

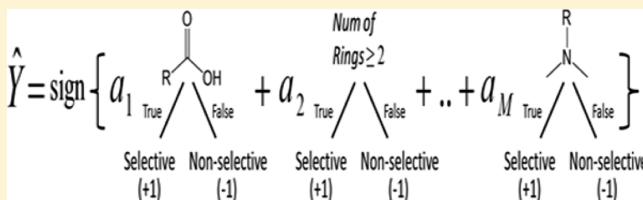
LiCABEDS II. Modeling of Ligand Selectivity for G-Protein-Coupled Cannabinoid Receptors

Chao Ma,^{†,‡,#} Lirong Wang,^{†,§,||,#} Peng Yang,^{†,§,||} Kyaw Z. Myint,^{†,‡} and Xiang-Qun Xie*,^{†,‡,§,||}

[†]Department of Pharmaceutical Sciences, School of Pharmacy, Computational Chemical Genomics Screening Center, [‡]Department of Computational and Systems Biology, [§]Center for Chemical Methodologies & Library Development (UPCMLD), ^{||}Drug Discovery Institute, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States

 Supporting Information

ABSTRACT: The cannabinoid receptor subtype 2 (CB2) is a promising therapeutic target for blood cancer, pain relief, osteoporosis, and immune system disease. The recent withdrawal of Rimonabant, which targets another closely related cannabinoid receptor (CB1), accentuates the importance of selectivity for the development of CB2 ligands in order to minimize their effects on the CB1 receptor. In our previous study, LiCABEDS (Ligand Classifier of Adaptively Boosting Ensemble Decision Stumps) was reported as a generic ligand classification algorithm for the prediction of categorical molecular properties. Here, we report extension of the application of LiCABEDS to the modeling of cannabinoid ligand selectivity with molecular fingerprints as descriptors. The performance of LiCABEDS was systematically compared with another popular classification algorithm, support vector machine (SVM), according to prediction precision and recall rate. In addition, the examination of LiCABEDS models revealed the difference in structure diversity of CB1 and CB2 selective ligands. The structure determination from data mining could be useful for the design of novel cannabinoid lead compounds. More importantly, the potential of LiCABEDS was demonstrated through successful identification of newly synthesized CB2 selective compounds.



1. INTRODUCTION

The whole drug discovery and development process typically takes 10 to 15 years and costs 500 million to 2 billion dollars¹ for a new drug. Despite significant investment in the discovery and optimization of “lead compound”, the failure rate in clinic phase II and phase III has been rising.² Lack of desired pharmacological and physicochemical properties frequently causes rejection of drug candidates from further development. The selectivity to specific drug targets has been a crucial pharmacological property of drug candidates.³ Since the advent of chemical genetics and chemical genomics,⁴ even more emphasis has been placed on ligand selectivity to study biological systems and explore mechanisms of action.

The endocannabinoid system is assumed to regulate psychological processes and is directly related to human mental and physical health. The first cannabinoid receptor was discovered in 1988,⁵ later named as the CB1 receptor. A second cannabinoid receptor, CB2 receptor, was identified in human peripheral organs in 1993.⁶ These two receptors share high sequence similarity, especially in transmembrane portions.⁷ Thus, selective cannabinoid ligands are desired to ensure minimal effect on the other receptor. Recent studies show that the cannabinoid CB2 receptor may serve as a potential target for many diseases,⁸ including neurodegenerative disorders,⁹ blood cancer,¹⁰ and osteoporosis.¹¹ On the other hand, a famous drug, Rimonabant, that targets the cannabinoid CB1 receptor to treat obesity, was withdrawn from the market due to its side psychotropic effects.¹² In spite of the therapeutic potential of the CB2 receptor, the suspension of

Rimonabant reemphasizes the importance of selectivity regarding the design of CB2 agonists and antagonists.

Felder et al. reported the first CB2 selective ligand, WIN-55,212-2, that showed a 19-fold higher binding affinity for CB2 than for CB1.¹³ Since then, the discovery of novel CB2 selective ligands has become the endeavor of many scientists.^{8,14} Nowadays, computer-aided drug design has been integrated into the pipeline of drug discovery to accelerate traditional experimental screening, which requires significant efforts. Thus, the search for CB2 selective ligands is an important application of drug design. For example, several reports were published about homology modeling and molecular docking to develop CB2 specific ligands.¹⁵ Besides structure-based drug design, machine learning and pattern recognition are gaining popularity for virtual screening and the prediction of various molecular properties. In the context of selectivity profiling, Wassermann et al predicted ligand selectivity by using support vector machine ranking.³

Ligand Classifier of Adaptively Boosting Ensemble Decision Stumps (LiCABEDS) is a generic ligand classification algorithm developed by Xie's laboratory¹⁶ for the prediction of categorical ligand properties. LiCABEDS was established on Freund and Schapire's ensemble learning framework.¹⁷ The underlying theory and its application on modeling ligand functionality have been outlined in our previous report. Briefly, the motivation of LiCABEDS is to build a “strong” classifier by integrating a set of

Received: August 20, 2012

Published: January 1, 2013



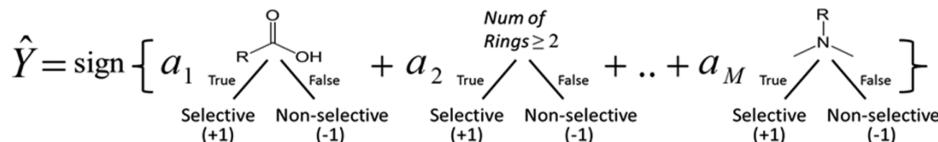


Figure 1. Graphical illustration of the constitution of a LiCABEDS classifier. Each decision stump outputs a categorical value (+1 or -1) that contributes to the selectivity label by examining the presence of a predefined structural pattern. The final prediction is the summation of the predictions of individual decision stumps weighted by constant a_i .

weighted “weak learners” that are defined in a specific hypothesis space. Unlike many other supervised learning algorithms, LiCABEDS minimizes training error by iteratively adding more “learners” into the classifier ensemble but still aims at margin maximization. The implementation of LiCABEDS is available at www.cbligand.org/LiCABEDS, including automated model training, cross validation, and predicting. In this study, we continued to explore the performance and applicability of LiCABEDS for the prediction of cannabinoid ligand selectivity. LiCABEDS was compared with a famous supervised learning algorithm, support vector machine (SVM), in order to evaluate their performance in recovering true selective ligands. Performance measures were given by precision, recall rate, the geometric mean of both, and ROC curve with area-under-curve (AUC). An ideal classifier would recover all selective ligands from a screening library (100% recall rate) and have no false positive error (100% precision). In addition, the investigation of LiCABEDS models revealed information on the structure diversity of CB1 selective and CB2 selective ligands. Supported by scaffold and fragment analysis, we hypothesized that CB2 selective ligands tended to be more structurally diverse than CB1 selective compounds among discovered cannabinoid ligands. More interestingly, the LiCABEDS model successfully identified a true CB2 selective ligand from 12 newly synthesized compounds. With these 12 compounds added to the training set, a new LiCABEDS model was capable of recovering four CB2 selective compounds from another batch of 26 novel testing compounds. In the following, we report data sets involved in model training and validation, a comprehensive computation protocol of LiCABEDS and SVM, and results of systematic calculations.

2. MATERIALS, METHODS, AND CALCULATIONS

2.1. Selectivity and Data Sets. To demonstrate the performance and robustness of machine learning algorithms in the prediction of cannabinoid ligand selectivity, 703 chemical structures and their bioactivity (K_i value) to CB1 and CB2 receptors were retrieved from the public cannabinoid ligand database (www.cbligand.org). A selective ligand is usually defined descriptively as its differing binding affinity to form a ligand–protein complex with different receptors. However, it lacks a quantitative definition. In this study, a ligand was regarded as CB1selective (or CB2 selective) as long as its ratio of CB1 K_i to CB2 K_i (CB2 K_i to CB1 K_i) was less than 0.1. In other words, a ligand bound selectively to cannabinoid receptor subtypes if it exhibited more than a 10-fold K_i difference. The rationality of this criterion was supported by an expert in selective cannabinoid ligand discovery, J.W. Huffman.¹⁸ He commented that O_2 -propano- Δ^8 -THC analogues¹⁹ exhibited modest CB2 selectivity. The CB1/CB2 K_i ratio of these analogues, reported by Reggio et al., ranged from 2.8 to 4.5. In addition, the 10-fold K_i threshold created selective and nonselective compound sets with reasonable sizes. Following this approach, 149 compounds were identified as CB1 selective; 147 compounds were identified as CB2 selective.

The remaining compounds were treated as nonselective. The goal of the study was to distinguish CB1 selective compounds from the CB1 nonselective (including CB2 selective and nonselective) and CB2 selective compounds from CB2 nonselective (including CB1 selective and nonselective).

2.2. LiCABEDS. Ligand Classifier of Adaptively Boosting Ensemble Decision Stumps (LiCABEDS) is a general-purpose ligand classification tool that is established on a famous ensemble learning algorithm, adaptive boosting.¹⁷ The performance, robustness, interpretability, and parameters of LiCABEDS were thoroughly discussed through modeling 5-HT_{1A} ligand functionality.¹⁶ Now we aim to demonstrate the wide applicability of LiCABEDS by modeling cannabinoid ligand selectivity. Briefly, the prediction in LiCABEDS is determined by weighted summation of a set of “weak” classifiers, i.e., decision stumps. Each decision stump outputs a predicted categorical label according to whether the testing sample possesses a specific compound fragment or structure pattern. Figure 1 visualizes the underlying mechanism of the LiCABEDS prediction model. Even though the performance of each individual decision stump is not necessarily much better than a random guess, ensemble learning theory shows that strong classifiers can be obtained by boosting these weak learners.

Thus, the responsibility of LiCABEDS training algorithms is to derive M decision stumps and weight assignment. Molecular fingerprints define a high-dimensional chemistry space from which decision stumps can be reasonably constructed. As LiCABEDS theory has been reported previously, the training algorithm is only outlined here. Let $y(\mathbf{x}, i, t) = 2I(\mathbf{x}_i = t) - 1$ represent a decision stump, where \mathbf{x} is molecular fingerprint vector, i is vector index, t is a parameter, and I is indicator function. Thus, the LiCABEDS model is formulated as

$$Y_M = \text{sign}(\sum_m^M a_m y_m(\mathbf{x}, i_m, t_m)) \quad (1)$$

The parameters, a_m , i_m , and t_m , for each decision stump m can be decided following the training algorithm:

1. Initialize the sample weights for each training compound n , $w_n = 1/N, n = 1, \dots, N$, and N is the total number of training compounds.
2. For each round of calculation, $m = 1, \dots, M$, find i_m, t_m for weak learner y_m by minimizing the weighted error function.

$$(i_m, t_m) = \arg \min_{i_m, t_m} \sum_{n=1}^N w_n I(y_m(\mathbf{X}_n, i_m, t_m) \neq l_n) \quad (2)$$

where “arg min” is the function to return the arguments that minimize the object function; \mathbf{X}_n is the descriptor vector for compound n ; $l_n = \pm 1$ is the selectivity label. The optimal values of i_m, t_m are found by enumerating all possible combinations of i_m, t_m .

