

## Article

# Discovery of Kinase and Carbonic Anhydrase Dual Inhibitors by Machine Learning Classification and Experiments

Min-Jeong Kim <sup>†</sup>, Sarita Pandit <sup>†</sup>  and Jun-Goo Jee <sup>\*</sup>

Research Institute of Pharmaceutical Sciences, College of Pharmacy, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 41566, Korea; dns01185@naver.com (M.-J.K.); saritapandit94@gmail.com (S.P.)

<sup>\*</sup> Correspondence: jjee@knu.ac.kr; Tel.: +82-53-950-8568

<sup>†</sup> These authors contributed equally to this work.

**Abstract:** A multi-target small molecule modulator is advantageous for treating complicated diseases such as cancers. However, the strategy and application for discovering a multi-target modulator have been less reported. This study presents the dual inhibitors for kinase and carbonic anhydrase (CA) predicted by machine learning (ML) classifiers, and validated by biochemical and biophysical experiments. ML trained by CA I and CA II inhibitor molecular fingerprints predicted candidates from the protein-specific bioactive molecules approved or under clinical trials. For experimental tests, three sulfonamide-containing kinase inhibitors, **5932**, **5946**, and **6046**, were chosen. The enzyme assays with CA I, CA II, CA IX, and CA XII have allowed the quantitative comparison in the molecules' inhibitory activities. While **6046** inhibited weakly, **5932** and **5946** exhibited potent inhibitions with 100 nM to 1  $\mu$ M inhibitory constants. The ML screening was extended for finding CAs inhibitors of all known kinase inhibitors. It found XMU-MP-1 as another potent CA inhibitor with an approximate 30 nM inhibitory constant for CA I, CA II, and CA IX. Differential scanning fluorimetry confirmed the direct interaction between CAs and small molecules. Cheminformatics studies, including docking simulation, suggest that each molecule possesses two separate functional moieties: one for interaction with kinases and the other with CAs.

**Keywords:** carbonic anhydrase; cheminformatics; docking simulation; dual inhibitor; fingerprint; kinase; machine learning; polypharmacology



**Citation:** Kim, M.-J.; Pandit, S.; Jee, J.-G. Discovery of Kinase and Carbonic Anhydrase Dual Inhibitors by Machine Learning Classification and Experiments. *Pharmaceuticals* **2022**, *15*, 236. <https://doi.org/10.3390/ph15020236>

Academic Editor: Paweł Kafarski

Received: 17 December 2021

Accepted: 12 February 2022

Published: 16 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

A molecule targeting a single disease-related protein often causes insufficient efficacy, even potent from the viewpoint of target engagement. Therapeutic strategies for treating cancer, neurodegenerative, cardiovascular, and infectious diseases often necessitate multiple target simultaneous modulation due to their complicated origins [1–3]. Targeting numerous targets may increase side effects. There is an increasing interest in developing therapeutics that selectively manipulate multiple desired pathways at once [4].

Two approaches are known for simultaneous control. The first method is a mixture of monotherapies, including a combination of drugs with an active ingredient and a single formula comprising multiple active ingredients. The second approach is a single molecule that can modulate several targets. Multi-target-directing molecules are advantageous because of the fewer side effects and toxicities [5,6]. Instead, finding and designing a molecule with multiple functions is not straightforward, necessitating an effective strategy [4,7].

One of the diseases where the multiprotein-targeting modulators are necessary is cancer [8]. Cancer cells have changed signal pathways. The altered pathways are the causes and effects of the disease. Drugs targeting several cancer pathways have been developed. Notably, more than 40 small molecule kinase inhibitors have been approved [9]. Meanwhile, cancer cells require more control of pH. Carbonic anhydrases (CAs) catalyze the production of carbonate ion from carbon dioxide, and mono- and di-carbonate are part

of the central buffer systems in our bodies. The increased modulation of pH requires the hyper-functioning CAs. Indeed, the relation between the decreased survival rates and the elevated CA IX and CA XII in triple-negative breast cancer has been reported [10]. Reliance on the role of CAs by cancer cells has made CA a target for cancer therapy. Of accumulated CA inhibitors, dual inhibitors with kinase have been reported [11–13]. Nevertheless, the dual inhibitors are limited, and there is no established strategy to find new dual inhibitors.

Accumulating chemical and biological data enables the prediction of small bioactive cellular target molecules. Computational methods have helped to decode off-targets and polypharmacological effects with the data. The similarity ensemble approach (SEA) is a representative example [14–16]. Successful SEA applications have been reported for predicting off-targets of FDA-approved drugs and excipients [17–19]. Besides the SEA, several computational approaches have been reported [20–24]. Of them, some employ machine learning (ML) [24].

ML is an approach to finding the hidden rule among the accumulated data. Classification, regression, and clustering are three areas applied by ML. If the data are significant, one can apply deep learning using neural networks. Two crucial factors for ML are the features to train and the algorithm. How to select and optimize the pair determines the performance of ML. Popular features to employ ML for small molecule research are molecular fingerprints. Several applications have been reported for discovering new inhibitors against a specific target, including the inhibitor screenings for JAK2 and FGFR4 kinases [25,26]. However, the applications to discover multiple modulators or find off-targets are less known. The coupled experimental validation with the prediction is especially limited.

This study reports the dual inhibitors of kinases and CAs using ML and experiments. The small molecule inhibitors of CA I and CA II were gathered and treated as active. Their physicochemically matched but topologically different decoys were generated for use as inactive molecules. The conversion of each molecule into a molecular fingerprint and the trainings using the ML classifiers followed. Use of an external dataset allowed selection of the best pair of features and classifiers. The selected ML classifier screened the database of experimentally validated protein-specific modulators. The successive biochemical and biophysical experiments confirmed the inhibitory activities of the predicted kinase inhibitors against CAs. The ML classifier was extended for screening all known kinase inhibitors. The cheminformatics, including docking simulation, also provide the chemical features of the dual inhibitors.

## 2. Results

### 2.1. The Pair of ECFP4 Fingerprint and Logistic Regression Classifier Was Selected for Screening

We retrieved the inhibitors for human CA I (UniProt ID: P00915) and CA II (P00918) from BindingDB [27]. Here, molecules smaller than 500 Da and more potent than 500 nM in terms of inhibitory constant ( $K_i$ ) or half-maximal inhibitory concentration ( $IC_{50}$ ) were considered. Once clustering the molecules with the cutoff of 0.6 Tanimoto coefficient (TC), the most potent ones in each cluster were representatively chosen as active. This is to minimize the bias in the chemical spaces of the ligands. Then, the DUD-E server [28] generated 40 decoys per inhibitor, which were treated as inactive. No decoy shares similarity higher than 0.6 to inhibitors and 0.8 to each other in TC. The respective numbers of actives (inactives) for CA I and CA II were 167 (6627) and 302 (11,968). We then digitized the SMILES of each chemical with three fingerprints [29,30], extended-connectivity fingerprint 4 (ECFP4), extended-connectivity fingerprint 6 (ECFP6), and molecular access system keys (MACCS), and used them as the features for ML classification. For ML classification, we used seven methods, which are K-Nearest Neighbor (KNN), Naïve Bayesian (NB), Logistic Regression (Logit), Decision Tree (DT), Random Forest (RF), Multi-Layer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost).

Table 1 indicates the result metrics by ML classification with three fingerprints and seven models in two proteins of CA I and CA II (21 cases per protein). The metrics for judging the results by each ML classifier are accuracy (ACC), the area under the curve of

receiver operating characteristics curve (AUC), and the Matthews correlation coefficient (MCC). The values of ACC and AUC were comparable with the values approximating 1, which makes them less helpful in discriminating the models. Meanwhile, MCC values differed depending on the combination. MCC is a suited metric reflecting true and false positives and negatives of varying sizes [31]. The abundance of inactive data in screening can bring about substantial false positives when the classification is inaccurate. In this situation, MCC can be the metric for reflecting the reliability of the ML classifier. Overall, the ECFP4 fingerprint exhibited better results than those with ECFP6 and MACCS. Indeed, many publications have used ECFP4 as a fingerprint [14,17–19,24,25,32,33]. The models of Logit and MLP revealed comparable and higher MCC values than other methods. Therefore, we focused on the two methods.

**Table 1.** The metrics of machine learning classifiers with fingerprints.

CA I	ECFP4			ECFP6			MACCS		
	ACC	AUC	MCC	ACC	AUC	MCC	ACC	AUC	MCC
KNN	0.970	0.950	0.632	0.952	0.932	0.532	0.984	0.916	0.720
NB	0.979	0.697	0.483	0.964	0.660	0.291	0.812	0.868	0.284
Logit	0.994	0.994	0.863	0.993	0.994	0.849	0.989	0.988	0.771
DT	0.988	0.889	0.757	0.985	0.858	0.703	0.985	0.864	0.703
RF	0.992	0.988	0.826	0.990	0.985	0.782	0.990	0.988	0.771
MLP	0.995	0.993	0.883	0.993	0.990	0.843	0.990	0.991	0.794
XGBoost	0.992	0.977	0.832	0.990	0.985	0.797	0.991	0.994	0.815
CA II	ECFP4			ECFP6			MACCS		
	ACC	AUC	MCC	ACC	AUC	MCC	ACC	AUC	MCC
KNN	0.981	0.973	0.734	0.972	0.965	0.659	0.984	0.929	0.739
NB	0.932	0.779	0.331	0.872	0.779	0.268	0.840	0.900	0.321
Logit	0.994	0.998	0.878	0.994	0.998	0.869	0.990	0.993	0.793
DT	0.989	0.899	0.777	0.988	0.868	0.747	0.986	0.862	0.716
RF	0.992	0.997	0.831	0.991	0.996	0.801	0.992	0.995	0.815
MLP	0.994	0.996	0.872	0.993	0.995	0.848	0.991	0.996	0.812
XGBoost	0.992	0.994	0.836	0.992	0.995	0.839	0.990	0.995	0.797

The input dataset comprises 167 actives and 6627 inactive molecules for CA I, and 302 and 11,968 for CA II. The bits for ECFP4, ECFP6, and MACCS fingerprints are 1024, 1024, and 166, respectively. In ML classifiers, KNN stands for K-Nearest Neighbor, NB for Naïve Bayesian, Logit for Logistic Regression, DT for Decision Tree, RF for Random Forest, MLP for Multi-Layer Perceptron, and XGBoost for Extreme Gradient Boosting. The metrics are accuracy (ACC), the area under the curve (AUC) of the receiver operating characteristic curve, and Matthews correlation coefficient (MCC). The values through five-fold cross-validation are averaged in each.

