment of the run.

In turn, for RSK, the interaction profile was dominated by hydrophobic and stacking interactions. Notably, water bridges were also present, again with R489. A crucial detail, however, is that RSK presented an inverse binding; i.e., a general orientation toward the engagement niche and R489 and the phenylalanine triad. Uzunova et al., found similar arrangements for TRP using tomographic docking [53]. Perhaps this can be attributed to the absence of the carboxylic acid moiety, esterified in RSK.

Of note, both RSK and related picolinate Arylex are readily metabolized to carboxylates; it has been suggested that their high potency is due to this activation step [82]. In our case, the preservation of the ester was deliberate, precisely to test the interaction profile and prior observations. The transition network shows a putative basin with transient behavior; still, as the simulation continues, microstates are highly transient and the basin is no longer explored. In this regard, the concurrence map suggests that overall RSK remains very stable during most of the run. Therefore, this is yet another case where water bridges may be the reason behind this pattern.

3.3.4. Coarse Metadynamics

In order to have an additional criterion, for cases where consensus analyses could prove ambiguous, we conducted coarse metadynamics. Briefly, metadynamics belongs to enhanced sampling methods, intended to favor rare events. For this case, a bias is introduced as specific degrees of freedom (CVs), which are subject to Gaussian potentials. These potentials help to fill the minima defined by CVs; in the context of small molecule binding, a coarse approach has been proposed and validated elsewhere [49]. While it is true that metadynamics has been successfully applied to determine binding mode stability, a major challenge is CV selection.

Herein, we used the distance between centers of mass as it has proven to be a very robust choice in previous studies. Now, for the ligand's degrees of freedom, we settled on the WHIM descriptor due to its conceptualization and the overall positive results found during machine learning. Commonly, metadynamics simulations are made from docking poses. Therefore, in order to probe the engagement niche, we carried out docking in the vicinity of this region and used metadynamics to "force" ligands toward the binding pocket. Figure 9 shows free energy surfaces (FES) for DCF.
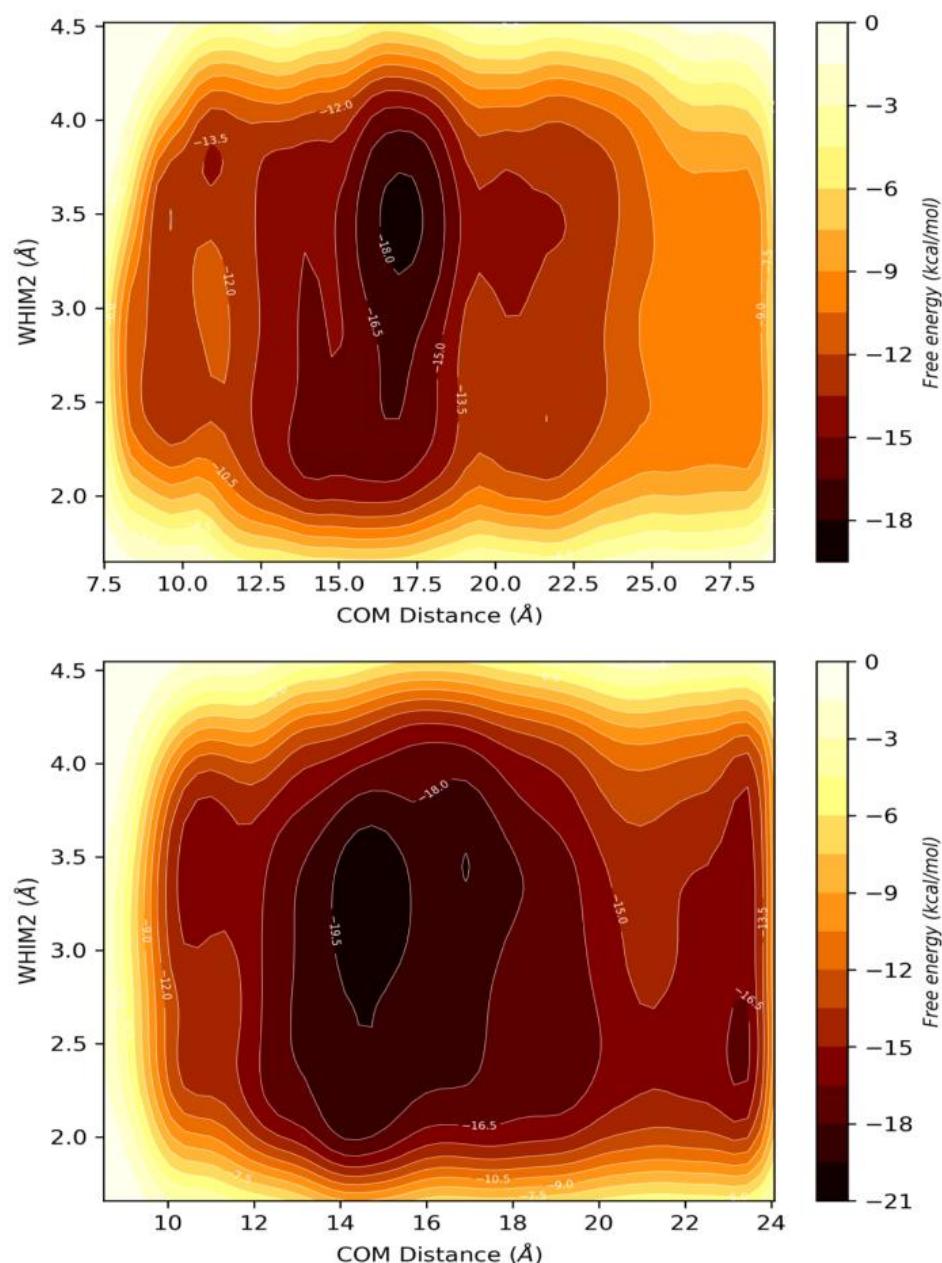
Figure 9. Free energy surfaces computed from well-tempered metadynamics for the DCF-TIR1 complexes with ligand positioned within the binding site (**top**) and positioned within the engagement niche (**bottom**).