Then evaluate the quantities

$$\varepsilon = \frac{\sum_{n=1}^N w_n I(y_m(\mathbf{x}_n, i_m, t_m) \neq l_n)}{\sum_{n=1}^N w_n}; a_m = \ln \frac{1 - \varepsilon}{\varepsilon} \quad (3)$$

a_m becomes the weight for the “decision stump” m . Then update the weights of training compounds for next round of calculation:

$$w_n \leftarrow w_n \exp(a_m I(y_m(\mathbf{x}_n, i_m, t_m) \neq l_n)) \quad (4)$$

2.3. Support Vector Machine. Support vector machine (SVM) shares some similarity with adaptive boosting regarding margin maximization. Both algorithms have a native margin maximization mechanism to control overfitting, but margins in these two algorithms are expressed in different norms. The binary classification in standard SVM setting is formulated by function $f_{w,b}(\mathbf{x}) = \text{sign}(\mathbf{w}^T \mathbf{x} + b)$, where \mathbf{w} is the normal vector to the decision hyperplane. The motivation of the support vector machine is that the optimal decision surface has the largest distance to the nearest data points of both categories, i.e., margin maximization, and still classifies training samples correctly by satisfying $y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \forall i$ where y_i is label of training data. Given the same training data sets $\{\mathbf{x}_i, y_i\}$, $i = 1$ to n , the problem of margin optimization can be transformed into constraint optimization

$$\text{Minimize}_{w,b} \|\mathbf{w}\|^2$$

$$\text{subject to } y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \forall i, i = 1, \dots, n$$

Among all the constraints, the training samples that are relevant to the determination of margin are called “support vectors”. More details about SVM algorithm can be found in the original publication²⁰ and its application in chemoinformatics³ as well as our SHT-1a classification paper.²¹

2.4. Calculations. Four types of molecular fingerprints, including Maccs key,²² Unity (www.tripos.com), FP2,²³ and Molprint 2D,^{24,25} were generated as descriptors to represent each cannabinoid ligand in machine learning algorithms. Maccs key (166 bit) and FP2 (1024 bit) were calculated using Open Babel, and Unity fingerprints (992 bit) were generated in Tripos Sybyl. The whole cannabinoid ligand data set produced 2530 three-layer Molprint 2D features from Bender et al.’s program. These features were mapped to binary vector according to previously published protocol.¹⁶ To systematically evaluate prediction accuracy, 50% CB1 selective and 50% CB1 non-selective compounds were randomly selected as training set in order to build a prediction model. The remaining compounds, which were not present to learning algorithms, were used to test the model. The calculation strategy was also applied to CB2 selectivity modeling. Two machine learning algorithms (LiCABEDS and SVM) combined with four types of fingerprint and two receptor subtypes resulted in 16 calculation settings. Each calculation setting was repeated 20 times on different randomly selected training and testing samples to assess stability and reliability.

All calculations regarding LiCABEDS were automated with our published program (www.cbligand.org/LiCABEDS). The SVM-based selectivity modeling method in this study was conducted using library “e1071”, which included a R wrapper of LIBSVM.²⁶ This article also reports the effect of cross validation on the prediction performance of testing data sets. With cross validation, 30% of the training data was omitted from the

cross-validation set in order to choose optimal training parameters. For LiCABEDS, the training parameter was the number of “decision stumps”, or training iterations, while the parameter of SVM was the constant C that regularized the trade-off of training error and margin size. All the training data then were used to train a model according to the optimal parameter specified during cross validation. During this study, a LiCABEDS model was developed using all the labeled compounds to predict CB2 selectivity of 12 newly synthesized compounds.

The following performance metrics were calculated for each calculation setting

$$\text{Precision} = \frac{\#\text{True Selectives}}{\#\text{True Selectives} + \#\text{False Selectives}}$$

True Positive Rate (TPR)

$$= \frac{\#\text{True Selectives}}{\#\text{True Selectives} + \#\text{False Nonselectives}}$$

False Positive Rate (FPR)

$$= \frac{\#\text{False Selectives}}{\#\text{False Selectives} + \#\text{True Nonselectives}}$$

Precision measured the percentage of correctly identified selective compounds and Recall Rate (or True Positive Rate) depicted the capability of the prediction model to retrieve or recover selective compounds. Geometric mean of Precision and Recall Rate could serve as a single criterion to rate prediction performance. ROC (receiver operating characteristic) curve plotting TPR versus FPR showed the enrichment of true selective compounds with varying decision threshold.

Structural skeletons of cannabinoid ligands were generated and compared for the exploration of structure diversity in selective CB1 and CB2 ligands. Briefly, compounds were reduced to carbon skeletons by deleting all non-ring substituent except linkers between ring systems, replacing all heteroatoms with carbon atoms, and converting all bond orders to single bonds.²⁷ Following these procedures, generic compound scaffolds were generated by shrinking the linker chain from a sequence of two or more CH2s to only one CH2. Figure 2 exemplifies the procedure of scaffold generation with a CB1 selective ligand, SR141716.

Furthermore, a LiCABEDS model was trained to predict the CB2 selectivity of newly synthesized compounds with a novel scaffold. In this case, the model was developed with all annotated cannabinoid ligands as training data and Molprint 2D fingerprint as descriptor. The binding affinities of these compounds to CB1 and CB2 receptors were determined by radioligand competition binding assays. The LiCABEDS prediction was compared with experimental CB2 selectivity following the same 10-fold K_i criteria. The compound structures and experimental details can be found in Part 1 of the Supporting Information.

2.5. Model Interpretation. The interpretation of a LiCABEDS model may help us to understand the underlying classification mechanism and to discover significant features related to ligand properties. The models developed out of the 20 rounds of calculations are used to demonstrate the interpretation of CB1 selectivity prediction. As presented in the

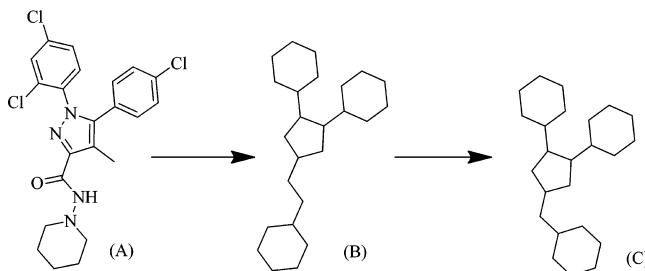


Figure 2. SR141716, a CB1 selective ligand, as an example to show the procedure of scaffold generation. (A) Original structure of SR141617. (B) Carbon skeleton after deleting the side chains, such as the $=\text{O}$, $-\text{CH}_3'$ and $-\text{Cl}$ groups, replacing all non-carbon atoms, in this case nitrogen atoms, with carbon atoms and converting all bond orders to single bonds. (C) General carbon skeleton is produced by shrinking the linker chain of the carbon skeleton from two or more CH_2 s to one CH_2 .

Materials, Methods, and Calculations section, each “decision stump” contributes to the final prediction according to its weight, a_m in eq 1. The highly weighted “decision stumps” imply that they contribute significantly to the prediction outcome. We extract such decision stumps together with related features and analyze their occurrences among the CB1 selective and non-CB1 selective compounds. In this demonstration, we use Molprint 2D descriptors as predictive features. Each Molprint 2D feature depicts a central atom and its atom environment up to a specific topological distance, and the atom environment in Molprint 2D is defined as the distribution of heavy atoms surrounding the central atom. Heavy atoms are distinguished by Sybyl atom types. The same analogy applies to the analysis of CB2 selectivity models.

3. RESULTS AND DISCUSSION

This section is focused on the analysis of LiCABEDS, SVM, and different molecular fingerprints for their capability of distinguishing cannabinoid selective compounds from nonselective ones. The results are quantitatively supported by various performance metrics that depict aspects of prediction outcomes. The effect of cross validation on LiCABEDS and SVM is also discussed. In addition, the cross-validation studies reveal that the LiCABEDS model complexity is positively correlated with the recall rate of CB2 selective compounds but not CB1 selective ones. A possible explanation for this is provided in terms of structure diversity, which gives us a deeper understanding of the LiCABEDS mechanism. A case study shows how LiCABEDS could direct drug discovery and chemical modification by predicting the selectivity of newly synthesized compounds.

3.1. LiCABEDS and SVM in Default Settings. LiCABEDS was originally designed as a general purpose ligand classifier for the prediction of categorical ligand properties. The theoretical framework of LiCABEDS shows that LiCABEDS is not necessarily a linear classification algorithm. However, when the feature space is represented in binary molecular fingerprints, the training algorithm described in the Materials, Methods, and Calculations section produces a linear classifier. Given that a type of fingerprint defines a pattern set S and that a function $f: S \rightarrow N$ maps each structure pattern to a unique index. There are $2|S|$ possible decision stumps $y(\mathbf{x}_i, t) = 2I(\mathbf{x}_i = t) - 1$ because $i \in \{id: id = f(s), s \in S\}$ and $t \in \{0, 1\}$. Due to the sample space of t in the context of binary fingerprint, the indicator function, I , can be omitted by expressing a decision stump as $y(\mathbf{x}_i, t) = k(2\mathbf{x}_i - 1)$, $k \in \{+1, -1\}$. $k = +1$ if $t = 1$, $k = -1$ otherwise. Therefore,

the LiCABEDS prediction function $Y_M = \text{sign}(\sum_m^M a_m y_m(\mathbf{x}, i_m, t_m))$ can be updated as $Y_M = \text{sign}(\sum_m^M a'_m (2\mathbf{x}_{i_m} - 1))$, $a'_m = a_m k_m$. For a specific i , define a set A_i that contains any a'_m associated with \mathbf{x}_i .

Thus, $Y_M = \text{sign}(\sum_{i=1}^K (\sum_{a'_k \in A_i} a'_k (2\mathbf{x}_i - 1))) = \text{sign}(\sum_{i=1}^K \beta_i \mathbf{x}_i + \beta_0)$. This proves the linearity of LiCABEDS with decision stumps created in a binary feature space.

In our previous studies, LiCABEDS was compared with naive Bayes classifier and classification tree (or recursive partitioning method). Support vector machine (SVM) is another popular classification algorithm for building linear and nonlinear models. SVM has been widely applied in cheminformatics and ligand-based drug design. For the linear nature of LiCABEDS models here, we present a comparison between LiCABEDS and linear SVM. Note that nonlinear SVM models may exhibit advantages over linear ones in certain circumstances.²¹ Nevertheless, the selection of nonlinear kernel functions and parameters is computationally expensive and could become a risk factor for overfitting.

