

## Article

# Identification of Novel Arachidonic Acid 15-Lipoxygenase Inhibitors Based on the Bayesian Classifier Model and Computer-Aided High-Throughput Virtual Screening

Yinglin Liao <sup>1,†</sup>, Peng Cao <sup>2,†</sup> and Lianxiang Luo <sup>3,4,\*</sup> 

- <sup>1</sup> The First Clinical College, Guangdong Medical University, Zhanjiang 524023, China  
<sup>2</sup> Department of Pharmacy, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China  
<sup>3</sup> The Marine Biomedical Research Institute, Guangdong Medical University, Zhanjiang 524023, China  
<sup>4</sup> The Marine Biomedical Research Institute of Guangdong Zhanjiang, Zhanjiang 524023, China  
\* Correspondence: luolianxiang321@gdmu.edu.cn; Tel.: +86-150-1301-7454; Fax: +86-0759-238-8163  
† These authors contributed equally to this work.

**Abstract:** Ferroptosis is an iron-dependent lipid peroxidative form of cell death that is distinct from apoptosis and necrosis. ALOX15, also known as arachidonic acid 15-lipoxygenase, promotes ferroptosis by converting intracellular unsaturated lipids into oxidized lipid intermediates and is an important ferroptosis target. In this study, a naive Bayesian machine learning classifier with a structure-based, high-throughput screening approach and a molecular docking program were combined to screen for three compounds with excellent target-binding potential. In the absorption, distribution, metabolism, excretion, and toxicity characterization, three candidate molecules were predicted to exhibit drug-like properties. The subsequent molecular dynamics simulations confirmed their stable binding to the targets. The findings indicated that the compounds exhibited excellent potential ALOX15 inhibitor capacity, thereby providing novel candidates for the treatment of inflammatory ischemia-related diseases caused by ferroptosis.

**Keywords:** ferroptosis; ALOX15; homology modeling; machine learning; virtual screening; molecular dynamic simulation



**Citation:** Liao, Y.; Cao, P.; Luo, L. Identification of Novel Arachidonic Acid 15-Lipoxygenase Inhibitors Based on the Bayesian Classifier Model and Computer-Aided High-Throughput Virtual Screening. *Pharmaceuticals* **2022**, *15*, 1440. <https://doi.org/10.3390/ph15111440>

Academic Editor: Krzysztof Marciniak

Received: 8 October 2022

Accepted: 16 November 2022

Published: 20 November 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

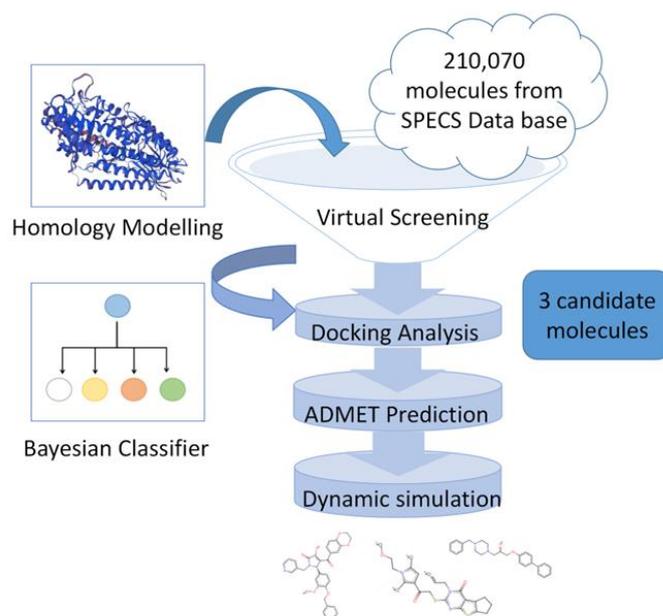
Unlike apoptosis, necrosis, or autophagy, ferroptosis is a newly identified mode of cell death that occurs in RAS-mutated tumor cells [1]. Ferroptosis is typically triggered by the accumulation of intracellular iron, which induces intracellular lipid peroxidation. The accumulation of intracellular reactive oxygen species (one of the critical features of ferroptosis) leads to organelle and biofilm damage and cell death [2,3]. Ferroptosis can efficiently kill many types of tumors. Numerous drug molecules, mainly Erastin and RSL3 [4], and a number of compounds targeting other key pathway targets that induce tumor ferroptosis have been widely used [5]. However, ferroptosis plays an equally critical role in inflammation and injury. The inhibition of ferroptosis can effectively ameliorate ischemic and degenerative tissue damage in vivo [6]. Therefore, targeting the induction or inhibition of ferroptosis in tumor cells is a novel strategy for the targeted therapy of many diseases, particularly cancer and spontaneous tissue damage [7–10].

The corresponding protease encoded by the ALOX15 gene, arachidonate 15-lipoxygenase, is an oxygenated lipase with both pro-oxidative and esterifying effects that catalyzes the multimerization of polyunsaturated fatty acids (PUFAs) on cell membranes [11]. In 2016, Yang et al. reported that the ALOX enzyme family can promote the onset of cellular ferroptosis. Subsequent cellular-level studies have revealed that ALOX15 promotes ferroptosis in various tumor cells such as HT-1080 and PANC1 cells [12,13]. Thus, the inhibition of ALOX family enzymes, particularly ALOX15, is essential for the treatment of injury and degenerative diseases

triggered by ferroptosis. Novel ALOX15 inhibitors with clinical therapeutic implications have been discovered. For example, Ma et al. (2022) discovered that soya sapogenins can improve ischemia-induced myocardial injury [14]. Cepharanthine, a p53-ALOX pathway blocker developed by Gao et al. [15], reduced ischemia-reperfusion injury in the brain. Although several inhibitors of ALOX15 have been identified to date [16–18], various inhibitor molecules with strong clinical therapeutic effects remain undiscovered because of the long lead time of the drug development process and the tedious procedures of chemical extraction and synthesis. Therefore, potential inhibitor molecules should be identified from established commercial compounds to inspire a drug screening effort against ALOX15 targets.

With the rapid development of electronic programming technologies, computer-aided drug design (CADD) is widely being used in drug development [19]. Many assignment functions and components that can calculate the binding energy between small molecules and target proteins and predict the possible types of interactions with the target based on the molecular structure have been developed and used with good results in numerous studies [20]. Computer-dependent drug prediction methods are more cost-effective than conventional experimental methods. [21]. Machine learning models have been used frequently for predicting compound libraries because of their high accuracy in addressing the classification and identification problems based on large amounts of data [22]. Classification models developed based on machine learning methods have been applied to solve numerous compound property prediction problems [23–25]. Combining machine learning methods in the typical molecular docking process can help identify compounds with drug potential in a targeted manner.

In this study, a coherent computer-aided screening process (Figure 1) was designed to obtain compounds with ALOX15 inhibitory potential from a database of purchasable compounds. The stereospecific protein structure of ALOX15 was established from *Homo sapiens* using a homologous protein modeling approach that combined both structure-based high-throughput screening and Bayesian machine learning models to find three potential compounds. ADMET (absorption, distribution, metabolism, excretion, and toxicity) property prediction and molecular dynamics simulations were used to confirm the drug-forming potential of the candidate compounds. Subsequently, based on existing studies on ALOX15 inhibitors, the molecular features that may play a key role in target binding were analyzed to facilitate the efficient exploration of this protease inhibitor.



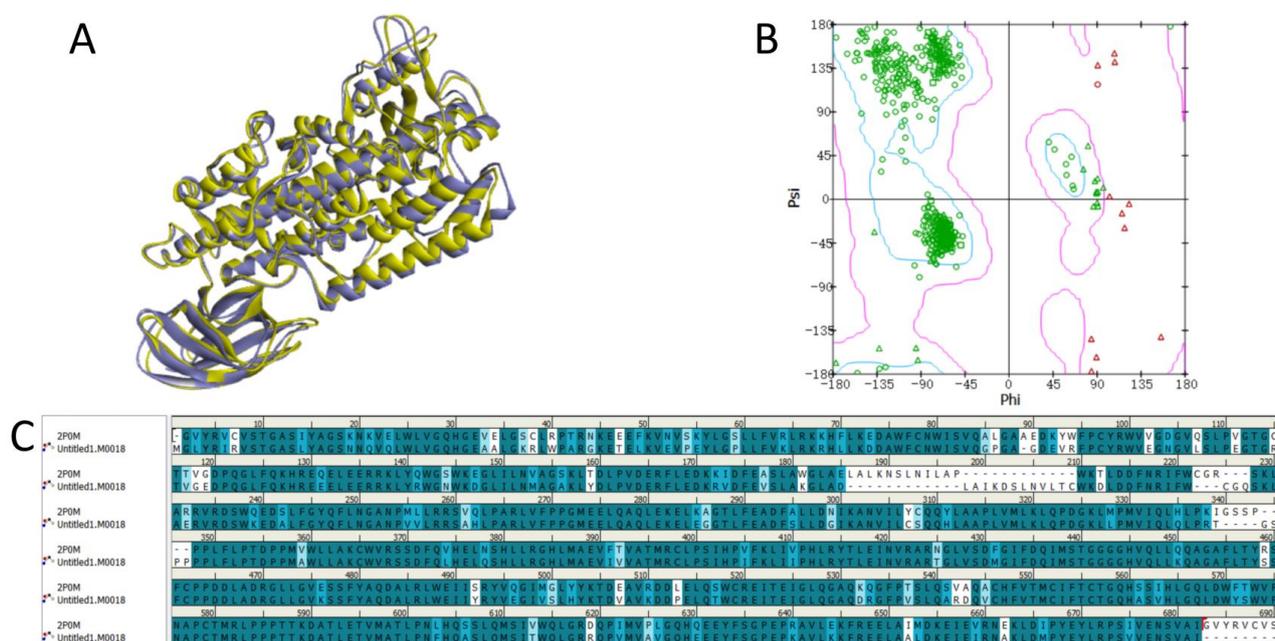
**Figure 1.** Workflow for this study.

## 2. Results

### 2.1. Protein Homology Modeling

In this study, the BLAST search algorithm was used to determine amino acid sequences of proteins with homology to ALOX15 from various biological sources in the UniProt nr\_clustered experimental database based on the rabbit-derived ALOX15 amino acid sequence in the PDB database (Table S1). We selected the sequence of ALOX15 derived from *Homo sapiens*, which has 81.12% homology to the PDB parent template sequence. Twenty possible structural models of the ALOX15 protein were constructed based on this sequence, and the best model, M0018, was selected based on an automatically generated evaluation by MODELLER.

The quality of the model was assessed using the profile-3D function. The results revealed that the verified score for M0018 was 278.82, which was considerably higher than the expected low score (136.417). However, this score was close to the expected high score (303.15), which indicated the high quality of the M0018 model (Table 1). The PDB structures of the M0018 and template sequence source proteins were superimposed and the root mean square deviation (RMSD) values of their spatial structures were calculated. To minimize the chance of error resulting from the overlay of various structures, two calculations were performed using the two protein structures as templates. The results are presented in Table 2 and Figure 2A,C. The difference in the spatial position overlap between the two structures was minimal, and the sequence amino acid repeat was sufficiently high. The Ramachandran plots revealed that of the 662 total residues, 87.6% (580) were located in favorable spatial regions, 10.6% (70) in permissive spatial regions, and 1.8% (12) in non-permissive spatial regions, indicating that the protein model was of sufficiently high quality. Therefore, the M0018 protein model was used for the subsequent studies.



