

# Computational Analysis of LOX1 Inhibition Identifies Descriptors Responsible for Binding Selectivity

**Gousiadou, Chrysoula; Kouskoumvekaki, Eirini**

*Published in:*
ACS Omega

*Link to article, DOI:*
10.1021/acsomega.7b01622

*Publication date:*
2018

*Document Version*
Publisher's PDF, also known as Version of record

ACS OMEGA

Article

Cite This: ACS Omega 2018, 3, 2261−2272

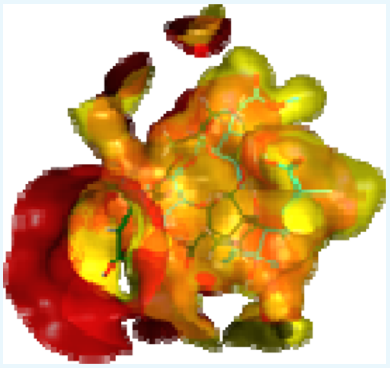
# Computational Analysis of LOX1 Inhibition Identifies Descriptors Responsible for Binding Selectivity

Chrysoula Gousiadou* and Irene Kouskoumvekaki

Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Lyngby, Denmark

Supporting Information

**ABSTRACT:** Lipoxygenases are a family of cytosolic, peripheral membrane enzymes, which catalyze the hydroperoxidation of polyunsaturated fatty acids and are implicated in the pathogenesis of major human diseases. Over the years, a substantial number of scientific reports have introduced inhibitors active against one or another subtype of the enzyme, but the selectivity issue has proved to be a major challenge for drug design. In the present work, we assembled a dataset of 317 structurally diverse molecules hitherto reported as active against 15S-LOX1, 12S-LOX1, and 15S-LOX2 and identified, using supervised machine learning, a set of structural descriptors responsible for the binding selectivity toward the enzyme 15S-LOX1. We subsequently incorporated these descriptors in the training of QSAR models for LOX1 activity and selectivity. The best performing classifiers are two stacked models that include an ensemble of support vector machine, random forest, and k-nearest neighbor algorithms. These models not only can predict LOX1 activity/inactivity but also can discriminate with high accuracy between molecules that exhibit selective activity toward either one of the isozymes 15S-LOX1 and 12S-LOX1.



## 1. INTRODUCTION

Human lipoxygenases are a structurally related family of cytosolic, peripheral membrane enzymes, which catalyze the hydroperoxidation of polyunsaturated fatty acids producing leukotrienes, lipoxins, and/or hydroxy fatty acids (arachidonic acid cascade).[1−4] These products play important roles in the development of inflammation, and over the years, an accumulating number of scientific reports emphatically involves LOXs in the pathogenesis of almost all the diseases with major health relevance (bronchial asthma, atherosclerosis, cancer, obesity, osteoporosis, and neurodegenerative disorders).[5−13] As a result, lipoxygenase (LOX) research is a vital scientific area today with more than 500 new articles published annually.[2] Corresponding to the genes of the human ortholog, LOXs are named ALOX15, ALOX15B, ALOX12, ALOX12B, and ALOX5.[1] ALOX12B and ALOX15B are mainly expressed in the skin and other epithelial cells, whereas ALOX15, ALOX12, and ALOX5 are expressed in hematopoietic/immune cells.[13] LOX enzymes have considerable molecular mass (75−81 kDa) and share highly conserved structural features, as well as the unique topology of the catalytic (C-terminal) domain. The C-terminal domain contains both the catalytically active nonheme iron and the substrate-binding cavity.[14] Studies of various complexes with different inhibitors have found the latter in this location.[15−21] The natural substrate for human LOXs is arachidonic acid.[14,22] With respect to their stereo and positional specificity of arachidonic acid oxygenation, the conventional nomenclature classifies human LOXs as 5S-LOX1 (ALOX5), 12R-LOX1 (ALOX12B), 12S-LOX1 (ALOX12), 15S-LOX1 (ALOX15), and 15S-LOX2 (ALOX15B).[14,22] Currently, research is focused on the biological relevance of the different LOX-isoforms.[23] Functional and isoform multiplicity as well as heterogeneity of the different isoenzymes are confusing issues.[2,23] As a consequence, the arising "selectivity" theme has made the discovery of LOX inhibitors increasingly challenging for drug design. Structural studies on these proteins reveal that although their three-dimensional (3D) structure is highly conserved, their sequences share low identity and show both similarities and striking differences at the active site.[19,24] These differences are held responsible for substrate selectivity and are believed to be the key for the design of isoform-selective inhibitors.[23,24] To date, there is only one FDA-approved drug on the market (Zileuton, 5-LOX1).[25]

In the present investigation, our main focus has been the inhibition of 15S-LOX1. This enzyme is reported to play a crucial role in obesity, since 15S-LOX1 expression is directly related with the proliferation and hypertrophy of adipose cells.[50] Furthermore, 15S-LOX1 promotes cancer by amplifying PPARγ transcription activity, and therefore, 15-LOX1 inhibitors may be suitable chemotherapy agents in the near future.[50] Over the last years, both industrial and academic laboratories have been working intensely toward the discovery of potent and selective 15S-LOX1 inhibitors,[5,29−34] but it is noteworthy that to date, not one of these compounds has been approved for therapeutic usage.[50] Computational methods[26−28] have been employed for the discovery and design of suitable inhibitors,

Received: October 23, 2017
Accepted: February 12, 2018
Published: February 26, 2018






DOI: 10.1021/acsomega.7b01622
ACS Omega 2018, 3, 2261−2272

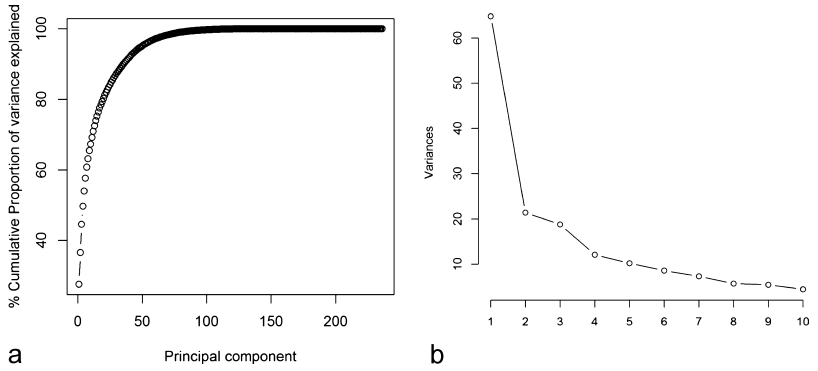



**Figure 1.** (a) Scree plot explaining the % percentage of the cumulative variance in the data. The first five principal components explain 54% of the total variation. (b) Breaks in the data corresponding to components 1−5.

natural[5,29,30] or synthetic[31−34] molecules interacting strongly and specifically with the protein. A relatively small number of data-driven models have been generated, concerning mostly the inhibition of 5*S*-LOX1, using small datasets of compounds with structural similarity.[35−38] These models, albeit statistically valid, have a limited scope and are unlikely to accurately predict chemical structures different from those with which they have been trained.

In this work, we introduce the training of stacked classification models built on a dataset of structurally diverse molecules hitherto reported as active against three LOX subtypes (15*S*-LOX1, 12*S*-LOX1, and 15*S*-LOX2). The chemical structures were retrieved from review papers[50] and reports[51−79] as well as publicly available data sets[80−91] (hundreds to thousands of compounds tested for inhibition on LOX isozymes). The resulting dataset contains a large number of compounds with a wide range of molecular weight (165−597) belonging to different chemical families. Thus, the algorithms were provided with (a) a sufficient size of samples, which is essential for any machine learning analysis, and (b) a large amount of biological activity end points for a wide range of chemical compounds.

Most importantly, the inhibitory activity toward two other human LOX subtypes besides 15-LOX1 (12*S*-LOX1 and 15*S*-LOX2) has been taken into consideration.[84−87] The resulting stacked quantitative structure−activity relationship (QSAR) models have high discriminatory ability and yield statistically valid predictions on the selective interaction of various chemicals with 15*S*-LOX1.

## 2. RESULTS AND DISCUSSION

### 2.1. Data Preprocessing and Variable Selection.

Initially, noninformative descriptors were removed, that is, all variables with missing values and zero variance (zero values for all ligands). This process reduced the number of descriptors to 236. The dataset was split randomly into an explicit training set (75%, 239 molecules) and a validation set (25%, 78 molecules) used to evaluate the performance of the models on unseen data. Each one of the subsets was a balanced representation of both the chemical structures and the classes contained in the initial dataset. An exploratory analysis using unsupervised methods was carried out on the training set to acquire information necessary for deciding the best modeling strategy. As many of the descriptors (134) were highly correlated (>0.75), attempts to include them all in the analysis led to models with poor performance. Ideally, it is desirable to have a reduced set of uncorrelated, nonredundant, and informative descriptors that would allow us to build interpretable prediction models. Therefore, we undertook to reduce the initial number of variables and evaluate the effects on the accuracy of the classifiers.

First, we employed principal component analysis.[96] A preprocessing of the data was performed, and the descriptor variables were mean centered and scaled to unit variance. The scree plot (Figure 1a), where the percentage of the cumulative variance is explained, shows that the first five principal components explained 54% of the total variation in the data. The scatterplot matrix for components 1−5, where the points are colored by class (Figure 2), reveals a poor separation

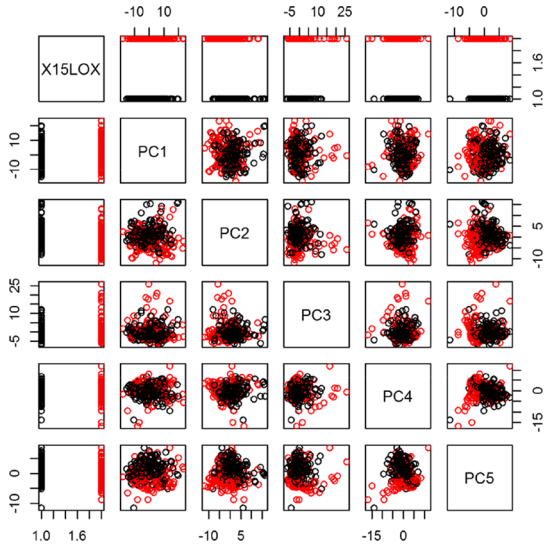



**Figure 2.** Scatterplot matrix by class for principal components 1−5: the plot reveals a poor separation between the classes and reflects the low percentage of variance (54%) explained by these components.

between the classes and reflects the low percentage of variance explained by these components. On the whole, the exploratory analysis revealed that the separation between the two classes would be challenging. This challenge we decided to address with powerful, highly nonlinear models as will be described in the following section (QSAR modeling). By applying a criterion of a minimum value (80%) of cumulative percent of variance accounted for in our data, we selected the first 20 principal components to be used as variables for building a series of machine learning models for classifying the dataset into actives and inactives. These classifiers had moderate performance, and



DOI: 10.1021/acsomega.7b01622
*ACS Omega* 2018, 3, 2261−2272

it became clear that alternative methods for variable reduction were needed to increase the accuracy and to reduce the complexity of the models. To this end, we employed two supervised methods, which evaluated the importance of the descriptors and subsequently selected a suitable subset of variables to be used for the modeling.

