An Algorithm to Identify Target-Selective Ligands – A Case Study of 5-HT7/5-HT1A Receptor Selectivity

Rafał Kurczab[1]*, **Vittorio Canale**[2], **Paweł Zajdel**[2], **Andrzej J. Bojarski**[1]

1 Department of Medicinal Chemistry, Institute of Pharmacology Polish Academy of Sciences, 12 Smętna Street, 31–343, Kraków, Poland, **2** Department of Medicinal Chemistry, Jagiellonian University Medical College, 9 Medyczna Street, 30–688, Kraków, Poland

* kurczab@if-pan.krakow.pl

Abstract

A computational procedure to search for selective ligands for structurally related protein targets was developed and verified for serotonergic 5-HT$_7$/5-HT$_{1A}$ receptor ligands. Starting from a set of compounds with annotated activity at both targets (grouped into four classes according to their activity: selective toward each target, not-selective and not-selective but active) and with an additional set of decoys (prepared using DUD methodology), the SVM (Support Vector Machines) models were constructed using a selective subset as positive examples and four remaining classes as negative training examples. Based on these four component models, the consensus classifier was then constructed using a data fusion approach. The combination of two approaches of data representation (molecular fingerprints vs. structural interaction fingerprints), different training set sizes and selection of the best SVM component models for consensus model generation, were evaluated to determine the optimal settings for the developed algorithm. The results showed that consensus models with molecular fingerprints, a larger training set and the selection of component models based on MCC maximization provided the best predictive performance.

Introduction

The identification of ligands that display a high affinity for one protein target and that are significantly less active for another or a group of closely related members of a given family is of high relevance for modern drug discovery. Apart from using selective ligands as leads in drug design workflows, they can also be applied as molecular probes for studying, e.g., cellular functions [1]. Because the validation of compound selectivity requires significant experimental efforts and financial resources, fast and accurate computational methods to predict ligand selectivity are highly desirable.

In recent years, diverse computational ligand- and/or structure-based approaches to explain the molecular mechanism of selectivity and/or to predict compound selectivity have been

developed. The most prominent example reported on molecular dynamic simulations combined with free energy calculations to study mechanisms underlying the selectivity of tyrosine phosphatases PTP1B/TCPTP/SHP-2 [2], phosphatidylinositol-3-kinases PI3Kα/PI3Kγ [3] and phosphodiesterase PDE5/PDE6 [4]. Other studies have described QSAR modeling to predict the ligand selectivity for serotonin $5\text{-}HT_{1E}/5\text{-}HT_{1F}$[5] or dopamine D_2/D_3 receptors [6] and for a panel of 45 different kinases [7]. Yet other investigations used machine learning (ML) methods to construct selectivity prediction models, e.g., ML based on neural networks to generate structure-selectivity relationship models [8], the binary classification SVM (Support Vector Machines) algorithm to solve multiclass predictions and compound ranking to distinguish between selective, active but non-selective, and inactive compounds [9], and the LiCABEDS (Ligand Classifier of Adaptively Boosting Ensemble Decision Stumps) algorithm to model cannabinoid CB_1/CB_2 selectivity [10].

Among fourteen 5-HT receptor (5-HTR) subtypes, $5\text{-}HT_7R$ represents the most recent addition to a subfamily of G-protein-coupled receptors (GPCRs). Distribution studies revealed a correlation between the localization of $5\text{-}HT_7Rs$ in the CNS (especially in the suprachiasmatic nucleus) and their function, suggesting that they are involved in the regulation of circadian rhythm, learning and memory processes, as well as in pathological processes such as affective disorders, neurodegenerative diseases, and cognitive decline [11]. A large body of evidence has demonstrated that the clinically established antidepressant effects of atypical antipsychotics, i.e., amisulpiride, lurasidone and aripiprazole, are mediated by antagonism at $5\text{-}HT_7Rs$ [12,13]. Several preclinical studies support the hypothesis that $5\text{-}HT_7R$ antagonists may produce beneficial pro-cognitive effects and ameliorate negative symptoms of schizophrenia in animal models [14–17]. On the other hand, potential application for $5\text{-}HT_7R$ agonists has been proposed for the treatment of dysfunctional memory in age-related decline and Alzheimer's disease [18], diabetic neuropathy and neuropathic pain [19,20]. Moreover, recent preclinical findings have demonstrated novel therapeutic applications of $5\text{-}HT_7R$ agonists for the treatment of fragile X syndrome, ADHD and other attention deficit-related diseases [21,22].