**Figure 2.** Schematic of the protein homology modeling results: (A) structural overlay of homology model M0018 (in yellow) and the template protein (PDB ID: 2P0M, in purple); (B) a residue spatial position rationality analysis of the M0018 Ramachandran plot; (C) a sequence comparison of the homology model M0018 with the template protein.

**Table 1.** Verified score for the homologous protein model M0018.

Name	Verify Score	Verify Expected High Score	Verify Expected Low Score
M0018	278.82	303.15	136.417

**Table 2.** Alignment cluster root mean square deviation (RMSD) values for 2P0M and M0018 (unit: Angstrom).

Protein	2P0M	M0018
2P0M	–	1.1630
M0018	1.2990	–

### 2.2. Structure-Based, High-Throughput Virtual Screening

The LibDock module of the Discovery Studio platform was used to perform a virtual screening of 210,070 small molecules within the SPECS compound library by using the ALOX15 model based on a homology approach, which resulted in the generation of 46,127 dockable conformations. The top 1% of compounds with the highest LibDock scores (450) were studied based on the assumption that higher scores correspond to superior activity. To avoid a possible chance bias arising from the calculation function of the docking function module, only the molecules that produced two or more interacting conformations with the target were considered in the compound selection.

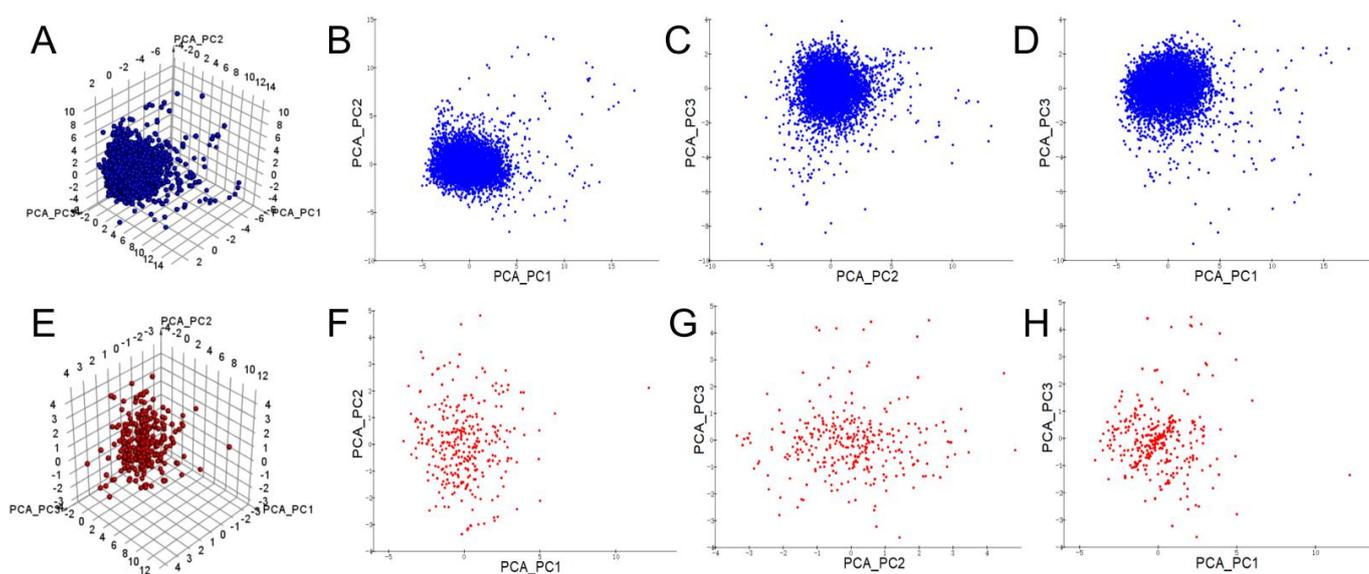
### 2.3. Chemical Spatial Distribution Analysis

The discrimination ability of machine learning classification models depends heavily on whether the chemical spatial distribution of the compounds in the training dataset is sufficiently diverse [26]. Therefore, the chemical spatial diversity of 55 calculated descriptors (Table S2) from the entire dataset was investigated using a principal component analysis (PCA) and Tanimoto similarity analysis. The PCA of the molecular descriptors yielded eight key molecular descriptors, namely ES\_Count\_aasN, ES\_Count\_aaO, ES\_Count\_dCH2, Num\_AromaticRings, Molecular\_FractionalPolarSurfaceArea, IsChiral, Num\_Rings, and QED. Five principal components (PCs) were obtained by assigning distinct weighting factors to these descriptors. The first three most critical principal components (PC1–PC3), whose component descriptors and coefficient correlation equations are presented in Table S3) were selected for the chemical spatial analysis. As shown in Figure 3A–D, for the three PCs of the dataset, most of the molecules were clustered in a limited spatial range, probably because of their inherently even descriptor properties and the large number of molecules in the training set. To analyze the chemical similarity between compounds in numerical terms, the fingerprint distance and Tanimoto similarity coefficient ( $T_c$ ) between compounds were calculated based on ECFP\_2. The results revealed an overall  $T_c$  value of 0.09968 for our training set of molecules, which indicated the overall low chemical structural similarity of the molecules. The same chemical spatial analysis was performed on the test set of 450 LibDock-screened compounds selected from the SPECS compounds. These test set molecules were widely distributed between the three PCs (Figure 3E–H), with a similarity coefficient of 0.13856 calculated from the molecular fingerprints. This result indicated the wide chemical space of these compounds to be screened. Both our training and test sets exhibited a wide distribution of the chemical space, which can be used for machine learning modeling to achieve an important prerequisite for accurate classification.

### 2.4. Machine Learning Classifier Models

Bayesian classification models were constructed using 5360 active/inactive ALOX15-targeted compounds obtained from the ChEMBL database. The key molecular descriptors obtained from the dimensionality reduction in the PCA, as well as the molecular fingerprint ECFP\_n, were applied to the proposed model construction process. Various fingerprint-descriptor combinations were attempted during the model construction, whereas the area under the ROC curve (AUC) values of the obtained models were considered as indicators representing their classification ability. The best-performing Bayesian model was selected for a further analysis, in which an active flag was used for classification. The naive Bayesian model is based on a binary (i.e., ‘yes or no’) approach to determining whether a compound has target inhibitory activity, with the classification process using the model’s activity

cut-off value ( $-2.366$  in our model) as a benchmark, along with a Bayesian score for each compound. When the absolute value of the Bayesian score is greater than the cut-off value, the compound is classified as ‘active’, and the opposite is classified as “inactive” (Table S4). Subsequently, ECFP\_6 and eight key descriptors were applied (ES\_Count\_aasN, ES\_Count\_aaO, Num\_AromaticRings, Molecular\_FractionalPolarSurfaceArea, IsChiral, Num\_Rings, ES\_Count\_dCH2, and QED). The training result metrics of this model are presented in Table 3, and its ROC curve is presented in Figure 4A. The Naive Bayesian (NB) model revealed excellent classification recognition ability. For the classification of active molecules, the true positive rate (recall) reached 92.5%; for inactive molecules, the prediction rate of the model for true positives (inactive molecules correctly classified) was 91.3%. The model had an AUC of 0.895, close to 1, which indicated its excellent classification ability. However, combining the tests and the confusion matrix of the active/inactive molecules obtained from the model differentiation (Table 4) revealed that the classifier model was slightly better at identifying inactive molecules than it was at differentiating active molecules. This result implied that more false positive examples may be obtained using the model for classification. However, the risk of incorrectly excluding active compounds can be avoided. Because we used more screening and validation methods throughout the screening process, a slightly higher false positive rate is acceptable.



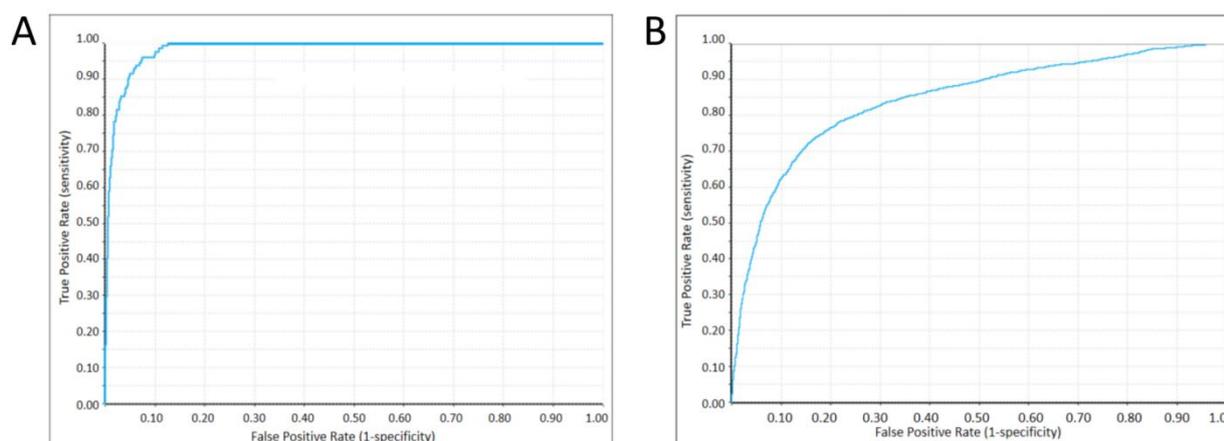
**Figure 3.** Schematic of the spatial distribution of the principal components (PC1–PC3) of the compounds: (A–D) chemical spatial distribution of the test set molecules used as model constructs; (E–H) chemical spatial distribution of the 450 test set molecules from the LibDock screening program.

**Table 3.** Training metrics for the best naive Bayesian (NB) model.

Class	Precision	Recall	F-Measure	AUC	MCC
Active	0.847	0.925	0.884		
Inactive	0.958	0.913	0.935	0.895	0.838
Weighted Avg	0.903	0.919	0.910		

**Table 4.** Training activity/inactivity confusion matrix for the best NB model.

Classification	Predicted Active	Predicted Inactive
Active	1697	138
Inactive	305	3220



**Figure 4.** Plots of validated ROC curves for NB models: (A) training ROC curve of the model; (B) ROC curve obtained from the ten-fold validation of the model.

The constructed Bayesian model was validated using the ten-fold cross-validation method. The results are presented in Table 5, and the ROC plot is displayed in Figure 4B. The ten-fold validation results revealed a similar tendency to the discriminatory ability of the model; that is, the model's ability to identify inactive molecules is superior to when identifying active molecules. The true positive rates for the classification of active molecules and the discrimination of inactive molecules were 85.5 and 87.7%, respectively. Under the ten-fold validation condition, the classification precision metric of the model was significantly higher for inactive examples than that for active examples (Table 6). Compared with the discriminatory ability recognition metrics of the model, all metrics of the ten-fold model validation results decreased but remained within a good range, which confirmed the reliability of the algorithm used to construct the NB model.