Our first approach was to use the area under the receiver operating characteristic (ROC) curve to quantify the relevance of the descriptors (*caret* package, *varImp* algorithm).[97] The descriptor variables were used as inputs into the ROC curve. If a descriptor could perfectly separate the classes, there would be a cutoff for that descriptor that would achieve sensitivity and specificity of 1, and the area under the curve would be one. The query resulted in a set of 20 uncorrelated descriptors ranked according to their importance (Figure 3).

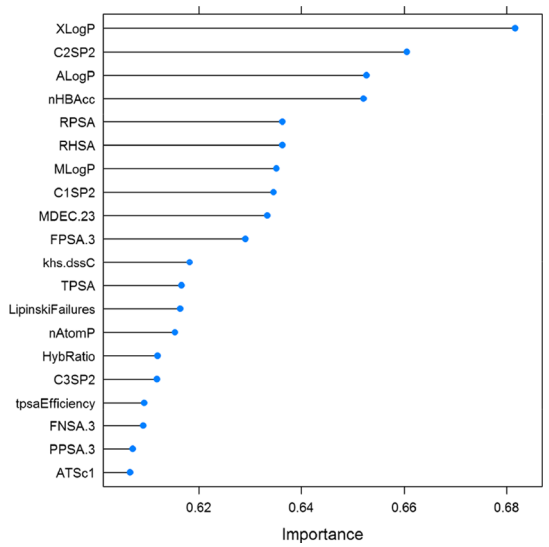



**Figure 3.** Variable selection using the area under the ROC curve: a set of 20 uncorrelated descriptors are ranked according to their importance.

Our second approach was a simple backward selection of descriptors, that is,—recursive feature elimination with random forest (RF)[98] (*caret* package, RFE algorithm). RF applied a resampling method of 10-fold cross-validation for the selection of the descriptors and produced a set of 84 variables ranked according to accuracy. The top 5 variables were HybRatio, XLog*P*, nHBAcc, BCUTc.1h, and ALog*P*. Both approaches took into consideration the response variable (active = 1 and inactive = −1) when evaluating the importance of the selected descriptors (supervised method).

There was a high consensus between the two methods with regard to the relevance of the descriptors, as 18 out of 20 descriptors that resulted from the ROC curve were included in the set of 84 proposed by RF, albeit in different order. Using as a starting point the descriptors common to both methods, we created a series of subsets by gradually including more variables from the list of 84. With these subsets we built a number of classifiers and compared their accuracy. A careful evaluation of the results highlighted a subset of 37 most relevant descriptors, which were subsequently used for the analysis. The selected variables optimized the accuracy of the models. Any attempt to increase the number of descriptors beyond that point had a negative effect on the model performance.

**2.2. QSAR Modeling.** Initially, a number of machine learning models were built on the training set using different algorithms with default parameters, to choose those that would fit our data best (Table 1). For this process, we used the *caret* package in R. We chose both linear and nonlinear algorithms on the basis of their diversity of learning style, which included classification and regression trees (CARTs),[99] linear discriminant analysis (LDA),[100] support vector machines (SVMs) with radial basis function,[101] k-nearest neighbors (KNNs),[102] RFs,[103] and gradient boosting machines (GBMs).[104] The evaluation metrics used were "accuracy" and "kappa". The generated models had different performance characteristics. A 10-fold cross-validation resampling method with 20 repeats was employed to get an estimate of the accuracy with which each model could predict unseen data. A summary table was created containing the evaluation metrics for each model (Table 1). As can be seen, the mean accuracy across the board was rather low, which implied that the classes in the dataset could not be easily predicted. SVMs and RFs showed comparable performance and had the highest accuracy on this classification problem (68%), whereas KNNs were the weakest classifiers (56%). Both SVMs and RFs are powerful modeling methods and highly nonlinear functions of the descriptor variables.

**Table 1. Summary Table Containing the Evaluation Metrics of Diverse Linear and Nonlinear Algorithms Used for the Data Analysis**

| | min. | mean | max. |
|---|---|---|---|
| | **Accuracy** | | |
| CART | 0.375 | 0.600 | 0.870 |
| LDA | 0.468 | 0.671 | 0.848 |
| SVM | 0.468 | 0.679 | 0.906 |
| KNN | 0.322 | 0.557 | 0.741 |
| RF | 0.515 | 0.680 | 0.838 |
| GBM | 0.451 | 0.668 | 0.870 |
| | **Kappa** | | |
| CART | −0.250 | 0.198 | 0.742 |
| LDA | −0.062 | 0.342 | 0.695 |
| SVM | −0.062 | 0.356 | 0.812 |
| KNN | −0.364 | 0.113 | 0.483 |
| RF | 0.029 | 0.359 | 0.677 |
| GBM | −0.100 | 0.335 | 0.741 |

model correlation

| | CART | LDA | SVM | KNN | RF | GBM |
|---|---|---|---|---|---|---|
| CART | 1.000 | 0.365 | 0.449 | 0.097 | 0.379 | 0.401 |
| LDA | 0.365 | 1.000 | 0.521 | 0.045 | 0.500 | 0.429 |
| SVM | 0.449 | 0.521 | 1.000 | 0.135 | 0.672 | 0.530 |
| KNN | 0.097 | 0.045 | 0.135 | 1.000 | 0.180 | 0.126 |
| RF | 0.379 | 0.500 | 0.672 | 0.180 | 1.000 | 0.578 |
| GBM | 0.401 | 0.429 | 0.530 | 0.126 | 0.578 | 1.000 |

The density plot of the distribution of the estimated accuracy (Figure 4) showed an overlap in the behavior of the algorithms. Differences in the spread and peaks of the distributions provided further information on the model performance. The distributions of the classifiers KNN, SVM, and RF were depicted as normal distributions (bell-shaped). In the cases of SVM (light green) and RF (orange), the bells were steep (small standard deviation), indicating that the data were tightly clustered around the mean accuracy value (∼68%, Table 1). This practically meant that the largest part of the data would be predicted with accuracy close to the mean value. On the other



DOI: 10.1021/acsomega.7b01622
*ACS Omega* 2018, 3, 2261−2272

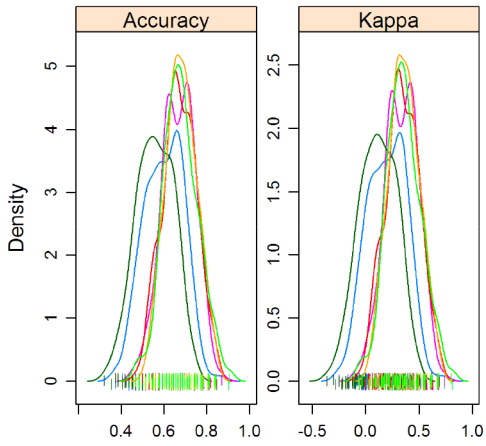



**Figure 4.** Density plot showing the distribution of the estimated accuracy of various algorithms as depicted in Table 1. KNN (dark green) is the weakest classifier with the lowest mean accuracy (55%), whereas RF (orange) and SVM (light green) are the strongest with mean accuracy 68 & 67%, respectively.

hand, the normal distribution of KNN (dark green) was depicted as a flattened bell curve (large standard deviation), indicating that the data were spread out, and therefore, a large part would be predicted with accuracy away from the mean value (55%, Table 1).

Based on the acquired information, we selected the more promising SVM and RF methods to build a new series of models with improved accuracy. To this end, parameters were properly tuned for the two algorithms. SVMs are particularly sensitive to the values of "gamma" (the width of the kernel function used for mapping the data into the high-dimensional space) and "C" (the error penalty parameter).[105,106] We optimized the parameters by performing a grid search. RF is a less sensitive machine learning technique, protected against overfitting,[103] but good parameter choices give robust models. The parameters with the biggest effect on the final accuracy of RF are "ntree" (the number of trees to grow) and "mtry" (the number of variables randomly sampled as candidates at each split). For calculating these parameters, we crafted a parameter search by creating a series of RF models and comparing their accuracy.

The whole process resulted in a shortlist of SVM and RF classifiers with an improved performance (Table 2). For estimating the accuracy of the models, we applied a 10-fold cross-validation with 20 repeats, and the observed total error was within a range of 11−23%. Subsequently, the ability to predict unseen data was evaluated for each model using the external validation dataset of 78 molecules. The total error of the models increased, ranging from 22 to 27% with a corresponding accuracy range of 78 to 73% (Table 2).