Despite a great interest in $5\text{-}HT_7R$ since the 1990s, its function remains incompletely understood. Apart from fundamental criteria for the classification of receptors, i.e., primary amino acid sequence and signal transduction (G-protein, β-arrestin or MAPK/ERK pathways), $5\text{-}HT_7R$ displays structural features that are similar to those of $5\text{-}HT_{1A}R$ [23–26]. Although this similarity hampers the design of selective ligands of $5\text{-}HT_7R$ [27,28], the situation appears to be even more complicated when considering the co-localization and functional interplay between $5\text{-}HT_7$ and $5\text{-}HT_{1A}Rs$ (i.e., homo/hetero dimerization, receptor desensitization and/or internalization) [23,29].

Considering the aforementioned findings regarding the clinical significance of $5\text{-}HT_7R$, the elaboration of new algorithms to support the design of selective $5\text{-}HT_7R$ agents (as an alternative to those reported in the literature—Fig 1) appears to be critical to obtain a more detailed understanding of the pharmacological function of $5\text{-}HT_7Rs$.

Here, we developed and investigated the algorithm (based on SVM [38] classification models of ligands showing different affinity/selectivity relationships for $5\text{-}HT_7/5\text{-}HT_{1A}$ receptors and a data fusion approach) for its application to predict ligand selectivity between both targets (Fig 2). The performance of this algorithm was compared to a simple ranking strategy and the best in-class component SVM models. Furthermore, ligand- (molecular fingerprints) and structure-based (Structural Interaction Fingerprint, SIFt) data representations, as well as performance metrics (AUC and MCC), were evaluated to select the best SVM models.
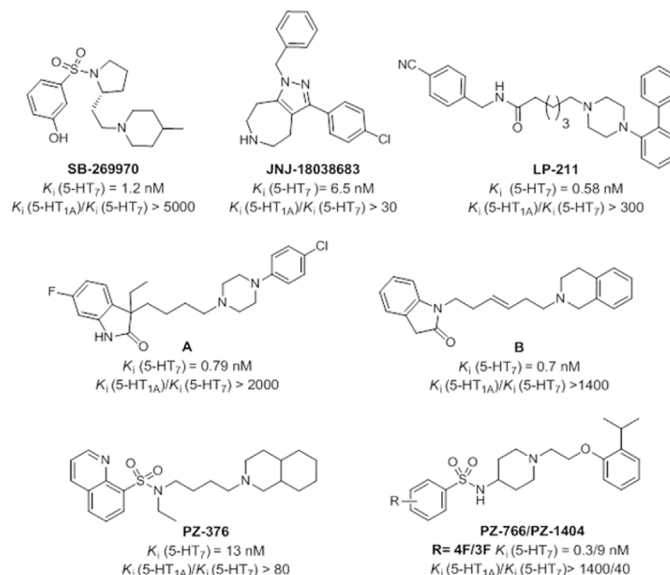
Fig 1. Chemical structure of different chemical classes of selective 5-HT$_7$R ligands [30–37].

doi:10.1371/journal.pone.0156986.g001

Materials and Methods

Data sets and definition of training classes

The compounds with activity determined for both 5-HT$_{1A}$ and 5-HT$_7$ receptors were retrieved from the ChEMBL v17 database [39]. The parameters (K_i, IC$_{50}$ and pK_i) describing the ligand affinity were used to define four subsets of compounds (Table 1), i.e., selective toward 5-HT$_7$R (Selective) or 5-HT$_{1A}$R (Revsel), not-selective but active (Nselbact) and not-selective (Notsel). The list of compound ChEMBL IDs belonging to a given subset is provided in the Supporting Information (S2 File).

The pK_i and IC$_{50}$ values were recalculated to the K_i using the following expressions: $K_i = 10^{-pKi}$ and $K_i = IC_{50}/2$. The conversion factor of 2 was suggested by Kalliokoski et al. [40] and has been applied successfully in similar studies [41,42]. In addition, for each selective ligand, 50 decoys with similar 1D physicochemical properties to remove bias (e.g., molecular weight, logP) and a dissimilar 2D topology to be likely non-binders, were generated using DUD-e service [43]. Accordingly defined sets were further used to construct component (class-specific) SVM models by combining the Selective subset (positive learning examples) with DUD, Revsel, Notsel and Nselbact (negative learning examples).

Data representation

Two approaches, i.e., ligand-based and structure-based approaches were tested to identify the optimal way for constructing selectivity prediction models. In the ligand-based approach, the structure of a molecule was encoded by three different molecular fingerprints (FP): hashed FP [44] (CDK FP, 1024 bits), Klekota-Roth FP [45] (KlekFP, 4860 bits) and MACCS FP [46] (MACCSFP, 166 bits), which were calculated using PaDEL-Descriptor software [47].