Using an external dataset can help choose the better ML model [34]. The discriminating powers in the pairs of ECFP4 and Logit, and ECFP4 and MLP were compared using an external dataset of FDA-approved drugs. The 1318 small molecules registered as “Approved drugs” in the International Union of Basic and Clinical Pharmacology database (IUPHAR) and having a molecular weight smaller than 700 Da were extracted for the validation [35]. The family of diuretic drugs includes CA inhibitors. All the analogs showing the similarity to FDA-approved drugs with  $TC \geq 0.6$  were excluded from the ligands in the training dataset for a fair comparison. Their corresponding decoys were also omitted from inactive molecules. Logit and MLP classifiers trained with the remaining CA I dataset retrieved seven molecules as common candidates. An extensive literature search confirmed that these seven molecules are direct binders to at least one of the CAs (Table 2 and Figure S1). The CA inhibitors are valdecoxib (IUPHAR ID: 2894), sulpiride (5501 and 958), acetazolamide (6792), diclofenamide (6807), methazolamide (6828), and zonisamide (7047). Logit and MLP classifiers trained using CA II dataset predicted six common true positives, which are celecoxib (2892), valdecoxib (2894), acetazolamide (6792), diclofenamide (6807), ethoxzolamide (6814), and methazolamide (6828). These have also been known to directly interact with CAs. When compared to those from Logit, the results from MLP classifiers were worse, predicting more non-binders as candidates both in CA I

and CA II cases (Table 2 and Figure S1). Altogether, it was judged that Logit classifier is more suited for our study. Notably, Logit classifier is simpler than MLP in terms of the parameters' numbers. A simpler model can decrease the concerns on overfitting if the discriminative powers are comparable.

**Table 2.** Validation of machine learning classifiers using the approved drug database.

Protein	Classifier	TP1	TP2	FP	Sum
CA I	Logit	7		2	9
	MLP	7	1	4	12
CA II	Logit	6		5	11
	MLP	6		8	14

True positives 1 (TP1) indicates the known CA targeting drugs predicted both by Logit and MLP classifiers with ECFP4 fingerprint. True positives 2 (TP2) means additional CA-targeting drugs by MLP. The molecules of false positives (FP) are those unrelated to CA inhibitors by the literature search. These data are prepared by ML classifiers trained with other datasets that exclude the actives similar ( $TC \geq 0.6$ ) to the approved drugs and their corresponding inactive molecules.

## 2.2. Screening Known Protein-Specific Modulators by ML Classified the Candidate Inhibitors for CA

IUPHAR-registered “Synthetic organics” 6975 molecules smaller than 700 Da were screened using the two paired ECFP4 and Logit classifiers trained with CA I and CA II datasets. IUPHAR database contains only experimentally validated protein-specific modulators, mostly FDA-approved or under clinical trials [35]. CA I and CA II cases led to 42 and 49 candidates with  $p \geq 0.5$ , respectively (Table 3 and Figure S2). Of them, 23 molecules are common in CA I and CA II predictions. Their on-targets are various, but the primary targets are CAs in 13 molecules (Table 3). It covers 72% (=13/18) of the total 18 CA inhibitors in the database (Figure S3). Interestingly, the candidates include **4836** ( $p = 0.897$ , hydrochlorothiazide) and **7197** (0.509, hydroflumethiazide). They are unassigned as CA inhibitors in the IUPHAR database. However, the literature search found that they inhibit CAs [36], demonstrating the performance of our ML-based prediction.

Our proof-of-concept study focused on the kinase inhibitors of unverified CA inhibitor candidates. The literature search identified four potent kinase inhibitors, **5932** (JNJ-7706621) [37], **5946** (CDK1/2 inhibitor III) [37], **8146** [38], and **6046** [39]. The molecules' primary targets are cyclin-dependent kinases (CDKs) and mammalian sterile 20-like kinase 3 (MST3) for **5932** ( $IC_{50}$  for CDK2: 2 nM) and **5946** ( $IC_{50}$  for CDK2: 0.5 nM), TXK tyrosine kinase for **8146** ( $IC_{50}$  for TXK: 0.4 nM), and spleen associated tyrosine kinase (SYK) for **6046** ( $IC_{50}$  for SYK: 15.8 nM). All target proteins are related to cancers. We chose **5932**, **5946**, and **6046** for experimental tests considering commercial availability. A literature search revealed that pazopanib (IUPHAR ID: **5698**) is inhibitory in both CAs and kinases such as CSF1R [13], although ML could not list it as a candidate. Acetazolamide (**6792**) and pazopanib were used as positive controls. There are a number of notable points. First, active molecules for training did not include **5932**, **5946**, and **6046**. The closest compounds of **5932**, **5946**, and **6046** in the training data were ZINC28389402, ZINC28389402, and ZINC13804313 with the TC values of 0.460, and 0.407, and 0.270, respectively (Table 3). Second, all the molecules have a sulfonamide moiety needed to chelate with the zinc atom in the catalytic core. Third, **5932** was not listed by the CA II model, despite the shared similarity to **5946** with a TC of 0.685.



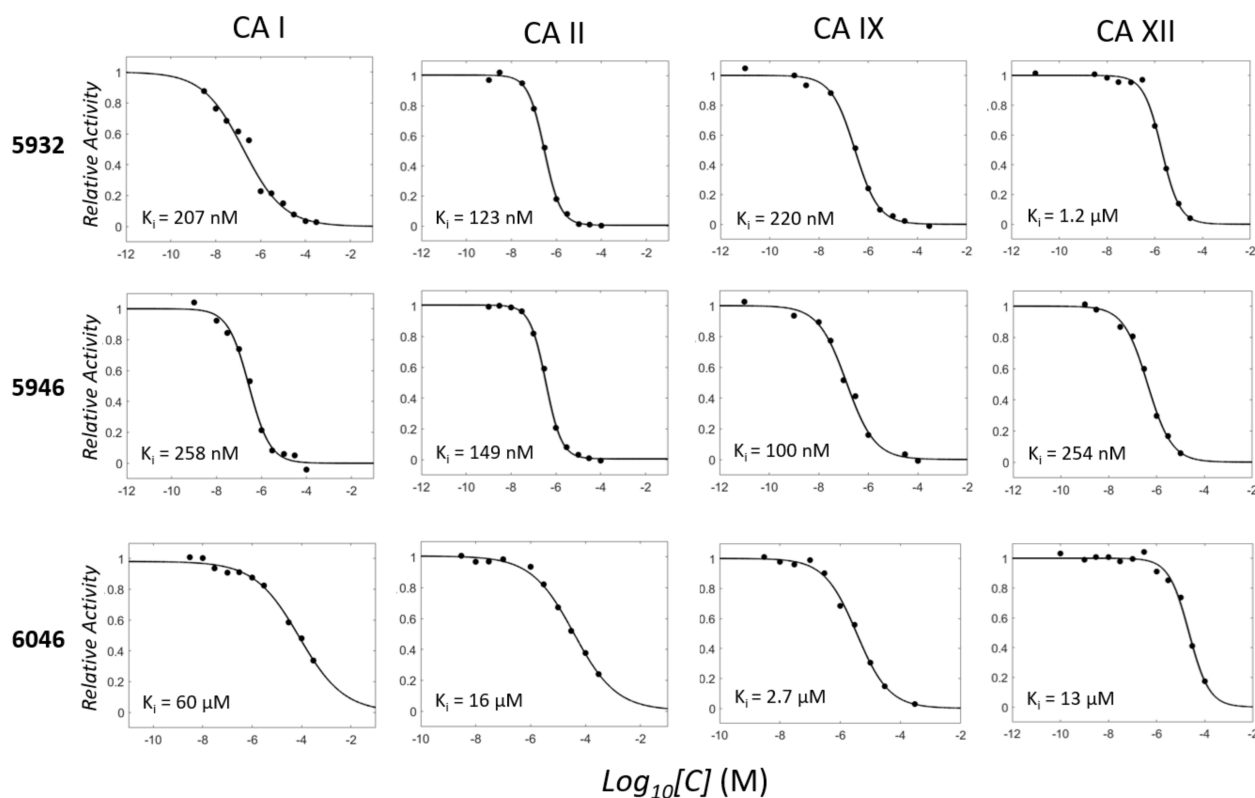
**Table 3.** Candidates from the screen of the known bioactive molecules.

CA I (19)				CA I and CA II (23)				CA II (26)			
ID	Prob	Max TC	ZINC ID	ID	Prob	Max TC	ZINC ID	ID	Prob	Max TC	ZINC ID
3463	0.982	0.394	13800463	<b>6920</b>	0.999	0.737	28526472	9753	0.935	0.475	13780724
10070	0.935	0.274	28526472	<b>6814</b>	0.995	1.000	56721	3962	0.931	0.290	13612907
4840	0.900	0.288	13650488	<b>6828</b>	0.994	1.000	8415468	3986	0.928	0.464	64527862
<b>7047</b>	0.838	0.283	58569258	<b>7046</b>	0.992	0.422	184018	10483	0.910	0.264	34717999
1376	0.834	0.342	1099	4839	0.990	0.404	13800448	10664	0.903	0.353	87722919
10175	0.801	0.426	131170	<b>6792</b>	0.979	1.000	3813042	4836	0.897	0.700	34799864
8821	0.762	0.227	13650488	10017	0.978	0.270	28388514	4742	0.850	0.328	27638369
980	0.751	0.328	22198192	<b>5946</b>	0.959	0.407	<b>28389402</b>	10242	0.785	0.474	95586265
9092	0.740	0.192	2101	<b>10149</b>	0.948	0.639	84759371	11205	0.776	0.241	64526424
932	0.734	0.283	595377	<b>8146</b>	0.912	0.413	<b>13800465</b>	7289	0.757	0.340	13800448
4702	0.724	0.464	1099	2894	0.909	0.435	58569258	8316	0.714	0.375	40917210
960	0.685	0.324	95591272	<b>6807</b>	0.905	0.405	13800446	4055	0.680	0.341	26387397
958	0.685	0.324	95591272	<b>6810</b>	0.884	1.000	1530622	11201	0.641	0.227	64526424
5501	0.685	0.324	95591272	<b>6849</b>	0.838	0.245	34717916	4357	0.623	0.367	27636999
10321	0.589	0.391	84670597	9513	0.830	0.361	84670374	6574	0.599	0.195	13472881
8378	0.548	0.244	1099	7028	0.817	0.380	84652324	4835	0.585	0.360	34799864
<b>5932</b>	0.546	0.460	<b>28389402</b>	4837	0.800	0.317	84670374	9129	0.583	0.243	64526424
7409	0.535	0.408	16525334	<b>6797</b>	0.722	0.542	1530622	11174	0.580	0.404	27635960
1394	0.509	0.239	5159179	<b>10433</b>	0.694	0.541	13829485	10693	0.578	0.406	27741075
				4635	0.661	0.212	13800446	7893	0.572	0.471	16525334
				4084	0.661	0.212	13800446	<b>2892</b>	0.571	0.509	95586265
				11220	0.601	0.250	26387397	6514	0.570	0.357	84669523
				7125	0.584	0.318	28349861	<b>6046</b>	0.565	0.270	<b>13804313</b>
								7870	0.562	0.425	131170
								7197	0.509	0.690	34799864
								6648	0.507	0.273	27644927

Logistic Regression classification models for CA I and CA II screened IUPHAR “Synthetic organics” comprising 6975 molecules. The candidates are arranged into three criteria: CA I only (the number of candidates: 19), both CA I and CA II (23), and CA II only (26). The ID in IUPHAR, the probability by Logistic Regression classifier, and TC (Max TC) and the name (ZINC ID) of the closest molecule in the training dataset are presented. Those colored in red are known CA inhibitors. Known kinase inhibitors are colored in gray.