This compound only exhibited one basin in each metadynamics run. Interestingly, the distance and affinity values for these were close. Nonetheless, when binding poses from the basins were inspected, an interesting picture emerged (Figure 10). When DCF starts within the pocket and is pulled outward, it gets stuck in the vicinity of F351 and R489. In contrast, when DCF starts from the engagement niche, it is able to enter the auxin pocket and adopts a canonical orientation toward H78, F82, and 380, indicative of true binding. It can be argued that in both arrangements, DCF is enclosed by several water molecules. Yet, this could be due, to the noisy behavior of short metadynamics. Now, taking all prior results and the computed affinity value (~−19 kcal/mol) confirms the role of DCF as a true binder. Nevertheless, the significance of this value remains relative to that of other auxins calculated with the same protocol, discussed below.
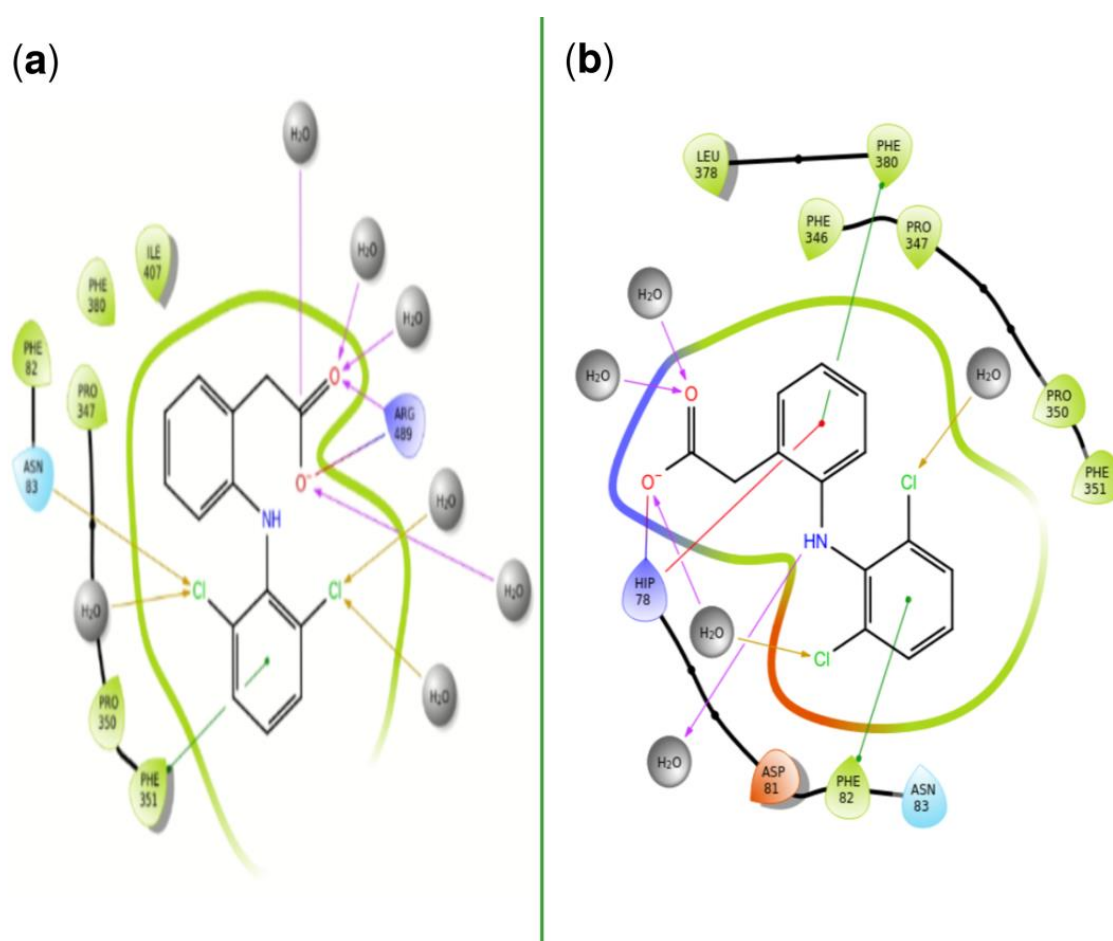
Figure 10. Representative DCF conformations found at basins from metadynamics runs with ligands positioned within the binding site (**a**) and positioned within the engagement niche (**b**).

Recovered FES for QCL are shown in Figure 11. For this compound, a clear unbinding event took place. Akin to DCF, a single basin was found, when a visual inspection of the binding mode was carried out (Figure 12), QCL was outside the neighboring region of the engagement niche. Contrary to DCF, the synthetic auxin was not able to effectively bind to any region during its exit. When QCL starts from the engagement niche, based on COM, distance values suggest that binding did occur. Following visual inspection of the binding mode, it would seem that a true binding was indeed achieved, as QCL is interacting with H78. Then again, this is only via hydrogen bonding, there is no interaction with F79, 82, or 380 nor any contact or presence of any arginine residue. Even if the suggested affinity is rather high (−32 kcal/mol), this binding mode is not characteristic of any auxin under study. Also, this value can be the result of low sampling of the observed basin. As stated, even affinity estimates from metadynamics are qualitative in nature, more so when short runs are produced. Given enough time, this value is likely to diminish toward a more representative one.