In an attempt to boost model accuracy, classifiers from the list were combined in ensemble predictions, resulting in high-order models that best combined the predictions of the base classifiers. To this end, we included in the shortlist the weak classifier KNN1, and we employed the method of stacking algorithms.[107] We used RF algorithm to combine the predictions. Combining KNN1 with the SVM and RF submodels resulted in an impressive improvement of the evaluation metrics of the stacked RFs (Table 2). As can be seen (Table 1, Figure 5), the base classifiers were not highly

**Table 2. Evaluation of the Model Performance**[a]

| A: evaluation of the model performance with *k*-fold cross-validation [training set, class: 122 ("−1"), 117 ("1")] | | | | |
|---|---|---|---|---|
| models | false positive | false negative | total error % | model summary |
| SVM1 | 16 | 25 | 17.2 | *k*-fold = 3, nrepeat = 20, cost = 4, gamma = 0.01, support vectors (sv) = 174 (86, 88) |
| SVM3 | 1 | 54 | 23.0 | *k* = 10, nrepeat = 20, cost = 16, gamma = 0.01, sv = 174 (103,71), class.weights ("1" = 0.3, "−1" = 0 0.7) |
| SVM3(b) | 33 | 12 | 18.8 | *k* = 10, nrepeat = 20, cost = 4, gamma = 0.01, sv = 181 (103,78), class.weights ("1" = 0.3, "−1" = 0.7) |
| SVM3(c) | 20 | 7 | 11.3 | *k* = 10, nrepeat = 20, cost = 16, gamma = 0.01, sv = 168 (75, 93), class.weights ("1" = 0.3, "−1" = 0.7) |
| KNN1 | 37 | 48 | 35.5 | *k*-fold = 10, nrepeat = 20, k-neighbors = 9 |
| RF1 | 0 | 0 | OOB: 35.2 | importance = 148, mtry = 4, ntree = 1000 |
| RF2 | 0 | 0 | OOB: 34.7 | importance = 148, mtry = 3, ntree = 2000 |

| B: evaluation of the model performance with an external validation set [class: 40 ("−1"), 38 ("1")] | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| models | accuracy | sensitivity | specificity | false positive | false negative | total error % | kappa | *P*-value | AUC % | 95% CI (DeLong) |
| SVM1 | 0.769 | 0.762 | 0.778 | 8 | 10 | 23.1 | 0.538 | $2.129 \times 10^{-5}$ | 80.72 | 70.66−90.79 |
| SVM3 | 0.782 | 0.850 | 0.710 | 6 | 11 | 21.8 | 0.562 | $8.692 \times 10^{-7}$ | 81.71 | 71.94−91.48 |
| SVM3(b) | 0.731 | 0.732 | 0.729 | 10 | 11 | 26.9 | 0.461 | 0.0001688 | 78.75 | 68.40−89.10 |
| SVM3(c) | 0.756 | 0.800 | 0.721 | 12 | 7 | 24.4 | 0.514 | 0.0001442 | 79.51 | 69.30−89.71 |
| KNN1 | 0.666 | 0.675 | 0.658 | 13 | 13 | 33.3 | 0.333 | 0.004299 | 66.91 | 54.74−79.07 |
| RF1 | 0.756 | 0.775 | 0.737 | 9 | 10 | 24.4 | 0.512 | $8.921 \times 10^{-6}$ | 83.06 | 73.91−92.21 |
| RF2 | 0.756 | 0.800 | 0.710 | 8 | 11 | 24.4 | 0.511 | $8.921 \times 10^{-6}$ | 83.03 | 73.90−92.15 |
| **Stacked Models (*)** | | | | | | | | | | |
| RF4* | 0.795 | 0.900 | 0.684 | 4 | 12 | OBB: 20.5 | 0.587 | $2.44 \times 10^{-7}$ | 80.46 | 70.94−89.98 |
| RF5* | 0.795 | 0.925 | 0.658 | 3 | 13 | OBB: 23.0 | 0.587 | $2.44 \times 10^{-7}$ | 81.12 | 71.69−90.55 |
| RF7* | 0.808 | 0.825 | 0.789 | 7 | 8 | OBB: 25.6 | 0.615 | $6.353 \times 10^{-8}$ | 81.74 | 72.44−91.05 |
| RF8* | 0.820 | 0.875 | 0.763 | 5 | 9 | OBB: 26.9 | 0.639 | $1.529 \times 10^{-8}$ | 86.22 | 77.82−94.61 |
| RF9* | 0.833 | 0.925 | 0.737 | 3 | 10 | OBB: 23.0 | 0.665 | $3.384 \times 10^{-9}$ | 84.14 | 75.04−93.25 |
| RF10* | 0.846 | 0.875 | 0.816 | 5 | 7 | OBB: 24.3 | 0.692 | 6.861e-10 | 87.99 | 80.12−95.87 |

[a]**RF4***: SVM1 + RF1 **RF5***: SVM1 + RF2 **RF7***: SVM1 + SVM3 + RF1 + RF2 **RF8***: SVM1 + SVM3 + RF1 + RF2 + KNN1 **RF9***: SVM1 + SVM3(b) + RF1 + RF2 + KNN1 **RF10***: RF2 + SVM1 + SVM3(c) + SVM3(b) + KNN1.



DOI: 10.1021/acsomega.7b01622
ACS Omega 2018, 3, 2261−2272

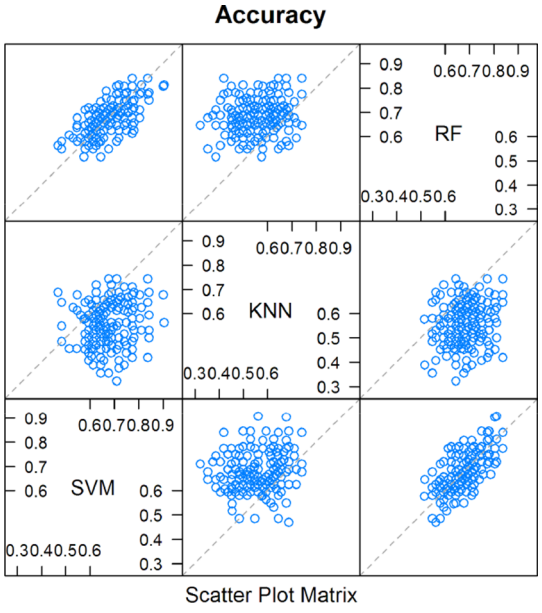



**Figure 5.** Scatterplot matrix correlating the predictions from the algorithms KNN, SVM, and RF used in the analysis (Table 1, model correlation). The classifiers are not strongly correlated (<0.75); an indication of them being informative in different ways and, therefore, suited to be combined in ensemble models.

correlated (<0.75), an indication of them being informative in different ways, allowing the high-order learner to get the best of each model. Two powerful new learners, RF9 and RF10, were thus created (Table 2), (Supporting Information, Table S2). The ROC curve (Figure 6) illustrating the diagnostic ability of the stacked model RF10 clearly shows optimized performance compared to the submodels combined to build it.

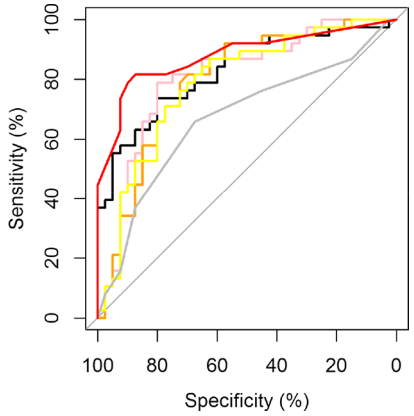



**Figure 6.** ROC curves illustrating the diagnostic abilities of the high-order learner RF10 (red) and the basic classifiers RF2 (black), SVM1 (pink), SVM3(c) (orange), SVM3(b) (yellow), and KNN1 (gray). RF10 is clearly shown to have an optimized performance compared to the submodels combined to build it.

The two learners RF9 and RF10 can be used in combination for the 15*S*-LOX1 classification problem (Table 2), (Supporting Information, Table S2). RF9 is a strong classifier with very high sensitivity (0.9250) reflected in the small number of molecules incorrectly classified as active (3 false positive). This would imply that should this classifier be used for any future prediction on the biological activity of molecules toward 15*S*-LOX1, the chances for inactive molecules to pass the filter would be very low. On the other hand, RF9 has a somewhat lower specificity (0.7368) and falsely classifies as inactive more molecules compared to RF10. Model RF10 is a stronger classifier and highly successful in identifying true 15*S*-LOX1 inhibitors. Most importantly, both learners are able to discriminate accurately between molecules showing selectivity toward either one of the enzymes 15*S*-LOX1 and 12*S*-LOX1 (Supporting Information, Table S2). This discrimination ability of the stacked models was decisively helpful in our attempt to explore the selectivity issue later on.

**2.3. Interpretability of the Models.** In this investigation, apart from the attempt to model chemical structures that are active against the LOX enzymes in general, the focus has been to explore the underlying properties (and their combinations) responsible for selective activity toward 15*S*-LOX1. The topology of the catalytic domain is shared by all LOX isozymes. The binding pocket is lined by hydrophobic residues except for second shell residues near the catalytic iron.[19] In a previous study, where in an attempt to explore LOX1 inhibition, we generated a pharmacophore model, we found that the assembly of the pharmacophore features strongly pointed toward a hydrophobic active site, which is indeed the case for all LOX subtypes.[2,3,108]

However, although the 3D structure of the enzyme's subtypes is highly conserved, LOX isozymes share low sequence identity.[19,24] A sequence alignment of the five human LOX isozymes generated with UniProt[109] revealed that the highest identity between sequences is observed for 15*S*-LOX1 and 12*S*-LOX1 (65.5%), whereas 15*S*-LOX2, 12*R*-LOX1, and 5*S*-LOX1 share an identity of 39% or lower with 15*S*-LOX1 (Supporting Information). As a result of the low sequence identity, apart from the apparent similarities, these enzymes exhibit striking differences at the active site. The size and nature of the residues present at the active site are crucial for determining both the volume of the binding pocket and the type of interactions that would ensure the binding of any chemical with the protein.[19] In positions 410, 415, 593, 597, and 716 of the sequence alignment, we can see the five residues that coordinate the iron with octahedral geometry (iron ligands) (Supporting Information). Four of them are conserved for all the enzymes (H410, H415, H593, and I716). Different amino acids occupy the position 597 in the sequences of 15*S*-LOX1 and 12*S*-LOX1, that is, the electrically charged histidine (H) and the polar uncharged asparagine (N), respectively. Mutation experiments have indicated the vital importance of these five iron ligands for the enzymes' activity [UniProtKB (P16050, O15296, P18054, O75342, and P09917)].[109]

Numerous scientific reports have also considered the significance of crucial amino acids in other positions, as is the case with the nonconserved residues in positions 467 and 468 of 15*S*-LOX1 and 12*S*-LOX1 sequences, respectively.[19] The residues A467/V468 that occupy these positions in 12*S*-LOX1 are considerably smaller in size compared to I467/M468 in 15*S*-LOX1. This difference is responsible for both the increase of volume (~6%) in the binding pocket of 12*S*-LOX1 and for the change in shape at the bottom of the active site.[19] In conclusion, based on the studies concerning the structure and the positional specificity of the LOX enzymes, we should expect that differences in the nature and geometry of crucial residues would affect the selective binding of ligands with the protein.

In this work, we undertook to explain the biological activity of the molecules with the help of their structural descriptors.