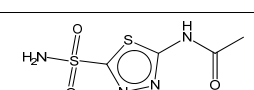
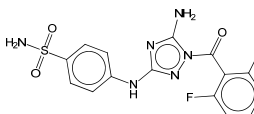
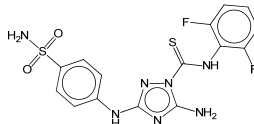
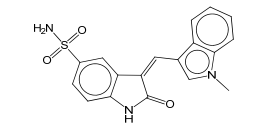
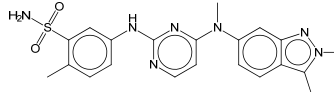
### 2.3. Two Candidates, 5932 and 5946, Were Highly Inhibitory for CAs

Two candidates, **5932** and **5946**, exhibited potent inhibition against CAs, but the inhibition by **6046** was moderate (Figure 1 and Figure S3, and Table 4). The  $K_i$  values for **5932**, **5946**, and **6046** were 123 nM, 149 nM, and 16  $\mu$ M in CA II, whereas they were 207 nM, 258 nM, and 60  $\mu$ M in CA I, respectively. We also quantified the inhibition toward two anticancer therapeutic targets, CA IX and CA XII. The  $K_i$  values for CA IX were 220 nM, 100 nM, and 2.7  $\mu$ M for **5932**, **5946**, and **6046**, while CA XII revealed 1.2  $\mu$ M, 254 nM, and 13  $\mu$ M values. Except for CA XII, **5932** and **5946** showed similar inhibitory constants. It is noteworthy that all the reaction mixtures contain 0.01% Triton X-100 to avoid the aggregation-based enzyme inhibition by small molecules [40–43]. The direct bindings of **5932**, **5946**, and **6046** to CA I and CA II were checked using differential scanning fluorimetry (DSF) (Figure 2). The measured  $\Delta T_m$  from the apo CA I were in the descending order of **6792** (4.5 degrees), **5698** (3.8), **5932** (2.0), **5946** (1.3), and **6046** (0.1). The degrees of  $\Delta T_m$  were more distinct in CA II. The order was **6792** (9.9), **5632** (4.4), **5946** (3.2), **5698** (2.3), and **6046** (0.2). All shifted beyond  $3 \times$  standard deviation from the reference except for **6046**, reflecting the significance of the change and thus direct interactions. The data qualitatively agree with enzyme assay, supporting the consistency in the orthogonal experimental confirmation.

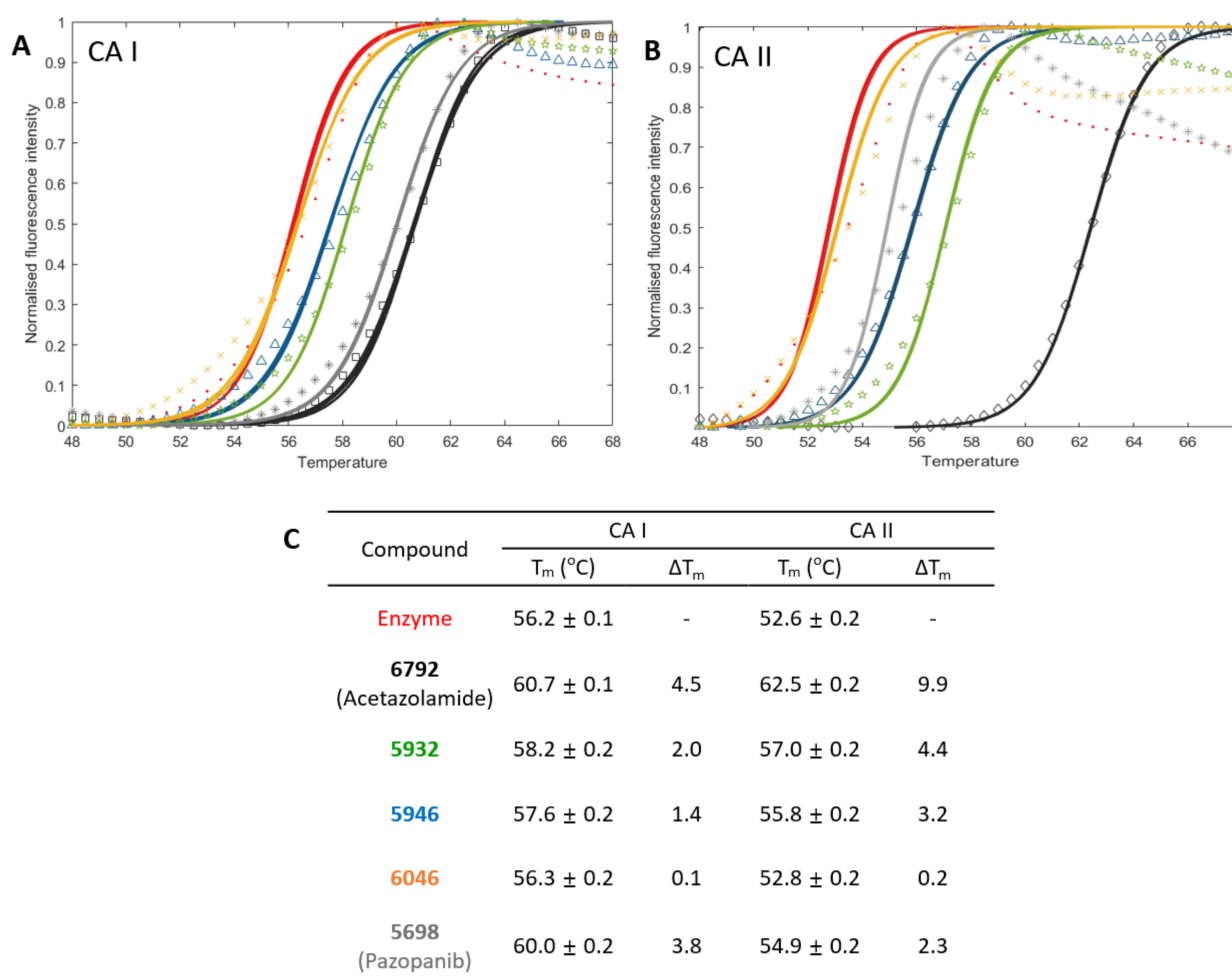


**Figure 1.** Concentration-dependent inhibitory profiles. Enzyme activities are scaled to have a relative value in the range of 0–1 on the y-axis. The enzyme activity in the absence of an inhibitor is expressed as one. The molar concentrations in each small molecule ( $[C]$ ) are expressed as  $\text{Log}_{10}[C]$  on the x-axis. The  $K_i$  values that are converted from the fitted  $\text{IC}_{50}$  values are written with units.

**Table 4.** Quantified compound activities.

Compound	2D Structure	CA I	CA II	CA IX	CA XII
6792 (Acetazolamide)		$197 \pm 18$	$3 \pm 1$	$22 \pm 5$	$23 \pm 10$
5932		$207 \pm 66$	$123 \pm 16$	$220 \pm 32$	$1.2 \pm 0.2$ ( $\mu\text{M}$ )
5946		$258 \pm 58$	$149 \pm 21$	$100 \pm 28$	$254 \pm 40$
6046		$60 \pm 9$ ( $\mu\text{M}$ )	$16 \pm 5$ ( $\mu\text{M}$ )	$2.7 \pm 0.3$ ( $\mu\text{M}$ )	$13 \pm 4$ ( $\mu\text{M}$ )
5698 (Pazopanib)		$67 \pm 10$	$242 \pm 32$	$1.7 \pm 0.5$ ( $\mu\text{M}$ )	$0.8 \pm 0.3$ ( $\mu\text{M}$ )