Now, for the additional auxins and nonbinders, their FES diagrams and representative binding poses from basins can be found in the supplementary information (Figures S20–S31). Herein, we just provide some details and highlights but we think that the results are self-explanatory.

IAA serves to establish a baseline for affinity values. Akin to experimental reports, it showed the highest value of all true binders (−24 kcal/mol) when initiating from within the binding pocket, followed closely by NAA isomers (−22 kcal/mol), also consistent with experimental affinity, which is usually one order of magnitude below IAA [15,40,83].

For these cases, metadynamics could not distinguish between isomers just with affinity estimates. Nonetheless, only NAA showed true binding modes in both metadynamics runs.
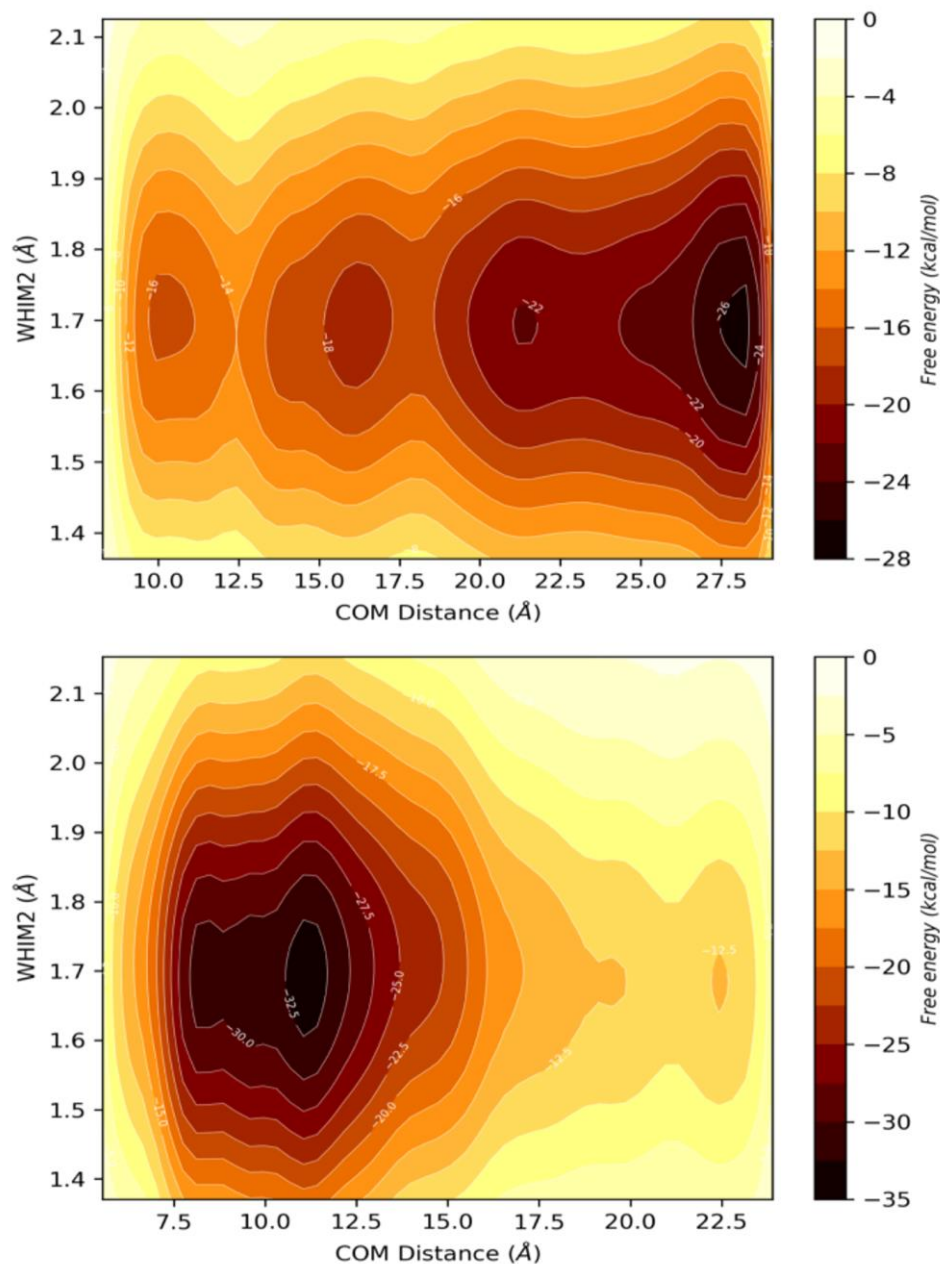
Figure 11. Free energy surfaces computed from well-tempered metadynamics for the QCL-TIR1 complexes with ligand positioned within the binding site (**top**) and positioned within the engagement niche (**bottom**).

TRP was correctly classified as a nonbinder, even when a quasicanonical arrangement was found, the affinity estimate was very low (−16 kcal/mol). RSK showed even lower values (−14 kcal/mol); notably, this was consistent between runs and its tendency to remain inversely placed within the pocket.