DOI: 10.1021/acsomega.7b01622
*ACS Omega* 2018, 3, 2261−2272

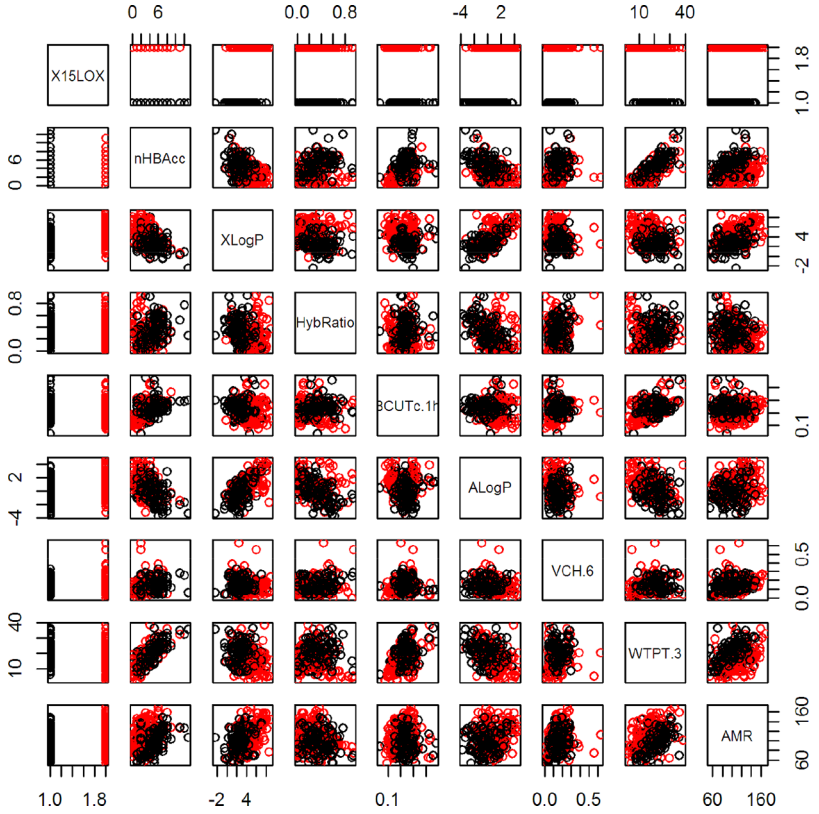



**Figure 7.** Scatter plot matrix by class of the descriptors with highest relevance for the LOX classification analysis. Descriptor ranking was performed by the RF classifier RF2.

This might seem like a challenge difficult to address without including in the analysis chemical information on the protein structure, as would be the case with a proteochemometric approach.[110] Nevertheless, previous knowledge of the nature of the enzyme proved to be a key ingredient for the decisions about model development because it allowed for an informed use of the data. It laid the foundation for predictive models with good accuracy that helped clarify certain aspects of this problem. The high-order learner RF10 was able to classify correctly the molecules as active (1) and inactive (−1) toward 15*S*-LOX1 to a high degree (Supporting Information, Table S2). Furthermore, it has been possible to address the selectivity issue, since RF10 could successfully discriminate those molecules which, although being active toward other LOX isozymes (12*S*-LOX1), were inactive toward 15*S*-LOX1. We found these results intriguing, and therefore, we attempted to link the descriptors of the highest discriminatory affinity with the molecules' ability to bind in a certain pocket and not in another.

As RF10 was a stacked model combining the predictions of submodels, primarily SVMs and RFs, further knowledge of the influence of the variables on the submodels was highly desirable. However, a straightforward interpretation of SVM and RF models is not possible. Both are "black boxes" that do not reveal how they relate the descriptors to the response variable. Nevertheless, RF has an ensemble nature (bagging trees) and combines the predictions of a large number of decision trees that are used as base learners. This allows for a degree of transparency that enabled us to have a measure of the impact of the descriptors in the base classifiers RF1 and RF2. Therefore, for these classifiers only, it was possible to gain some insights into the importance of every variable in classifying the data, the decrease in accuracy of the RF models after the exclusion of a single variable (mean decrease accuracy), and the average gain of purity by splits of every given variable (mean decrease Gini) (Supporting Information, Table S3).

Across the two supervised methods employed for evaluating the relevance of descriptors, the following were ranked very high: the number of hydrogen bond acceptors (nHBAcc), descriptors related to hydrophobicity (the octanol/water partition coefficients Xlog *P* and Alog *P*), two topological descriptors related to carbon hybridization (HybRatio and C2SP2), and a molecular descriptor related to partial charges (BCUTc.1h). Hydrophobicity and hydrogen bond acceptors were prominent pharmacophore features, and we had found this to be true in our previous investigation.[108] A molecule must be hydrophobic and possess lone electron pairs to bind with the protein. These results were generally in accord with the evaluation of descriptors performed later by the RF algorithms (RF1 & RF2) as part of the learning process, while the actual models were being generated (Supporting Information, Table S3), (Figure 7).

The classifiers identified the aforementioned variables as the most important, although in different order. In addition, three descriptors were ranked very highly by RF1 and RF2: (a) the topological weighted path descriptor WTPT.3, which is based on the molecular identification number and characterizes molecular branching. Descriptor WTPT.3 gives the sum of weighted paths starting from heteroatoms and[111] (b) the two valence chain descriptors VCH.6 and VCH.7.[95]

Shown in Table 3 are the values of the 10 top descriptors for a number of compounds from the validation set, which were correctly classified by RF10 (chemical structures of selective 15*S*-LOX1 ([A]) and 12*S*-LOX1 ([B]) inhibitors from Table 3 are



DOI: 10.1021/acsomega.7b01622
*ACS Omega* 2018, 3, 2261−2272

Table 3. Values of the 10 Top Descriptors (as Evaluated by RF2) for Selected Compounds from the Validation Set Correctly Classified by RF10[b]

| index[a] | nHBAcc | XLogP | ALogP | AMR | HybRatio | WTPT.3 | BCUTc.1h | VCH.6 | VCH.7 | nAtomP | class |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6[A] | 2 | 8.531 | −1.655 | 76.42 | 0.850 | 5.316 | 0.261 | 0.068 | 0.048 | 3 | 1 |
| 7[A] | 7 | 2.885 | −0.892 | 108.5 | 0.608 | 19.34 | 0.296 | 0.315 | 0.413 | 6 | 1 |
| 29[A] | 3 | 1.030 | 0.948 | 59.93 | 0.400 | 14.35 | 0.229 | 0.277 | 0.514 | 10 | 1 |
| 50[A] | 3 | 3.679 | 1.816 | 82.65 | 0.133 | 11.69 | 0.338 | 0.086 | 0.112 | 18 | 1 |
| 72[A] | 5 | 4.780 | 2.936 | 120.3 | 0.200 | 24.47 | 0.317 | 0.150 | 0.209 | 12 | 1 |
| 75[A] | 4 | 4.042 | 1.639 | 121.4 | 0.200 | 17.23 | 0.214 | 0.074 | 0.120 | 9 | 1 |
| 99[A] | 1 | 4.290 | 3.701 | 73.67 | 0.600 | 2.404 | 0.080 | 0 | 0 | 2 | 1 |
| 112[A] | 1 | 3.811 | 0.952 | 122.0 | 0.375 | 17.47 | 0.175 | 0.082 | 0.175 | 10 | 1 |
| 113[A] | 2 | 5.916 | 1.736 | 74.86 | 0.941 | 5.323 | 0.302 | 0.558 | 1.267 | 3 | 1 |
| 232[A] | 2 | 3.574 | 0.445 | 78.62 | 0 | 10.85 | 0.223 | 0.068 | 0.079 | 19 | 1 |
| 103[B] | 6 | 4.081 | 0.899 | 119.2 | 0.136 | 24.37 | 0.261 | 0.141 | 0.179 | 14 | −1 |
| 127[B] | 5 | 2.989 | 0.405 | 123.9 | 0.095 | 26.26 | 0.248 | 0.117 | 0.135 | 20 | −1 |
| 225[B] | 1 | 3.230 | −0.173 | 82.13 | 0 | 21.20 | 0.276 | 0.060 | 0.107 | 23 | −1 |
| 69[C] | 5 | 2.975 | −0.262 | 100.3 | 0.235 | 21.09 | 0.249 | 0.265 | 0.361 | 19 | 1 |
| 106[C] | 0 | 3.193 | 0.631 | 76.21 | 0 | 10.07 | 0.126 | 0.048 | 0.070 | 18 | 1 |
| 168[C] | 3 | 7.323 | 0.216 | 146.0 | 0.285 | 17.45 | 0.240 | 0.157 | 0.209 | 16 | 1 |
| 172[C] | 3 | 5.293 | 1.527 | 135.9 | 0.120 | 17.83 | 0.240 | 0.205 | 0.278 | 18 | 1 |
| 228[C] | 0 | 2.785 | 0.216 | 82.54 | 0.250 | 10.98 | 0.139 | 0.078 | 0.118 | 9 | 1 |
| 234[C] | 0 | 5.773 | 1.375 | 103.1 | 0.636 | 8.498 | 0.129 | 0.122 | 0.453 | 11 | 1 |
| 236[C] | 0 | 4.817 | 1.991 | 87.99 | 0 | 15.76 | 0.135 | 0.044 | 0.167 | 15 | 1 |
| 257[D] | 5 | 4.322 | −2.362 | 92.33 | 0.555 | 21.42 | 0.210 | 0.181 | 0.147 | 11 | −1 |
| 259[D] | 5 | 5.337 | 1.957 | 125.6 | 0.272 | 23.87 | 0.211 | 0.118 | 0.146 | 9 | −1 |
| 260[D] | 5 | 2.712 | −0.550 | 80.64 | 0.333 | 19.52 | 0.163 | 0.139 | 0.138 | 10 | −1 |
| 261[D] | 2 | 2.535 | 1.013 | 63.79 | 0 | 8.046 | 0.194 | 0.142 | 0.091 | 14 | −1 |
| 262[D] | 5 | 0.973 | −2.504 | 53.37 | 0.600 | 22.90 | 0.318 | 0.101 | 0.053 | 9 | −1 |
| 263[D] | 2 | 1.885 | −0.117 | 78.01 | 0.153 | 14.19 | 0.241 | 0.082 | 0.085 | 16 | −1 |
| 264[D] | 6 | 3.087 | −0.344 | 105.9 | 0.470 | 30.51 | 0.216 | 0.275 | 0.265 | 13 | −1 |
| 268[D] | 6 | 2.238 | −3.768 | 81.98 | 0.625 | 22.92 | 0.242 | 0.172 | 0.122 | 14 | −1 |
| 277[D] | 8 | 3.873 | −1.060 | 131.4 | 0.592 | 27.47 | 0.264 | 0.228 | 0.250 | 11 | −1 |
| 281[D] | 7 | 2.635 | −2.279 | 131.1 | 0.333 | 23.31 | 0.255 | 0.140 | 0.127 | 9 | −1 |
| 282[D] | 7 | 2.527 | −3.503 | 120.0 | 0.583 | 31.18 | 0.232 | 0.262 | 0.203 | 13 | −1 |
| 286[D] | 5 | 2.224 | −0.838 | 97.04 | 0.187 | 24.75 | 0.269 | 0.048 | 0.062 | 21 | −1 |
| 287[D] | 7 | 3.143 | −2.273 | 108.6 | 0.666 | 27.23 | 0.223 | 0.147 | 0.092 | 10 | −1 |
| 293[D] | 8 | 1.711 | −3.031 | 125.8 | 0.461 | 22.98 | 0.222 | 0.158 | 0.081 | 9 | −1 |
| 297[D] | 4 | 5.833 | 2.786 | 119.9 | 0.318 | 18.96 | 0.217 | 0.217 | 0.360 | 11 | −1 |
| 303[D] | 7 | 0.706 | 1.014 | 91.94 | 0.500 | 23.03 | 0.277 | 0.240 | 0.332 | 10 | −1 |
| 312[D] | 5 | 2.059 | −2.657 | 80.60 | 0.625 | 19.27 | 0.153 | 0.177 | 0.173 | 6 | −1 |
| 313[D] | 5 | 3.122 | −1.484 | 100.5 | 0.333 | 24.33 | 0.218 | 0.079 | 0.064 | 12 | −1 |
| 314[D] | 8 | 0.977 | −2.283 | 111.8 | 0.333 | 32.95 | 0.243 | 0.131 | 0.128 | 15 | −1 |