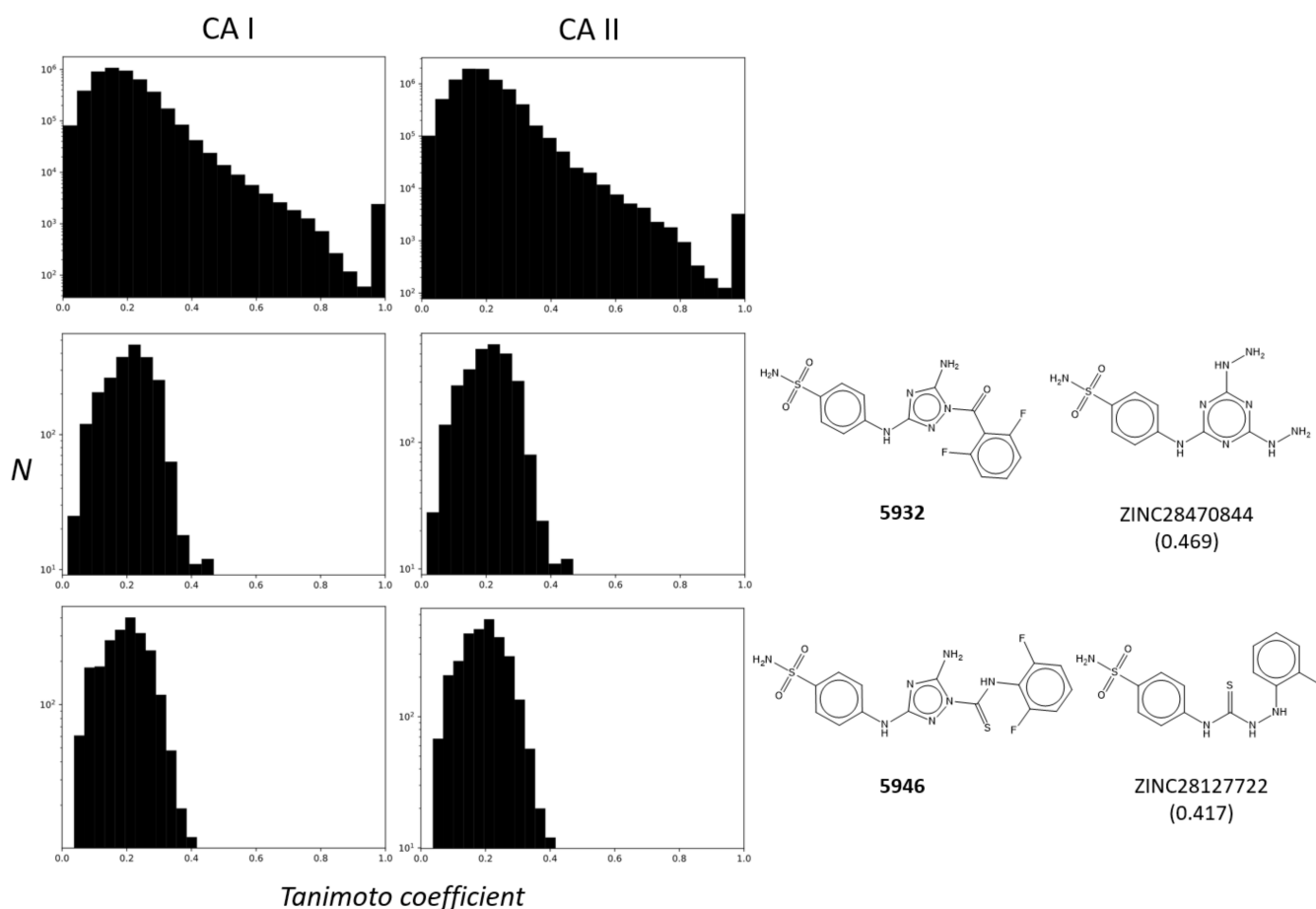
Enzyme inhibitory activities against CAs by compounds are measured and quantified under a series of small molecule concentrations. The values are inhibitory constant,  $K_i$ . 6792 (acetazolamide) and 5698 (pazopanib) are known CA inhibitors and positive controls in this study. Except for those explicitly labeled, the other values have units of nM. The value following  $\pm$  is standard deviation that Monte-Carlo simulation of 100 cycles calculates assuming at least 5% uncertainty in experimental data.



**Figure 2.** Profiles of differential scanning fluorimetry. Temperature-dependent changes of fluorescence from SYPRO® Orange in the mixture of protein and inhibitor are fitted for calculating the mid-temperature ( $T_m$ ) of denaturation. Fluorescence intensities are normalized with relative fluorescence units in the range of 40 °C–68 °C. Markers and lines indicate the raw and fitted data in each case, respectively. The DSF profiles (A,B) and the name (C) in each compound are prepared to have identical colors. Three independent experiments were repeated for calculating means and standard deviations.

#### 2.4. Cheminformatics Demonstrated Unique Features Compared to Other CA Inhibitors

Are the chemical scaffolds of **5932** and **5946** unique compared to the other known CA inhibitors? To answer this question, we inspected the distribution of TCs between **5932** and **5946**, and other known inhibitors in CA I and CA II. In this SEA, the distribution's biased shape toward one on the x-axis reflects the shared chemical similarities between two groups of inhibitors [14]. Compared to the cases of CA I and CA II inhibitors, the distributions between **5932** and **5946** and the known inhibitors were constrained, with values mostly smaller than 0.4 (Figure 3). The closest similarities were found in ZINC28470844 with a TC of 0.469 for **5932** and ZINC28127722 with 0.417 for **5946**. Indeed, the apparent similarities are limited, except for benzenesulfonamide moiety (Figure 3). This data can qualitatively support the scaffolds' novelty in **5932** and **5946**.



**Figure 3.** Cheminformatics using similarity ensemble approach. Distributions of TCs between the pairs in a set of known CA I and CA II inhibitors (**upper**) are expressed using histogram as references. TCs of two new inhibitors, **5932** (**middle**) and **5946** (**bottom**), and the known CA I and CA II inhibitors, are computed for comparison. The closest known CA inhibitors registered in BindingDB [25] are drawn with the TCs in the parentheses.

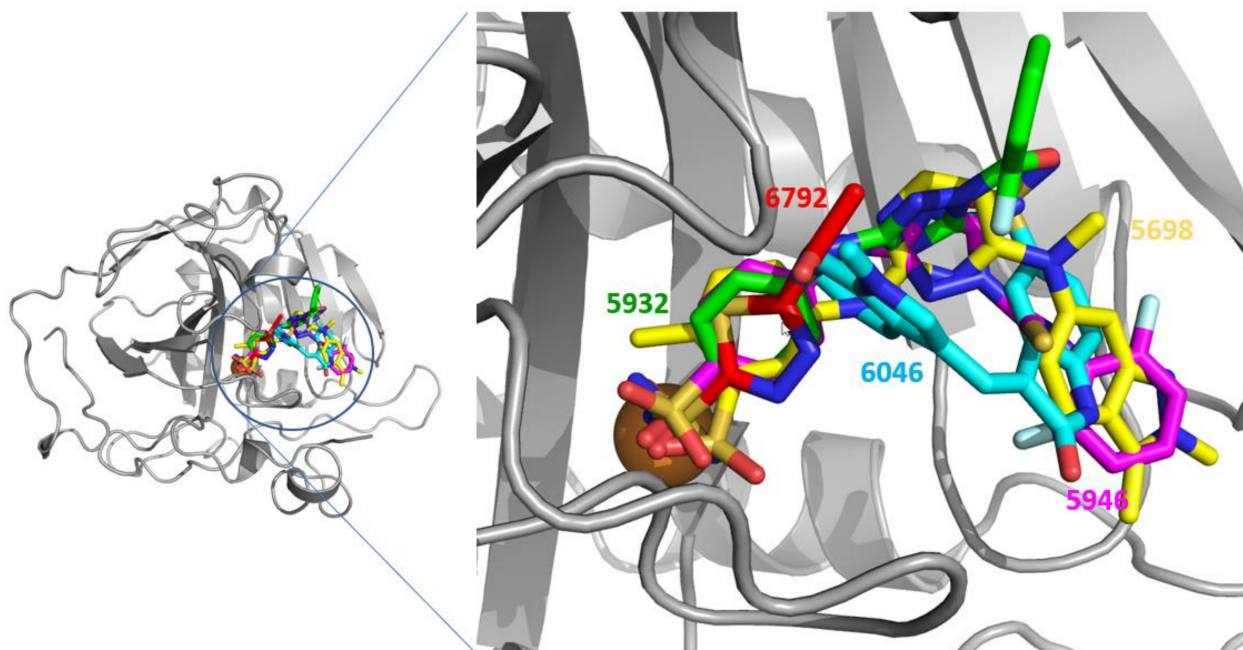
### 2.5. Docking Simulation Predicted the Binding Modes of the Dual Inhibitors

To obtain detailed information at the atomic level, we performed a docking simulation. Direct inhibitors targeting metalloenzymes usually require a metal-binding group (MBG) to interact with the catalytic metals [44]. Sulfonamide is the representative MBG for CAs. Considering the shared sulfonamide, one can infer that **5932** and **5946** interact with CA with the same moiety. The protein database contained 58 coordinates of CA I from 33 crystal structures in the absence or presence of inhibitors. Complex structures with sulfonamide-containing inhibitors have shown a lack of a proton in the amide. It makes the moiety have the net charge of  $-1$ , enabling the chelation between the nitrogen and the zinc metal. The proton's exclusion was approximated by setting pH as 14.0 in the ligand preparation.

The pair of Glide-SP [45,46] and 4WUP-B (4WUP Chain B) of CA I [47] was chosen as a docking engine and a template coordinate, respectively. Two key factors considered for judging structure-based docking are pose reproduction and early enrichment of true positives over false positives. We first tested which software reproduced the crystal poses most faithfully of DOCK3.7 [48,49], Fred [50,51], and Glide-SP [45,46]. Once aligning 58 coordinates with an identical direction, 24 inhibitors from the crystal structures were docked into the structures using each software. The runs using Glide-SP showed the best results with reasonable energies and geometries. Then, the total of 960 decoys ( $40$  per ligand,  $960 = 24 \times 40$ ) was prepared with the identical method applied for ML. The docking simulations using Glide-SP into each coordinate followed with a set of inhibitors

and decoys. The early enrichment of true positives in the docking with 4WUP-B was shown as the best with AUC and LogAUC [48] values of 86.3 and 20.8 (Figure S4). Note that the preparation of small molecules by setting pH as 14.0 is indispensable to obtain the desirable profiles.

Docked poses indicated that **6792**, **5932**, **5946**, and **5698** contacted the zinc atoms through the sulfonamide. Particularly, **6792** had an identical pose to that of the crystal structure. However, **6046** displayed entirely different poses, not contacting the zinc (Figure 4). These data are consistent with the experimental observation that **6046** is less inhibitory for CAs. It shows that the presence of the sulfonamide moiety is not a sufficient condition for inhibiting CAs. Both the existence of MBG and the geometrical fitness of other parts are equally essential.