The case for FXY is a rather interesting one as an unbinding event occurred and it remained tightly bound to the engagement niche surface. Given the previous results, we cannot provide an explanation for this discrepancy. Indeed, this case presses for further study while also highlighting the need for additional experimental information.
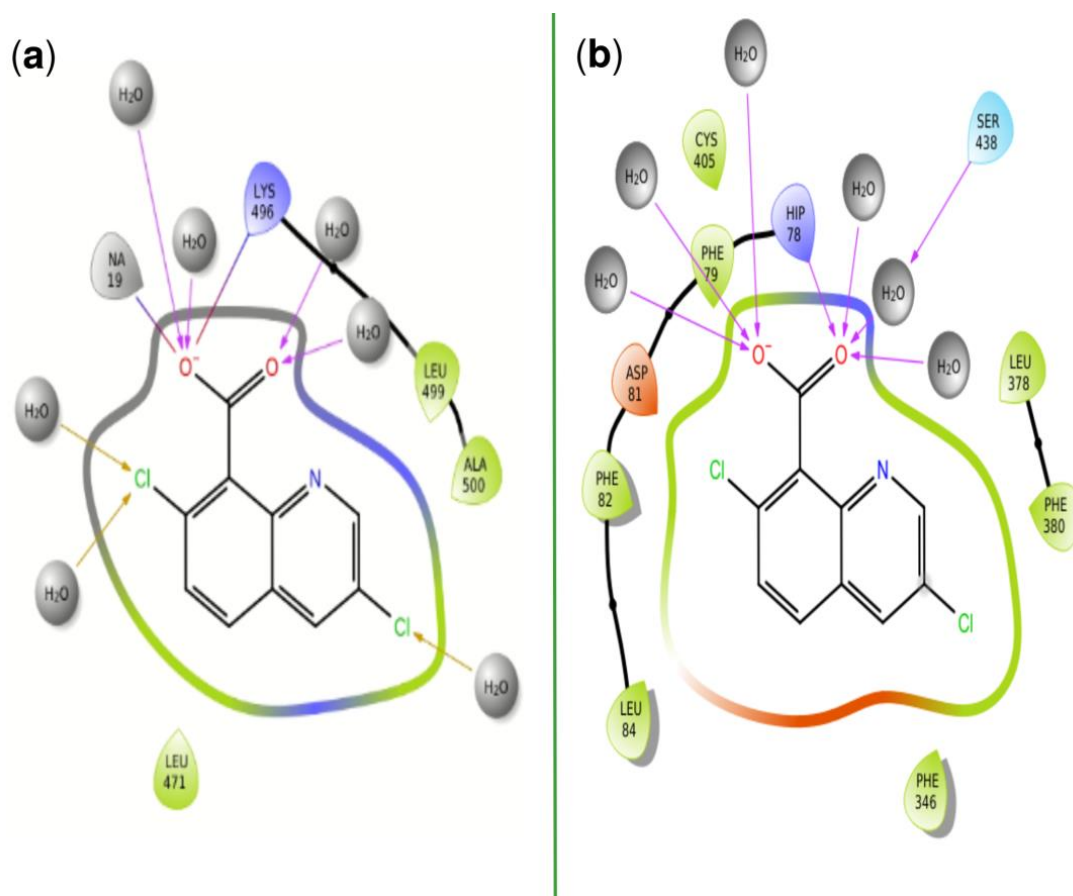
Figure 12. Representative QCL conformations found at basins from metadynamics runs with ligands positioned within the binding site (**a**) and positioned within the engagement niche (**b**).

4. Conclusions

Herbicide development is an endeavor of high interest from both an economic and chemical standpoint. Currently, there are worldwide efforts to replace commonly used agents such as glyphosate or 2,4-D. Herein, we presented a computational characterization of TIR1, one of the molecular targets associated with auxin pathways. With the aid of structure-based methods, we analyzed a series of compounds in silico; we found that water plays an important role in binding and were able to characterize significant protein–ligand interactions.

Using de novo design, we proposed a library of putative binders and molecular decoys. Chemical space showed a good superposition between them, a trait further supported by machine learning results. Herein, nonlinear methods were more consistent and performed reasonably while conveying enough information on true binders. Our results also showed that current data make correct classification of false negatives difficult. This behavior was recurrent in this and other studies.

Hence, we conducted a structure-based characterization of the TIR1 pocket to determine major traits that could provide a robust protocol for true binder assessment. We found that the inclusion of water-mediated interactions gave a more robust description of local minima and metastability. Also, we found a persistent water bridge between ligands and R436, which, to the best of our knowledge, has not been described as significant.

In addition, our results suggest that the TIR1 structures best suited for molecular modeling are 2P1Q and 2P1N as these are able to retain subtle solvent effects, even under conventional protocols.

We also probed the engagement niche, confirming its role in molecular recognition and the overall plasticity of the auxin-binding pocket. Changes in ligand polarity make

a significant impact on pocket dynamics and the orientation of hydrophobic residues involved in stacking interactions.

Using coarse metadynamics, we were able to further assess the affinity of known auxins and TIR1. Overall, the collective variables used in this study provide a robust description of ligand degrees of freedom and provide yet another criterion for virtual screening campaigns.

The database and the analyses presented in this work will be a valuable starting point for the development of molecular generators or virtual screening campaigns.