This section presents the performance of LiCABEDS and SVM models trained with default parameters (parameter tuning will be discussed later) to show the robustness of these two algorithms. The only parameter in linear SVM is the “C”—constant of regularization. Its default value is 1, meaning that training error and margin size are equally important. The only parameter in LiCABEDS training is the number of training iterations or decision stumps, M . A reasonable default value of M is $2|S|$, so that every possible decision stump could be incorporated in the ensemble classifier, even though only some relevant decision stumps contribute to the prediction in real-world problems. By following this principle, the number of LiCABEDS training iterations for Molprint 2D, FP2, Unity, and Maccs fingerprints are 5000, 2000, 2000, and 400, respectively.

The results of a systematic evaluation of cannabinoid-subtype selectivity prediction are summarized in Figure 3 and Table 1. Figure 3 plots the distribution of precision and recall rate of different computational methods in 20 rounds of calculation. The performance metrics of LiCABEDS and SVM that are trained with different fingerprints are shown. Correspondingly, Table 1 lists the average, standard deviation, geometric mean of precision, and recall rates. Overall, satisfactory results are achieved. LiCABEDS + Unity outperform other combinations for CB1 selectivity prediction, with the highest geometric mean, 67.7. SVM + FP2 lead the CB2 selectivity prediction, with the highest geometric mean, 65.4. Regardless of fingerprint types, the range of precision and recall of LiCABEDS models covers 56.1–77.9% and 49.3–80.0% for CB1 selectivity and 52.1–74.1% and 33.8–74.3% for CB2 selectivity. At the same time, SVM models yield precision of 57.7–76.4% and recall of 54.7–80.0% for CB1 selectivity and precision of 55.7–76.3% and recall of 43.2–77.0% for CB2 selectivity. In spite of the variability, the precision and recall rate stay above 50% in most cases.

In general, all the fingerprints exhibit decent prediction accuracy in LiCABEDS and SVM. According to Table 1, Unity fingerprint is the optimal choice for screening CB1 selective compounds because it achieves a 67% precision rate in both machine learning algorithms. FP2 fingerprints produce the highest geometric mean of precision and recall for CB2 selective compounds (63.5% for LiCABEDS and 65.4%). Also, the best recall rate of CB2 models is established on FP2 fingerprints (63.4% for LiCABEDS and 64.5% for SVM). Figure 3 reveals consistent high precision rate of Molprint 2D fingerprint for both CB1 and CB2 selectivity prediction. Maccs

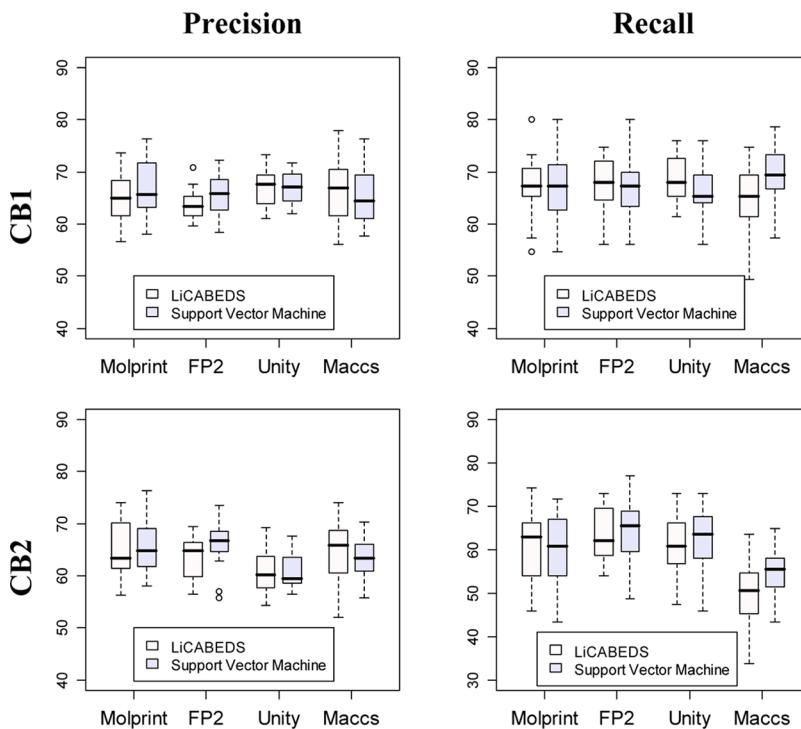


Figure 3. Boxplots display precision and recall rate after 20 rounds of calculation using either LiCABEDS or SVM. LiCABEDS and SVM models are trained with four types of fingerprints (indicated along the X-axis) and default training parameters for the prediction of either CB1 selective or CB2 selective ligands. The Y-axis shows performance metrics in percentage values.

Table 1. Average and Standard Deviation of Precision and Recall Rate after 20 Rounds of alculation^a

molecular fingerprint	classification algorithm	CB1 selectivity			CB2 selectivity		
		precision (%)	recall (%)	GM	precision (%)	recall (%)	GM
Molprint	LiCABEDS	64.8 ± 4.5	67.4 ± 5.7	66.1	64.9 ± 5.1	61.2 ± 7.6	63.0
	SVM	66.8 ± 5.1	66.6 ± 6.3	66.7	65.3 ± 5.1	60.3 ± 7.8	62.8
FP2	LiCABEDS	63.8 ± 2.9	67.8 ± 5.2	65.8	63.7 ± 3.7	63.4 ± 5.9	63.5
	SVM	65.6 ± 3.8	67.2 ± 5.7	66.4	66.3 ± 4.5	64.5 ± 7.0	65.4
Unity	LiCABEDS	67.1 ± 3.9	68.3 ± 4.6	67.7	61.0 ± 4.6	60.7 ± 6.6	60.8
	SVM	67.0 ± 3.0	65.8 ± 5.5	66.4	60.9 ± 3.6	62.6 ± 7.0	61.7
Maccs	LiCABEDS	66.0 ± 6.1	64.6 ± 6.5	65.3	64.4 ± 6.0	50.3 ± 7.0	56.9
	SVM	65.6 ± 5.3	69.5 ± 6.0	67.5	63.4 ± 3.9	54.7 ± 5.2	58.9

^aThe numbers are reported for each combination of machine learning algorithm (SVM or LiCABEDS), molecular fingerprint (Molprint, FP2, Unity, or Maccs), and selectivity type (CB1 or CB2 selective). All the SVM and LiCABEDS models are trained with default parameters. GM: geometric mean. GM = $\sqrt{\text{precision} \times \text{recall}}$.

key is weak at recovering CB2 selective but is surprisingly sufficient for CB1 selective. Altogether, Molprint 2D, FP2, and Unity have subtle different in terms of performance metrics, leaving Maccs key slightly behind.

With four types of fingerprints, the geometric mean (GM) gap of LiCABEDS and SVM models falls in the interval (-2.2, 1.3). Table 1 suggests that SVM generally outperforms LiCABEDS. However, the difference is insignificant when visualized in Figure 3 or compared with the standard deviation in Table 1. Another interesting fact is that the precision and recall rate of CB1 models are uniformly higher than those of CB2 models, even if the computation protocol is identical. The Training Iterations of LiCABEDS section will explain and discuss this in depth.

The evaluation and comparison of modeling strategies are quantified according to precision, recall rate, and geometric mean. Precision should be emphasized when the cost of validating a predicted selective is high. On the other hand, recall rate catches

our attention when many lead compounds are desired at an early stage of virtual screening. However, precision and recall rate are usually negatively correlated. An astringent strategy guarantees a high precision rate by selecting high-confident samples but sacrifices recall rate by ignoring potential true positives. Even if the geometric mean is not perfect as single-number performance measure, it remains a reasonable way to rank screening strategies.

These analyses assume correct selectivity label of training and testing compounds. Because of systematic and random error in bioassays, the reality obviously deviates from this assumption. For example, Huffman et al. pointed out that the CB2/CB1 affinity ratio of a famous cannabinoid, WIN-55,212-2, was reported in a range from 0.6 to 30.¹⁸ In our study, this compound could be CB2 selective or nonselective. The inconsistent selectivity label in training and testing compounds causes confusion and uncertainty. To minimize this effect, we tried to extract bioaffinity values of a cannabinoid ligand from the same literature and use the K_i values reported for the same cell line.

Table 2. Performance of LiCABEDS and SVM Combined with Different Fingerprints after Running Cross Validation

molecular fingerprint	classification algorithm	CB1 selectivity			CB2 selectivity		
		precision (%)	recall (%)	GM	precision (%)	recall (%)	GM
Molprint	LiCABEDS	69.1 ± 5.0	68.5 ± 6.5	68.8	68.4 ± 4.5	58.2 ± 8.9	63.1
	SVM	70.7 ± 4.0	72 ± 7.8	71.3	68.5 ± 3.3	64.5 ± 7.9	66.5
FP2	LiCABEDS	68.0 ± 4.0	68.5 ± 5.8	68.2	69.2 ± 4.1	61.8 ± 6.5	65.4
	SVM	67.5 ± 5.0	70.8 ± 7.4	69.1	71.1 ± 4.0	60.9 ± 6.7	65.8
Unity	LiCABEDS	70.7 ± 4.8	69.2 ± 4.7	69.9	65.0 ± 3.6	59.2 ± 7.4	62.0
	SVM	69.8 ± 3.6	72.2 ± 5.3	71.0	65.2 ± 3.6	60.9 ± 7.8	63.0
MacCs	LiCABEDS	68.0 ± 5.5	66.1 ± 6.2	67.0	63.8 ± 6.5	47.7 ± 8.4	55.2
	SVM	69.0 ± 5.3	72.9 ± 5.6	70.9	65.9 ± 6.6	48.3 ± 10.4	56.4

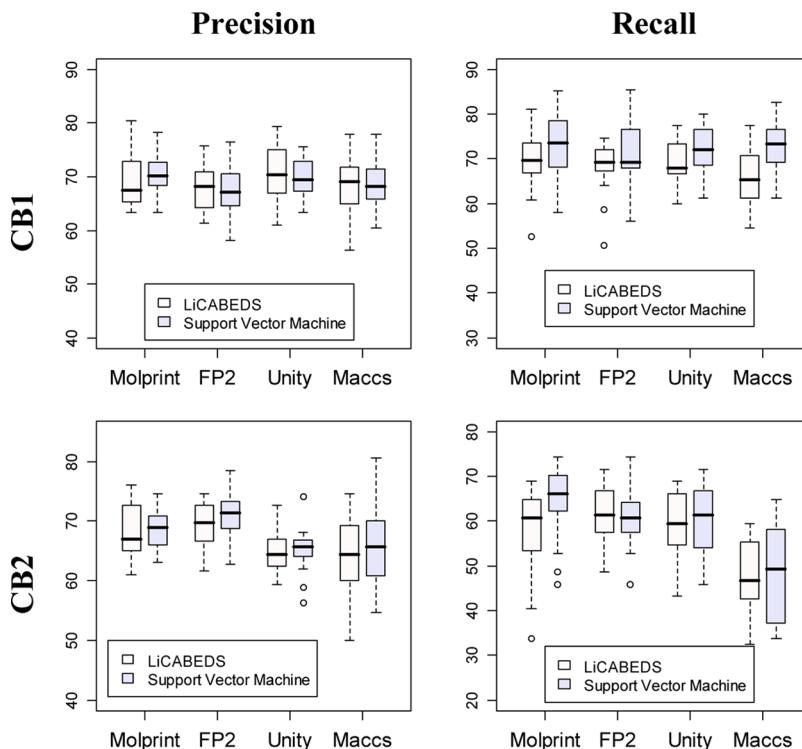


Figure 4. Boxplots display distribution of precision and recall rate after parameter tuning. Parameters of LiCABEDS and SVM are tuned by cross validation.