[a]Index numbers as they appear in the Supporting Information (Table S1) [b]Chemical structures of selective 15S-LOX1 ([A]) and 12S-LOX1 ([B]) inhibitors from this table are shown in Chart 1.

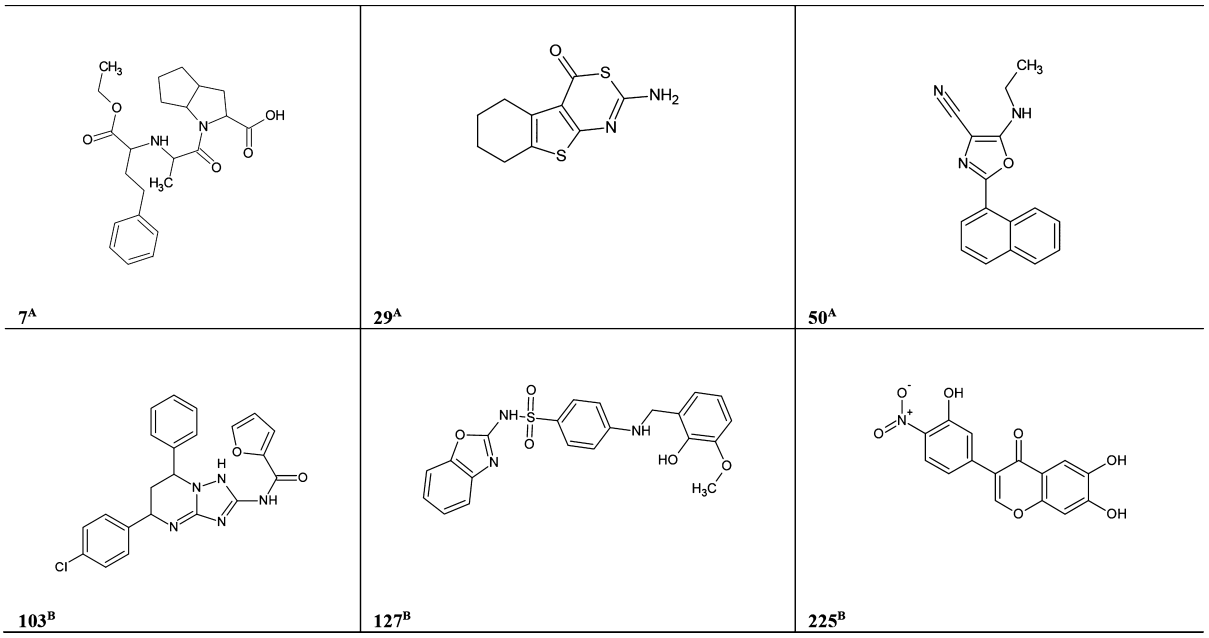
shown in Chart 1). The ranking of the descriptors was performed by RF2, a submodel of RF10. The compounds are divided in four groups according to their inhibitory activity toward LOX, that is, selective 15S-LOX1 inhibitors (A), selective 12S-LOX1 inhibitors (B), nonselective inhibitors with strong preference for the 15S-LOX1 pocket (C), and inactive molecules (D) (Supporting Information, Table S1). At a first glance, an interesting observation could be made in regard to differences between the groups A and B, concerning the values of descriptors HybRatio and WTPT.3. Low values of HybRatio in combination with high values of WTPT.3 seemed to favor selectivity toward 12S-LOX1. On the other hand, ALogP seemed to be significant for the discrimination between the inhibitors in groups A, B, and C, and the inactive molecules in the group D. Large negative ALogP values were in most cases associated with inactivity toward the protein.

However, as already mentioned, it is impossible to elucidate how RFs actually combined the descriptors and linked them to the response variable. This is even more the case for the support vector machine models used as base learners for the ensemble model RF10. Nevertheless, since our aim has been to acquire some elemental insights into how decisions were made, without necessarily knowing every detail of the full model, a simpler explanation could be sufficient. To this end, we used the *rpart* algorithm to create a single decision tree on the validation set using the 10 top descriptors as ranked by RF2. The decision path clarified which features were associated with every decision (Figure 8). We see that 3 of the top 10



DOI: 10.1021/acsomega.7b01622
ACS Omega 2018, 3, 2261−2272

**Chart 1. Chemical Structures of Selective 15*S*-LOX1 ([A]) and 12*S*-LOX1 ([B]) Inhibitors from Table 3**



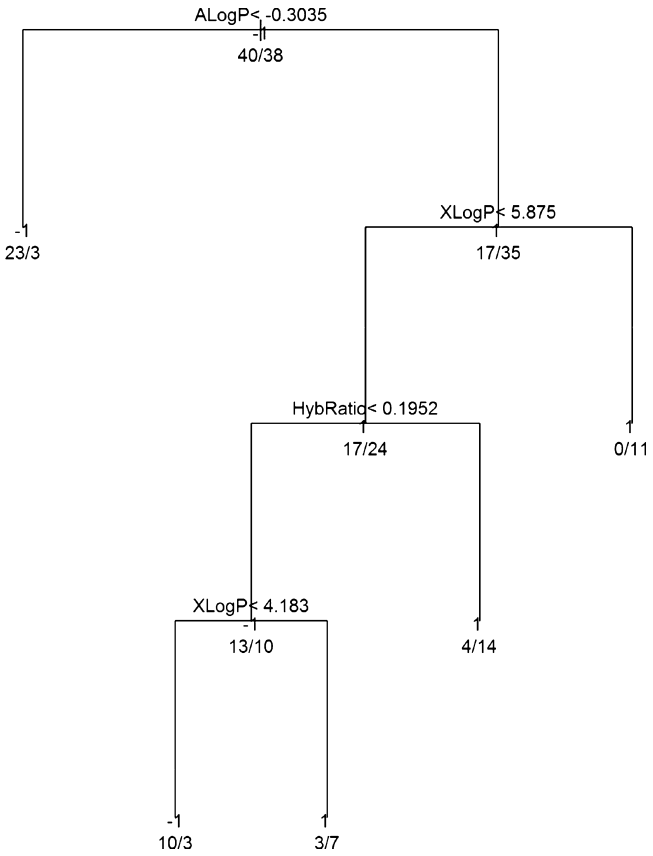




**Figure 8.** A single decision tree created on the validation set using the 10 top descriptors as ranked by the classifier RF2. The decision path clarifies which features are associated with every decision as well as the threshold values of the top descriptors that are responsible for a molecule being classified as active/inactive against 15-LOX1.

descriptors are the same, although with different ordering. ALog*P* is the top descriptor for the decision tree but is ranked 6th by RF2. There is a consensus regarding XLog*P* (second), whereas HybRatio is ranked third by the decision tree and fourth by RF. Interestingly, RF2 identifies nHBAcc as the top descriptor, whereas the decision tree ignores it. The differences observed in the ranking of descriptors between the RF and the single decision tree are not atypical as they are attributed to the greediness of the single tree.[112]

## 3. CONCLUSIONS

In the present study, we have created a dataset of structurally diverse molecules hitherto known as active against three LOX isoenzymes (15*S*-LOX1, 12*S*-LOX1, and 15*S*-LOX2). Subsequently, we considered a binary classification problem with the aim of exploring the selective activity of chemicals toward the enzyme 15*S*-LOX1. We identified a set of 37 structural descriptors of high discriminatory affinity, which are responsible for binding selectivity toward 15*S*-LOX1. Based on these descriptors, we created two highly accurate stacked models (RF9 and RF10) as ensembles of SVM, RF, and KNN algorithms that predict the selective interactions of various chemicals with 15*S*-LOX1. The models successfully classify the compounds as active/inactive against LOX1 and most importantly, they discriminate the molecules with selective activity toward either one of the isozymes 15*S*-LOX1 and 12*S*-LOX1. These QSAR classifiers can be used in drug discovery as computational filters in the virtual screening of novel LOX1 inhibitors, whereas the selected descriptors can be used as a guide for the design and synthesis of new 15*S*-LOX1 inhibitors.

## 4. METHODS

**4.1. Binary Classification and Model Generation.** The present work is concerned with a binary classification problem studied in supervised machine learning.[39] Therefore, the response (output) variable is categorical, and the two categories are predefined according to the biological activity in vitro of various chemicals toward the enzyme 15*S*-LOX1, that is, active (1) and inactive (−1).