Supplementary Materials: The following Supplementary Information can be downloaded at: https://www.mdpi.com/article/10.3390/computation12050094/s1, 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); Table S1. Parameters used for OpenGrowth in the present study; Figure S2. Performance assessment of logistic regression classifier trained with physicochemical descriptors; Table S2. Molecular descriptors selected by the Boruta algorithm; Figure S3. Performance assessment of logistic regression classifier trained with physicochemical and structural descriptors; Table S3. Loading values from principal component analyses for the Boruta defined space and RO5 space; Figure S4. Performance assessment of supporting vectors classifier trained with physicochemical descriptors; Table S4. Hyperparameter values determined by grid search for the trained classifiers only with physicochemical descriptors; Figure S5. Performance assessment of supporting vectors classifier trained with physicochemical and structural descriptors; Table S5. Hyperparameter values determined by grid search for the trained classifiers with physicochemical and structural descriptors; Figure S6. Performance assessment of random forest classifier trained with physicochemical descriptors; Figure S7. Performance assessment of random forest classifier trained with physicochemical and structural descriptors; Figure S8. Performance assessment of extreme gradient boost classifier trained with physicochemical descriptors; 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.

Author Contributions: Conceptualization, F.D.P.-M.; methodology, F.D.P.-M. and J.M.-C.; software, F.D.P.-M. and K.M.-M.; validation, F.D.P.-M.; formal analysis, F.D.P.-M. and J.M.-C.; investigation, F.D.P.-M. and J.M.-C.; resources, K.M.-M.; data curation, F.D.P.-M. and J.M.-C.; writing—original draft preparation, F.D.P.-M. and J.M.-C.; writing—review and editing, F.D.P.-M., J.M.-C. and K.M.-M.; supervision, K.M.-M.; project administration, K.M.-M.; funding acquisition, K.M.-M. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Ritchie, H.; Rosado, P.; Roser, M. Agricultural Production. Our World in Data, January 2023. Available online: https://ourworldindata.org/agricultural-production (accessed on 17 March 2024).
2. 2.1 Food Security Indicators—Latest Updates and Progress towards Ending Hunger and Ensuring Food Security. Available online: https://www.fao.org/3/cc0639en/online/sofi-2022/food-security-nutrition-indicators.html (accessed on 17 March 2024).
3. Carvalho-Moore, P.; Rangani, G.; Heiser, J.; Findley, D.; Bowe, S.J.; Roma-Burgos, N. PPO2 Mutations in Amaranthus palmeri: Implications on Cross-Resistance. *Agriculture* **2021**, *11*, 760. [CrossRef]
4. Shaner, D.L. Lessons Learned From the History of Herbicide Resistance. *Weed Sci.* **2014**, *62*, 427–431. [CrossRef]
5. Heap, I. The International Survey of Herbicide Resistant Weeds. Available online: http://www.weedscience.com/Home.aspx (accessed on 1 February 2024).
6. Gupta, P.K. Chapter 37—Herbicides and Fungicides. In *Reproductive and Developmental Toxicology*, 2nd ed.; Gupta, R.C., Ed.; Academic Press: Cambridge, MA, USA, 2017; pp. 657–679. ISBN 978-0-12-804239-7.
7. Panchal, D.; Mistry, J.; Amin, D. Auxin. In *Practical Handbook on Agricultural Microbiology*; Amaresan, N., Patel, P., Amin, D., Eds.; Springer: New York, NY, USA, 2022; pp. 251–255. ISBN 978-1-07-161724-3.
8. Teale, W.D.; Paponov, I.A.; Palme, K. Auxin in action: Signalling, transport and the control of plant growth and development. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 847–859. [CrossRef]
9. Hayashi, K.; Neve, J.; Hirose, M.; Kuboki, A.; Shimada, Y.; Kepinski, S.; Nozaki, H. Rational Design of an Auxin Antagonist of the SCFTIR1 Auxin Receptor Complex. *ACS Chem. Biol.* **2012**, *7*, 590–598. [CrossRef] [PubMed]
10. Grossmann, K. Mediation of Herbicide Effects by Hormone Interactions. *J. Plant Growth Regul.* **2003**, *22*, 109–122. [CrossRef]
11. Yu, H.; Moss, B.L.; Jang, S.S.; Prigge, M.; Klavins, E.; Nemhauser, J.L.; Estelle, M. Mutations in the TIR1 Auxin Receptor That Increase Affinity for Auxin/Indole-3-Acetic Acid Proteins Result in Auxin Hypersensitivity. *Plant Physiol.* **2013**, *162*, 295–303. [CrossRef]
12. Dharmasiri, N. Auxin signaling and regulated protein degradation. *Trends Plant Sci.* **2004**, *9*, 302–308. [CrossRef]
13. Wang, R.; Estelle, M. Diversity and specificity: Auxin perception and signaling through the TIR1/AFB pathway. *Curr. Opin. Plant Biol.* **2014**, *21*, 51–58. [CrossRef] [PubMed]
14. Todd, O.E.; Figueiredo, M.R.A.; Morran, S.; Soni, N.; Preston, C.; Kubeš, M.F.; Napier, R.; Gaines, T.A. Synthetic auxin herbicides: Finding the lock and key to weed resistance. *Plant Sci.* **2020**, *300*, 110631. [CrossRef]
15. Ma, Q.; Grones, P.; Robert, S. Auxin signaling: A big question to be addressed by small molecules. *J. Exp. Bot.* **2018**, *69*, 313–328. [CrossRef]
16. Chéron, N.; Jasty, N.; Shakhnovich, E.I. OpenGrowth: An Automated and Rational Algorithm for Finding New Protein Ligands. *J. Med. Chem.* **2016**, *59*, 4171–4188. [CrossRef]
17. Imrie, F.; Bradley, A.R.; Deane, C.M. Generating property-matched decoy molecules using deep learning. *Bioinformatics* **2021**, *37*, 2134–2141. [CrossRef]
18. Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A molecular descriptor calculator. *J. Cheminform.* **2018**, *10*, 4. [CrossRef]
19. Kursa, M.B.; Rudnicki, W.R. Feature selection with the Boruta package. *J. Stat. Softw.* **2010**, *36*, 1–13. [CrossRef]
20. Sander, T.; Freyss, J.; von Korff, M.; Rufener, C. DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *J. Chem. Inf. Model.* **2015**, *55*, 460–473. [CrossRef]
21. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
22. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794.
23. Prieto-Martínez, F.D.; Fernández-de Gortari, E.; Medina-Franco, J.L.; Espinoza-Fonseca, L.M. An in silico pipeline for the discovery of multitarget ligands: A case study for epi-polypharmacology based on DNMT1/HDAC2 inhibition. *Artif. Intell. Life Sci.* **2021**, *1*, 100008. [CrossRef]
24. Landrum, G.; Tosco, P.; Kelley, B.; Ric; Cosgrove, D.; Sriniker; Vianello, R.; gedeck; NadineSchneider; Jones, G.; et al. *rdkit/rdkit: 2024_03_1 (Q1 2024) Release Beta*; Zenodo: Geneva, Switzerland, 2024. [CrossRef]
25. Feng, H.; Zhang, Y.; Bos, P.H.; Chambers, J.M.; Dupont, M.M.; Stockwell, B.R. K-RasG12D Has a Potential Allosteric Small Molecule Binding Site. *Biochemistry* **2019**, *58*, 2542–2554. [CrossRef]