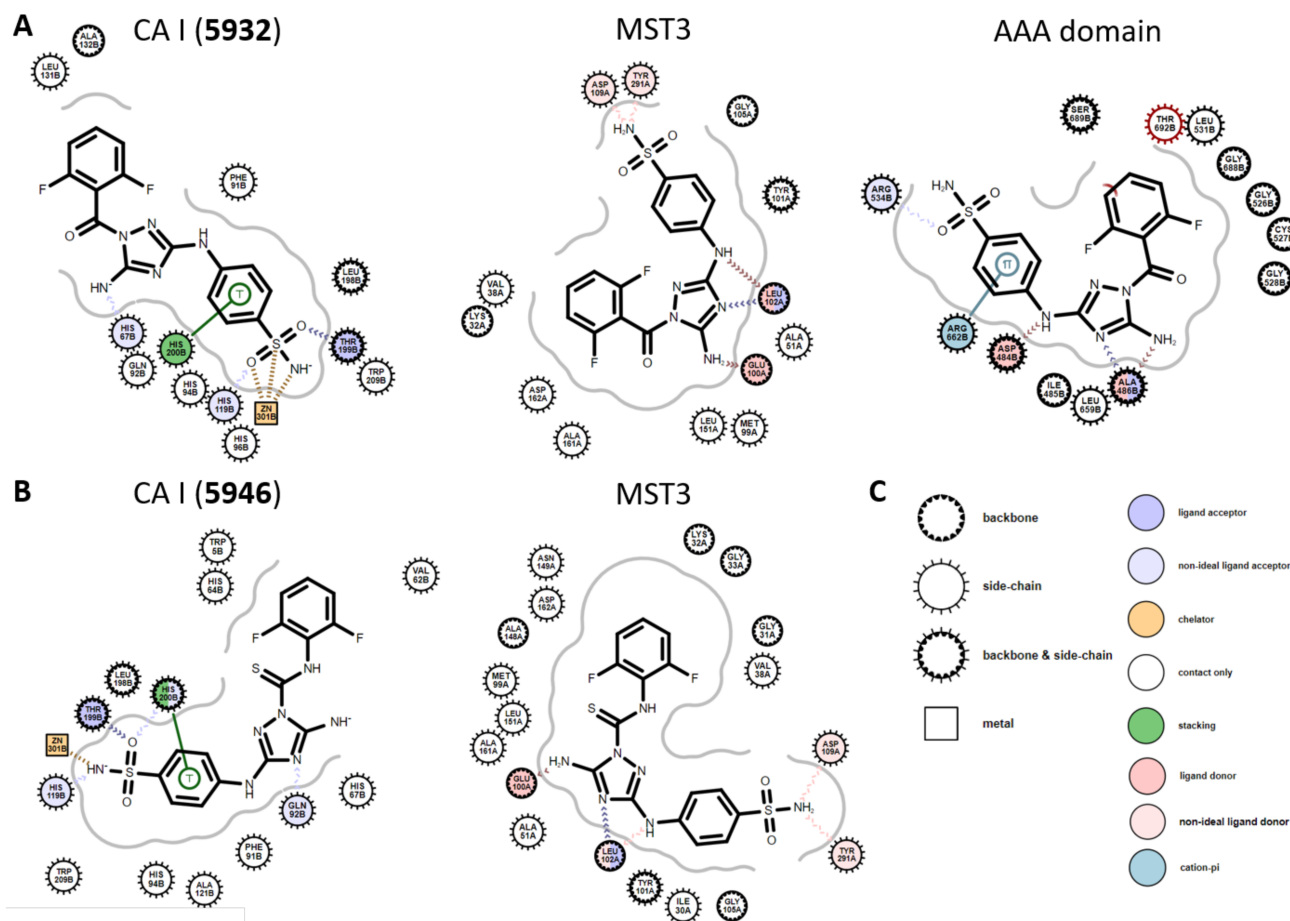
**Figure 4.** Docking simulation. The docked poses between CA I and the molecules in this study are prepared using Glide and the coordinate of 4WUP-B. The molecules in red, green, magenta, cyan, and yellow indicate **6792** (acetazolamide), **5932**, **5946**, **6046**, and **5698** (pazopanib), respectively. The sphere of brown is the zinc ion. The ligands for docking were prepared setting pH as 14 using LigPrep of Schrödinger package.

Discriminating the functional moieties for recognizing CAs and kinases is valuable. Comparing the docked poses of **5932** and **5946** in CA I to those in crystal kinase structures would allow a detailed understanding of the multiple and specific inhibition by **5932** and **5946**. The protein databank includes 13 crystal structures with **5932** as a ligand named SKE (<http://www.rcsb.org/ligand/SKE>, accessed on 1 April 2021). Eleven belong to kinase or pseudokinase domains (PDB IDs: 3AMA, 4QMU, 5DPV, 5DR6, 5DR9, 5DT0, 5OBR, 5USY, 5USZ, 5WIN, and 6DRW), and two cases are spastin AAA domain (6P10 and 6P11). The kinases are MST3 and Aurora A kinases. Additionally, 10 crystal structures with **5946** are found to have the name of DKI (<http://www.rcsb.org/ligand/DKI>, accessed on 1 April 2021). The PDB IDs are 2CHL, 2J51, 2JFL, 2W4O, 2WU6, 3HMI, 3LJ1, 4AAA, 4FZF, and 4QMP. All proteins belong to the kinase family, including ABL2, casein kinase, CLK3, and MST3. Notably, the kinase structures complexed with **5932** and **5946** have a DFGin conformation at the kinase activation loop [52].

In MST3 (4QMU) and AAA domain (6P10) structures, the sulfonamides of **5932** exist on the solvent-accessible surface, forming contacts with the side chains of aspartate and tyrosine. This interaction seems not to be the key contact compared with other kinase



inhibitors. AAA domain additionally uses arginine to interact with the aromatic through intermolecular  $\pi$ -cation contact. Meanwhile, the docked models indicate that the sulfonamide mediates the chelation with the catalytic zinc atom in the core of CAs. The existence of the next aromatic group to sulfonamide is also important by forming intermolecular  $\pi$ - $\pi$  stacking with histidine. Both kinase and AAA domains use amino-triazole for intermolecular hydrogen bonds (Figure 5A). The poses of **5946** were similar to those of **5932** in CA I and MST3, respectively. The benzenesulfonamide plays a crucial role in intermolecular interaction with CA I. Triazole diamine is more vital for interacting with kinase (Figure 5B).

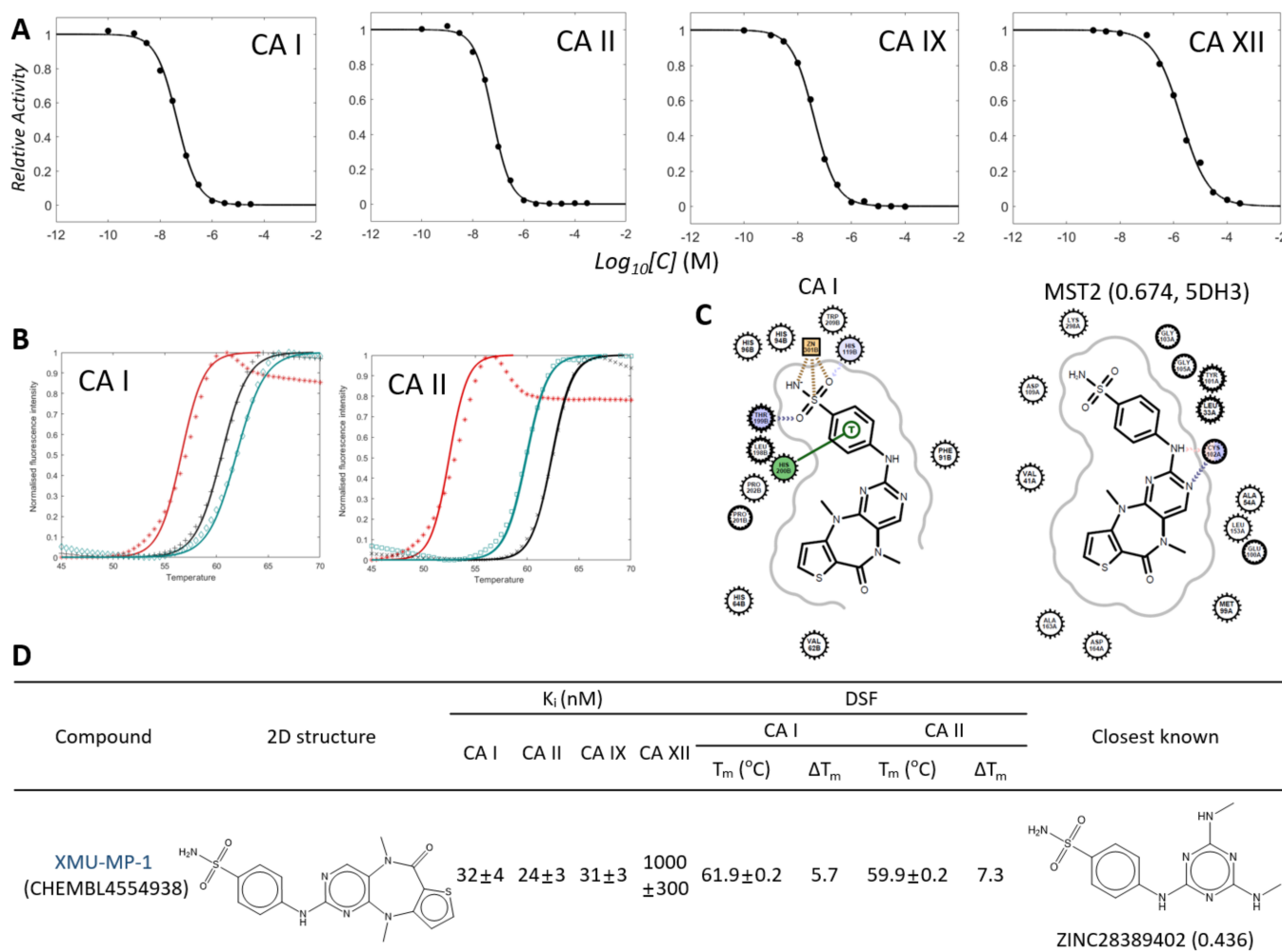


**Figure 5.** Comparison of poses in carbonic anhydrase and other proteins. (A) 2D diagrams of intermolecular interactions of **5932** with CA I (docked model), MST3 kinase (PDB ID: 4QMU), and AAA domain (6P10). (B) 2D diagrams of intermolecular interactions of **5946** with CA I (docked model) and MST3 kinase (4QMP). (C) Residue and interaction styles used in (A,B) are defined with the shapes and colors of circles according to the OpenEye Grapheme Toolkit.

## 2.6. ML Screening of the Known Kinase Inhibitors Led to Discovering a New Potent CA Inhibitor, XMU-MP-1

A meaningful question is whether our approach is extensible to a larger database for finding dual kinase and CA inhibitors. We extracted 35,455 kinase inhibitors of 437 kinases from the ChEMBL database [53]. The molecules smaller than 450 Da and more potent than 500 nM in  $K_i$  or  $IC_{50}$  were considered. We then applied the ML classifier of the CA I model. It retrieved 192 kinase inhibitors with  $p \geq 0.6$  as candidates for CA inhibition. Most molecules have a sulfonamide moiety, but those without the moiety also exist (Figure S5). Interestingly, the candidates include eight molecules whose complex structures with kinases are available. Besides **5932** (ChEMBL191003) and **5946** (ChEMBL261720), the other six molecules were ChEMBL2377825 ( $p = 0.965$ , PDB ligand name: 106, and PDB ID:

1FVT), CHEMBL231950 (0.923, X64, and 3QXP), CHEMBL233149 (0.893, C85, and 2UZD), CHEMBL319467 (0.847, 4SP, and 1H1S/5M57/6BSS), CHEMBL585367 (0.752, ZZF, and 2WOU), and CHEMBL4554938 (0.674, 5BS, and 5DH3) (Figure 6 and Figure S6). Notably, the IUPHAR database does not include these six molecules.