**Table 5.** Parameters of the ten-fold validation results for the NB model.

Class	Precision	Recall	F-Measure	AUC	MCC
Active	0.782	0.855	0.817		
Inactive	0.920	0.876	0.897	0.745	0.478
Weighted Avg	0.851	0.866	0.857		

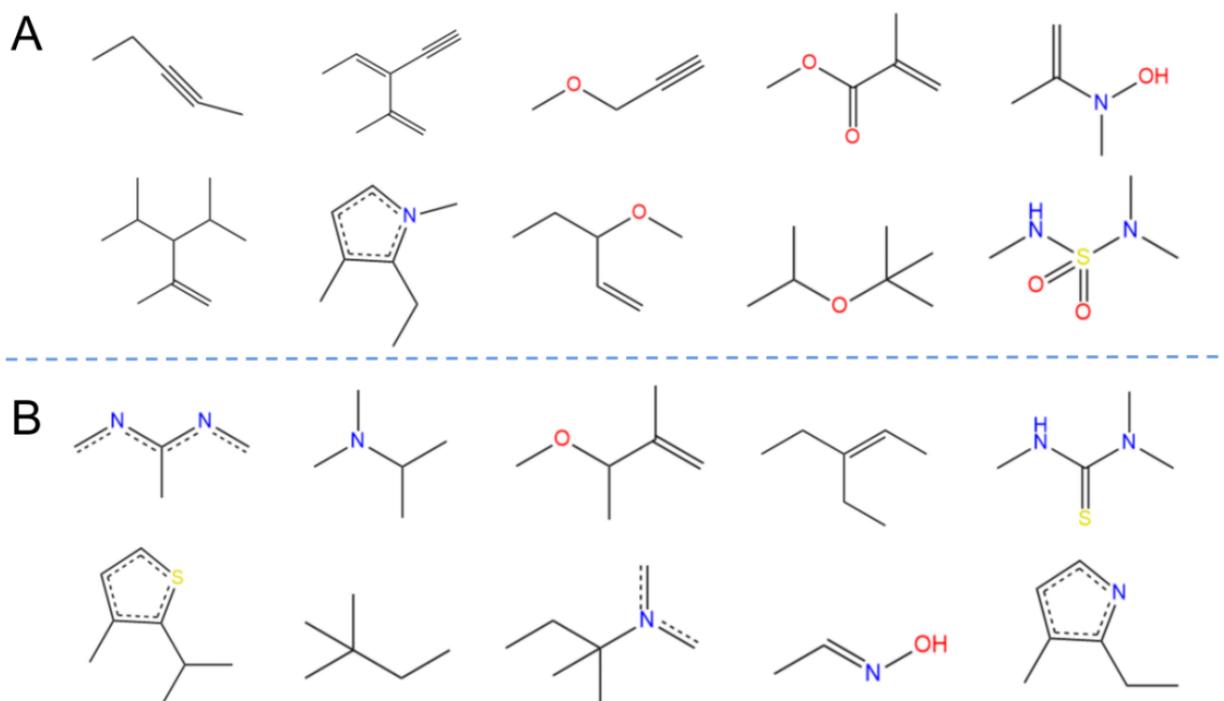
**Table 6.** Active/inactive confusion matrix for ten-fold validation of the NB model.

Classification	Predicted Active	Predicted Inactive
Active	1569	266
Inactive	437	3088

The results obtained by the model in the 10-fold validation test were significantly worse than the trained classification results, which indicated that there was a certain degree of over-fit in our model training, which is often affected by the number of features and the complexity of the algorithm. Overfitting makes the model more susceptible to noise and reduces the accuracy of the predictions, but it is also unavoidable. However, in our 10-fold validation results, the model also showed a good ROC–AUC index, so although the Bayesian model we developed may have some overfitting, it is still considered to have good distinguishability.

We calculated the high-frequency good/bad feature fragments (GF/BF) based on ECFP<sub>6</sub> fingerprinting for all compounds classified as active (171) and inactive (279). Because the frequency of molecular fingerprints of all 171 compounds classified as “active” by the Bayesian classifier was considered in the calculation of the favorable fragments,

the fragments in Figure 5 represent only the statistical results based on a large amount of data for the structure–activity relationship study of potential ALOX15 inhibitors. For the favorable fragments GF1–GF10 calculated via molecular fingerprinting, a larger proportion of fragments contained ether, carboxyl, or ester bonds.

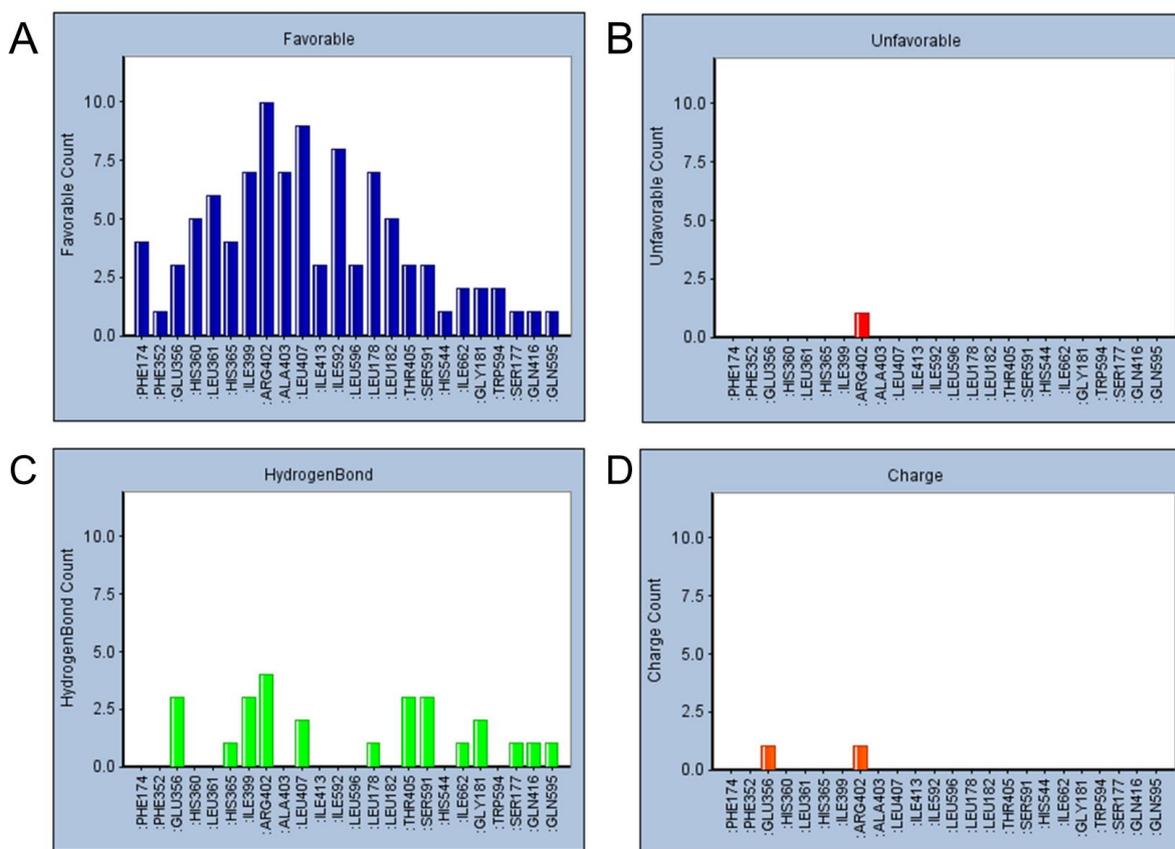


**Figure 5.** Favorable/unfavorable active fragments calculated based on ECFP<sub>6</sub>: (A) top 10 good fingerprints (GF1–GF10) calculated based on ECFP<sub>6</sub> fingerprints; (B) top 10 bad fingerprints (BF1–BF10) calculated based on ECFP<sub>6</sub> fingerprints.

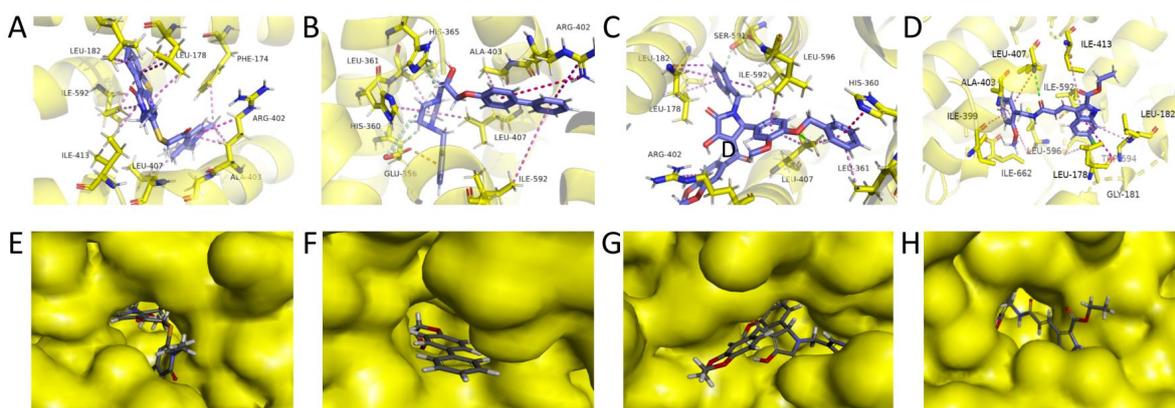
### 2.5. Refined Molecular Docking Analysis

In the previous high-throughput docking and classifier predictions, 10 compounds that were predicted to have a good effect were selected. To analyze how these compounds interact with the targets, compound–target precision docking was conducted using the CDOCKER module. First, the docking residues and types of action of the 10 candidate compounds were counted. As displayed in Figure 6, six residues, namely Leu178, Ile399, Arg402, Ala403, Leu407, and Ile592, produced the most favorable interactions with the ligands. Most of these interactions were hydrophobic interactions. The study by Meng et al. in 2018 revealed Phe174, His360, Ile399, Arg402, Ala403, Leu407, and Leu596 as the key ALOX15 residues, consistent with the statistical results. In 2020, Cruz et al. revealed that residue 596 plays a key role in maintaining the structural stability of ALOX15 [27].

Guo et al. obtained a novel ALOX15 inhibitory compound “i472” as a control [28] by using the CDOCKER docking fraction (including both CDOCKER ENERGY and CDOCKER INTERACTION ENERGY). The three most promising compounds (the first three compounds presented in Table 7) were identified in terms of their CDOCKER ENERGY (which is based on the intermolecular interaction energy, whereas in the CDOCKER INTERACTION ENERGY, the intramolecular energy is considered), type of interaction, and their interactions with key residues. These compounds were labeled C1–C3. The three-dimensional (3D) docking schematics are displayed in Figure 7A–D.



**Figure 6.** Heat map of favorable interaction types and residue statistics for the ten available active compounds: (A) total counts of favorable interactions formed between the top 10 best available compounds and target residues; (B) counts of hydrophobic interactions formed between the top 10 best available compounds and target residues; (C) counts of hydrogen bonds formed between the top 10 best available compounds and target residues; (D) counts of electrostatic charge interactions formed between the top 10 best available compounds and target residues.



**Figure 7.** Schematic 3D interactions of the best compounds C1–C3 with the target: (A) schematic of the interaction of compound C1 with residues within the target-binding pocket of ALOX15; (B) schematic of the interaction of compound C2 with residues within the target-binding pocket; (C) schematic of the interaction of compound C3 with residues within the target-binding pocket; (D) schematic of the interaction of compound i472 with residues within the target-binding pocket; (E) mosaic binding of compound C1 within the target-binding pocket; (F) mosaic binding of compound C2 within the target-binding pocket; (G) mosaic binding of compound C3 within the target-binding pocket; (H) mosaic binding of compound i472 in the target-binding pocket.