26. Ghanakota, P.; Carlson, H.A. Driving Structure-Based Drug Discovery through Cosolvent Molecular Dynamics: Miniperspective. *J. Med. Chem.* **2016**, *59*, 10383–10399. [CrossRef]

27. Barrera-Téllez, F.J.; Prieto-Martínez, F.D.; Hernández-Campos, A.; Martínez-Mayorga, K.; Castillo-Bocanegra, R. In Silico Exploration of the Trypanothione Reductase (TryR) of L. mexicana. *Int. J. Mol. Sci.* **2023**, *24*, 16046. [CrossRef]

28. Ladbury, J.E. Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* **1996**, *3*, 973–980. [CrossRef] [PubMed]

29. Shadrick, W.R.; Slavish, P.J.; Chai, S.C.; Waddell, B.; Connelly, M.; Low, J.A.; Tallant, C.; Young, B.M.; Bharatham, N.; Knapp, S.; et al. Exploiting a water network to achieve enthalpy-driven, bromodomain-selective BET inhibitors. *Bioorganic Med. Chem.* **2018**, *26*, 25–36. [CrossRef]

30. Ge, Y.; Wych, D.C.; Samways, M.L.; Wall, M.E.; Essex, J.W.; Mobley, D.L. Enhancing Sampling of Water Rehydration on Ligand Binding: A Comparison of Techniques. *J. Chem. Theory Comput.* **2022**, *18*, 1359–1381. [CrossRef] [PubMed]

31. Lukauskis, D.; Samways, M.L.; Aureli, S.; Cossins, B.P.; Taylor, R.D.; Gervasio, F.L. Open Binding Pose Metadynamics: An Effective Approach for the Ranking of Protein–Ligand Binding Poses. *J. Chem. Inf. Model.* **2022**, *62*, 6209–6216. [CrossRef] [PubMed]

32. Bowers, K.; Chow, E.; Xu, H.; Dror, R.; Eastwood, M.; Gregersen, B.; Klepeis, J.; Kolossvary, I.; Moraes, M.; Sacerdoti, F.; et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In Proceedings of the ACM/IEEE SC 2006 Conference (SC'06), Tampa, FL, USA, 11–17 November 2006.

33. Predescu, C.; Bergdorf, M.; Shaw, D.E. Midtown splines: An optimal charge assignment for electrostatics calculations. *J. Chem. Phys.* **2020**, *153*, 224117. [CrossRef]

34. McGibbon, R.T.; Harrigan, M.; Beauchamp, K.; Klein, C.; Swails, J.; Hernández, C.; Peastman; Scherer, M.K.; Schwantes, C.; Pfrstg; et al. *Mdtraj/mdtraj: MDTraj 1.9 2017*; Zenodo: Geneva, Switzerland, 2019. [CrossRef]

35. Aldeghi, M.; Ross, G.A.; Bodkin, M.J.; Essex, J.W.; Knapp, S.; Biggin, P.C. Large-scale analysis of water stability in bromodomain binding pockets with grand canonical Monte Carlo. *Commun. Chem.* **2018**, *1*, 19. [CrossRef] [PubMed]

36. De Ruiter, A.; Oostenbrink, C. Free energy calculations of protein-ligand interactions. *Curr. Opin. Chem. Biol.* **2011**, *15*, 547–552. [CrossRef]

37. Samways, M.L.; Bruce Macdonald, H.E.; Essex, J.W. grand: A Python Module for Grand Canonical Water Sampling in OpenMM. *J. Chem. Inf. Model.* **2020**, *60*, 4436–4441. [CrossRef] [PubMed]

38. Nittinger, E.; Flachsenberg, F.; Bietz, S.; Lange, G.; Klein, R.; Rarey, M. Placement of Water Molecules in Protein Structures: From Large-Scale Evaluations to Single-Case Examples. *J. Chem. Inf. Model.* **2018**, *58*, 1625–1637. [CrossRef]

39. El Khoury, L.; Santos-Martins, D.; Sasmal, S.; Eberhardt, J.; Bianco, G.; Ambrosio, F.A.; Solis-Vasquez, L.; Koch, A.; Forli, S.; Mobley, D.L. Comparison of affinity ranking using AutoDock-GPU and MM-GBSA scores for BACE-1 inhibitors in the D3R Grand Challenge 4. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 1011–1020. [CrossRef]