QSAR modeling was carried out with R.[40] R is both a language and an environment for statistical computing and



DOI: 10.1021/acsomega.7b01622
*ACS Omega* 2018, 3, 2261−2272

graphics. Extended functionalities were added to R by installing a number of packages, which are machine learning algorithms implemented as third party libraries.The following R packages were used for the analysis: *rcdk*,[41] *SparseM*,[42] *randomForest*,[43] *caret*,[44] *e1071*,[45] *pROC*,[46] *rpart*,[47] *AppliedPredictiveModeling*,[48] and *caretEnsemble*.[49]

**4.2. Calculation of Molecular Descriptors.** A total of 317 structurally diverse molecules with recorded biological activity in vitro toward three LOX subtypes (15*S*-LOX1, 12*S*-LOX1, and 15*S*-LOX2) were used for the classification analysis (Supporting Information, Table S1). The molecules have been collected from available public sources: (a) reviews and reports[50−79] and (b) the PubChem Bioassay database, where we found several assays concerning the biological activity of series of molecules toward LOX.[80−91] Special consideration was taken to avoid structural redundancy. The ligands were drawn in ACD/ChemSketch[92] (Advanced Chemical Development) and a single 3D conformation was created for each structure with the Bioclipse software.[93,94] An SDF file containing the 3D coordinates of the molecules was imported in R, and the *rcdk* package was used to calculate automatically a number of descriptor variables. These descriptors are divided broadly into three main groups,[95] that is, atomic, bond, and molecular and belong to the specific categories "topological", "geometrical", "hybrid", "constitutional", and "electronic". The calculation resulted in 282 descriptors for each molecule.

**4.3. $IC_{50}$ Experimental Data.** It is a common practice in in vitro experiments to use three different LOX subtypes[50−79] to draw conclusions regarding biological activity toward human 15*S*-LOX1: The human 15*S*-LOX1, the soybean 13LOX1, and the mammalian (rabbit reticulocyte) 12-/15-LOX1. Such extrapolations are valid on the grounds that the enzymes have high structural similarity, most importantly in the area of the binding site, and the mode of action of the small molecule inhibitors is similar.[4,13,14,19] Therefore, for the collection of data for the training of the classification models, we adopted the same rationale.

Rabbit reticulocyte 12-/15-LOX1 has been characterized[19] and, as mentioned above, is often used as a standard for biochemical assays. Soybean 13LOX1 has been for many years a model system for understanding catalysis and structure of all lipoxygenases.[4] It is sufficiently stable, obtainable in large quantities, and relatively easy to purify.[4,19] In addition, the overall architecture of the soybean structure is similar to the mammalian, although the latter is much more compact.[19] Therefore, soybean 13LOX1 is used in bioassays for recording bioactivity toward 15*S*-LOX1, although the $IC_{50}$ values of a molecule may be different for the two enzymes.

In the present analysis, we have considered the molecules with $IC_{50} \geq 100$ $\mu$M as being inactive. Also, as already mentioned, we have included a number of molecules, which, although inactive against 15*S*-LOX1, exhibit bioactivity toward other LOX isozymes. These molecules were considered to be inactive. Our aim was to train our models to distinguish subtle differences among chemical structures and to be able to predict those that showed selectivity for 15*S*-LOX1.

In accordance to their inhibitory activity in vitro toward 15*S*-LOX1, the molecules were classified as follows: (a) active (1): selective 15*S*-LOX1 inhibitors and nonselective inhibitors with preference for the 15*S*-LOX1 pocket and (b) inactive (−1): molecules exhibiting selectivity toward two other LOX isozymes (12*S*-LOX1 and 15*S*-LOX2) but inactive against 15*S*-LOX1 and molecules inactive toward LOX in general. For all the molecules in the dataset, the $IC_{50}$ values were recorded (Supporting Information, Table S1).

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.7b01622.

Dataset of molecules with recorded inhibitory activity (IC50) against various LOX isoenzymes, comparison of the predictions of the stacked classifiers RF9 and RF10 on the validation set, impact of descriptors on the predictions of base classifiers RF1 and RF2, and sequence alignment of the five human LOX isozymes generated with UniProt (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: cgousiadou@gmail.com (C.G.).

### ORCID

Chrysoula Gousiadou: 0000-0002-7093-3055


### Notes

The authors declare no competing financial interest.

## ■ REFERENCES


(1) Horn, T.; Adel, S.; Schumann, R.; Sur, S.; Kakularam, K. R.; Polamarasetty, A.; Redanna, P.; Kuhn, H.; Heydeck, D. Evolutionary aspects of lipoxygenases and genetic diversity of human leukotriene signaling. *Prog. Lipid Res.* **2015**, *57*, 13−39.

(2) Ivanov, I.; Heydeck, D.; Hofheinz, K.; Roffeis, J.; O'Donnell, V. B.; Kuhn, H.; Walther, M. Molecular enzymology of lipoxygenases. *Arch. Biochem. Biophys.* **2010**, *503*, 161−174.

(3) Kuhn, H.; Saam, J.; Eibach, S.; Holzhütter, H.-G.; Ivanov, I.; Walther, M. Structural biology of mammalian lipoxygenases: Enzymatic consequences of targeted alterations of the protein structure. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 93−101.

(4) Prigge, S. T.; Boyington, J. C.; Faig, M.; Doctor, K. S.; Gaffney, B. J.; Amzel, L. M. Structure and mechanism of lipoxygenases. *Biochimie* **1997**, *79*, 629−636.

(5) Skrzypczak-Jankun, E.; Chorostowska-Wynimko, J.; Selman, S.; Jankun, J. Lipoxygaenases—a challenging problem in enzyme inhibition and drug development. *Curr. Enzyme Inhib.* **2007**, *3*, 119−132.

(6) Thomas, C. P.; Morgan, L. T.; Maskrey, B. H.; Murphy, R. C.; Kühn, H.; Hazen, S. L.; Goodall, A. H.; Hamali, H. A.; Collins, P. W.; O'Donnell, V. B. Phospholipid-esterified eicosanoids are generating agonist-activated human platelets and enhance tissue factor-dependent thrombin generation. *J. Biol. Chem.* **2010**, *285*, 6891−6903.

(7) Stavniichuk, R.; Drel, V. R.; Shevalye, H.; Vareniuk, I.; Stevens, M. J.; Nadler, J. L.; Obrosova, I. G. Role of 12/15-lipoxygenase in nitrosative stress and peripheral pre-diabetic and diabetic neuropathies. *Free Radical Biol. Med.* **2010**, *49*, 1036−1045.

(8) van Leyen, K. Lipoxygenase: an emerging target for stroke therapy. *CNS Neurol. Disord.: Drug Targets* **2013**, *12*, 191−199.

(9) Dobrian, A. D.; Lieb, D. C.; Cole, B. K.; Taylor-Fishwick, D. A.; Chakrabarti, S. K.; Nadler, J. L. Functional and pathological roles of the 12- and15-lipoxygenases. *Prog. Lipid Res.* **2011**, *50*, 115−131.

(10) Martínez-Clemente, M.; Clària, I.; Titos, E. The 5-lipoxygenase/leukotrienes pathway in obesity, insulin resistance, and fatty liver disease. *Curr. Opin. Clin. Nutr. Metab. Care* **2011**, *14*, 347−353.

(11) Klil-Drori, A. J.; Ariel, A. 15-Lipoxygenases in cancer: A double-edged sword. *Prostaglandins Other Lipid Mediators* **2013**, *106*, 16−22.

(12) Kuhn, H. Biologic relevance of lipoxygenase isoforms in atherogenesis. *Expert Rev. Cardiovasc. Ther.* **2005**, *3*, 1099−1110.