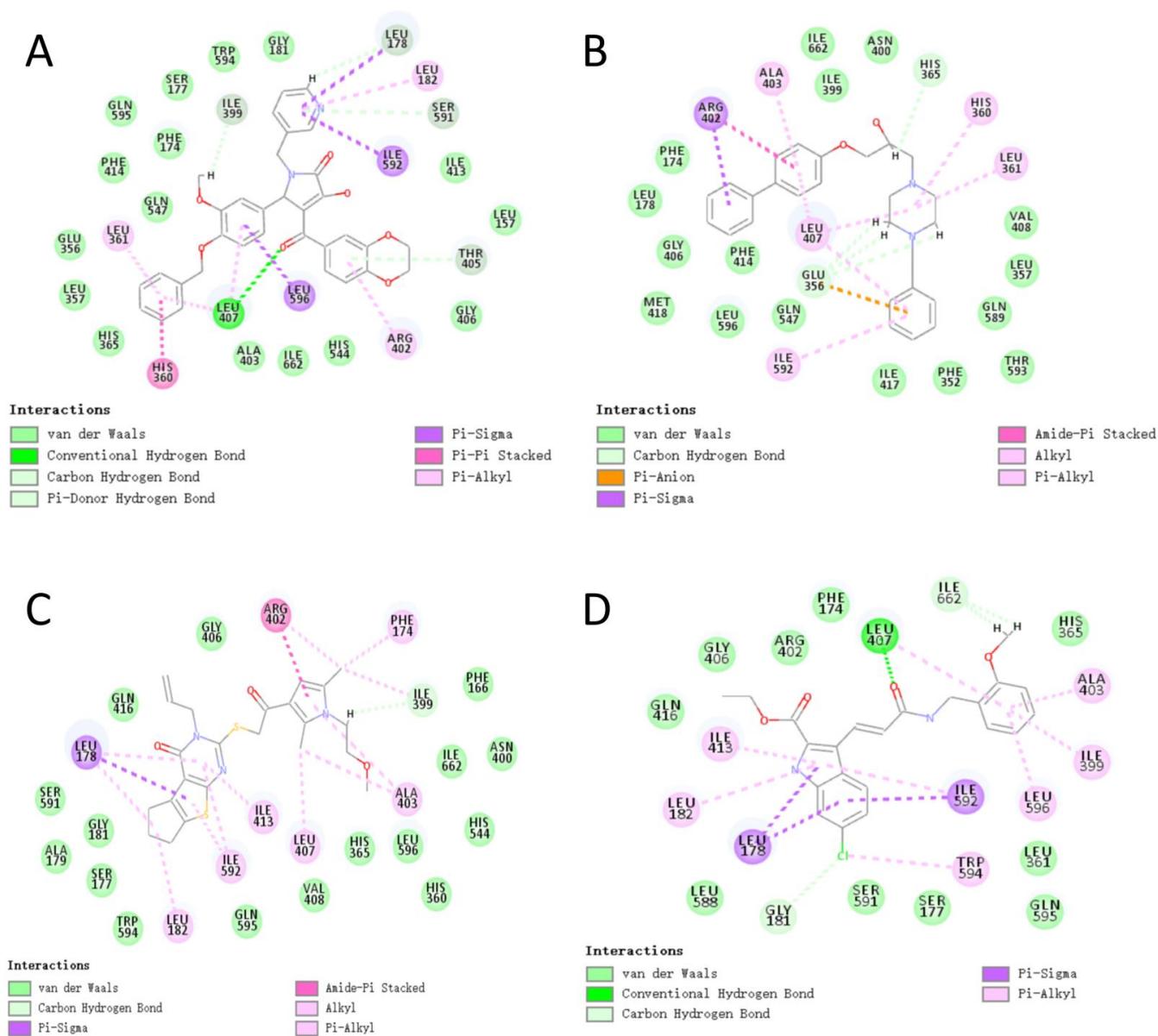
**Table 7.** Chemical names, molecular structures, and docking fractions of candidate compounds 1–10.

Index	Name	Structure	LibDock Score	CDOCKER Energy	CDOCKER Interaction Energy
1	5-[4-(benzyloxy)-3-methoxyphenyl]-4-(2,3-dihydro-1,4-benzodioxin-6-ylcarbonyl)-3-hydroxy-1-(3-pyridinylmethyl)-1,5-dihydro-2H-pyrrol-2-one (C1)		156.52	−43.897	−36.283
2	1-(4-benzylpiperazin-1-yl)-3-([1,1'-biphenyl]-4-ylloxy)propan-2-ol (R) (C2)		129.364	−32.940	−25.541
3	3-allyl-2-((2-[1-(2-methoxyethyl)-2,5-dimethyl-1H-pyrrol-3-yl]-2-oxoethyl)sulfanyl)-3,5,6,7-tetrahydro-4H-cyclopenta [4,5]thieno [2,3-d]pyrimidin-4-one (C3)		133.492	−35.653	−29.243
4	1-(7-acetyl-9-[4-(octyloxy)benzylidene]-9H-fluoren-2-yl)ethanone		127.735	−34.728	−20.403
5	2-chloro-N-[2-((2-oxo-2-[(1-phenylethyl)amino]ethyl)sulfanyl)-1,3-benzothiazol-6-yl]benzamide		118.925	−23.495	−15.965
6	4-bromo-N-[2-(2-cyclooctylidenehydrazino)-2-oxoethyl]-N-(4-ethoxyphenyl)benzenesulfonamide		121.455	−25.582	−8.780
7	6-(4-(3-fluoro-4-methoxyphenyl)-2-[[3-(trifluoromethyl)phenyl]imino]-1,3-thiazol-3(2H)-yl)-1-hexanol		117.047	−21.340	−13.951
8	4-((2,5-dioxo-1-phenyl-3-pyrrolidinyl)[2-(4-methoxyphenyl)ethyl]amino)-4-oxo-2-butenoic acid		120.938	−31.054	−19.390
9	ethyl 2-((4-[(dipropylamino)sulfonyl]benzoyl)amino)-5-ethyl-3-thiophenecarboxylate		123.222	−27.384	−19.071
10	4-tert-butyl-N-[2-((2-[[2-(4-chlorophenoxy)ethyl]amino]-2-oxoethyl)sulfanyl)-1,3-benzothiazol-6-yl]benzamide		118.664	−29.755	−20.632
11	i472		119.591	−37.382	−26.544

Considering the van der Waals forces between the ligands and the residues, all three of the selected compounds interact with the key residue 596, which revealed the excellent target-binding potential of the selected compounds.

For the best scoring compound C1, as displayed in Figure 8A, the target–ligand interactions were dominated by Pi interactions. Leu178, Leu182, His360, Leu361, Arg402, Leu407, Ile592, and Leu596 were among the key Pi-type residues. This result is consistent with those reported by Meng et al. According to the 3D structure of the complex, ligand C1 was

essentially fully embedded in the target active residue pocket, resulting in the formation of more van der Waals interactions with the receptor, which enhanced its target binding.



**Figure 8.** Two-dimensional (2D) diagrams of the interactions of the three best compounds and the control compound i472 with the target: (A) schematic of the plane of interaction of compound C1 with the target; (B) schematic of the plane of interaction of compound C2 with the target; (C) schematic of the plane of interaction of compound C3 with the target; (D) schematic of the plane of interaction of the control compound i472 with the target.

For compound C2, the formation of Pi bond interactions between its residues, Glu356, His360, Leu361, Arg402, Ala403, Leu407, and Ile592, and the targets, including Pi-anion, Pi-sigma, and Pi-alkyl types, account for the compound's hexameric ring structure (Figure 8B). Furthermore, Glu356 formed a strong carbon–hydrogen bond with the hydrogen atom on the diazohexa ring of the small molecule while forming a Pi-anion interaction with the ligand, contributing to the ligand–acceptor interaction. Taking the carbon atom connected to the hydroxyl group on the backbone of compound 2 as the stereogenic center, the R/S configuration of the two enantiomers of each other can be observed. We only considered the R configuration of compound 2 during the above screening. Therefore, we additionally

docked the chiral S configuration of compound 2 to the target. However, during docking, the S configuration fails to successfully bind to the protein structure of ALOX15. Therefore, we believe that the R conformation of compound 2 is promising.

The interaction of compound C3 with the target is displayed in Figure 8C. Numerous Pi interactions can contribute to the ligand–receptor interaction. Similarly, the key residues Phe174, Leu178, Arg402, Ala403, Leu407, and Ile592 formed Pi interactions with the ligand, which confirmed the reliability of the docking analysis.

The control compound i472 formed Pi-alkyl interactions with the known target key residues Ala403 and Leu596 (Figure 8D). Overall, the compound exhibited similarities to our three candidate compounds in the type and number of target interactions. However, i472 formed fewer van der Waals interactions with protein residues. This result could be attributed to the ability of the three alternative compounds to access the active residue pocket more deeply than i472 (Figure 7E–H). The fine-grained interaction force analysis allowed the identification of three candidate compounds with an ALOX15 targeting ability comparable to that of compound i472.

Comparing previous favorable fragments obtained from molecular fingerprinting calculations, the favorable fragments of the top 10 compounds were identified and are highlighted in light red in the molecular structure diagrams in Table 7. In the top 10 compounds and the control compound i472, the favorable fragment GF4 containing an ester bond, a carbon–carbon double-bond structure, and the nitrogen-containing five-membered ring structure GF7 was observed, whereas for our three selected alternative compounds C1–C3, the favorable fragment 7 (GF7) was observed in the C1 and C3 structures. No predicted favorable fragment was observed in compound C2. Considering that these three compounds have the highest docking scores and target-binding ability, GF7, a nitrogen-containing five-membered ring structure can be assumed to be of some significance for the identification of ALOX15 inhibitors. However, when the hydrogen atom that linked to the nitrogen atom on the five-membered ring of the favorable fragment 7 was replaced by other groups (bad fragment 10), the activity was adversely affected. This result can provide guidelines for the conformational studies of ALOX15 inhibitors.

## 2.6. ADMET Property Prediction

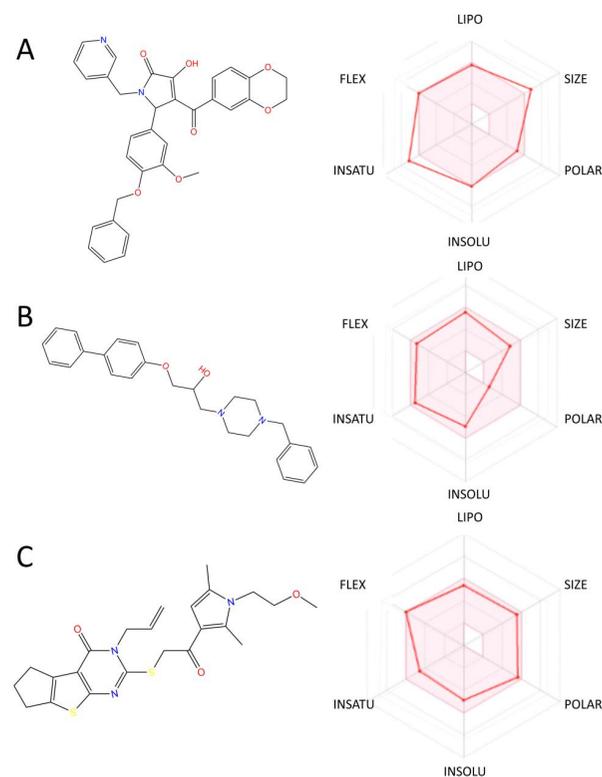
SwissADME was used to predict the absorption, in vivo distribution, metabolism, excretion, and toxicity properties of the three compounds that were identified as ALOX15 targets. First, the in vivo availability of these compounds, including the water solubility, lipid solubility and BBB penetration, gastrointestinal absorption (GI absorption), and inhibition of P-gp drug metabolism proteins, was investigated. The corresponding property indices for the compounds tested are presented in Figure 9 and Table 8.