40. Lee, S.; Sundaram, S.; Armitage, L.; Evans, J.P.; Hawkes, T.; Kepinski, S.; Ferro, N.; Napier, R.M. Defining Binding Efficiency and Specificity of Auxins for SCFTIR1/AFB-Aux/IAA Co-receptor Complex Formation. *ACS Chem. Biol.* **2014**, *9*, 673–682. [CrossRef] [PubMed]

41. Cho, M.; Kim, K. Diclofenac modified the root system architecture of Arabidopsis via interfering with the hormonal activities of auxin. *J. Hazard. Mater.* **2021**, *413*, 125402. [CrossRef]

42. Korb, O.; Bowden, S.; Olsson, T.; Frenkel, D.; Liebeschuetz, J.; Cole, J. Ensemble docking revisited. *J. Cheminform.* **2010**, *2*, P25. [CrossRef]

43. Korb, O.; Stützle, T.; Exner, T.E. Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84–96. [CrossRef] [PubMed]

44. Ren, X.; Shi, Y.-S.; Zhang, Y.; Liu, B.; Zhang, L.-H.; Peng, Y.-B.; Zeng, R. Novel Consensus Docking Strategy to Improve Ligand Pose Prediction. *J. Chem. Inf. Model.* **2018**, *58*, 1662–1668. [CrossRef]

45. Kurkinen, S.T.; Lätti, S.; Pentikäinen, O.T.; Postila, P.A. Getting Docking into Shape Using Negative Image-Based Rescoring. *J. Chem. Inf. Model.* **2019**, *59*, 3584–3599. [CrossRef]

46. ten Brink, T.; Exner, T.E. Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein−Ligand Docking Results. *J. Chem. Inf. Model.* **2009**, *49*, 1535–1546. [CrossRef]

47. Swenson, D.W.H.; Roet, S.; Faruk, N. *Dwhswenson/contact_map: Contact_map 0.7.0*; Zenodo: Geneva, Switzerland, 2020. [CrossRef]

48. Bussi, G.; Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* **2020**, *2*, 200–212. [CrossRef]

49. Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F.L. Exploring Complex Protein−Ligand Recognition Mechanisms with Coarse Metadynamics. *J. Phys. Chem. B* **2009**, *113*, 4807–4816. [CrossRef]

50. Provasi, D. Ligand-Binding Calculations with Metadynamics. In *Biomolecular Simulations: Methods and Protocols*; Bonomi, M., Camilloni, C., Eds.; Springer: New York, NY, USA, 2019; pp. 233–253. ISBN 978-1-4939-9608-7.

51. Raniolo, S.; Limongelli, V. Ligand binding free-energy calculations with funnel metadynamics. *Nat. Protoc.* **2020**, *15*, 2837–2866. [CrossRef]

52. Wang, J.; Ishchenko, A.; Zhang, W.; Razavi, A.; Langley, D. A highly accurate metadynamics-based Dissociation Free Energy method to calculate protein–protein and protein–ligand binding potencies. *Sci. Rep.* **2022**, *12*, 2024. [CrossRef] [PubMed]

53. Uzunova, V.V.; Quareshy, M.; del Genio, C.I.; Napier, R.M. Tomographic docking suggests the mechanism of auxin receptor TIR1 selectivity. *Open Biol.* **2016**, *6*, 160139. [CrossRef] [PubMed]

54. Leeson, P.D. Molecular inflation, attrition and the rule of five. *Adv. Drug Deliv. Rev.* **2015**, *101*, 22–33. [CrossRef] [PubMed]

55. Hopkins, A.L.; Keserü, G.M.; Leeson, P.D.; Rees, D.C.; Reynolds, C.H. The role of ligand efficiency metrics in drug discovery. *Nat. Rev. Drug Discov.* **2014**, *13*, 105–121. [CrossRef]

56. Maggiora, G.M. On Outliers and Activity CliffsWhy QSAR Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535. [CrossRef]

57. Stumpfe, D.; Hu, H.; Bajorath, J. Evolving Concept of Activity Cliffs. *ACS Omega* **2019**, *4*, 14360–14368. [CrossRef]

58. Stumpfe, D.; Hu, H.; Bajorath, J. Advances in exploring activity cliffs. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 929–942. [CrossRef]

59. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26. [CrossRef]

60. Shultz, M.D. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs. *J. Med. Chem.* **2019**, *62*, 1701–1714. [CrossRef]

61. Giblin, K.A.; Hughes, S.J.; Boyd, H.; Hansson, P.; Bender, A. Prospectively Validated Proteochemometric Models for the Prediction of Small-Molecule Binding to Bromodomain Proteins. *J. Chem. Inf. Model.* **2018**, *58*, 1870–1888. [CrossRef]

62. Skyner, R.E.; McDonagh, J.L.; Groom, C.R.; van Mourik, T.; Mitchell, J.B.O. A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6174–6191. [CrossRef]

63. Deshpande, M.; Kuramochi, M.; Karypis, G. Data Mining Algorithms for Virtual Screening of Bioactive Compounds. In *Data Mining in Biomedicine*; Pardalos, P.M., Boginski, V.L., Vazacopoulos, A., Eds.; Springer: Boston, MA, USA, 2007; pp. 59–90. ISBN 978-0-387-69319-4.