However, this systematic error cannot be fully eradicated. The CAS number, bioaffinity value, and reference of all the cannabinoid ligands involved in this study can be found in Part 2 of the Supporting Information.

3.2. LiCABEDS and SVM with Cross Validation. The ultimate goals of supervised learning algorithms are to minimize generalization error and to achieve optimal performance for prospective predictions. In this sense, LiCABEDS and SVM are not exceptions. The sole pursuit of reducing training error may lead to poor predictions for new testing samples (overfitting), while the overemphasis on margin maximization may result in lack-of-fit. This paradox is also known as “bias-variance trade-off” in statistics. Cross validation is an effective method to determine optimal model parameters. In this study, the parameters evaluated in cross validation are the number of training iterations in LiCABEDS (M) and C —constant in SVM. For LiCABEDS, $M \in [\text{minstep}, \text{maxstep}]$; maxstep is the default training iterations mentioned previously; $\text{minstep} = 50$ if MacCs is used as descriptor, otherwise $\text{minstep} = 100$. $C \in \{2^i : i = -16, -15, \dots, 15, 16\}$. Although the parameter space of LiCABEDS is much larger than SVM, dynamic programming

makes the cross validation in LiCABEDS as rapid as training a model. The M and C that perform best on cross-validation data are selected for model training. The performance of these models is further evaluated on the same testing sets, and results are shown in Table 2 and Figure 4.

Cross validation brings consistent improvement for both LiCABEDS and SVM, except for CB2 selectivity prediction using MacCs fingerprint. The geometric mean (GM) of LiCABEDS grows $\Delta 1.7$ – 2.7 for CB1 selectivity prediction, and $\Delta -1.7$ to 1.9 for CB2 selectivity prediction. Correspondingly, the geometric mean (GM) of SVM grows $\Delta 2.7$ – 4.6 for CB1 selectivity prediction, and $\Delta -2.5$ to 3.7 for CB2 selectivity prediction. Both algorithms are robust enough to deliver satisfactory predictions with default model parameters. Even if cross validation has a positive impact, the improvement is inconclusive when compared with performance variance. SVM seems to be more sensitive to the choice of parameters than LiCABEDS because the increment of its geometric mean is relatively higher.

SVM + Molprint 2D outperforms other models in both CB1 and CB2 selectivity prediction after parameter tuning for its highest GM (71.3 and 66.5). Figure 4 shows that the precision

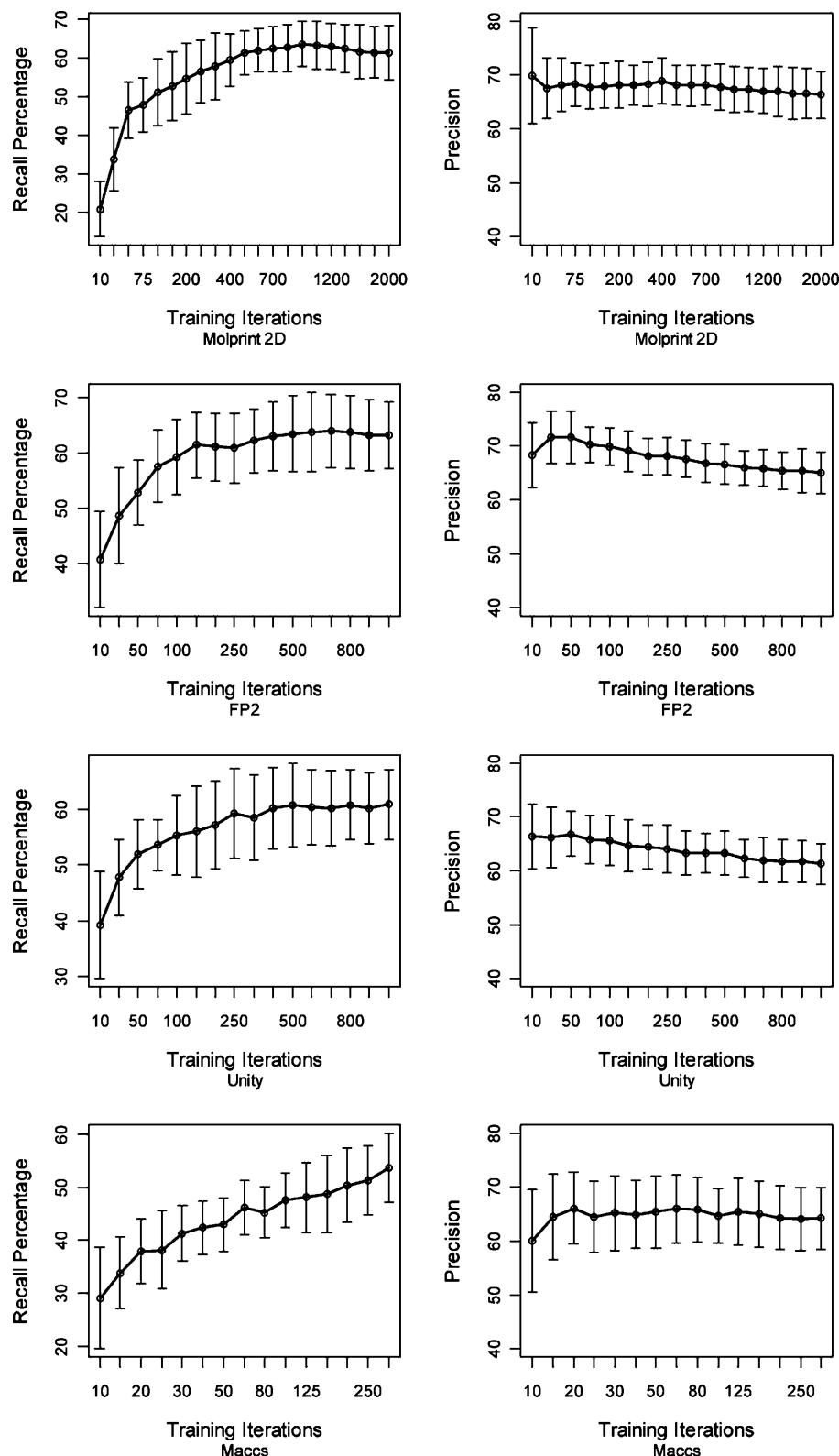


Figure 5. Boxplots show the average and standard deviation interval of recall and precision rate of CB2 selectivity models versus model training iterations. Average and standard deviation are calculated on the basis of 20 rounds of test calculation. Each row represents a specific fingerprint type. The Y-axis of each plot is the percentage. Values along the X-axis represent M . Note that the growth of X-axis values is not linear.

rates of LiCABEDS and SVM are similar among all fingerprints, but SVM is capable of retrieving more selective compounds than LiCABEDS, especially for CB1 selective compounds. In the end, the GM of SVM is 0.4–3.9 higher than that of LiCABEDS.

3.3. Training Iterations of LiCABEDS. Training iterations (M) of LiCABEDS are directly related to model complexity as every round of training adds one more “weak classifier” to the ensemble model. Parameter M was thoroughly discussed in our previous study,¹⁶ and the hypothesis was that large M produced

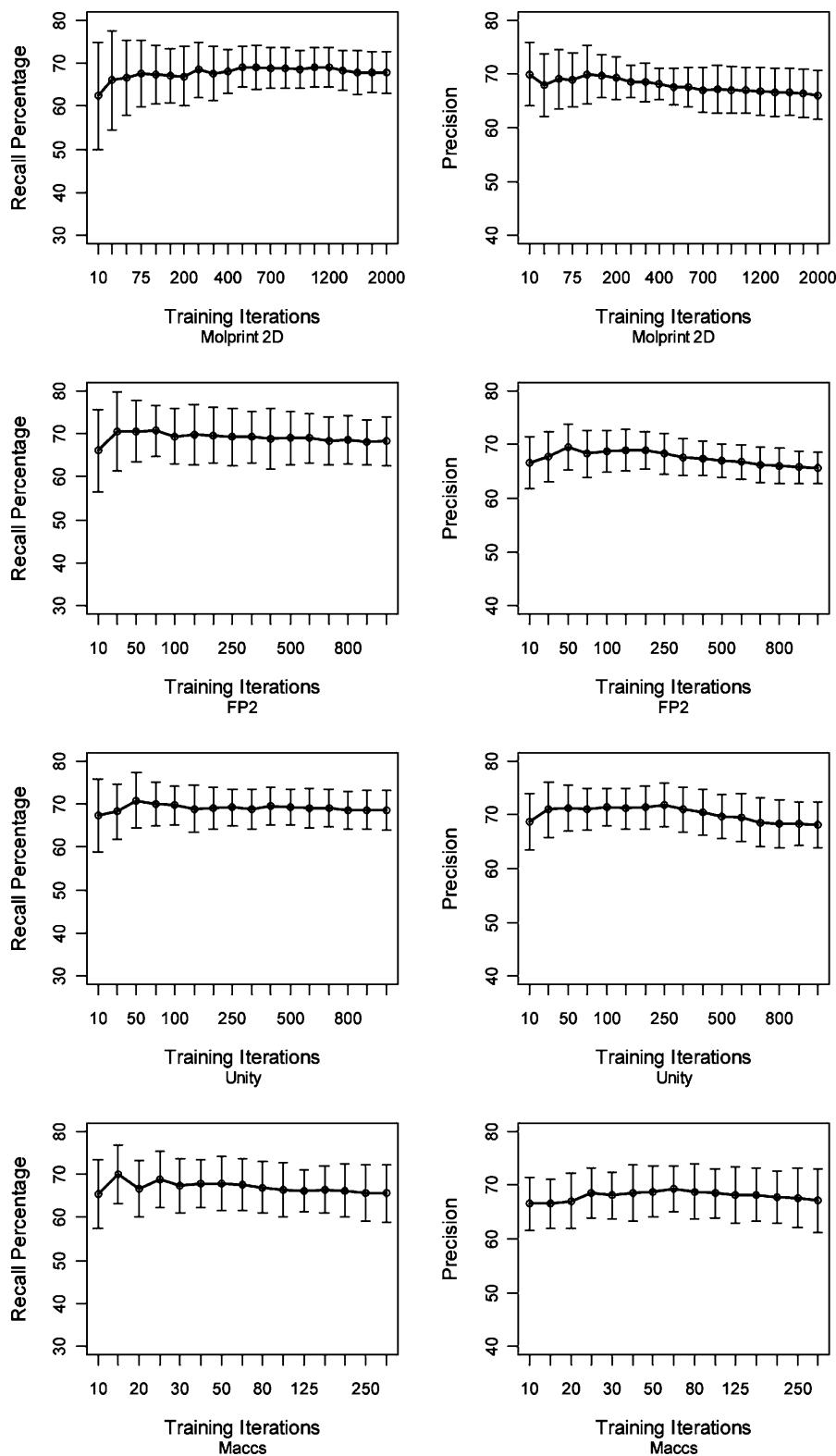


Figure 6. Average and standard deviation interval of recall and precision rate of CB1 selectivity models compared with model training iterations.

close-to-optimal models. Using cross validation might improve prediction performance, but its effect was statistically insignificant. The results in the LiCABEDS and SVM with Cross Validation section also support this hypothesis. Thus, assigning a relatively large value to M not only guarantees model convergence but also avoids costly cross-validation procedures and potential overfitting of cross-validation data.