**Figure 6.** Inhibition of carbonic anhydrase by XMU-MP-1. Profiles and uncertainty analyses of enzyme inhibition by XMU-MP-1 are prepared identically to Figures 1 and 2 and Table 4. (A) Enzyme inhibitory profiles. (B) Temperature-dependent fluorescence changes by DSF. Red, black, and navy lines correspond to those without inhibitor, with acetazolamide, and with XMU-MP-1, respectively. (C) 2D diagrams of intermolecular interactions of XMU-MP-1 in docked with CA I and in MST2 kinase structure. The values in the parentheses are probability by ML classifier and PDB ID. (D) Chemical structure, the values of  $K_i$  in enzyme activity, and  $T_m$  in DSF with XMU-MP-1 are presented. The closest molecule to XMU-MP-1 of the known CA I inhibitors is drawn with the TC in the parenthesis.

Of these molecules, considering the commercial availability and feasibility of a detailed comparison, we selected CHEMBL4554938 (XMU-MP-1) and tested its inhibition for CAs. XMU-MP-1 was developed to pharmacologically inhibit MST1/2 kinases [54]. The reported  $IC_{50}$  values for MST1 and MST2 are 71 and 31 nM. Remarkably, XMU-MP-1 was a potent inhibitor for CA I, CA II, and CA IX with the  $K_i$  values of 32, 24, and 31 nM, respectively (Figure 6A,D). The  $K_i$  for CA XII was 1  $\mu$ M. DSF also verified the direct interaction between XMU-MP-1 and CA I and CA II (Figure 6B,D). Docking simulation and crystal structure indicated that the intermolecular contacts found in the bound poses of XMU-MP-1 to CA I and kinase are similar to 5932 and 5946 cases (Figure 6C). While the benzenesulfonamide is

crucial for the interaction with the zinc atom of CA I, the key interaction with the hinge region of MST2 kinase uses 2-aminopyrimidine of XMU-MP-1 [54]. The activation loop of the kinase structure with XMU-MP-1 has a DFGin conformation too [52]. The chemical similarity to the known CA inhibitors is limited in the sulfonamide moiety. The closest known inhibitor is ZINC28389402 with a TC of 0.436 (Figure 6D).

### 3. Discussion

Rapidly accumulating chemical and biological data are expected to accelerate more accurate computational screens. However, imperfectness in the algorithm and the training dataset brings about vast false positives in computer-assisted screens [55,56], necessitating experimental confirmation. Quantitative prediction using ML regression is more challenging than classification. The optimization of ML regression relies on the loss function, expressed mainly as the mean squared error. Comparing the loss functions with k-fold cross-validation is the way to select the best model. By contrast, ML binary classification allows the use of various metrics. It is advantageous for judging several models simultaneously. In addition, ML classification is less sensitive to erroneous data than ML regression. A few errors can significantly influence the performance of ML regression. Our data showed that MCC [31] is the discriminative metric for selecting the more proper ML classification model, whereas ACC discriminates little. We even employed another approach to compensate for the imperfectness of ML, the validation using an external dataset. When the number of true positives is small, a direct comparison of true positives would be straightforward for selecting the best approach. Notably, our approach is readily extensible for discovering other types of dual inhibitors as long as the information for inhibitors is available.

A comprehensive understanding of the polypharmacology of a small molecule is crucial for developing selective drugs and for studying signaling pathways mediated through the target protein. Unidentified targets of a bioactive molecule may impact the interpretation made under the assumption of a single on-target effect. XMU-MP-1 is the chemical probe widely being used to study the Hippo pathway. Several publications have described the cellular roles of MST1/2 in the Hippo pathway with XMU-MP-1, assuming the molecule can pharmacologically inhibit the function of MST1/2 [54,57–63]. Our study showed that XMU-MP-1 is a comparably potent inhibitor for CA I, CA II, and CA IX to MST1/2. This raises the possibility that CAs inhibition influences the experimental observations. Indeed, the linkage between CAs and TEAD-YAP/TAZ in the Hippo pathway has been reported [64]. Similarly, the interpretation of cellular studies using 5932 and 5946 may require caution on off-target effects through CA inhibition.

For maximizing the therapeutic effects of kinase and CA dual inhibitors as anticancer agents, one must consider kinase type-specific and CA isoform-specific dual inhibitors. Our proof-of-concept study has mainly focused on the feasibility of the combined use of ML binary classification and experiments to discover dual inhibitors. The massive experimental tests for selective kinase and CA dual inhibitors are beyond the scope of this study. Nevertheless, our data provide meaningful quantitative information. XMU-MP-1 is much weaker in inhibiting CA XII by approximately 30-fold, and 5946 shows more potent inhibition for CA IX. Moreover, the 192 candidates (Figure S5) can hopefully trigger the research on selective kinase and CA dual inhibitors, particularly to find more selective CA IX and CA XII inhibitors.

Barelrier et al. classified the recognition of identical ligands in unrelated proteins into three interactions: identical, similar, and unshared [65]. The dual inhibition by 5932, 5946, and XMU-MP-1 for kinases and CAs belongs to the unshared recognition. The key contacts occur in the nonoverlapping parts of 5932, 5946, and XMU-MP-1. A synthetic modification for modulating polypharmacological effects such as selectivity might be more straightforward in these cases. Our docking protocol will help prepare the accurate poses of designed molecules.

Our approach minimally optimized the parameters of ML classification. Docking simulation was applied not for screening but for understanding the binding modes of active molecules. The physics-based docking screen can work synergistically with the informatics-based ML classification, particularly for screening the ultra-large library comprising more than several billion molecules [15,16]. ML is faster than conventional docking by several hundred-fold and suited for primary crude screening. Subsequent docking can decrease the ML-predicted candidates into manageable numbers for experimental tests. Of note, our optimized docking protocol for CA has the potential to rule out the weak inhibitor, **6046**, by checking the docked pose. The co-use of an optimized ML and docking screen would be a research area deserving further study [15,16]. Our data will be a notable addition to this direction.

#### 4. Materials and Methods

For the ML database, BindingDB [27] was used to extract inhibitors for CA I and CA II. The method of DUD-E server [28] was employed to generate physicochemically matched but topologically different decoys per ligand. The small molecules registered as “Approved drugs” and “Synthetic organics” in IUPHAR and with the molecular weight smaller than 700 were extracted for validation and screening [35], respectively.

The ML classification models were prepared using the Scikit-learn (version 0.22) library implemented in Python (version 3.7) by combining three fingerprints and seven models with default parameters. Setting the radii as 2 and 3 Å of Morgan circular fingerprints in RDKit (ver. 2020.03.3, <http://www.rdkit.org>, accessed on 1 April 2021) approximated ECFP4 and ECFP6 fingerprints comprising 1024 bits. MACCS fingerprint judges whether the predefined 166 patterns exist or not in a molecule, comprising 166 bits. Each bit has the values of zero or one dependent on the absence or presence of a functional moiety. The seven classifier models in test were K-Nearest Neighbor (KNN), Naïve Bayesian (NB), Logistic Regression (Logit), Decision Tree (DT), Random Forest (RF), Multi-Layer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost). Once randomly dividing the data into five blocks of equal sizes, five-fold cross-validation followed with five runs. Here, training and validation parts in a run consisted of four and one parts. Average and standard deviation was calculated for each metric. The final model for screening database was prepared using all the input data.

Binary classification of ML has the four resulting classes: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Three metrics for quantifying the qualities of the ML model are accuracy (ACC), the area under the curve (AUC) of receiver operating characteristic curve (ROC), and Matthews correlation coefficient (MCC). ACC and MCC are defined as follows:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$
$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

ROC is a two-dimensional expression of TP and FP ratios in the respective y- and x-axes. AUC is defined as the area in the ROC. AUC represents the enrichment of TP over FP.

All the reagents used for the biochemical experiments for enzyme inhibitions were bought from Cayman (Ann Harbor, MI, USA), Sigma–Aldrich (St. Louis, MO, USA), Sino Biological (Beijing, China), or Tokyo Chemical Industry (Tokyo, Japan). Inhibitory activities of small molecules on CAs were measured using 4-nitrophenyl acetate as a substrate. The conversion from the substrate into *p*-nitrophenol by CAs led to the optical density change at 405 nm. The changes were measured at 96-well microplates using Epoch2 from BioTek (Winooski, VT, USA). The slope in a time-dependent measurement was calculated by linear regression and interpreted as the quantified activity. Each 200 µL reaction solution comprised 5 nM enzyme and 1 mM substrate buffered using phosphate-buffered saline containing 5% dimethyl sulfoxide and 0.01% (*w/v*) Triton X-100 under the absence or the

presence of inhibitor. All reactions were performed at 37 °C. The substrate-only solution was used as a control to consider autohydrolysis of the substrate. The value of IC<sub>50</sub> was converted into K<sub>i</sub> using the Cheng–Prusoff equation [66].

The biophysical experiments involved the binding of small molecules to CA I or CA II, which were quantified using the T<sub>m</sub> of DSF. The profile was generated by CFX-connect reverse transcription-polymerase chain reaction system (BioRad, Hercules, CA, USA). The solution contained 5× SYPRO® Orange and 0.5 μM recombinant human CA I or CA II in the absence or presence of 200 μM inhibitor. By gradually elevating the solution's temperature 30 °C–90 °C with a ramping rate of 1 °C/min, the fluorescence signals of the dye at 610 nm were measured upon excitation at 492 nm. The patterns were nonlinearly fitted to the Boltzmann equation to calculate T<sub>m</sub> [67]. Only data in the range of 40 °C–68 °C were used for the calculation. Nonlinear curve fittings for K<sub>i</sub> in enzyme assay and T<sub>m</sub> in DSF were performed using MATLAB (MathWorks, Natick, MA, USA).

For the cheminformatics and docking simulation, RDKit was employed to calculate the pairwise TCs and to perform SEA using ECFP4 as a molecular fingerprint. For comparing the distributions in TCs, the data from ZINC15 [68] and BindingDB [27] were used. All the molecules more potent than 10 μM in K<sub>i</sub> or IC<sub>50</sub> and smaller than 500 Da were selected. Kinase inhibitors were extracted from the ChEMBL database [53]. DOCK3.7 [48,49], Fred from OpenEye (Santa Fe, NM, USA) [50,51], and Glide-SP from Schrödinger (New York, NY, USA) [45,46] were used as engines for docking with the default parameter setting. All the SMILES of ligands were converted into 3D coordinates considering tautomerization and electrostatic charge using LigPrep and Epik from Schrödinger. An in-house script, Automated pLatform for Integrative Structure-based DOCKing (ALIS-DOCK) automated the procedure [69,70]. OpenEye Grapheme was used to analyze the intermolecular interaction in the protein and ligand complexes in docking and PDB-deposited structures.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ph15020236/s1>, Figure S1: CA inhibitor candidates of FDA-approved small molecules by ML classifiers, Figure S2: CA inhibitor candidates from IUPHAR-registered small molecules by ML classifiers, Figure S3: Concentration-dependent inhibitory profiles of positive controls, Figure S4: Data in the pre-docking stage, Figure S5: ML-derived CA inhibitor candidates of known kinase inhibitors, Figure S6: CA inhibitor candidates reported as complex structures with kinases. References [35,45,46,48,53] are cited in the supplementary materials.