The LogSw value represents the water solubility coefficient of the drug. The lower this value than  $-4.0$ , the less soluble the compound is in water; LogSw  $< -6.0$  reveals that the compound is barely soluble in water, as detailed in Table 8. The predicted LogSw value for C1 was less than  $-6.0$ , whereas the remaining two molecules were poorly soluble in water. This result could be related to the molecular structure containing fewer ionizable groups. Furthermore, the compounds formed fewer hydrogen bonding interactions with the target protein and more Pi-polar interactions, which contributed to their low water solubility.

The LogP value represents the oil–water partition coefficient of a substance and is the logarithmic value of the ratio of the partition coefficients of a substance in n-octanol and water. Thus, a high LogP value indicates strong lipophilicity of the compound, whereas a low LogP value indicates weak hydrophilicity of the compound. C1–C3 showed similar lipophilicity levels, with consensus Log Po/w values of 3.66, 3.74, and 4.22, respectively, indicating that these alternative compounds have a partition ratio close to 4:1 in the organic and aqueous phases and are soluble in organic oil solvents. This result is consistent with their predicted low water solubility.

**Table 8.** ADME properties of candidate compounds C1–C3.

Molecule	Molecular Formula	Molecule Weight (g/mol)	Log $P_{o/w}$ (iLOGP)	Log S (ESOL)	Solubility	BBB Permeant	Number of Hydrogen Bond Acceptor	Number of Hydrogen Bond Donor	Number of Rotatable Bond	Bioavailability Score	GI Absorption	P-gp Substrate
C1	C33H28N2O7	564.58	3.25	−6.05	Poorly soluble	No	8	1	9	0.56	High	Yes
C2	C23H27N3O3S2	457.61	4.28	−4.87	Moderately soluble	No	4	0	9	0.55	High	No
C3	C26H30N2O2	402.53	4.06	−4.92	Moderately soluble	Yes	4	1	8	0.55	High	Yes

**Figure 9.** Distribution of the hexagonal range of ADME properties for candidates C1–C3: (A) distribution of ADME properties for compound C1; (B) distribution of ADME properties for compound C2; (C) distribution of ADME properties for compound C3.

P-glycoprotein 1, also known as permeable glycoprotein (P-gp), pumps foreign drugs out of cells in an ATP-dependent manner with broad substrate specificity [29]. When a drug is available as a substrate for P-gp, its *in vivo* availability is reduced. Compounds C1 and C3 are predicted to be substrates for P-gp; however, their predicted bioavailability values are not reduced compared with that of C2, which we assume to be related to the low water solubility of these compounds. The blood–brain barrier (BBB) permeability of small molecules depends on their lipid solubility, molecular complexity, polar surface area, and whether they are P-gp substrates. Crossing the BBB is inherently difficult for hydrophilic large-molecule drugs, whereas some lipophilic suitable molecular weight drugs can cross the BBB but are easily transported by efflux pumps, such as glycoproteins (Pgp), on the BBB. The predicted results revealed that only compound C3 had BBB permeability.

The Lipinski, Ghose and Veber, Egar, and Muegge rules were used in the module for the evaluation of molecules. The Lipinski rules for the evaluation of druggability are widely used [30] and include five main principles that should be satisfied if a compound exhibits formulation properties: a molecular weight of less than 500, hydrogen bond donor number of less than 5, hydrogen bond acceptor number of less than 10, lipid-water partition coefficient of less than 5, and rotatable bond number that does not exceed 10. As shown in the results displayed in Table 8, all compounds satisfied the drug-formulating rules, which revealed excellent drug properties, with the exception of compound C1, which violated one of the Lipinski rules because its molecular mass was slightly higher than 500.

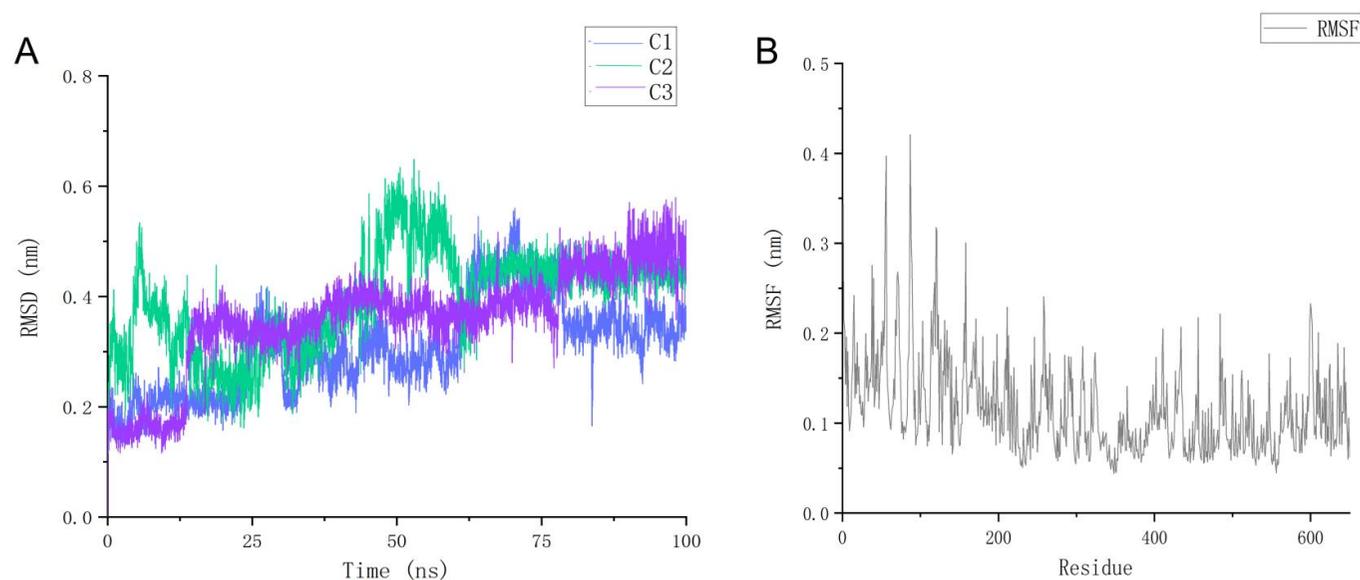
Finally, the potential toxicity of these candidate compounds was assessed using the quantitative structure–toxicity relationship model of the TOPKAT module. The results are presented in Table 9 and all three compounds were predicted to be non-AMES mutagenic. In the FDA-standard mouse model, both alternative compounds were predicted to be non-carcinogenic, with the exception of compound 3aI, which had a single possible carcinogenic toxicity to female mice.

### 2.7. Molecular Dynamics Simulation

To confirm whether the binding of the candidate compounds to the target is sustained and stable, we performed molecular dynamics simulations for 100 ns on three ligand–receptor complexes and obtained the RMSD values of the ligands and RMSF data for the protein residues. As displayed in Figure 10A, all three candidate ligands exhibited stable conformational RMSD fluctuations over the 100 ns duration. C1 exhibited a slight increase in RMSD values throughout the 100 ns period but remained generally flat. C2 showed non-significant differences in RMSD values at the beginning and end of the MD, but clear conformational peaks were observed at the 4–12 and 47–53 ns time periods. Subsequently, the RMSD values rapidly returned to their initial levels without large fluctuations, which still indicates that its stable ALOX15 binding properties mimic the *in vivo* solvent environment. For compound C3, its RMSD values maintained a steady upward trend throughout the MD process but did not exhibit significant fluctuations. By the end of 100 ns, the final RMSD values for the three candidate compounds were similar (C1: 0.382 nm; C2: 0.426 nm; C3: 0.424 nm), which indicated that all three complex systems reached a similar steady state within a short time.

**Table 9.** Predicted TOPKAT teratogenicity, in vivo carcinogenicity in mice, and hepatotoxicity of candidate compounds.

Molecule	Ames_Prediction	Ames_Probability	Ames_TOPKAT Score	Carcinogen_Prediction (Male/Female Mouse)	Carcinogen_Probability	Carcinogen_TOPKAT Score	Hepatotoxic_Prediction	Predicted Hepatotoxic Value
C1	Non-Mutagen	0.0764858	−18.3544	Non-Carcinogen/Non-Carcinogen	0.130807	−11.1785	false	−5.65242
C2	Non-Mutagen	0.655966	−3.35449	Non-Carcinogen/Single Carcinogen	0.211263	−4.29546	true	−2.59417
C3	Non-Mutagen	0.402773	−9.6271	Non-Carcinogen/Non-Carcinogen	0.18041	−6.31686	false	−11.614

**Figure 10.** Results of molecular dynamics calculations performed on the candidate compound–protein receptor complexes: (A) the changes in conformational RMSD values for the three candidate ligands during the 100 ns kinetic simulation; (B) the changes in conformational RMSF values for the receptor protein residues over 100 ns.

The RMSF of a protein residue represents the root mean square displacement of the residue in the protein conformation; that is, its flexibility. As displayed in Figure 10B, the RMSF values for the proteins ranged from 0.03 to 0.44 nm. The front approximately 220 residues exhibited high RMSF values, which indicated their high flexibility in salt solutions. Residues further back in the sequence (which also host the protein active site) exhibited low RMSF values throughout because of their more stable binding to the ligand. For the key active binding residues of the target, Arg402, Leu407, and Leu596 (represented here by the first three favorably acting ranked residues in the molecular docking results section), their RMSF values of 0.08, 0.06, and 0.06 nm were all low, indicating that they form stable interactions with the candidate compound molecules. This result confirmed the previous molecular docking results. The RMSD and RMSF data revealed that the three candidate compounds exhibited stable binding properties to their targets during the 100 ns aqueous environmental kinetic simulations.

### 3. Discussion

Ferroptosis is a critical cell-regulated death pathway that causes irreversible damage to body tissues. The peroxidation of unsaturated lipids (PUFAs) is a key mechanism for the induction of cellular ferroptosis. ALOX15 is a non-hemoglobin-dependent dioxygenase that directly oxidizes PUFAs in biological membranes and leads to the accumulation of a series of intracellular peroxidized lipids and ferroptosis [14]. Thus, ALOX15 has become a popular target in the study of many diseases, including tumors and degenerative diseases. The role played by ALOX15 in programmed cell death and abnormal metabolism has long been recognized. The role of ALOX family enzymes in ferroptosis was first reported in the study by Yang et al. in 2016 [11], which promoted research interest in ALOX15 enzyme inhibitors. During 2016–2019, numerous ALOX15 inhibitors were reported. The most representative model was the compound i472, designed and synthesized by Guo et al. based on the structure–activity relationship study approach [28], which was protective against macrophage death induced by external toxic factors. Golovanov et al. in February 2022 revealed that a combination of conventional chemical design–synthesis techniques and molecular dynamics simulations were applied to confirm the ALOX15 inhibition capacity of the N-substituted 5-(1 H-Indol-2-yl)-2-methoxyanilines family of compounds [17]. In March of the same year, Aghasizadeh et al. proved the excellent inhibitory effect of 8-Geranyloxycarbostyryl on ALOX15 in human PCa cells through cellular assays and in vitro mouse models [31]. Although many ALOX15 inhibitory compounds have been reported for experimental use, the true ALOX15 molecules that can be used for clinical treatment are yet to be discovered, and options and possibilities should be explored. With the rapid development of computer technology, CADD has been widely used to rapidly screen out molecules with excellent target-binding potential from a large number of compounds, which results in significant cost savings for numerous experimental consumables. A large-scale virtual screening based on molecular structure was conducted.