Apart from machine learning language, the meaning of M can be also interpreted in applied domain knowledge. This section is primarily focused on how M affects the recovery of selective ligands.

Naturally selective compounds are supposed to be overwhelmed by nonselective ones, which is also true for our training and testing sets. Thus, the LiCABEDS training algorithm

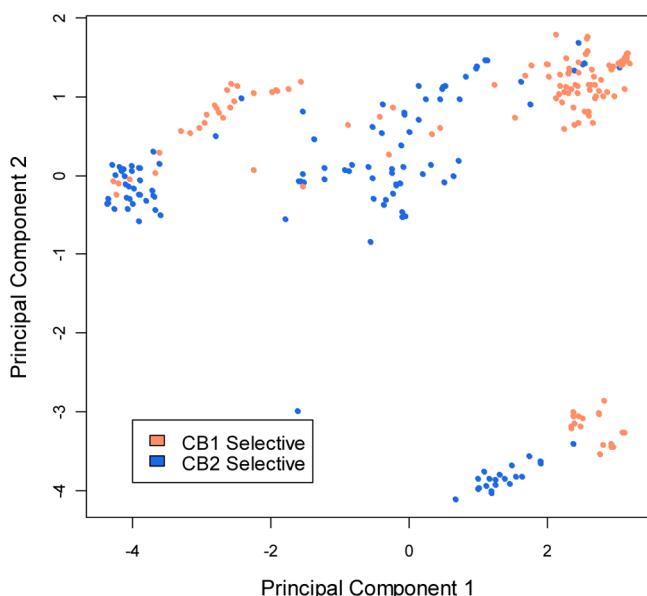


Figure 7. 2D scatter plot shows the spatial arrangement of selective ligands in the coordinates of two principal components. The principal components are solved according to Maccs fingerprints of all selective ligands. The X-axis and Y-axis represent the two most significant components. CB1 and CB2 selective ligands are represented in different colors.

faces unbalanced data. Treating each training sample equally at the initialization stage, the training algorithm first assures the correct classification of nonselective compounds even at the cost of misclassification of selective ones. As the majority of the samples are nonselective, this is an effective technique for minimizing training errors when the classification power of LiCABEDS is limited. Later when the model complexity grows, the training algorithm aims at recovering selective ligands from the training pool by picking up discriminative features and building “decision stumps” accordingly. This process is visualized in Figure 5.

Figure 5 plots the average and standard deviation of recall and precision rates on the testing data sets as a function of training iterations, M . As shown in the left column, the recall rate steadily increases as M grows. For example, with Molprint

2D fingerprint, the recall rate starts from 20%. It converges to a plateau (approximately 60%) when M is more than 500. A similar trend can be observed in other plots. Meanwhile, the precision is more stable compared to recall rate. The precision rate mainly fluctuates from 60% to 70% for Molprint 2D, FP2, Unity fingerprints, and from 50% to 60% for Maccs fingerprint. The precision rate seems to be negatively correlated with training iterations, meaning that a large M reduces precision of the prediction. To explain this, easy-to-classify selective ligands are correctly labeled at the beginning stage of training. As training continues, the algorithm tries to recover more selective ligands that are harder to classify. At the same time, more nonselective ligands also may be predicted as selective, which reduces precision. However, the positive effect of a large M on recall rate is obviously more significant than its negative effect on precision. Therefore, training error is effectively minimized in the early stage of training. As recall rate reaches the plateau, adding more “decision stumps” only impairs precision rate. The value that seeks a balanced trade-off between these two metrics can be determined by cross validation, which has been well explained in the previous section.

The precision and recall rate of CB1 selectivity models are very different from those of the CB2 selectivity models. Their values as a function of training iterations are displayed in Figure 6. One major difference is that the recall rate of CB1 selectivity models reaches the plateau when M is about 50 for Molprint 2D, FP2, and Unity fingerprints. The uptrend of recall rate is almost not observable with Maccs fingerprint. Figure 6 suggests that precision rate gradually reduces as M grows, which is similar to Figure 5. Thus, a large M is less favored compared to CB2 selectivity models. This also explains why cross validation is more beneficial to CB1 selectivity prediction than CB2 selectivity prediction. As mentioned in the LiCABEDS and SVM with Cross Validation section, cross validation enhances the geometric mean of CB1 selectivity models by $\Delta 1.7\text{--}2.7$, but $\Delta -1.7$ to 1.9 for CB2 selectivity models. Furthermore, the average of optimal M for CB1 selectivity prediction is 283 with Molprint 2D fingerprint. The average for CB2 selectivity prediction is 736 with the same fingerprint. A straightforward conclusion is that the complexity of CB2 selectivity models is higher than the CB1 models. It also indirectly suggests that the

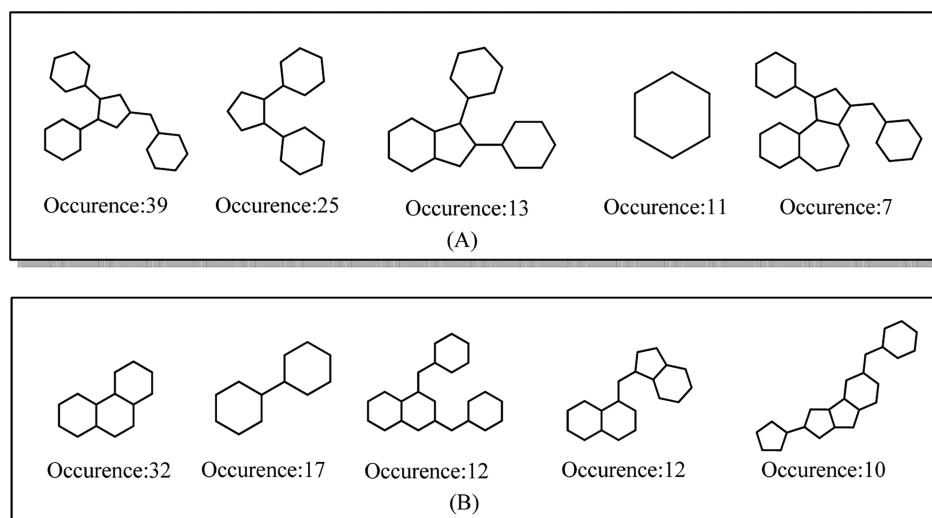


Figure 8. Top five scaffolds in CB1 selective compounds (A) and CB2 selective compounds (B).

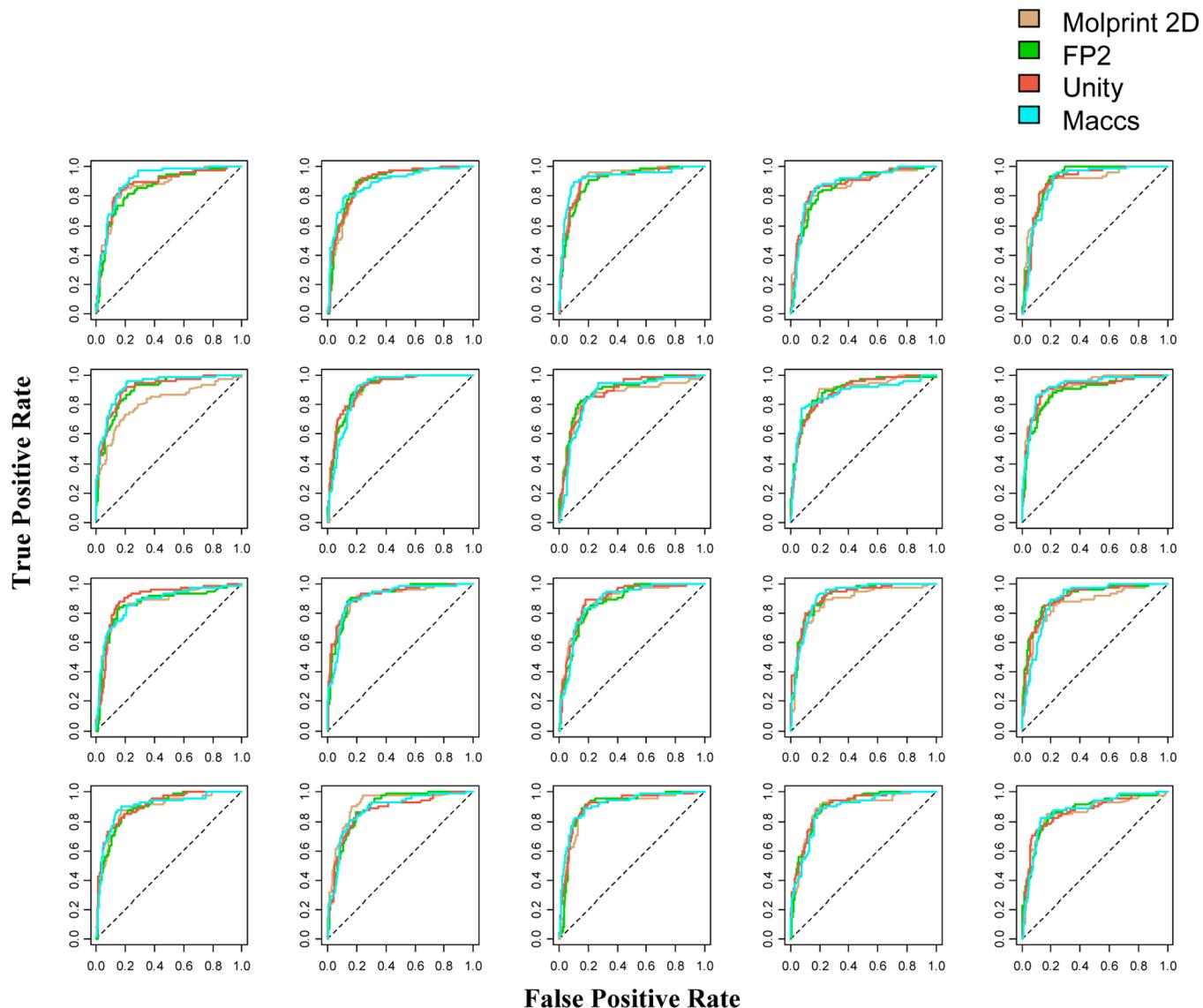


Figure 9. ROC curves of LiCABEDS models for the prediction of CB1 selectivity on 20 testing sets. Each individual plot contains four curves that represent different fingerprints.

structure of CB2 selective ligands is more diverse. More factors therefore need to be considered in the classifier.