DOI: 10.1021/acsomega.7b01622
*ACS Omega* 2018, 3, 2261−2272

(13) Mashima, R.; Okuyama, T. The role of lipoxygenases in pathophysiology; new insights and future perspectives. *Redox Biol.* **2015**, *6*, 297−310.
(14) Kuhn, H.; Saam, J.; Eibach, S.; Holzhütter, H.-G.; Ivanov, I.; Walther, M. Structural biology of mammalian lipoxygenases: Enzymatic consequences of targeted alterations of the protein structure. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 93−101.
(15) Skrzypczak-Jankun, E.; Borbulevych, O. Y.; Zavodszky, M. I.; Baranski, M. R.; Padmanabhan, K.; Petricek, V.; Jankun, J. Effect of crystal freezing and smallmolecule binding on internal cavity size in a large protein: X-ray and dockingstudies of lipoxygenases at ambient and low temperature at 2.0°A resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 766−775.
(16) Rådmark, O. Arachidonate 5-lipoxygenase. *J. Lipid Mediators Cell Signalling* **1995**, *12*, 171−184.
(17) Neau, D. B.; Bender, G.; Boeglin, W. E.; Bartlett, S. G.; Brash, A. R.; Newcomer, M. E. Crystal structure of a lipoxygenase in complex with substrate: the arachidonic acid-binding site of 8R-lipoxygenase. *J. Biol. Chem.* **2014**, *289*, 31905−31913.
(18) Yamamoto, S.; Suzuki, H.; Ueda, N. Arachidonate 12-lipoxygenases. *Prog. Lipid Res.* **1997**, *36*, 23−41.
(19) Gillmor, S. A.; Villaseñor, A.; Fletterick, R.; Sigal, E.; Browner, M. F. The structure of mammalian 15-lipoxygenase reveals similarity to the lipases and the determinants of substrate specificity. *Nat. Struct. Biol.* **1997**, *4*, 1003−1009.
(20) Kobe, M. J.; Neau, D. B.; Mitchell, C. E.; Bartlett, S. G.; Newcomer, M. E. The structure of human 15-lipoxygenase-2 with a substrate mimic. *J. Biol. Chem.* **2014**, *289*, 8562−8569.
(21) Skrzypczak-Jankun, E.; Bross, R. A.; Carroll, R. T.; Dunham, W. R.; Funk, M. O. Three-dimensional structure of a purple lipoxygenase. *J. Am. Chem. Soc.* **2001**, *123*, 10814−10820.
(22) Alexander, S. P. H.; Fabbro, D.; Kelly, E.; Marrion, N.; Peters, J. A.; Benson, H. E.; Faccenda, E.; Pawson, A. J.; Sharman, J. L.; Southan, C.; Davies, J. A.; CGTP Collaborators. The Concise Guide to PHARMACOLOGY 2015/16: Enzymes. *Br. J. Pharmacol.* **2015**, *172*, 6024−6109.
(23) Kuhn, H.; Banthiya, S.; van Leyen, K. Mammalian lipoxygenases and their biological relevance. *Biochim. Biophys. Acta* **2015**, *1851*, 308−330.
(24) Kobe, M. J.; Neau, D. B.; Mitchell, C. E.; Bartlett, G.; Newcomer, M. E. The structure of human 15-lipoxygenase-2 with a substrate mimic. *J. Biol. Chem.* **2014**, *289*, 8562−8569.
(25) McGill, K. A.; Busse, W. W. Zileuton. *Lancet* **1996**, *348*, 519−524.
(26) Aparoy, P.; Reddy, K. K.; Reddanna, P. Structure and ligand based drug design strategies in the development of novel 5-LOX inhibitors. *Curr. Med. Chem.* **2012**, *19*, 3763−3778.
(27) Charlier, C.; Hénichart, J.-P.; Durant, F.; Wouters, J. Structural insights into human 5-lipoxygenase inhibition: combined ligand-based and target-based approach. *J. Med. Chem.* **2006**, *49*, 186−195.
(28) Kenyon, V.; Chorny, I.; Carvajal, W. J.; Holman, T. R.; Jacobson, M. P. Novel human lipoxygenase inhibitors discovered using virtual screening with homology models. *J. Med. Chem.* **2006**, *49*, 1356−1363.
(29) Schneider, I.; Bucar, F. Lipoxygenase inhibitors from natural plant sources. Part1: medicinal plants with inhibitory activity on arachidonate 5-lipoxygenase and 5-lipoxygenase/cycloxygenase. *Phytother. Res.* **2005**, *19*, 81−102.
(30) Schneider, I.; Bucar, F. Lipoxygenase inhibitors from natural plant sources. Part2: medicinal plants with inhibitory activity on arachidonate 12-lipoxygenase, 15-lipoxygenase and leukotriene receptor antagonists. *Phytother. Res.* **2005**, *19*, 263−272.
(31) Rai, G.; Joshi, N.; Jung, J. E.; Liu, Y.; Schultz, L.; Yasgar, A.; Perry, S.; Diaz, G.; Zhang, Q.; Kenyon, V.; Jadhav, A.; Simeonov, A.; Lo, E. H.; van Leyen, K.; Maloney, D. J.; Holman, T. R. Potent and selective inhibitors of human reticulocyte12/15—lipoxygenase as anti-stroke therapies. *J. Med. Chem.* **2014**, *57*, 4035−4048.
(32) Luci, D. K.; Jameson, J. B.; Yasgar, A.; Diaz, G.; Joshi, N.; Kantz, A.; Perry, S.; Kuhn, N.; Yeung, J.; Kerns, E. H.; Schultz, L.; Holinstat, M.; Nadler, J. L.; Taylor-Fishwick, D. A.; Jadhav, A.; Simeonov, A.; Holman, T. R.; Maloney, D. J. Synthesis and structure−activity relationship studies of 4-((2-hydroxy-3-methoxybenzyl)amino) benzenesulfonamide derivatives as potent and selective inhibitors of12-lipoxygenase. *J. Med. Chem.* **2014**, *57*, 495−506.
(33) Pergola, C.; Werz, O. 5-Lipoxygenase inhibitors: a review of recent developments and patents. *Expert Opin. Ther. Pat.* **2010**, *20*, 355−375.
(34) Hoobler, E. K.; Rai, G.; Warrilow, A. G. S.; Perry, S. C.; Smyrniotis, C. J.; Jadhav, A.; Simeonov, A.; Parker, J. E.; Kelly, D. E.; Maloney, D. J.; Kelly, S. L.; Holman, T. R. Discovery of a novel dual fungal CYP51/Human 5-Lipoxygenase inhibitor: implications for antifungal therapy. *PLoS One* **2013**, *8*, No. e65928.
(35) Sharma, B. K.; Pilania, P.; Singh, P. Modeling of cyclooxygenase-2 and 5-lipooxygenase inhibitory activity of apoptosis-inducing agents potentially useful in prostate cancer chemotherapy: Derivatives of diarylpyrazole. *J. Enzyme Inhib. Med. Chem.* **2009**, *24*, 607−615.
(36) Ruiz, J.; Pérez, C.; Pouplana, R. QSAR study of dual cyclooxygenase and 5-lipoxygenase inhibitors 2,6-di-tert-butylphenol derivatives. *Bioorg. Med. Chem.* **2003**, *11*, 4207−4216.
(37) Zheng, M.; Zhang, Z.; Zhu, W.; Liu, H.; Luo, X.; Chen, K.; Jiang, H. Essential structural profile of a dual functional inhibitor against cyclooxygenase-2 (COX-2) and 5-lipoxygenase (5-LOX): molecular docking and 3D-QSAR analyses on DHDMBF analogues. *Bioorg. Med. Chem.* **2006**, *14*, 3428−3437.
(38) Pontiki, E.; Hadjipavlou-Litina, D. Lipoxygenase inhibitors: a comparative QSAR study review and evaluation of new QSARs. *Med. Res. Rev.* **2008**, *28*, 39−117.
(39) Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer Science+Business Media: New York, 2013; Chapter 11, pp 247−266.
(40) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017; http://www.R-project.org.
(41) Guha, R. Chemical Informatics Functionality in R. *J. Stat. Software* **2007**, *18*, .10.18637/jss.v018.i05
(42) Koenker, R.; Ng, P. SparseM: SparseM: Sparse Linear Algebra. R package version 1.7, 2015; https://CRAN.R-project.org/package=SparseM.
(43) Liaw, A.; Wiener, M. Classification and Regression by random Forest. *R News* **2002**, *2*, 18−22.
(44) Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Software*. **2008**, *28*, .10.18637/jss.v028.i05
(45) Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-7, 2015; https://CRAN.R-project.org/package=e1071.
(46) Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Müller, M. pROC: an open-source package for R and S + to analyze and compare ROC curves. *BMC Bioinf.* **2011**, *12*, 77.
(47) Therneau, T.; Atkinson, B.; Ripley, B. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-11, 2017; https://CRAN.R-project.org/package=rpart.
(48) Kuhn, M.; Johnson, K. AppliedPredictiveModeling: Functions and Data Sets for Applied Predictive Modeling. R package version 1.1-6, 2014; https://CRAN.R-project.org/package=AppliedPredictiveModeling.
(49) Zachary, A.; Deane-Mayer; Knowles, J. E. caretEnsemble: Ensembles of Caret Models. R package version 2.0.0, 2016; https://CRAN.R-project.org/package=caretEnsemble.
(50) Sadeghian, H.; Jabbari, A. 15-Lipoxygenase inhibitors: a patent review. *Expert Opin. Ther. Pat.* **2016**, *26*, 65−88.
(51) Hoobler, E. K.; Rai, G.; Warrilow, A. G. S.; Perry, S. C.; Smyrniotis, C. J.; et al. Discovery of a Novel Dual Fungal CYP51/Human 5-Lipoxygenase Inhibitor: Implications for Anti-Fungal Therapy. *PLoS One* **2013**, *8*, No. e65928.
(52) Luci, D. K.; Jameson, J. B.; Yasgar, A.; Diaz, G.; Joshi, N.; Kantz, A.; Markham, K.; Perry, S.; Kuhn, N.; Yeung, J.; Kerns, E. H.; Schultz, L.; Holinstat, M.; Nadler, J. L.; Taylor-Fishwick, D. A.; Jadhav, A.; Simeonov, A.; Holman, T. R.; Maloney, D. J. Synthesis and Structure−