In the structure-based approach (Fig 3), Structural Interaction Fingerprint profiles (SIFt-p) were used [48,49]. They were obtained by docking all the defined subsets of ligands to different conformations of 5-HT$_{1A}$ and 5-HT$_7$Rs homology models [26,50] with and without extracellular loops (EL). The 3-dimensional structures of the ligands were prepared using LigPrep v3.6 [51], and the appropriate ionization states at pH = 7.4 were assigned using Epik v3.4 [52]. The
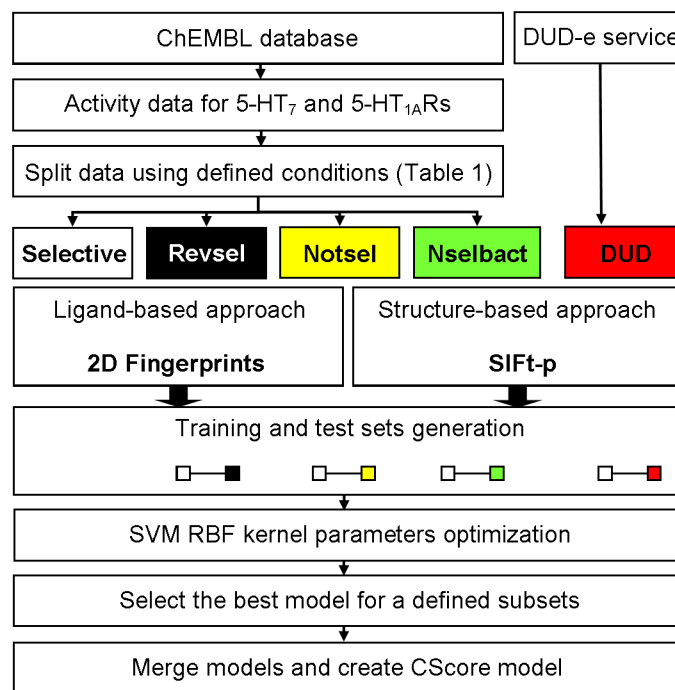
Fig 2. Schematic of the algorithm. The ChEMBL database was filtered out to extract the compounds with annotated activity for both 5-HT$_7$ and 5-HT$_{1A}$ receptors. Next, the obtained set of compounds was divided into four subsets using defined rules (Table 1). Additionally, using the DUD-e web service, the decoys for the Selective set were generated. The compounds from each subset were next encoded in binary string format using a set of molecular (ligand-based approach) and interaction fingerprints (SIFt-p, structure-based approach). Next, for each representation, the Selective subset was merged with one from the remaining sets to generate four groups for use in independent training and testing of RBF SVM models (10 trials per issue). The best in-class SVM models were selected based on AUC and MCC values. The final ranking was obtained by application of the SUM rule of data fusion, in which the scores of component SVM models were summed. Abbreviations used: Revsel—reversed selective, i.e., at least 5-fold more active for 5-HT$_{1A}$R than 5-HT$_7$; Nselbact—not selective but active, i.e., dual ligands; Notsel—not selective, i.e., the remaining compounds; SIFt-p—Structural Interaction Fingerprints profile (calculated by averaging the SIFts obtained for the docking of a given compound to three target conformations); SVM RBF—Support Vector Machines with radial basis function kernel.

doi:10.1371/journal.pone.0156986.g002

Table 1. Definitions of SVM training sets used for component model generation.

Set Name	Description/Conditions used to generate the set	Size	SVM class	Inter-class Tc similarity[a]
Selective	Selective ligands for 5-HT$_7$R:K_i(5-HT$_7$) < 100 nM & K_i(5-HT$_{1A}$)/K_i(5-HT$_7$) > 5	69	Positive	0.33
DUD	Decoy set generated for selective set using DUD-e service	5198	Negative	0.25
Revsel	Reverse selective ligands (selective ligands for 5-HT$_{1A}$R):K_i(5-HT$_{1A}$) < 100 nM & K_i(5-HT$_7$)/K_i(5-HT$_{1A}$) > 5	124	Negative	0.29
Nselbact	Active ligands for both receptors simultaneously:K_i(5-HT$_7$) & K_i(5-HT$_{1A}$) < 100 nM	89	Negative	0.31
Notsel	Ligands that do not meet the condition of selectivity:0.20 < K_i(5-HT$_{1A}$)/K_i(5-HT$_7$) < 5	440	Negative	0.29

[a] An average Tanimoto coefficient obtained by pairwise comparison (compounds encoded by CDK fingerprint) for each molecule of the Selective subset with each molecule of the remaining subsets. All average pairwise intra- and inter-class Tc are presented as a heat map in S1 Fig.