Experimentally extracted proteins can be subjected to enzymatic digestion and cryo-electron microscopy to obtain amino acid sequences and 3D structures that can be used in virtual research processes. However, for many proteins used as targets for drug research, such as ALOX15, limited accessible structure files have been developed for proteins of Homo sapiens origin, and the experimental dataset collected by the BLAST sequence database provides a solution to this problem. Given that homology modeling studies are yet to be conducted for human-derived ALOX15 proteins, this operation was performed using numerous computational tools. The structural models proved to be highly reliable and could provide the basis for subsequent CADD studies on this target.

Machine learning methods have also been increasingly applied in the drug discovery process, such as the differentiation of active and inactive molecules and the prediction of drug-like and non-drug-like compounds. Bayesian algorithms tend to exhibit excellently recognition under dichotomous scenarios. Based on a collection of high/low-activity

ALOX15 inhibitor molecules from previous studies, the Bayesian classifier was constructed to classify molecules from commercial compounds after initial structural screening. The PCA of the molecular descriptors revealed that the dataset used to construct the classifier model was broad and sufficiently dispersed to provide a basis for the reliability of the constructed model. The results of the test set and the 10-fold validation showed that the Bayesian model allows for the excellent discrimination of inactive molecules. However, this also resulted in a high false positive rate, i.e., the compounds we classified as “active” were likely to contain compounds that did not actually have target inhibitory activity. Given this situation, we went a step further and performed a precise docking analysis of the selected compounds to confirm whether the molecules identified by the classification model actually bind to the target and how well they interact with the target. On the other hand, for the ECFP fingerprints used to characterize the structure of the compounds, the fingerprint radius is also likely to have a significant impact on the classification performance of the model, as it relates to how large a molecular fragment the classification learning model will use as a training reference. We also present the top 10 most frequent fingerprints of all molecules classified as ‘active’ and ‘inactive’. For the compounds predicted to be ‘active’, structures with unsaturated ether and ester bonds were the most frequent fingerprints. The fragments containing nitrogenous five-membered rings also appear to be critical when considered in conjunction with the structures of the 10 compounds we selected as the best predictors. Overall, although we did not consider the classification efficacy of more machine learning models, the performance of the Bayesian classifier on the active training set was adequate for our screening.

In a precise molecular docking analysis, the CDOCKER ENERGY and CDOCKER INTERACTION ENERGY values were calculated, i.e., both the internal energy of the ligand molecule and the external target-binding energy were taken into account. The final three most promising compounds identified were very similar to the positive control molecule and to the key binding residues of ALOX15 in terms of the type of target-binding action. This highlights their potential as target inhibitors. Molecular dynamics studies of the ligand–target complexes confirmed the results in terms of fine docking; although the compounds showed “peak-like” conformational fluctuations over several small time intervals, their RMSD values eventually stabilized in similar intervals, confirming the stable interaction between the ligand and protein. Finally, we predicted the drug-likeness of the three candidate compounds C1–C3 using the ADMET prediction program, which showed excellent druggability for all candidates, as well as low teratogenicity and low carcinogenicity in mice. Although C1–C3 showed excellent dummy predictive properties, their true ALOX15 inhibitory activity has to be verified by further experimental data. Overall, these three compounds showed excellent target-binding potential and druggability and could be used as a reference for further development of ALOX15 inhibitors.

## 4. Materials and Methods

### 4.1. Protein Homology Modeling

For protein-based virtual screening processes, a known target protein structure is mandatory. As no existing human-derived ALOX15 protein structures were retrieved on the PDB database, we first searched to obtain the rabbit-derived reticulocyte ALOX15 protein sequence (PDB ID: 2P0M). By searching the nr\_clustered experimental database from the UniProt KB [32] database, we performed a balstp algorithm on the rabbit ALOX15 sequence to obtain the human-derived ALOX15 with protein sequences with high similarity. The search parameter was set to PDB\_nr95, i.e., the templates were searched in an information database with no redundancy at 95% sequence homology. The searched templates were sorted by the E value size, with lower E values indicating the higher reliability of the sequence alignment [33].

Subsequently, the human-derived ALOX15 protein model was constructed using the MODELLER module [34] on the Discovery Studio 4.5 platform. The accuracy of the model was calculated by calculating the model-verified score.

For pre-processing of the homologous model, we used the Prepare Ligand module of the DS platform for the energy and structural optimization of the protein model. The loops structure of the protein was defined using the SEQRES algorithm, while a CHARMM force field [35] was applied to the structure for energy minimization and protonation; the Dielectric constant was set to 10, while the pH for protonation was set to 7.4 in order to keep the protonation acid–base stable.

#### 4.2. Protein Model Validation

The quality and spatial distributions of the residues of the constructed protein models were evaluated using the DS platform's MODELLER and Profile-3D scoring functions and Ramachandran plots. For each model evaluated, an expected high score, expected low score, and verify score were assigned. When the verify score is higher than the expected low score, the reliability of the model is guaranteed; the closer the verify score is to the expected high score, the higher the refinement (quality) of the model. The Ramachandran plot [36] is used to evaluate the spatial distribution of two adjacent peptide unit backbones by the minimum non-bonded atomic contact distance. In general, a reasonable protein model should have no more than 5% of the total number of residues located in the irrational region.

#### 4.3. Structure-Based High-Throughput Molecular Screening

We selected 210,070 molecules from the SPECS commercial compound library and performed structure-based virtual screening based on homologous modelled proteins with the help of the LibDock [37] module of the Discovery Studio platform. The protein models were pre-corrected for structure by the Macromolecules module. The active site residues (Thr412, Arg415, Val420, Thr429, Ile602, and Trp606) were identified by searching the protein structure literature [38] with docking radius sphere 3D coordinates of  $-55.173933$ ,  $169.711977$ , and  $32.549350$ . In the LibDock module, 100 spatial hotspots are randomly generated in a docking radius sphere, and compound conformations are rapidly matched to each hotspot and analyzed for good interactions with protein structures. Molecules contained within the compound library were pre-hydrogenated and charge-balanced by the Prepare Ligand module and all possible docking conformations were generated to ensure diversity of docking poses and accuracy of results.

#### 4.4. Molecular Descriptor Calculation and Principal Component Analysis

Here, 5360 ALOX15 inhibitory compounds from the ChEMBL database were collected from the ExCAPE database (<https://solr.ideaconsult.net/search/excape/>, accessed on 13 August 2022) [39]. The compounds were first subjected to energy minimization and valence bond repair by applying the Prepare Molecule module in Discovery Studio. Subsequently, 55 molecular physicochemical property descriptors were calculated for the ALOX15 inhibitors collected from the ExCAPE database using the Calculate Molecular Properties function. In order to avoid duplicate intersections between the SPECS screening set and the ALOX15 inhibitor molecules, the DS Find Similar function was used to search for possible similar molecules to ensure that there were no duplicate components between the SPECS dataset and the ALOX15 inhibitor set. The module was used to calculate the principal components of these descriptors in order to perform data dimensionality reductions. The Pearson correlation analysis method [40] was used to calculate the activity correlation coefficient for each descriptor, setting the minimum variance explained to 0.9, i.e., excluding descriptors with activity correlation coefficients below 0.1 and retaining key descriptors with coefficients above 0.9. The minimum number of components was 3. The OPS analysis method was applied to the calculation of principal components. The correlation coefficients preceding each descriptor in the PC equation calculated from the principal component analysis reflect the magnitude of their contribution to the principal component. The descriptor with the largest contribution in aggregate was selected as the base property for constructing the machine learning classifier. The ECFP<sub>2</sub> fingerprint was also calculated for the dataset compounds to calculate the Tanimoto distance of the features [41]. Extended connectivity

fingerprinting (ECFP\_n) is a class of 2D ring fingerprint tool based on a variant of Morgan's algorithm that captures atomic information based on daylight atomic invariant rules [42]. It sets radius "n" as the number of iterations to calculate the atomic environment identifier. Compared to molecular physical and chemical property descriptors, ECFP fingerprints are more focused on the chemical backbone information of molecules. Thus, it is possible to provide machine learning with structural reference information for individual training set molecules. The fingerprint features were tracked during the computation and the minimum sample of each variable was set to SqrtEstimate. In the process of scaling the data, while all features being computed were averaged and centered.

#### 4.5. Machine Learning Classifier Construction

The 5360 ExCAPE compounds were pre-flagged as active and inactive (activity flag) based on PXC50 values: compounds with PXC50 values >5 were considered active in the definition given by the database itself, equating to IC50/EC50 dose-response values  $\leq 10 \mu\text{M}$ . The others were inactive (1835 active compounds and 3525 inactive compounds, an active/inactive ratio close to 1:2). The collected active molecule data were converted to the nominal form, normalized by the unsupervised metric filter in WEKA software [43], and saved in the arff format for use in the subsequent classifier construction.

The NB classifier is an efficient machine learning algorithm based on Bayes' theorem and is often used to classify training sets with a large number of samples because of its immunity to random noise during training [44].  $P(A)$  in Equation (1) represents the prior probability of independent event A occurring;  $P(B)$  is the prior probability of independent event B occurring;  $P(B|A)$  is the posterior probability of event B occurring when event A occurs;  $P(A|B)$  is the posterior probability of event A occurring if independent event B is known to occur. Thus, the Bayesian algorithm is essentially a modification of the prior probability of the occurrence of independent events, which after a suitable number of validation modifications ultimately gives the best binary validation model. In several earlier examples of drug research, Bayesian models have demonstrated excellent discriminatory power in dichotomous cases [45], which is why we decided to choose the NB model.

We used the key molecular descriptors calculated from the principal component analysis to assist in the construction of the model. At the same time, a series of ECFP fingerprints (ECFP\_0, ECFP\_2, ECFP\_4, ECFP\_6, ECFP\_8) were included to ensure that the structural features of the molecules were fully considered in the construction of the classification model. The classifier construction method was then tested using a 10-fold validation method. This method randomly divides the original dataset into 10 equally sized subsets, one of which is selected in turn as the validation set, until each subset is individually selected for validation. The 10-fold validation method has proven to be the most effective algorithm testing procedure in a large number of dataset testing examples [46]. The true positive rate, false positive rate, area under the ROC curve, precision, recall, F-measure, and Matthews correlation coefficient (MCC) values were used as metrics to evaluate the model.

The ROC curve (receiver operating characteristic curve) is often used to characterize the ability of the model to discriminate between positive and negative samples. It takes the true positive rate (TPR, also sensitivity) as the Y-axis and the false positive rate (FPR, also 1-specificity) as the X-axis, and gives a coordinate point representing the optimal classification threshold in the coordinate species; the smaller the distance between this coordinate point and the top left corner of the curve axis, the closer the area under the curve (AUC) value will be to 1, and the better the model's TPR (sensitivity) and specificity, which are two different dimensions of the ROC curve used to characterize the recognition ability of the model. The latter represents the ratio of inactive compounds correctly classified by the model for all actual inactive compounds, representing the classifier's ability to identify negative examples.