A universal quantitative definition of structure diversity does not exist because it can be evaluated according to aspects of criteria, such as molecular weight and number of rings. We attempt to illustrate the structure diversity of cannabinoid selective ligands in both fragments and scaffolds. Principal component analysis (PCA) is an unsupervised learning skill that transforms original data into a new orthogonal coordinate system in order to capture the maximum variance. In the new coordinate system, the first component has the largest possible variance; the second component has the second largest possible variance, and so on. Figure 7 displays the first two components of Maccs fingerprints of all selectivity ligands. PCA reduces the dimensionality of 166 bit fingerprint for visualization while maintaining minimum information loss. In Figure 7, CB1 selective ligands form three clusters that are approximately centered at (-3,1), (2,1), and (2,-3). Only a few points are scattered near the origin. Similarly, CB2 selective ligands also form three clusters that are approximately centered at (-4,0), (0,0), and (-3,-4). The radius of each cluster is associated with the

structure variance of the compounds in the cluster. Although the distributions of CB1 and CB2 selective ligands in the PCA plot are different, it is difficult to distinguish these two types of ligands according to only two principal components.

To assess structure diversity in a different aspect, the scaffolds of CB1 and CB2 selective compounds were generated according to the protocol described in the Materials, Methods, and Calculations section. Results showed that 149 CB1 selective compounds possessed 22 scaffolds. Meanwhile, 147 CB2 selective compounds were reduced to 38 scaffolds. Figure 8 lists the five most populated scaffolds in each compound category. An interesting finding is that the top three CB1 scaffolds have significant overlapping. The second scaffold in Figure 8(A) therefore is a substructure of the other two. On the other hand, CB2 scaffolds have relatively more variation. These observations also support our hypothesis that CB2 selective compounds could be more structurally diverse than CB1 selective compounds. However, the greater number of scaffolds for CB2 selective compounds is more likely to be attributable to the fact that companies and academic groups are actively looking for

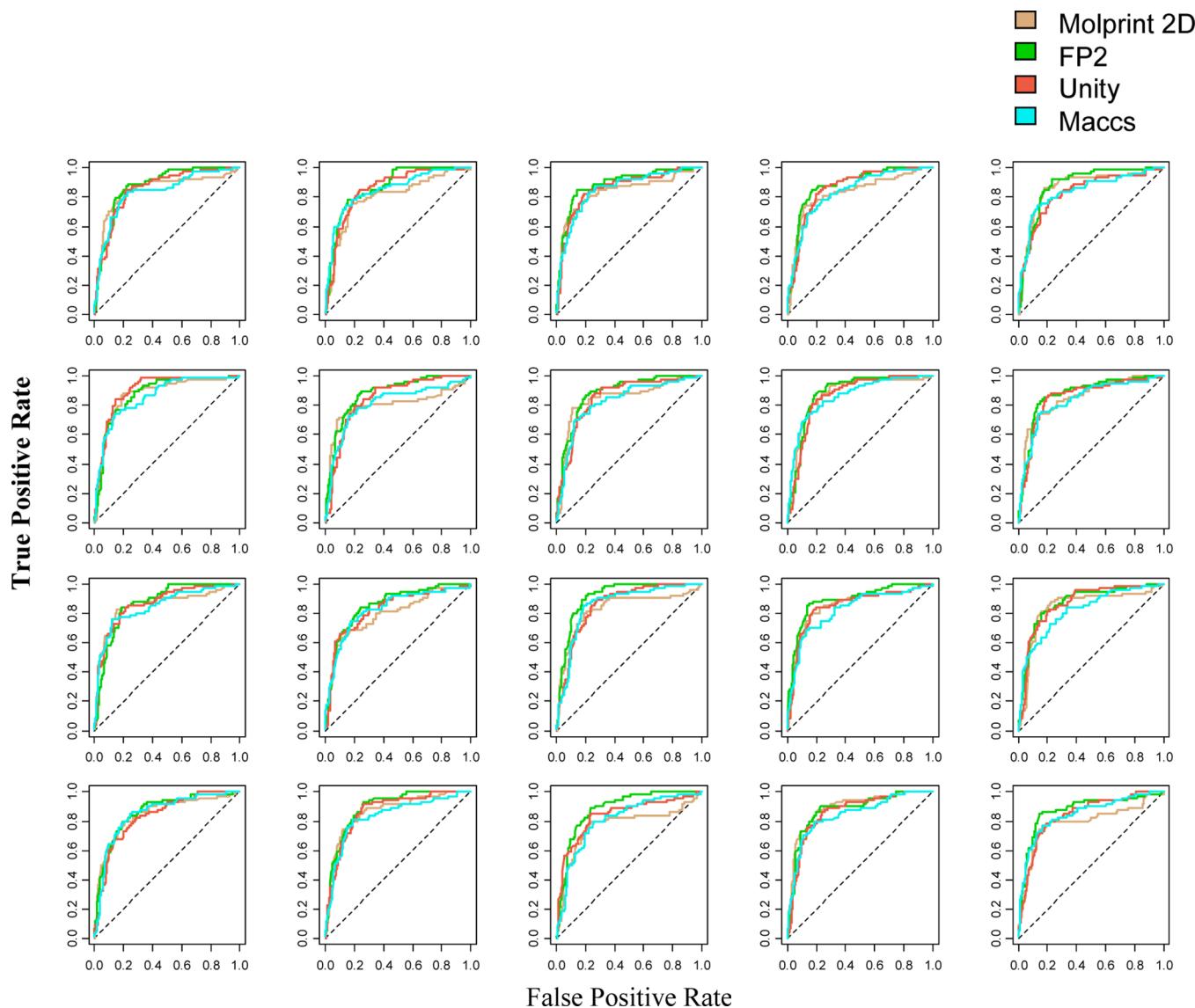


Figure 10. ROC curves of LiCABEDS models for the prediction of CB2 selectivity on 20 testing sets. Each individual plot contains four curves that represent different fingerprints.

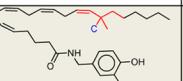
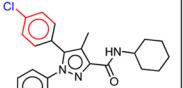
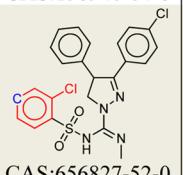
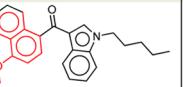
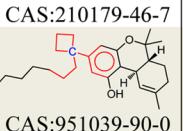
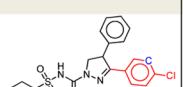
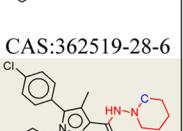
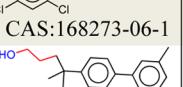
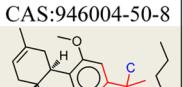
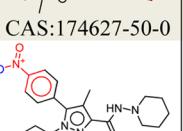
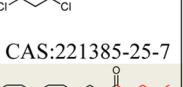
CB2 selective compounds. This scaffold information could provide useful clues for the design of CB selective compounds.

3.4. ROC Analysis of LiCABEDS Models. Sensitivity and specificity are classical statistical measures of the performance of binary classification. The ROC curve visualizes the true positive rate (sensitivity) as a function of false positive rate (1-specificity) by changing the classification boundary value. Visual inspection of enrichment of positive samples is straightforward by examining ROC curve. Besides the convenience of visualization, the area under curve (AUC) serves as a quantitative performance measure of enrichment of relevant samples. The ROC curve of a random guess function (curve in dashed line) yields an AUC of 0.5. A perfect classification function has an AUC of 1.

As discussed in our original LiCABEDS paper, $A = \sum_m a_m y_m(\mathbf{x}, i_m, t_m)$ indicates the degree of confidence in each prediction. The default classification boundary is $A = 0$. Thus, changing the decision criterion generates a set of TPR and FPR values and allows inspection of the enrichment of selective ligands. Figures 9 and 10 plot the ROC curves of LiCABEDS models for the prediction of CB1 and CB2 selectivity with four types of fingerprints. Out of 20 rounds of test calculation,

the averages for AUC of CB1 selectivity models (Figure 10) are 0.883, 0.893, 0.897, and 0.895 with Molprint 2D, FP2, Unity, and Maccs fingerprints, respectively. Correspondingly, the averages for AUC of CB2 selectivity models are 0.839, 0.880, 0.855, and 0.839 with the same set of fingerprints, respectively. Visualization of these two figures and AUC values both confirm that LiCABEDS effectively enriches selective ligands. In addition, the ROC curve also indicates that large $|A|$ suggests high probability of observing relevant samples because the left region of ROC curve corresponds to astringent decision criteria. Figures 9 and 10 reveal that the performance of fingerprints involved in LiCABEDS models is comparable, especially for CB1 selectivity models. Molprint 2D models produce relatively low AUC for CB2 selectivity prediction (average 0.839), which seems to be contradictory to our previous findings. In real world problems, a practical decision boundary corresponds to the left region of the ROC curve because large FPR is mostly disfavored and remains unexplored. The slope of the left part of ROC curve is more important than AUC. As shown in Figures 9 and 10, also by AUC calculation, the LiCABEDS models for CB1 selectivity prediction exhibit better enrichment than CB2 models.