DOI: 10.1021/acsomega.7b01622
*ACS Omega* 2018, 3, 2261−2272

Activity Relationship Studies of 4-((2-Hydroxy-3-methoxybenzyl)-amino) benzenesulfonamide Derivatives as Potent and Selective Inhibitors of 12-Lipoxygenase. *J. Med. Chem.* **2014**, *57*, 495−506.
(53) Bråthe, A.; Andresen, G.; Gundersen, L.-L.; Malterud, K. E.; Rise, F. Antioxidant Activity of Synthetic Cytokinin Analogues: 6-Alkynyl- and 6-Alkenylpurines as Novel 15-Lipoxygenase Inhibitors. *Bioorg. Med. Chem.* **2002**, *10*, 1581−1586.
(54) Bråthe, A.; Gundersen, L.-L.; Malterud, K. E.; Rise, F. 6-Substituted Purines as Inhibitors of 15-Lipoxygenase; a Structure-Activity Study. *Arch. Pharm. Chem. Life Sci.* **2005**, *338*, 159−166.
(55) Gundersen, L.-L.; Malterud, K. E.; Negussie, A. H.; Rise, F.; Teklu, S.; Østby, O. B. Indolizines as Novel Potent Inhibitors of 15-Lipoxygenase. *Bioorg. Med. Chem.* **2003**, *11*, 5409−5415.
(56) Teklu, S.; Gundersen, L.-L.; Larsen, T.; Malterud, K. E.; Rise, F. Indolizine 1-sulfonates as potent inhibitors of 15-lipoxygenase from soybeans. *Bioorg. Med. Chem.* **2005**, *13*, 3127−3139.
(57) Cornicelli, J. A.; Padia, J. K.; Tait, B. D.; Trivedi, B. K. Method for treating and preventing inflammation and atherosclerosis. U.S. Patent 5,972,980 A, 1999.
(58) Weinstein, D. S.; Liu, W.; Gu, Z.; Murugesan, N.; et al. Tryptamine and homotryptamine-based sulfonamides as potent and selective inhibitors of 15-lipoxygenase. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1435−1440.
(59) Connor, D. T.; Roark, W. H.; Sorenson, R. J. Indole and benzimidazole 15-lipoxygenase inhibitors. U.S. Patent 6,858,739 B2, 2005.
(60) Lin, B. B.; Morita, T.; Lin, Y.-S.; Chen, H.-L. A facile synthesis of 1-ethoxy-4-cyano-5-ethoxycarbonyl-3Hazuleno[1,2-c]pyran-3-one, a selective 15-lipoxygenase inhibitor. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 63−65.
(61) Weinstein, D. S.; Liu, W.; Ngu, K.; et al. Discovery of selective imidazole-based inhibitors of mammalian 15-lipoxygenase: highly potent against human enzyme within a cellular environment. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5115−5120.
(62) Assadieskandar, A.; Amini, M.; Salehi, M.; Sadeghian, H.; Alimardani, M.; Sakhteman, A.; Nadri, H.; Shafiee, A. Synthesis and SAR study of 4,5-diaryl-1H-imidazole-2(3H)-thione derivatives, as potent 15-lipoxygenase inhibitors. *Bioorg. Med. Chem.* **2012**, *20*, 7160−7166.
(63) Bakavoli, M.; Nikpour, M.; Rahimizadeh, M.; Saberi, M. R.; Sadeghian, H. Design and synthesis of pyrimido [4,5-b][1,4]-benzothiazine derivatives, as potent 15-lipoxygenase inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 2120−2126.
(64) Nikpour, M.; Mousavian, M.; Davoodnejad, M.; Alimardani, M.; Sadeghian, H. Synthesis of new series of pyrimido[4,5-b][1,4] benzothiazines as 15-lipoxygenase inhibitors and study of their inhibitory mechanism. *Med. Chem. Res.* **2013**, *22*, 5036−5043.
(65) Ngu, K.; Weinstein, D. S.; Liu, W.; et al. Pyrazole-based sulfonamide and sulfamides as potent inhibitors of mammalian 15-lipoxygenase. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 4141−4145.
(66) Rai, G.; Kenyon, V.; Jadhav, A.; Schultz, L.; Holman, T. R.; et al. Discovery of potent and selective inhibitors of human reticulocyte 15-lipoxygenase-1. *J. Med. Chem.* **2010**, *53*, 7392−7404.
(67) Tehrani, M. B.; Emami, S.; Asadi, M.; Shafiee, A.; et al. Imi0064azo[2,1-b]thiazole derivatives as new inhibitors of 15-lipoxygenase. *Eur. J. Med. Chem.* **2014**, *87*, 759−764.
(68) Rai, G.; Joshi, N.; Jung, J. E.; Holman, T. R.; et al. Potent and selective inhibitors of human reticulocyte 12/15- lipoxygenase as anti-stroke therapies. *J. Med. Chem.* **2014**, *57*, 4035−4048.
(69) Pelcman, B.; Sanin, A.; Nilsson, P.; Kromman, H. Triazole compounds as lipoxygenase inhibitors. U.S. Patent 20,090,186,918 A1, 2009.
(70) Tait, B. D.; Dyer, R. D.; Auerbach, B. J.; et al. Catechol-based inhibitors of 15-lipoxygenase. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 93−96.
(71) Whitman, S.; Gezginci, M.; Timmermann, B. N.; Holman, T. R. Structure−activity relationship studies of nordihydroguaiaretic acid inhibitors toward soybean, 12-human, and 15- human lipoxygenase. *J. Med. Chem.* **2002**, *45*, 2659−2661.
(72) Malterud, K. E.; Rydland, K. M. Inhibitors of 15-lipoxygenase from orange peel. *J. Agric. Food Chem.* **2000**, *48*, 5576−5580.
(73) Wangensteen, H.; Miron, A.; Alamgir, M.; Rajia, S.; Samuelsen, A. B.; Malterud, K. E. Antioxidant and 15-lipoxygenase inhibitory activity of rotenoids, isoflavones and phenolic glycosides from Sarcolobus globosus. *Fitoterapia* **2006**, *77*, 290−295.
(74) Vasquez-Martinez, Y.; Ohri, R. V.; Kenyon, V. A.; Holman, T. R.; Sepúlveda-Boza, S. Structure−activity relationship studies of flavonoids as potent inhibitors of human platelet 12-hLO, reticulocyte 15-hLO-1, and prostate epithelial 15-hLO-2. *Bioorg. Med. Chem.* **2007**, *15*, 7408−7425.
(75) Amagata, T.; Whitman, S.; Johnson, T. A.; Stessman, C. C.; Loo, C. P.; Lobkovsky, E.; Clardy, J.; Crews, P.; Holman, T. R. Exploring sponge-derived terpenoids for their potency and selectivity against 12-human, 15-human, and 15-soybean lipoxygenases. *J. Nat. Prod.* **2003**, *66*, 230−235.
(76) Segraves, E. N.; Shah, R. R.; Segraves, N. L.; Johnson, T. A.; Whitman, S.; Sui, J. K.; Kenyon, V. A.; Cichewicz, R. H.; Crews, P.; Holman, T. R. Probing the activity differences of simple and complex brominated aryl compounds against 15-soybean, 15-human, and 12-human lipoxygenase. *J. Med. Chem.* **2004**, *47*, 4060−4065.
(77) Cichewicz, R. H.; Kenyon, V. A.; Whitman, S.; Morales, N. M.; Arguello, J. F.; Holman, T. R.; Crews, P. Redox inactivation of human 15-lipoxygenase by marine-derived meroditerpenes and synthetic chromanes: archetypes for a unique class of selective and recyclable inhibitors. *J. Am. Chem. Soc.* **2004**, *126*, 14910−14920.
(78) Gousiadou, C.; Gotfredsen, C. H.; Matsa, M.; Hadjipavlou-Litina, D.; Skaltsa, H. Minor Iridoids from *Scutellaria albida* ssp. *albida*. Inhibitory potencies on lipoxygenase, linoleic acid lipid peroxidation and antioxidant activity of iridoids from *Scutellaria* sp. *J. Enzyme Inhib. Med. Chem.* **2013**, *28*, 704−710.
(79) Schneider, I.; Bucar, F. Lipoxygenase inhibitors from natural plant sources. Part2: medicinal plants with inhibitory activity on arachidonate 12-lipoxygenase, 15-lipoxygenase and leukotriene receptor antagonists. *Phytother. Res.* **2005**, *19*, 263−272.
(80) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202−D1213.
(81) Wang, Y.; Bryant, S. H.; Cheng, T.; Wang, J.; Gindulyte, A.; Shoemaker, B. A.; Thiessen, P. A.; He, S.; Zhang, J. PubChem BioAssay: 2017 update. *Nucleic Acids Res.* **2017**, *45*, D955−D963.
(82) National Center for Biotechnology Information. PubChem BioAssay Database; AID=887. https://pubchem.ncbi.nlm.nih.gov/bioassay/887 accessed Sept 21, 2017.
(83) National Center for Biotechnology Information. PubChem BioAssay Database; AID=2157. https://pubchem.ncbi.nlm.nih.gov/bioassay/2157 accessed Sept 21, 2017.
(84) National Center for Biotechnology Information. PubChem BioAssay Database; AID=881. https://pubchem.ncbi.nlm.nih.gov/bioassay/881 accessed Sept 21, 2017.
(85) National Center for Biotechnology Information. PubChem BioAssay Database; AID=2537. https://pubchem.ncbi.nlm.nih.gov/bioassay/2537 accessed Sept 21, 2017.
(86) National Center for Biotechnology Information. PubChem BioAssay Database; AID=1452. https://pubchem.ncbi.nlm.nih.gov/bioassay/1452 accessed Sept 21, 2017.
(87) National Center for Biotechnology Information. PubChem BioAssay Database; AID=2162. https://pubchem.ncbi.nlm.nih.gov/bioassay/2162 accessed Sept 21, 2017.
(88) National Center for Biotechnology Information. PubChem BioAssay Database; AID=242641. https://pubchem.ncbi.nlm.nih.gov/bioassay/242641 accessed Sept 21, 2017.
(89) National Center for Biotechnology Information. PubChem BioAssay Database; AID=1250681. https://pubchem.ncbi.nlm.nih.gov/bioassay/1250681 accessed Sept 21, 2017.
(90) National Center for Biotechnology Information. PubChem BioAssay Database; AID=260125. https://pubchem.ncbi.nlm.nih.gov/bioassay/260125 accessed Sept 21, 2017.



DOI: 10.1021/acsomega.7b01622
*ACS Omega* 2018, 3, 2261−2272

(91) National Center for Biotechnology Information. PubChem BioAssay Database; AID=260124. https://pubchem.ncbi.nlm.nih.gov/bioassay/260124 accessed Sept 21, 2017.
(92) ACD/ChemSketch; Advanced Chemistry Development, Inc.: Toronto, ON, Canada. www.acdlabs.com, 2015.
(93) Spjuth, O.; Helmus, T.; Willighagen, E. L.; Kuhn, S.; Eklund, M.; Wagener, J.; Murray-Rust, P.; Steinbeck, C.; Wikberg, J. E. S. Bioclipse: An open source workbench for chemo- and bioinformatics. *BMC Bioinf.* **2007**, *8*, 59.
(94) Spjuth, O.; Alvarsson, J.; Berg, A.; Eklund, M.; Kuhn, S.; Mäsak, C.; Torrance, G.; Wagener, J.; Willighagen, E. L.; Steinbeck, C.; Wikberg, J. E. S. Bioclipse 2: A scriptable integration platform for the life sciences. *BMC Bioinf.* **2009**, *10*, 397.
(95) http://sysbiolab.bio.ed.ac.uk/wiki/index.php/CDK_Small_Molecule_Descriptors accessed Sept 21, 2017.
(96) Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer Science+Business Media: New York, 2013; Chapter 3, pp 35−40.
(97) Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer Science+Business Media: New York, 2013; Chapter18, pp 468−471.
(98) Guyon, I.; Weston, J.; Barnhill, S.; et al. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **2002**, *46*, 389.
(99) Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; Chapman and Hall: New York, 1984.
(100) Welch, B. L. Note on Discriminant Functions. *Biometrika* **1939**, *31*, 218−220.
(101) Vapnik, V. *The Nature of Statistical Learning Theory*; Springer, 2010.
(102) Coomans, D.; Massart, D. L. Alternative k-nearest neighbour rules in supervised pattern recognition : Part 1. k-Nearest neighbour classification by using alternative voting rules. *Anal. Chim. Acta* **1982**, *136*, 15−27.
(103) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.
(104) Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorob.* **2013**, 7−21.
(105) Chapelle, O.; Vapnik, V.; Bousquet, O.; Mukherjee, S. Choosing multiple parameters for support vector machines. *Mach. Learn.* **2002**, *46*, 131−159.
(106) Cawley, G. C.; Talbot, N. L. C. Over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079−2107.
(107) Ozay, M.; Yarman Vural, F. T. A New Fuzzy Stacked Generalization Technique and Analysis of its Performance. **2013**, arXiv:1204.0171 [cs.LG].
(108) Gousiadou, C.; Kouskoumvekaki, I. LOX1 inhibition with small molecules. *J. Mol. Graphics Modell.* **2016**, *63*, 99−109.
(109) The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158−D169.
(110) Kontijevskis, A.; Komorowski, J.; Wikberg, J. E. S. Generalized Proteochemometric Model of Multiple Cytochrome P450 Enzymes and Their Inhibitors. *J. Chem. Inf. Model.* **2008**, *48*, 1840−1850.
(111) Randic, M. On molecular identification numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164−175.
(112) Norouzi, M.; Collins, M. D.; Johnson, M.; Fleet, D. J.; Kohli, P. Efficient non-greedy optimization of decision trees. **2015**, arXiv:1511.04056 [cs.LG].



DOI: 10.1021/acsomega.7b01622
*ACS Omega* 2018, 3, 2261−2272