The recall value, also known as the sensitivity, represents the percentage of true positives that are correctly classified (Equation (2)), while the specificity value, in contrast

to the recall value, represents the percentage of true negatives that are correctly classified (Equation (3)). The F-measure, also known as the F-score, is the weighted summed average of the precision and recall values (Equation (5)). When the equation parameter is set to 1 ( $\alpha = 1$ ), the resulting F-measure is also known as the F-1 score, which gives the same weight to the precision and recall values and is often used to evaluate the strength of a model's classification; higher F-1 values tend to indicate higher model validity.

The MCC (Matthews correlation coefficient) is a Pearson product moment correlation coefficient (Equation (6)) based on a weighting matrix that takes into account the true positive, true negative, false positive, and false negative rates of the classification, and is often used as a measure of the classification performance of a dichotomous model. The metric has a wide range of applicability and is also good for assessing samples with uneven distribution [47]. The value range is  $[-1, 1]$ , with values closer to 1 indicating that the classifier is more accurate, values of 0 indicating that the classifier's predicted results are worse than those predicted by random classification, and values closer to  $-1$  suggesting that the predicted classification results deviate more from the actual classification.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (1)$$

$$\text{Sensitivity} = \text{Recall} = \frac{TP}{(TP + FN)} \quad (2)$$

$$\text{Specificity} = \frac{TN}{(FP + TN)} \quad (3)$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (4)$$

$$F - \text{Measure} = \frac{(1 + \alpha^2) * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (5)$$

$$\text{MCC} = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (6)$$

#### 4.6. High Accuracy Molecular Docking

To further explore the interactions between molecules selected by the high-throughput structure screen and ML classifier and target residues, we applied the CDOCKER module [48] of the DS platform to perform a docking analysis of alternative inhibitor molecules. Similar to the high-throughput structural screen we ran previously, the protein models obtained from the homology modelling were used as docking targets. During the fine docking, the CHARMM force field was applied and the program generated 10 random body images for each docked molecule through short time duration (1000 steps) molecular dynamics simulations. The maximum temperature of the kinetic process was 1000 K. During this process, the electrostatic interactions formed between the small molecule and the target were also taken into account. Meanwhile, the orientation for refining was set to 10 to ensure the diversity of the docking conformations, while the maximum number of bad orientations was set to 800 and the Momary–Rone method was used as the ligand partial charge method. In the final step of the ligand minimization process, full potential was supplied, but no gradient tolerance of minimization was allowed. The grid extension was set to 8.0.

#### 4.7. ADMET Property Prediction

In addition to the ability of the small molecules to bind to their targets, stable in vivo metabolic properties and low compound toxicity are also important indicators of a compound's potential to become a drug. Therefore, we need to further determine the ADMET (i.e.,) properties of compounds in order to exclude molecules with poor drug-forming potential from the alternative compounds. Conventional methods for the evaluation of

ADMET properties are performed with the aid of cellular or animal experiments, but such methods do not meet efficiency and cost optimization considerations for the large number of compounds to be tested [49]. In the present study, we used the SwissADME (<http://www.swissadme.ch/>, accessed on 10 September 2022) [50] online tool developed by Antoine et al. to predict the ADMET properties of the compounds to be determined. The website predicts the drug-generating properties by importing compound smile files, including the physicochemical parameters, pharmacokinetic characteristics, and drug-likeness. Compounds identified after the previous steps are generated using the Openbabel software with the corresponding smile files and imported into SwissADME to generate the drug-generating properties of the compounds. Additionally, the teratogenicity, carcinogenicity, and hepatotoxicity of the alternative compounds were predicted using the TOPKAT toxicity prediction module [51] from the DS platform.

#### 4.8. Molecular Kinetic Simulations

A molecular dynamics simulation (MD) is a molecular simulation technique based on Newtonian mechanics theory that has been applied to the monitoring of thermodynamic reactions within a single system. MD methods are widely used in computer-assisted drug development studies to examine whether a pre-drug compound can bind stably to a specified target. To examine the target-binding ability of the three candidate molecules identified by the above process, we ran a 100 ns aqueous solution environmental kinetic simulation based on the ligand–receptor complexes previously obtained via accurate docking using the GROMACS 2019.1 (sourced by Mark Abraham et al. from Uppsala University, Stockholm University and the Royal Institute of Technology, Sweden) software package [52]. First, the complex system was constructed. The PDB format files of the ALOX15 homologous protein model and candidate ligand molecules were obtained via transformation using the Openbabel tool [53]. The AMBER99SB-ILDN force field was used to construct the topological system of the protein. For the ligands, we submitted a GAFF force field-based topology scheme from the ACPYPE online server [54] (<https://www.bio2byte.be/acpype/>, access on 19 September 2022) to obtain the force field parameters of the compounds. A cubic box with a radius of 1.2 nm was constructed and centered on the integrated resulting complex system, to which the SPC216 water model was added to simulate an aqueous solution environment. To ensure the total charge neutrality of the simulated system, corresponding amounts of sodium and chloride ions were added to the system to replace the water molecules. Periodic boundary conditions (PBC) were applied to the three directions of the spatial coordinates of the complex system.

The energy minimization of the system was carried out in 50,000 steps at a simulated temperature of 300 K, provided that the whole complex system had been prepared. After correction for positional constraints, equilibrium for the receptor, ligand, and solvent is achieved through a constant temperature and volume (NVT) and a constant temperature and pressure (NPT) conditioning process, respectively. Finally, MD simulations of 100 ns in duration were run.

## 5. Conclusions

In this study, we identified three compounds with excellent ALOX15 target-binding potential through computer-assisted screening processes. In the screening phase, a machine learning classifier constructed based on the Bayes method was combined with a conventional structure-based, high-throughput screening approach to accurately select suitable molecules. To validate the drug-forming potential of these selected molecules, a fine-grained docking analysis was applied to the process and the three compounds had a target-binding pattern similar to that of the positive control molecule. The stability of the binding was also confirmed by MD simulations. The ADMET property predictions indicated that the drug-forming potential of the compounds was desirable. The three newly identified ALOX15 inhibitory potential compounds in this study exhibit the potential to

inhibit the development of cellular ferroptosis and provide a reference for the treatment of tissue-damaging diseases.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/xxx/s1>, Table S1: Top 10 protein names, source organism, PDB/BLAST accession, sequence length and E-Value evaluation values obtained from BLASTp search; Table S2: 55 molecular descriptors computed for machine learning model building; Table S3: Formulae for the composition of the top 3 key principal components; Table S4: Predicted target inhibitory activity of the Bayesian classifier for the 450 molecules with top Libdock scores.

**Author Contributions:** Funding acquisition, L.L.; investigation, Y.L.; methodology, Y.L. and P.C.; project administration, L.L.; software, Y.L.; supervision, L.L.; validation, P.C.; visualization, Y.L.; writing—original draft, Y.L. and L.L.; writing—review and editing, P.C. and L.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Key Discipline Construction Project of Guangdong Medical University (4SG22004G), the Science and technology program of Guangdong Province (2019B090905011), the National Natural Science Foundation of China (grant no. 81903901), and the Dawning Program of Wuhan Knowledge Innovation Special Project (grant no. 2022020801020467).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data is contained within the article and Supplementary Material.

**Acknowledgments:** We thank the Public Service Platform of South China Sea for R&D Marine Biomedicine Resources for their support.

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

## References

1. Dixon, S.J.; Lemberg, K.M.; Lamprecht, M.R.; Skouta, R.; Zaitsev, E.M.; Gleason, C.E.; Patel, D.N.; Bauer, A.J.; Cantley, A.M.; Yang, W.S.; et al. Ferroptosis: An iron-dependent form of nonapoptotic cell death. *Cell* **2012**, *149*, 1060–1072. [[CrossRef](#)]
2. Basit, F.; van Oppen, L.M.; Schockel, L.; Bossenbroek, H.M.; van Emst-de Vries, S.E.; Hermeling, J.C.; Grefte, S.; Kopitz, C.; Heroult, M.; Hgm Willems, P.; et al. Mitochondrial complex I inhibition triggers a mitophagy-dependent ROS increase leading to necroptosis and ferroptosis in melanoma cells. *Cell Death Dis.* **2017**, *8*, e2716. [[CrossRef](#)]
3. Park, E.; Chung, S.W. ROS-mediated autophagy increases intracellular iron levels and ferroptosis by ferritin and transferrin receptor regulation. *Cell Death Dis.* **2019**, *10*, 822. [[CrossRef](#)]
4. Sui, X.; Zhang, R.; Liu, S.; Duan, T.; Zhai, L.; Zhang, M.; Han, X.; Xiang, Y.; Huang, X.; Lin, H.; et al. RSL3 Drives Ferroptosis Through GPX4 Inactivation and ROS Production in Colorectal Cancer. *Front. Pharmacol.* **2018**, *9*, 1371. [[CrossRef](#)]
5. Xu, G.; Wang, H.; Li, X.; Huang, R.; Luo, L. Recent progress on targeting ferroptosis for cancer therapy. *Biochem. Pharmacol.* **2021**, *190*, 114584. [[CrossRef](#)]
6. Sun, Y.; Chen, P.; Zhai, B.; Zhang, M.; Xiang, Y.; Fang, J.; Xu, S.; Gao, Y.; Chen, X.; Sui, X.; et al. The emerging role of ferroptosis in inflammation. *Biomed. Pharmacol.* **2020**, *127*, 110108. [[CrossRef](#)]
7. Mou, Y.; Wang, J.; Wu, J.; He, D.; Zhang, C.; Duan, C.; Li, B. Ferroptosis, a new form of cell death: Opportunities and challenges in cancer. *J. Hematol. Oncol.* **2019**, *12*, 34. [[CrossRef](#)]
8. Li, D.; Li, Y. The interaction between ferroptosis and lipid metabolism in cancer. *Signal Transduct. Target. Ther.* **2020**, *5*, 108. [[CrossRef](#)]
9. Jiang, X.; Stockwell, B.R.; Conrad, M. Ferroptosis: Mechanisms, biology and role in disease. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 266–282. [[CrossRef](#)]
10. Qiu, Y.; Cao, Y.; Cao, W.; Jia, Y.; Lu, N. The Application of Ferroptosis in Diseases. *Pharmacol. Res.* **2020**, *159*, 104919. [[CrossRef](#)]
11. Yang, W.S.; Kim, K.J.; Gaschler, M.M.; Patel, M.; Shchepinov, M.S.; Stockwell, B.R. Peroxidation of polyunsaturated fatty acids by lipoxygenases drives ferroptosis. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E4966–E4975. [[CrossRef](#)]
12. Coffey, R.; Nam, H.; Knutson, M.D. Microarray analysis of rat pancreas reveals altered expression of Alox15 and regenerating islet-derived genes in response to iron deficiency and overload. *PLoS ONE* **2014**, *9*, e86019. [[CrossRef](#)]
13. Shintoku, R.; Takigawa, Y.; Yamada, K.; Kubota, C.; Yoshimoto, Y.; Takeuchi, T.; Koshiishi, I.; Torii, S. Lipoxygenase-mediated generation of lipid peroxides enhances ferroptosis induced by erastin and RSL3. *Cancer Sci.* **2017**, *108*, 2187–2194. [[CrossRef](#)]
14. Ma, X.H.; Liu, J.H.; Liu, C.Y.; Sun, W.Y.; Duan, W.J.; Wang, G.; Kurihara, H.; He, R.R.; Li, Y.F.; Chen, Y.; et al. ALOX15-launched PUFA-phospholipids peroxidation increases the susceptibility of ferroptosis in ischemia-induced myocardial damage. *Signal Transduct. Target. Ther.* **2022**, *7*, 288. [[CrossRef](#)]