Table 3. Important Features for the Classification of CB1 Selectivity

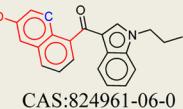
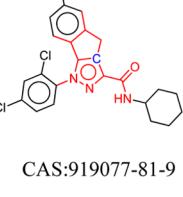
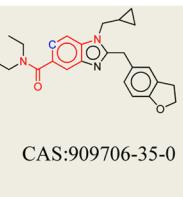
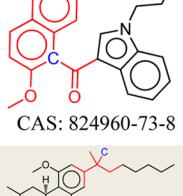
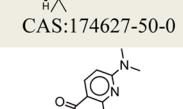
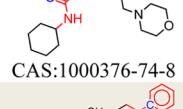
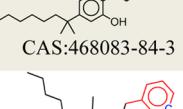
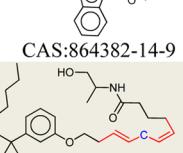
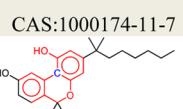
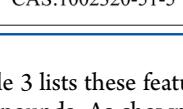
Feature	Index	Average Weight in 20 models	Number of CB1 selective ligands having this feature	Number of Non-CB1 selective ligands having this feature	Representative compound	
					Structure	CB1/CB2 activity and reference
0;0-1-0;1-2-0;1-1-1;2-1-0;2-1-1;14;0-1-2;1-2-2;2-2-2;	697	3.58	6	1	 CAS:322399-51-9	261.8nM/>1000nM PMID:11181068
2;0-2-2;1-2-2;2-2-2-2;2-1-14;	1446	2.27	107	62	 CAS:158940-64-8	2.46nM/228nM PMID:12061874
2;0-2-2;0-1-7;1-1-0;1-3-2;2-4-2;	99	2.11	20	4	 CAS:656827-52-0	75.4nM/>1000nM PMID:14736243
0;0-3-0;0-1-2;1-3-0;1-2-2;2-3-0;2-2-2;	546	1.98	1	14	 CAS:210179-46-7	1.2nM/12.4nM PMID:10940540
2;0-2-2;1-2-2;2-1-1-14;2-1-1;2-2-2;	1329	1.92	2	7	 CAS:951039-90-0	2.7nM/52.3nM PMID:17672444
2;0-2-2;1-2-2;2-1-1-14;2-1-1;2-2-2;	1809	1.81	24	10	 CAS:362519-28-6	52.6nM/>1000nM PMID:14736243
0;0-1-0;0-1-4;1-2-0;1-1-27;2-2-0;2-1-1;7;0-1-0;1-1-0;2-1-0;	1561	-1.54	30	21	 CAS:168273-06-1	6.18nM/313nM PMID:12061874
0;0-1-0;1-2-0;1-1-2;2-1-0;2-2-2;	1974	-1.60	2	8	 CAS:946004-50-8	1100nM/56nM PMID:17507224
31;0-1-18;1-1-2;1-1-8;2-2-2;	849	-1.61	2	129	 CAS:174627-50-0	1043nM/6.4nM PMID:10188977
0;0-2-0;0-1-7;1-1-27;2-1-1;	430	-2.28	0	4	 CAS:221385-25-7	57.5nM/252nM PMID:10052983
0;0-2-0;0-1-7;1-1-27;2-1-1;	1691	-2.29	0	4	 CAS:157182-47-3	20nM/12.5nM PMID:9774174

This finding also agrees with the previous hypothesis that the structure patterns of CB2 selective ligands are more diverse.

3.5. Important Features for CB1 and CB2 Selective Ligands.

To demonstrate further the advantage of LiCABEDS,

Table 4. Important Features for the Classification of CB2 Selectivity

Feature	Index	Average Weight in 20 models	Number of CB2 selective ligands having this feature	Number of Non-CB2 selective ligands having this feature	Representative Compound	
					Structure	CB1/CB2 activity and Reference
2;0-2- 2;1-3- 2;2-1- 1;2-4- 2;2-1-7;	2089	3.11	4	0	 CAS:824961-06-0	2358nM/138nM PMID:15582455
2;0-1- 0;0-2- 2;1-1- 1;1-2- 2;1-2- 10;2-5- 2;2-1- 8;2-2- 10;2-1- 27;	1365	2.42	9	4	 CAS:919077-81-9	900nM/7.6nM PMID:17149879
2;0-2- 2;1-1- 1;1-2- 2;2-2- 2;2-1- 8;2-1- 10;2-1- 27;	94	2.35	31	9	 CAS:909706-35-0	>5000nM/84nM PMID:2006:66518 9
2;0-1- 1;0-2- 2;1-4- 2;1-1- 7;1-1- 8;2-1- 0;2-6-2;	1337	2.08	3	0	 CAS: 824960-73-8	3788nM/80nM PMID:15582455
0;0-1- 0;1-2- 0;1-1- 2;2-1- 0;2-2-2;	849	1.99	43	88	 CAS:174627-50-0	1043nM/6.4nM PMID:10188977
8;0-1- 1;1-1- 2;1-1- 27;2-1- 0;2-2-2;	309	1.79	20	7	 CAS:1000376-74-8	>5600nm/94nM PMID:17942307
2;0-2- 2;1-1- 0;1-2- 2;2-2- 0;2-2-2;	1622	1.68	9	3	 CAS:468083-84-3	144nM/9nM PMID:11961073
2;0-2- 2;1-2- 2;1-1- 15;2-1- 0;2-2-2;	2002	-1.14	3	8	 CAS:864382-14-9	39nM/76nM PMID:16005223
0;0-2- 1;1-2- 1;2-2-0;	98	-1.15	1	61	 CAS:1000174-11-7	8.55nM/21.48nM PMID:17827022
2;0-3- 2;1-4- 2;1-2- 7;2-2- 0;2-4-2;	28	-1.20	0	7	 CAS:1002320-51-5	2.6nM/4.8nM PMID:18038967

we analyzed the important features that contributed significantly to distinguishing the selective and nonselective compounds.

Table 3 lists these features for the classification of CB1 selective compounds. As shown in Table 3, features 697, 1446, 99, 546,

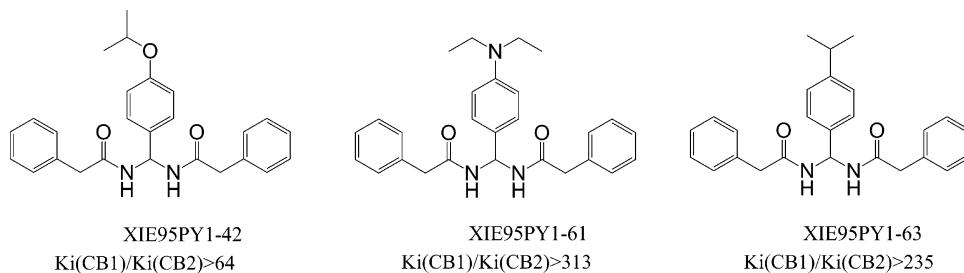


Figure 11. Structures of newly synthesized CB2 selective ligands.

1329, and 1809 are favored by CB1 selective ligands. For example, feature 697 (0;0-1-0;1-2-0;1-1-1;2-1-0;2-1-1) has an average weight (a_m) of 3.58 in 20 LiCABEDS models for CB1 selectivity. This feature can be translated to a substructure of a central sp^3 carbon atom (C.3, colored blue in the example compound with CAS ID 322399-51-9) neighbored by one sp^3 carbon atom and surrounded by two sp^3 carbon atoms, one sp^2 (C.2) carbon atom located two bonds away, and one sp^2 and one sp^3 atom located three bonds away (colored red). Six of 149 CB1 selective compounds have this feature, while only 1 of 554 non-CB1 selective compounds has this feature. Such features may be useful for guiding ligand optimization. On the other hand, features 1561, 1974, 849, 430, 1691, 1365, and 1935 are disfavored by CB1 selective ligands. For example, feature 1935 (2;0-2-2;1-2-2;2-3-2) has average a_m as -2.91 and can be translated to a substructure of a central aromatic carbon atom (C.ar, colored blue in the example compound with CAS ID 1000376-80-6) neighbored by 2 aromatic carbon atoms and surrounded by 2 and 3 aromatic carbon atoms located two and three bonds away, respectively (colored red). Five of 149 CB1 selective compounds have this feature, while 157 of 554 non-CB1 selective compounds has this feature.

Table 4 listed the most important features for characterizing of CB2 selective compounds. As shown in Table 4, features 2089, 1446, 99, 546, 1329, and 1809 are favored by CB2 selective ligands. For example, feature 2089 (2;0-2-2;1-3-2;2-1-1;2-4-2;2-1-7) is assigned an average a_m of 3.11 in 20 LiCABEDS models for CB2 selectivity. This feature can be translated to a substructure of a central aromatic carbon atom (C.ar, colored blue in the example compound with CAS ID 824961-06-0) neighbored by two aromatic carbon atoms and surrounded by three aromatic carbon atoms located two bonds away, one sp^2 , 4 aromatic carbon atoms, and one sp^3 oxygen atom located three bonds away (colored red). Four of 147 CB2 selective compounds have this feature, while none of 556 non-CB2 selective compounds has this feature. On the other hand, features 2002, 98, 28, 2403, 1148, and 1840 are nonfavored by CB2 selective ligands. For example, feature 1840 (2;0-2-2;1-1-0;1-2-2;2-3-0;2-3-2;) is assigned an average a_m of -1.29 and can be translated to a substructure of a central aromatic carbon atom (C.ar, colored blue in the example compound with CAS ID 174627-59-9) neighbored by two aromatic carbon atoms and surrounded by 1 sp^3 carbon atom, two aromatic carbon atoms located two bonds away, 3 sp^3 carbon atoms, and two aromatic carbon atoms located three bonds away (colored red). Six of 147 CB2 selective compounds have this feature, while 33 of 556 non-CB2 selective compounds have this feature. If we compare the features in Tables 3 and 4, feature 849 and feature 1365 are favored by CB2 selective compounds, while they are disfavored by CB1 selective ligands.

The results again suggest that those features might be related to ligand selectivity and ligand–protein interaction. The interpretability of LiCABEDS models is rooted in the explanation of each “decision stump”, especially the highly weighted ones. Therefore, LiCABEDS models can be understood and interpreted easily. This finding could potentially guide chemical modification to achieve a better pharmacological or physicochemical profile.

3.6. Prediction of Selectivity of Novel Compounds.

Besides retrospective validation, a LiCABEDS model was developed to predict the CB2 selectivity of 12 novel compounds¹⁴ (Calculations section) before they were experimentally tested. The training set consisted of all the CB2 selective and CB2 non-selective compounds in our cannabinoid database. The training iteration was set to 2000 with Molprint 2D as the molecular descriptor. The LiCABEDS model predicted XIE95PY1-63 (Figure 11) as CB2 selective, while predicting the other 11 compound as CB2 nonselective. Later, the experimental validation showed that the three structures in Figure 11 (XIE95PY1-42, XIE95PY1-61, and XIE95PY1-63) were CB2 selective, and that the remaining compounds were either nonselective or inactive. The structures of the 12 compounds and their bioactivity profiles are summarized in Table S1 of the Supporting Information. Therefore, LiCABEDS successfully discovered XIE95PY1-63 but missed the other two potential selective compounds. These two compounds contained a novel scaffold that was not present in training data. However, the prediction model still recovered a true selective without reporting any false positive. In the next step, these compounds were added to the training data to derive a new CB2 selectivity model. The new model was used to predict the CB2 selectivity of another 26 compounds. The structures of the 26 compounds, their experimental CB2 selectivity and predicted CB2 selectivity are listed in Table S2 of the Supporting Information. LiCABEDS identified 5 CB2 selective compounds out of 26. Among them, four compounds were true CB2 selective. In addition, there were a total of seven CB2 selective compounds according to a binding assay. Thus, LiCABEDS achieved a precision of 4/5 and a true positive rate of 4/7, which was consistent with previous retrospective validation. Because the first batch of 12 compounds were added into training data, the information regarding the novel scaffold was encoded in prediction model, and the prediction performance was improved. This result also suggested that high-quality training information is the prerequisite of a supervised learning algorithm. Even though the number of testing compounds was limited, this case study illustrated that the expense of drug discovery projects could be minimized by introducing calculation of in silico ligand profiling.