**Author Contributions:** J.-G.J. conceived the research project; M.-J.K., S.P., and J.-G.J. designed and carried out the experiments, and analyzed the data; J.-G.J. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Bio and Medical Technology Development Program of the National Research Foundation funded by the Ministry of Science and ICT (2017M3A9G8083382), the 4TH BK21 project (Educational Research Group for Platform development of management of emerging infectious disease) of the Korean ministry of education (5199990614732), and a grant from the Ministry of Oceans and Fisheries' R&D project, Korea (2021633).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data is contained within the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Roth, B.L.; Sheffler, D.J.; Kroeze, W.K. Magic shotguns versus magic bullets: Selectively non-selective drugs for mood disorders and schizophrenia. *Nat. Rev. Drug Discov.* **2004**, *3*, 353–359. [[CrossRef](#)]
2. Knight, Z.A.; Lin, H.; Shokat, K.M. Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer* **2010**, *10*, 130–137. [[CrossRef](#)] [[PubMed](#)]
3. Brotz-Oesterhelt, H.; Brunner, N.A. How many modes of action should an antibiotic have? *Curr. Opin. Pharmacol.* **2008**, *8*, 564–573. [[CrossRef](#)]



4. Besnard, J.; Ruda, G.F.; Setola, V.; Abecassis, K.; Rodriguiz, R.M.; Huang, X.P.; Norval, S.; Sassano, M.F.; Shin, A.I.; Webster, L.A.; et al. Automated design of ligands to polypharmacological profiles. *Nature* **2012**, *492*, 215–220. [[CrossRef](#)]
5. Zhou, J.; Jiang, X.; He, S.; Jiang, H.; Feng, F.; Liu, W.; Qu, W.; Sun, H. Rational Design of Multitarget-Directed Ligands: Strategies and Emerging Paradigms. *J. Med. Chem.* **2019**, *62*, 8881–8914. [[CrossRef](#)]
6. Benek, O.; Korabecny, J.; Soukup, O. A Perspective on Multi-target Drugs for Alzheimer's Disease. *Trends Pharmacol. Sci.* **2020**, *41*, 434–445. [[CrossRef](#)]
7. Proschak, E.; Stark, H.; Merk, D. Polypharmacology by Design: A Medicinal Chemist's Perspective on Multitargeting Compounds. *J. Med. Chem.* **2018**, *62*, 420–444. [[CrossRef](#)]
8. Raghavendra, N.M.; Pingili, D.; Kadasi, S.; Mettu, A.; Prasad, S. Dual or multi-targeting inhibitors: The next generation anticancer agents. *Eur. J. Med. Chem.* **2018**, *143*, 1277–1300. [[CrossRef](#)] [[PubMed](#)]
9. Roskoski, R., Jr. Properties of FDA-approved small molecule protein kinase inhibitors: A 2021 update. *Pharmacol. Res. Off. J. Ital. Pharmacol. Soc.* **2021**, *165*, 105463. [[CrossRef](#)] [[PubMed](#)]
10. Pastorekova, S.; Zatovicova, M.; Pastorek, J. Cancer-associated carbonic anhydrases and their inhibition. *Curr. Pharm. Des.* **2008**, *14*, 685–698. [[CrossRef](#)] [[PubMed](#)]
11. Georgey, H.H.; Manhi, F.M.; Mahmoud, W.R.; Mohamed, N.A.; Berrino, E.; Supuran, C.T. 1,2,4-Trisubstituted imidazolinones with dual carbonic anhydrase and p38 mitogen-activated protein kinase inhibitory activity. *Bioorg. Chem.* **2019**, *82*, 109–116. [[CrossRef](#)]
12. Olesen, S.H.; Zhu, J.Y.; Martin, M.P.; Schonbrunn, E. Discovery of Diverse Small-Molecule Inhibitors of Mammalian Sterile20-like Kinase 3 (MST3). *ChemMedChem* **2016**, *11*, 1137–1144. [[CrossRef](#)]
13. Winum, J.Y.; Maresca, A.; Carta, F.; Scozzafava, A.; Supuran, C.T. Polypharmacology of sulfonamides: Pazopanib, a multitargeted receptor tyrosine kinase inhibitor in clinical use, potently inhibits several mammalian carbonic anhydrases. *Chem. Commun.* **2012**, *48*, 8177–8179. [[CrossRef](#)]
14. Keiser, M.J.; Roth, B.L.; Armbruster, B.N.; Ernsberger, P.; Irwin, J.J.; Shoichet, B.K. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* **2007**, *25*, 197–206. [[CrossRef](#)]
15. Yang, Y.; Yao, K.; Repasky, M.P.; Leswing, K.; Abel, R.; Shoichet, B.K.; Jerome, S.V. Efficient Exploration of Chemical Space with Docking and Deep Learning. *J. Chem. Theory Comput.* **2021**, *17*, 7106–7119. [[CrossRef](#)]
16. Gentile, F.; Agrawal, V.; Hsing, M.; Ton, A.T.; Ban, F.; Norinder, U.; Gleave, M.E.; Cherkasov, A. Deep Docking: A Deep Learning Platform for Augmentation of Structure Based Drug Discovery. *ACS Cent. Sci.* **2020**, *6*, 939–949. [[CrossRef](#)]
17. Pottel, J.; Armstrong, D.; Zou, L.; Fekete, A.; Huang, X.P.; Torosyan, H.; Bednarczyk, D.; Whitebread, S.; Bhatarai, B.; Liang, G.; et al. The activities of drug inactive ingredients on biological targets. *Science* **2020**, *369*, 403–413. [[CrossRef](#)] [[PubMed](#)]
18. Lounkine, E.; Keiser, M.J.; Whitebread, S.; Mikhailov, D.; Hamon, J.; Jenkins, J.L.; Lavan, P.; Weber, E.; Doak, A.K.; Cote, S.; et al. Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **2012**, *486*, 361–367. [[CrossRef](#)] [[PubMed](#)]
19. Keiser, M.J.; Setola, V.; Irwin, J.J.; Laggner, C.; Abbas, A.I.; Hufeisen, S.J.; Jensen, N.H.; Kuijter, M.B.; Matos, R.C.; Tran, T.B.; et al. Predicting new molecular targets for known drugs. *Nature* **2009**, *462*, 175–181. [[CrossRef](#)] [[PubMed](#)]
20. Bosc, N.; Atkinson, F.; Felix, E.; Gaulton, A.; Hersey, A.; Leach, A.R. Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. *J. Cheminform.* **2019**, *11*, 4. [[CrossRef](#)]
21. Mathai, N.; Kirchmair, J. Similarity-Based Methods and Machine Learning Approaches for Target Prediction in Early Drug Discovery: Performance and Scope. *Int. J. Mol. Sci.* **2020**, *21*, 3585. [[CrossRef](#)]
22. Paricharak, S.; Cortes-Ciriano, I.; AP, I.J.; Malliavin, T.E.; Bender, A. Proteochemometric modelling coupled to in silico target prediction: An integrated approach for the simultaneous prediction of polypharmacology and binding affinity/potency of small molecules. *J. Cheminform.* **2015**, *7*, 15. [[CrossRef](#)] [[PubMed](#)]
23. Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H. PharmMapper server: A web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Res.* **2010**, *38*, W609–W614. [[CrossRef](#)] [[PubMed](#)]
24. Awale, M.; Reymond, J.L. Polypharmacology Browser PPB2: Target Prediction Combining Nearest Neighbors with Machine Learning. *J. Chem. Inf. Modeling* **2019**, *59*, 10–17. [[CrossRef](#)] [[PubMed](#)]
25. Yang, M.; Tao, B.; Chen, C.; Jia, W.; Sun, S.; Zhang, T.; Wang, X. Machine Learning Models Based on Molecular Fingerprints and an Extreme Gradient Boosting Method Lead to the Discovery of JAK2 Inhibitors. *J. Chem. Inf. Modeling* **2019**, *59*, 5002–5012. [[CrossRef](#)] [[PubMed](#)]
26. Chen, X.; Xie, W.; Yang, Y.; Hua, Y.; Xing, G.; Liang, L.; Deng, C.; Wang, Y.; Fan, Y.; Liu, H.; et al. Discovery of Dual FGFR4 and EGFR Inhibitors by Machine Learning and Biological Evaluation. *J. Chem. Inf. Modeling* **2020**, *60*, 4640–4652. [[CrossRef](#)]
27. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R.N.; Gilson, M.K. BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198–D201. [[CrossRef](#)]
28. Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594. [[CrossRef](#)]
29. Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Modeling* **2010**, *50*, 742–754. [[CrossRef](#)]
30. Durant, J.L.; Leland, B.A.; Henry, D.R.; Nourse, J.G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273–1280. [[CrossRef](#)]

31. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [[CrossRef](#)] [[PubMed](#)]
32. Irwin, J.J.; Gaskins, G.; Sterling, T.; Mysinger, M.M.; Keiser, M.J. Predicted Biological Activity of Purchasable Chemical Space. *J. Chem. Inf. Modeling* **2018**, *58*, 148–164. [[CrossRef](#)]
33. Krishnan, S.R.; Bung, N.; Bulusu, G.; Roy, A. Accelerating De Novo Drug Design against Novel Proteins Using Deep Learning. *J. Chem. Inf. Modeling* **2021**, *61*, 621–630. [[CrossRef](#)]
34. Sun, G.; Fan, T.; Sun, X.; Hao, Y.; Cui, X.; Zhao, L.; Ren, T.; Zhou, Y.; Zhong, R.; Peng, Y. In Silico Prediction of O(6)-Methylguanine-DNA Methyltransferase Inhibitory Potency of Base Analogs with QSAR and Machine Learning Methods. *Molecules* **2018**, *23*, 2892. [[CrossRef](#)] [[PubMed](#)]
35. Armstrong, J.F.; Faccenda, E.; Harding, S.D.; Pawson, A.J.; Southan, C.; Sharman, J.L.; Campo, B.; Cavanagh, D.R.; Alexander, S.P.H.; Davenport, A.P.; et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: Extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. *Nucleic Acids Res.* **2020**, *48*, D1006–D1021. [[CrossRef](#)]
36. Carta, F.; Supuran, C.T. Diuretics with carbonic anhydrase inhibitory action: A patent and literature review (2005–2013). *Expert Opin. Ther. Pat.* **2013**, *23*, 681–691. [[CrossRef](#)] [[PubMed](#)]
37. Lin, R.; Connolly, P.J.; Huang, S.; Wetter, S.K.; Lu, Y.; Murray, W.V.; Emanuel, S.L.; Gruninger, R.H.; Fuentes-Pesquera, A.R.; Rugg, C.A.; et al. 1-Acyl-1H-[1,2,4]triazole-3,5-diamine analogues as novel and potent anticancer cyclin-dependent kinase inhibitors: Synthesis and evaluation of biological activities. *J. Med. Chem.* **2005**, *48*, 4208–4211. [[CrossRef](#)]
38. Bamborough, P.; Angell, R.M.; Bhamra, I.; Brown, D.; Bull, J.; Christopher, J.A.; Cooper, A.W.; Fazal, L.H.; Giordano, I.; Hind, L.; et al. N-4-Pyrimidinyl-1H-indazol-4-amine inhibitors of Lck: Indazoles as phenol isosteres with improved pharmacokinetics. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4363–4368. [[CrossRef](#)]
39. Lai, J.Y.; Cox, P.J.; Patel, R.; Sadiq, S.; Aldous, D.J.; Thurairatnam, S.; Smith, K.; Wheeler, D.; Jagpal, S.; Parveen, S.; et al. Potent small molecule inhibitors of spleen tyrosine kinase (Syk). *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3111–3114. [[CrossRef](#)]
40. Duan, D.; Torosyan, H.; Elnatan, D.; McLaughlin, C.K.; Logie, J.; Shoichet, M.S.; Agard, D.A.; Shoichet, B.K. Internal Structure and Preferential Protein Binding of Colloidal Aggregates. *ACS Chem. Biol.* **2017**, *12*, 282–290. [[CrossRef](#)] [[PubMed](#)]
41. McLaughlin, C.K.; Duan, D.; Ganesh, A.N.; Torosyan, H.; Shoichet, B.K.; Shoichet, M.S. Stable Colloidal Drug Aggregates Catch and Release Active Enzymes. *ACS Chem. Biol.* **2016**, *11*, 992–1000. [[CrossRef](#)] [[PubMed](#)]
42. Owen, S.C.; Doak, A.K.; Wassam, P.; Shoichet, M.S.; Shoichet, B.K. Colloidal aggregation affects the efficacy of anticancer drugs in cell culture. *ACS Chem. Biol.* **2012**, *7*, 1429–1435. [[CrossRef](#)]
43. Coan, K.E.; Shoichet, B.K. Stoichiometry and physical chemistry of promiscuous aggregate-based inhibitors. *J. Am. Chem. Soc.* **2008**, *130*, 9606–9612. [[CrossRef](#)] [[PubMed](#)]
44. Chen, A.Y.; Adamek, R.N.; Dick, B.L.; Credille, C.V.; Morrison, C.N.; Cohen, S.M. Targeting Metalloenzymes for Therapeutic Intervention. *Chem. Rev.* **2019**, *119*, 1323–1455. [[CrossRef](#)]
45. Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759. [[CrossRef](#)]
46. Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. [[CrossRef](#)]
47. Zubriene, A.; Smirnoviene, J.; Smirnov, A.; Morkunaite, V.; Michailoviene, V.; Jachno, J.; Juozapaitiene, V.; Norvaisas, P.; Manakova, E.; Grazulis, S.; et al. Intrinsic thermodynamics of 4-substituted-2,3,5,6-tetrafluorobenzenesulfonamide binding to carbonic anhydrases by isothermal titration calorimetry. *Biophys. Chem.* **2015**, *205*, 51–65. [[CrossRef](#)] [[PubMed](#)]
48. Mysinger, M.M.; Shoichet, B.K. Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Modeling* **2010**, *50*, 1561–1573. [[CrossRef](#)] [[PubMed](#)]
49. Irwin, J.J.; Shoichet, B.K.; Mysinger, M.M.; Huang, N.; Colizzi, F.; Wassam, P.; Cao, Y. Automated docking screens: A feasibility study. *J. Med. Chem.* **2009**, *52*, 5712–5720. [[CrossRef](#)] [[PubMed](#)]
50. McGann, M. FRED and HYBRID docking performance on standardized datasets. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 897–906. [[CrossRef](#)]
51. McGann, M. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Modeling* **2011**, *51*, 578–596. [[CrossRef](#)]
52. Modi, V.; Dunbrack, R.L., Jr. Defining a new nomenclature for the structures of active and inactive kinases. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6818–6827. [[CrossRef](#)] [[PubMed](#)]
53. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrian-Uhalte, E.; et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954. [[CrossRef](#)] [[PubMed](#)]
54. Fan, F.; He, Z.; Kong, L.L.; Chen, Q.; Yuan, Q.; Zhang, S.; Ye, J.; Liu, H.; Sun, X.; Geng, J.; et al. Pharmacological targeting of kinases MST1 and MST2 augments tissue repair and regeneration. *Sci. Transl. Med.* **2016**, *8*, 352ra108. [[CrossRef](#)]
55. Zhavoronkov, A.; Ivanenkov, Y.A.; Aliper, A.; Veselov, M.S.; Aladinskiy, V.A.; Aladinskaya, A.V.; Terentiev, V.A.; Polykovskiy, D.A.; Kuznetsov, M.D.; Asadulaev, A.; et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **2019**, *37*, 1038–1040. [[CrossRef](#)] [[PubMed](#)]
56. Irwin, J.J.; Shoichet, B.K. Docking Screens for Novel Ligands Conferring New Biology. *J. Med. Chem.* **2016**, *59*, 4103–4120. [[CrossRef](#)]

57. Okuyama, M.; Jiang, W.; Yang, L.; Subramanian, V. Mst1/2 Kinases Inhibitor, XMU-MP-1, Attenuates Angiotensin II-Induced Ascending Aortic Expansion in Hypercholesterolemic Mice. *Circ. Rep.* **2021**, *3*, 259–266. [[CrossRef](#)]
58. Tian, Q.; Fan, X.; Ma, J.; Han, Y.; Li, D.; Jiang, S.; Zhang, F.; Guang, H.; Shan, X.; Chen, R.; et al. Resveratrol ameliorates lipopolysaccharide-induced anxiety-like behavior by attenuating YAP-mediated neuro-inflammation and promoting hippocampal autophagy in mice. *Toxicol. Appl. Pharmacol.* **2020**, *408*, 115261. [[CrossRef](#)]
59. Seeneevassen, L.; Giraud, J.; Molina-Castro, S.; Sifre, E.; Tiffon, C.; Beauvoit, C.; Staedel, C.; Megraud, F.; Lehours, P.; Martin, O.C.B.; et al. Leukaemia Inhibitory Factor (LIF) Inhibits Cancer Stem Cells Tumorigenic Properties through Hippo Kinases Activation in Gastric Cancer. *Cancers* **2020**, *12*, 2011. [[CrossRef](#)]
60. Mitchell, E.; Mellor, C.E.L.; Purba, T.S. XMU-MP-1 induces growth arrest in a model human mini-organ and antagonises cell cycle-dependent paclitaxel cytotoxicity. *Cell Div.* **2020**, *15*, 11. [[CrossRef](#)]
61. Liu, L.Y.; Shan, X.Q.; Zhang, F.K.; Fan, X.F.; Fan, J.M.; Wang, Y.Y.; Liu, S.F.; Mao, S.Z.; Gong, Y.S. YAP activity protects against endotoxemic acute lung injury by activating multiple mechanisms. *Int. J. Mol. Med.* **2020**, *46*, 2235–2250. [[CrossRef](#)] [[PubMed](#)]
62. Faizah, Z.; Amanda, B.; Ashari, F.Y.; Triastuti, E.; Oxtoby, R.; Rahaju, A.S.; Aziz, M.A.; Lusida, M.I.; Oceandy, D. Treatment with Mammalian Ste-20-like Kinase 1/2 (MST1/2) Inhibitor XMU-MP-1 Improves Glucose Tolerance in Streptozotocin-Induced Diabetes Mice. *Molecules* **2020**, *25*, 4381. [[CrossRef](#)] [[PubMed](#)]
63. Triastuti, E.; Nugroho, A.B.; Zi, M.; Prehar, S.; Kohar, Y.S.; Bui, T.A.; Stafford, N.; Cartwright, E.J.; Abraham, S.; Oceandy, D. Pharmacological inhibition of Hippo pathway, with the novel kinase inhibitor XMU-MP-1, protects the heart against adverse effects during pressure overload. *Br. J. Pharmacol.* **2019**, *176*, 3956–3971. [[CrossRef](#)] [[PubMed](#)]
64. Wang, M.; Dai, M.; Wang, D.; Xiong, W.; Zeng, Z.; Guo, C. The regulatory networks of the Hippo signaling pathway in cancer development. *J. Cancer* **2021**, *12*, 6216–6230. [[CrossRef](#)]
65. Barelier, S.; Sterling, T.; O'Meara, M.J.; Shoichet, B.K. The Recognition of Identical Ligands by Unrelated Proteins. *ACS Chem. Biol.* **2015**, *10*, 2772–2784. [[CrossRef](#)]
66. Cheng, Y.; Prusoff, W.H. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099–3108. [[CrossRef](#)] [[PubMed](#)]
67. Choi, J.; Lee, Y.-M.; Jee, J.-G. Thiopurine drugs repositioned as tyrosinase inhibitors. *Int. J. Mol. Sci.* **2018**, *19*, 77. [[CrossRef](#)]
68. Sterling, T.; Irwin, J.J. ZINC 15–Ligand Discovery for Everyone. *J. Chem. Inf. Modeling* **2015**, *55*, 2324–2337. [[CrossRef](#)]
69. Kim, H.H.; Hyun, J.S.; Choi, J.; Choi, K.E.; Jee, J.G.; Park, S.J. Structural ensemble-based docking simulation and biophysical studies discovered new inhibitors of Hsp90 N-terminal domain. *Sci. Rep.* **2018**, *8*, 368. [[CrossRef](#)] [[PubMed](#)]
70. Choi, J.; Choi, K.E.; Park, S.J.; Kim, S.Y.; Jee, J.G. Ensemble-Based Virtual Screening Led to the Discovery of New Classes of Potent Tyrosinase Inhibitors. *J. Chem. Inf. Modeling* **2016**, *56*, 354–367. [[CrossRef](#)]