15. Gao, S.; Zhou, L.A.-O.; Lu, J.; Fang, Y.; Wu, H.; Xu, W.; Pan, Y.; Wang, J.; Wang, X.; Zhang, J.A.-O.; et al. Cepharanthine Attenuates Early Brain Injury after Subarachnoid Hemorrhage in Mice via Inhibiting 15-Lipoxygenase-1-Mediated Microglia and Endothelial Cell Ferroptosis. *Oxid. Med. Cell Longev.* **2022**, *2022*, 4295208. [[CrossRef](#)]
16. Zhuravlev, A.; Golovanov, A.; Toporkov, V.; Kuhn, H.; Ivanov, I. Functionalized Homologues and Positional Isomers of Rabbit 15-Lipoxygenase RS75091 Inhibitor. *Med. Chem.* **2022**, *18*, 406–416. [[CrossRef](#)]
17. Golovanov, A.; Zhuravlev, A.; Cruz, A.; Aksenov, V.; Shafiullina, R.; Kakularam, K.R.; Lluch, J.M.; Kuhn, H.; Gonzalez-Lafont, A.; Ivanov, I. N-Substituted 5-(1H-Indol-2-yl)-2-methoxyanilines Are Allosteric Inhibitors of the Linoleate Oxygenase Activity of Selected Mammalian ALOX15 Orthologs: Mechanism of Action. *J. Med. Chem.* **2022**, *65*, 1979–1995. [[CrossRef](#)]
18. Hosseinymehr, M.; Matin, M.M.; Sadeghian, H.; Bahrami, A.R.; Kaseb-Mojaver, N. 8-Farnesyloxycoumarin induces apoptosis in PC-3 prostate cancer cells by inhibition of 15-lipoxygenase-1 enzymatic activity. *Anti-Cancer Drugs* **2016**, *27*, 854–862. [[CrossRef](#)]
19. Gomeni, R.; Bani, M.; D’Angeli, C.; Corsi, M.; Bye, A. Computer-assisted drug development (CADD): An emerging technology for designing first-time-in-man and proof-of-concept studies from preclinical experiments. *Eur. J. Pharm. Sci.* **2001**, *13*, 261–270. [[CrossRef](#)]
20. Li, J.; Fu, A.; Zhang, L. An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking. *Interdiscip. Sci.* **2019**, *11*, 320–328. [[CrossRef](#)]
21. Usha, T.; Shanmugarajan, D.; Goyal, A.K.; Kumar, C.S.; Middha, S.K. Recent Updates on Computer-aided Drug Discovery: Time for a Paradigm Shift. *Curr. Top. Med. Chem.* **2017**, *17*, 3296–3307. [[CrossRef](#)]
22. Zhao, L.; Ciallella, H.L.; Aleksunes, L.M.; Zhu, H. Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling. *Drug Discov. Today* **2020**, *25*, 1624–1638. [[CrossRef](#)]
23. Sun, H.; Pan, P.; Tian, S.; Xu, L.; Kong, X.; Li, Y.; Dan, L.; Hou, T. Constructing and Validating High-Performance MIEC-SVM Models in Virtual Screening for Kinases: A Better Way for Actives Discovery. *Sci. Rep.* **2016**, *6*, 24817. [[CrossRef](#)]
24. Kapsiani, S.; Howlin, B.J. Random forest classification for predicting lifespan-extending chemical compounds. *Sci. Rep.* **2021**, *11*, 13812. [[CrossRef](#)]
25. Dalecki, A.G.; Zorn, K.M.; Clark, A.M.; Ekins, S.; Narmore, W.T.; Tower, N.; Rasmussen, L.; Bostwick, R.; Kutsch, O.; Wolschen-dorf, F. High-throughput screening and Bayesian machine learning for copper-dependent inhibitors of *Staphylococcus aureus*. *Metallomics* **2019**, *11*, 696–706. [[CrossRef](#)]
26. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **2019**, *18*, 463–477. [[CrossRef](#)]
27. Cruz, A.; Di Venere, A.; Mei, G.; Zhuravlev, A.; Golovanov, A.; Stehling, S.; Heydeck, D.; Lluch, J.M.; Gonzalez-Lafont, A.; Kuhn, H.; et al. A role of Gln596 in fine-tuning mammalian ALOX15 specificity, protein stability and allosteric properties. *Biochim. Biophys. Acta Mol. Cell. Biol. Lipids* **2020**, *1865*, 158680. [[CrossRef](#)]
28. Guo, H.; Verhoek, I.C.; Prins, G.G.H.; van der Vlag, R.; van der Wouden, P.E.; van Merkerk, R.; Quax, W.J.; Olinga, P.; Hirsch, A.K.H.; Dekker, F.J. Novel 15-Lipoxygenase-1 Inhibitor Protects Macrophages from Lipopolysaccharide-Induced Cytotoxicity. *J. Med. Chem.* **2019**, *62*, 4624–4637. [[CrossRef](#)]
29. Rathod, S.; Desai, H.; Patil, R.; Sarolia, J. Non-ionic Surfactants as a P-Glycoprotein(P-gp) Efflux Inhibitor for Optimal Drug Delivery-A Concise Outlook. *AAPS PharmSciTech* **2022**, *23*, 55. [[CrossRef](#)]
30. Chagas, C.M.; Moss, S.; Alisarai, L. Drug metabolites and their effects on the development of adverse reactions: Revisiting Lipinski’s Rule of Five. *Int. J. Pharm.* **2018**, *549*, 133–149. [[CrossRef](#)]
31. Aghasizadeh, M.; Moghaddam, T.; Bahrami, A.R.; Sadeghian, H.; Alavi, S.J.; Matin, M.M. 8-Geranyloxycarbostyryl as a potent 15-LOX-1 inhibitor showed great anti-tumor effects against prostate cancer. *Life Sci.* **2022**, *293*, 120272. [[CrossRef](#)] [[PubMed](#)]
32. Wang, Y.; Wang, Q.; Huang, H.; Huang, W.; Chen, Y.; McGarvey, P.B.; Wu, C.H.; Arighi, C.N.; UniProt, C. A crowdsourcing open platform for literature curation in UniProt. *PLoS Biol.* **2021**, *19*, e3001464. [[CrossRef](#)] [[PubMed](#)]
33. Hulsen, T.; de Vlieg, J.; Leunissen, J.A.; Groenen, P.M. Testing statistical significance scores of sequence comparison methods with structure similarity. *BMC Bioinform.* **2006**, *7*, 444. [[CrossRef](#)]
34. Janson, G.; Grottesi, A.; Pietrosanto, M.; Ausiello, G.; Guarguaglini, G.; Paiardini, A. Revisiting the “satisfaction of spatial restraints” approach of MODELLER for protein homology modeling. *PLoS Comput. Biol.* **2019**, *15*, e1007219. [[CrossRef](#)] [[PubMed](#)]
35. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690. [[CrossRef](#)]
36. Ho, B.K.; Thomas, A.; Brasseur, R. Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the alpha-helix. *Protein Sci.* **2003**, *12*, 2508–2522. [[CrossRef](#)]
37. Rao, S.N.; Head, M.S.; Kulkarni, A.; LaLonde, J.M.; LaLonde, J.M. Validation studies of the site-directed docking program LibDock. *J. Chem. Inf. Model* **2007**, *47*, 2159–2171. [[CrossRef](#)]
38. Ivanov, I.; Cruz, A.; Zhuravlev, A.; Di Venere, A.; Nicolai, E.; Stehling, S.; Lluch, J.M.; González-Lafont, À.; Kuhn, H. Conformational Heterogeneity and Cooperative Effects of Mammalian ALOX15. *Int. J. Mol. Sci.* **2021**, *22*, 3285. [[CrossRef](#)]
39. Sun, J.; Jeliaskova, N.; Chupakin, V.; Golib-Dzib, J.F.; Engkvist, O.; Carlsson, L.; Wegner, J.; Ceulemans, H.; Georgiev, I.; Jeliaskov, V.; et al. ExCAPE-DB: An integrated large scale dataset facilitating Big Data analysis in chemogenomics. *J. Cheminformatics* **2017**, *9*, 17. [[CrossRef](#)]

40. Adler, J.; Parmryd, I. Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytom. A* **2010**, *77*, 733–742. [[CrossRef](#)]
41. Vogt, M.; Bajorath, J. Modeling Tanimoto Similarity Value Distributions and Predicting Search Results. *Mol. Inf.* **2017**, *36*, 1600131. [[CrossRef](#)] [[PubMed](#)]
42. Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model* **2010**, *50*, 742–754. [[CrossRef](#)] [[PubMed](#)]
43. Smith, T.C.; Frank, E. Introducing Machine Learning Concepts with WEKA. *Methods Mol. Biol.* **2016**, *1418*, 353–378. [[CrossRef](#)] [[PubMed](#)]
44. Hemmer, P.; Tauber, S.; Steyvers, M. Moving beyond qualitative evaluations of Bayesian models of cognition. *Psychon. Bull. Rev.* **2015**, *22*, 614–628. [[CrossRef](#)] [[PubMed](#)]
45. Kidwell, K.M.; Roychoudhury, S.; Wendelberger, B.; Scott, J.; Moroz, T.; Yin, S.; Majumder, M.; Zhong, J.; Huml, R.A.; Miller, V. Application of Bayesian methods to accelerate rare disease drug development: Scopes and hurdles. *Orphanet. J. Rare Dis.* **2022**, *17*, 186. [[CrossRef](#)]
46. Krstajic, D.; Buturovic, L.J.; Leahy, D.E.; Thomas, S. Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminformatics* **2014**, *6*, 10. [[CrossRef](#)]
47. 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](#)]
48. Wu, Y.; Brooks, C.L., III. Flexible CDOCKER: Hybrid Searching Algorithm and Scoring Function with Side Chain Conformational Entropy. *J. Chem. Inf. Model.* **2021**, *61*, 5535–5549. [[CrossRef](#)]
49. Ferreira, L.L.G.; Andricopulo, A.D. ADMET modeling approaches in drug discovery. *Drug Discov. Today* **2019**, *24*, 1157–1165. [[CrossRef](#)]
50. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717. [[CrossRef](#)]
51. Venkatapathy, R.; Moudgal, C.J.; Bruce, R.M. Assessment of the oral rat chronic lowest observed adverse effect level model in TOPKAT, a QSAR software package for toxicity prediction. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1623–1629. [[CrossRef](#)] [[PubMed](#)]
52. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [[CrossRef](#)] [[PubMed](#)]
53. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminformatics* **2011**, *3*, 33. [[CrossRef](#)] [[PubMed](#)]
54. Sousa da Silva, A.W.; Vranken, W.F. ACPYPE-AnteChamber Python Parser interface. *BMC Res. Notes* **2012**, *5*, 367. [[CrossRef](#)] [[PubMed](#)]