4. CONCLUSION

When introduced as a general purpose ligand classifier, LiCABEDS demonstrated its application for the classification

of 5-HT_{1A} ligand functionality.¹⁶ This article reports a follow-up study of LiCABEDS algorithm in the prediction of cannabinoid ligand selectivity. In-depth discussions are presented on LiCABEDS theoretical framework, performance measures compared with SVM, vulnerability to overfitting, choice of training parameter, and prospective validation. The results show that LiCABEDS models effectively recover 60–70% selective compounds from testing data sets, while maintaining satisfactorily low false positive rates. All fingerprints deliver decent performance metrics, thereby showing the flexibility of LiCABEDS to adapt assorted hypothesis space. More importantly, LiCABEDS successfully identifies a CB2 selective compound out of 12 newly synthesized ones without any false positive error. Another advantage of LiCABEDS is straightforward parameter setting. Default training iteration significantly reduces calculation time by omitting a costly cross-validation procedure but conveys close-to-optimal models. LiCABEDS is not a black-box method because models are interpretable by examining individual “decision stumps” that are components of ensemble models. Furthermore, the investigation of LiCABEDS models provides insight into structure diversity of cannabinoid selective ligands. The correlation between performance metrics and model complexity reveals that reported CB2 selective ligands are more diverse than CB1 selective ligands. This hypothesis is later supported by principal component analysis of fragments and analysis of compound scaffolds. It is also consistent with some performance metrics, such AUC of ROC curve. This suggests that more binding modes form ligand–protein complexes for CB1 than for CB2 receptors. We conclude that LiCABEDS has potential to model the pharmacological properties that are not well addressed by traditional QSAR methodology. LiCABEDS also could guide screening strategy in early stages of drug development.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is divided into two parts: Part 1 lists the structures of newly-synthesized compounds and their binding profile to CB1 and CB2 receptors; the structures, bioactivity, and reference of all cannabinoid ligands involved in retrospective validation are given in Part 2. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: xix15@pitt.edu. Tel.: +1-412-383-5276. Fax: +1-412-383-7436.

Notes

The authors declare no competing financial interest.

Author Contributions

[#]These authors contributed equally.

■ ACKNOWLEDGMENTS

This project is supported by Grants NIH R01 DA025612 and NIGMS P50-GM067082.

■ REFERENCES

- (1) (a) Adams, C. P.; Brantner, V. V. Estimating the cost of new drug development: Is it really 802 million dollars? *Health Aff. (Millwood)* **2006**, *25* (2), 420–428. (b) DiMasi, J. A. The value of improving the productivity of the drug development process: Faster times and better decisions. *Pharmacoeconomics* **2002**, *20* (Suppl 3), 1–10. (c) DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The price of innovation: New estimates of drug development costs. *J. Health Econ.* **2003**, *22* (2), 151–185.
- (2) (a) Arrowsmith, J. Trial watch: Phase II failures: 2008–2010. *Nat. Rev. Drug Discovery* **2011**, *10* (5), 328–329. (b) Arrowsmith, J. Trial watch: Phase III and submission failures: 2007–2010. *Nat. Rev. Drug Discovery* **2011**, *10* (2), 87–87.
- (3) Wassermann, A. M.; Geppert, H.; Bajorath, J. Searching for target-selective compounds using different combinations of multiclass support vector machine ranking methods, kernel functions, and fingerprint descriptors. *J. Chem. Inf. Model.* **2009**, *49* (3), 582–592.
- (4) (a) Spring, D. R. Chemical genetics to chemical genomics: Small molecules offer big insights. *Chem. Soc. Rev.* **2005**, *34* (6), 472–482. (b) Williams, R.; Brown, K. Chemical genetics and genomics and drug discovery. *Drug News Perspect.* **2005**, *18* (4), 285–288.
- (5) Devane, W. A.; Dysarz, F. A., 3rd; Johnson, M. R.; Melvin, L. S.; Howlett, A. C. Determination and characterization of a cannabinoid receptor in rat brain. *Mol. Pharmacol.* **1988**, *34* (5), 605–613.
- (6) Munro, S.; Thomas, K. L.; Abu-Shaar, M. Molecular characterization of a peripheral receptor for cannabinoids. *Nature* **1993**, *365* (6441), 61–65.
- (7) Xie, X. Q.; Chen, J. Z.; Billings, E. M. 3D structural model of the G-protein-coupled cannabinoid CB2 receptor. *Proteins: Struct., Funct., Bioinf.* **2003**, *53* (2), 307–319.
- (8) Yang, P.; Wang, L.; Xie, X. Q. Latest advances in novel cannabinoid CB2 ligands for drug abuse and their therapeutic potential. *Future* **2012**, *4* (2), 187–204.
- (9) Fernandez-Ruiz, J.; Pazos, M. R.; Garcia-Arencibia, M.; Sagredo, O.; Ramos, J. A. Role of CB2 receptors in neuroprotective effects of cannabinoids. *Mol. Cell. Endocrinol.* **2008**, *286* (1–2), S91–96.
- (10) McKallip, R. J.; Lombard, C.; Fisher, M.; Martin, B. R.; Ryu, S.; Grant, S.; Nagarkatti, P. S.; Nagarkatti, M. Targeting CB2 cannabinoid receptors as a novel therapy to treat malignant lymphoblastic disease. *Blood* **2002**, *100* (2), 627–634.
- (11) Ofek, O.; Karsak, M.; Leclerc, N.; Fogel, M.; Frenkel, B.; Wright, K.; Tam, J.; Attar-Namdar, M.; Kram, V.; Shohami, E.; Mechoulam, R.; Zimmer, A.; Bab, I. Peripheral cannabinoid receptor, CB2, regulates bone mass. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (3), 696–701.
- (12) Christensen, R.; Kristensen, P. K.; Bartels, E. M.; Bliddal, H.; Astrup, A. Efficacy and safety of the weight-loss drug rimonabant: A meta-analysis of randomised trials. *Lancet* **2007**, *370* (9600), 1706–1713.
- (13) Felder, C. C.; Joyce, K. E.; Briley, E. M.; Mansouri, J.; Mackie, K.; Blond, O.; Lai, Y.; Ma, A. L.; Mitchell, R. L. Comparison of the pharmacology and signal transduction of the human cannabinoid CB1 and CB2 receptors. *Mol. Pharmacol.* **1995**, *48* (3), 443–450.
- (14) Yang, P.; Myint, K. Z.; Tong, Q.; Feng, R.; Cao, H.; Almehizia, A. A.; Alqarni, M. H.; Wang, L.; Bartlow, P.; Gao, Y.; Gertsch, J.; Teramachi, J.; Kurihara, N.; Roodman, G. D.; Cheng, T.; Xie, X. Q. Lead discovery, chemistry optimization, and biological evaluation studies of novel biamide derivatives as CB(2) receptor inverse agonists and osteoclast inhibitors. *J. Med. Chem.* **2012**, *55* (22), 9973–9987.
- (15) (a) Ashton, J. C.; Wright, J. L.; McPartland, J. M.; Tyndall, J. D. Cannabinoid CB1 and CB2 receptor ligand specificity and the development of CB2-selective agonists. *Curr. Med. Chem.* **2008**, *15* (14), 1428–1443. (b) Chen, J. Z.; Wang, J.; Xie, X. Q. GPCR structure-based virtual screening approach for CB2 antagonist search. *J. Chem. Inf. Model.* **2007**, *47* (4), 1626–1637. (c) Raduner, S.; Majewska, A.; Chen, J.-Z.; Xie, X.-Q.; Hamon, J.; Faller, B.; Altmann, K.-H.; Gertsch, J. Alkylamides from echinacea are a new class of cannabinomimetics: Cannabinoid type 2 receptor-dependent and -independent immunomodulatory effects. *J. Bio. Chem.* **2006**, *281* (20), 14192–14206.
- (16) Ma, C.; Wang, L.; Xie, X. Q. Ligand classifier of adaptively boosting ensemble decision stumps (LiCABEDS) and its application on modeling ligand functionality for 5HT-subtype GPCR families. *J. Chem. Inf. Model.* **2011**, *51* (3), 521–531.
- (17) Freund, Y.; Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139.

- (18) Huffman, J. W. The search for selective ligands for the CB2 receptor. *Curr. Pharm. Des.* **2000**, *6* (13), 1323–1337.
- (19) Reggio, P. H.; Wang, T.; Brown, A. E.; Fleming, D. N.; Seltzman, H. H.; Griffin, G.; Pertwee, R. G.; Compton, D. R.; Abood, M. E.; Martin, B. R. Importance of the C-1 substituent in classical cannabinoids to CB2 receptor selectivity: Synthesis and characterization of a series of O₂-propano-delta8-tetrahydrocannabinol analogs. *J. Med. Chem.* **1997**, *40* (20), 3312–3318.
- (20) Vapnik, V. N. *The Nature of Statistical Learning Theory*; Springer: New York, 1995.
- (21) Wang, L.; Ma, C.; Wipf, P.; Xie, X.-Q. Linear and nonlinear support vector machine for the classification of human 5-HT1A ligand functionality. *Mol. Inf.* **2012**, *31* (1), 85–95.
- (22) Menon, P.; Yin, G.; Smolock, E. M.; Zuscik, M. J.; Yan, C.; Berk, B. C. GPCR kinase 2 interacting protein 1 (GIT1) regulates osteoclast function and bone mass. *J. Cell Physiol.* **2010**, *225* (3), 777–785.
- (23) Hutchison, G. Open Babel: File Translation for Computational Chemistry and Nanoscience, *NNIN/CNF Fall Workshop*, Ithaca, NY, 2005.
- (24) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Molecular similarity searching using atom environments, information-based feature selection, and a naive bayesian classifier. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (1), 170–178.
- (25) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): Evaluation of performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (5), 1708–1718.
- (26) Chang, C.-C.; Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2* (3), 1–27.
- (27) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39* (15), 2887–2893.