64. Lima, A.N.; Philot, E.A.; Trossini, G.H.G.; Scott, L.P.B.; Maltarollo, V.G.; Honorio, K.M. Use of machine learning approaches for novel drug discovery. *Expert Opin. Drug Discov.* **2016**, *11*, 225–239. [CrossRef]

65. Patlewicz, G.; Wambaugh, J.F.; Felter, S.P.; Simon, T.W.; Becker, R.A. Utilizing Threshold of Toxicological Concern (TTC) with high throughput exposure predictions (HTE) as a risk-based prioritization approach for thousands of chemicals. *Comput. Toxicol.* **2018**, *7*, 58–67. [CrossRef] [PubMed]

66. Vogel, S.M.; Bauer, M.R.; Boeckler, F.M. DEKOIS: Demanding Evaluation Kits for Objective in Silico Screening—A Versatile Tool for Benchmarking Docking Programs and Scoring Functions. *J. Chem. Inf. Model.* **2011**, *51*, 2650–2665. [CrossRef]

67. Stein, R.M.; Yang, Y.; Balius, T.E.; O'Meara, M.J.; Lyu, J.; Young, J.; Tang, K.; Shoichet, B.K.; Irwin, J.J. Property-Unmatched Decoys in Docking Benchmarks. *J. Chem. Inf. Model.* **2021**, *61*, 699–714. [CrossRef] [PubMed]

68. Li, G.-B.; Yang, L.-L.; Wang, W.-J.; Li, L.-L.; Yang, S.-Y. ID-Score: A New Empirical Scoring Function Based on a Comprehensive Set of Descriptors Related to Protein–Ligand Interactions. *J. Chem. Inf. Model.* **2013**, *53*, 592–600. [CrossRef] [PubMed]

69. Lu, J.; Hou, X.; Wang, C.; Zhang, Y. Incorporating Explicit Water Molecules and Ligand Conformation Stability in Machine-Learning Scoring Functions. *J. Chem. Inf. Model.* **2019**, *59*, 4540–4549. [CrossRef] [PubMed]

70. Mobley, D.L.; Dill, K.A. Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get". *Structure* **2009**, *17*, 489–498. [CrossRef]

71. Liu, K.; Kokubo, H. Exploring the Stability of Ligand Binding Modes to Proteins by Molecular Dynamics Simulations: A Cross-docking Study. *J. Chem. Inf. Model.* **2017**, *57*, 2514–2522. [CrossRef]

72. Brenke, R.; Kozakov, D.; Chuang, G.-Y.; Beglov, D.; Hall, D.; Landon, M.R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621–627. [CrossRef]

73. Ghanakota, P.; Van Vlijmen, H.; Sherman, W.; Beuming, T. Large-Scale Validation of Mixed-Solvent Simulations to Assess Hotspots at Protein–Protein Interaction Interfaces. *J. Chem. Inf. Model.* **2018**, *58*, 784–793. [CrossRef] [PubMed]

74. Hao, G.-F.; Yang, G.-F. The Role of Phe82 and Phe351 in Auxin-Induced Substrate Perception by TIR1 Ubiquitin Ligase: A Novel Insight from Molecular Dynamics Simulations. *PLoS ONE* **2010**, *5*, e10742. [CrossRef]

75. Tan, X.; Calderon-Villalobos, L.I.A.; Sharon, M.; Zheng, C.; Robinson, C.V.; Estelle, M.; Zheng, N. Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **2007**, *446*, 640–645. [CrossRef] [PubMed]

76. Mockaitis, K.; Estelle, M. Auxin receptors and plant development: A new signaling paradigm. *Annu. Rev. Cell Dev. Biol.* **2008**, *24*, 55–80. [CrossRef]

77. Juárez-Mercado, K.E.; Prieto-Martínez, F.D.; Sánchez-Cruz, N.; Peña-Castillo, A.; Prada-Gracia, D.; Medina-Franco, J.L. Expanding the Structural Diversity of DNA Methyltransferase Inhibitors. *Pharmaceuticals* **2021**, *14*, 17. [CrossRef]

78. Swinburne, T.D.; Wales, D.J. Defining, Calculating, and Converging Observables of a Kinetic Transition Network. *J. Chem. Theory Comput.* **2020**, *16*, 2661–2679. [CrossRef]

79. Ross, G.A.; Morris, G.M.; Biggin, P.C. One Size Does Not Fit All: The Limits of Structure-Based Models in Drug Discovery. *J. Chem. Theory Comput.* **2013**, *9*, 4266–4274. [CrossRef] [PubMed]

80. Anighoro, A.; Bajorath, J. Three-Dimensional Similarity in Molecular Docking: Prioritizing Ligand Poses on the Basis of Experimental Binding Modes. *J. Chem. Inf. Model.* **2016**, *56*, 580–587. [CrossRef]

81. Quareshy, M.; Prusinska, J.; Li, J.; Napier, R. A cheminformatics review of auxins as herbicides. *J. Exp. Bot.* **2018**, *69*, 265–275. [CrossRef]

82. Epp, J.B.; Alexander, A.L.; Balko, T.W.; Buysse, A.M.; Brewster, W.K.; Bryan, K.; Daeuble, J.F.; Fields, S.C.; Gast, R.E.; Green, R.A.; et al. The discovery of Arylex^{TM} active and Rinskor^{TM} active: Two novel auxin herbicides. *Bioorganic Med. Chem.* **2016**, *24*, 362–371. [CrossRef] [PubMed]

83. Prusinska, J.; Uzunova, V.; Schmitzer, P.; Weimer, M.; Bell, J.; Napier, R.M. The differential binding and biological efficacy of auxin herbicides. *Pest Manag. Sci.* **2023**, *79*, 1305–1315. [CrossRef] [PubMed]