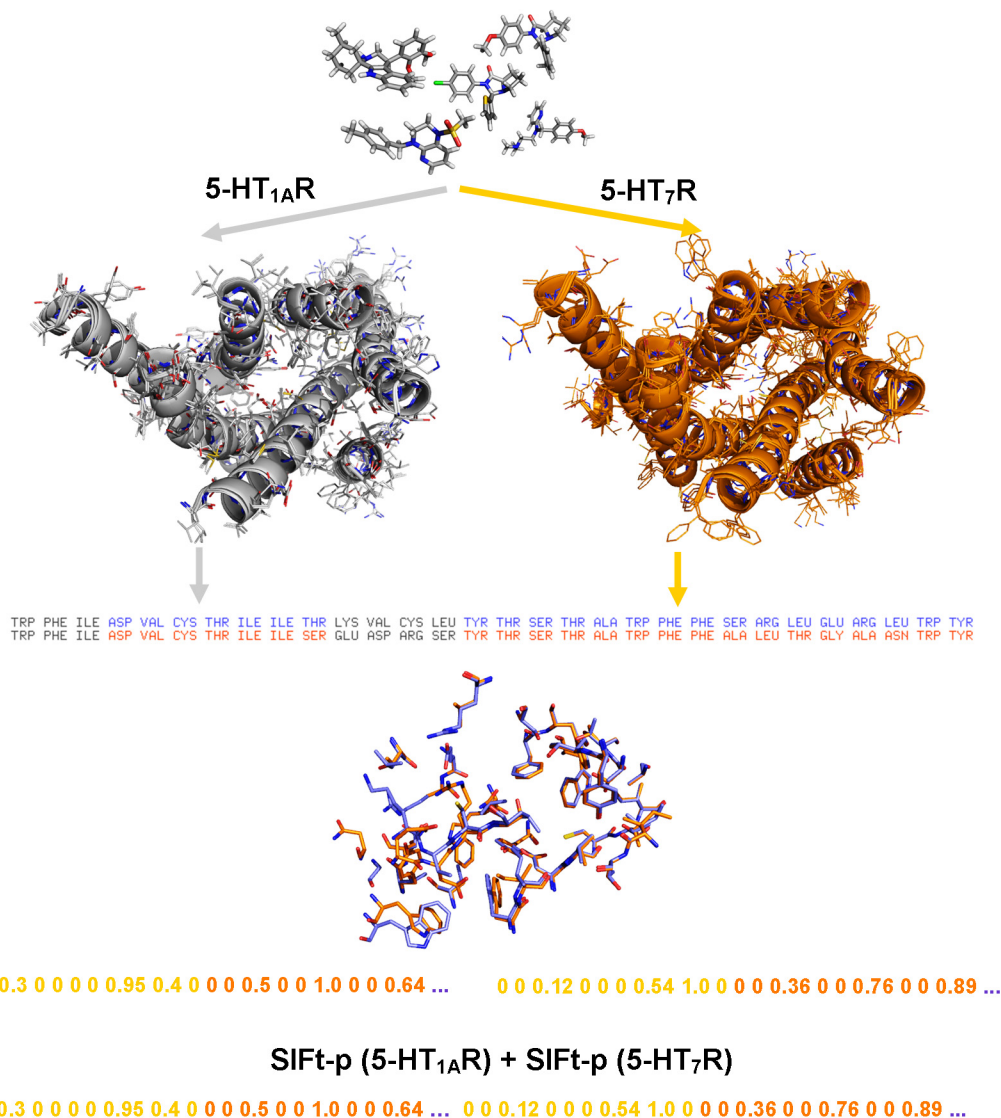doi:10.1371/journal.pone.0156986.t001

Fig 3. Schematic of the structure-based approach. In the first stage, all subsets of compounds were docked (Glide SP mode) to the three conformations of the 5-HT$_{1A}$R and 5-HT$_7$R homology models (generated on 5-HT$_{1B}$R and D$_3$R templates). Next, the interaction fingerprints (SIFt) were calculated for all obtained ligand-receptor complexes. The interaction analysis resulted in the selection of 34 common amino acids that formed any type of interaction with the ligands. For a given compound, the SIFts were recalculated (independently for each receptor and template) and averaged (SIFt-p). Finally, for each compound, the reduced SIFt profile (concatenating dockings to 5-HT$_{1A}$ and 5-HT$_7$ receptors to a single vector) was used as an input to SVM.

doi:10.1371/journal.pone.0156986.g003

Protein Preparation Wizard was used to assign the bond orders, appropriate amino acid ionization states and to check for steric clashes. The receptor grid was generated (OPLS_2005 force field) by centering the grid box with a size of 15 Å on Asp3.32. Automated docking was performed using Glide v6.9 at the SP level with the flexible docking option turned on [53]. Five poses per ligand were generated, but only the one with the best Glide Score was used for SIFt generation.

The SIFt encodes the 3D ligand-protein complex in the form of a 1D binary string, in which a nine-bit pattern is used to describe the interaction type: any contact, backbone, side chain, polar, aromatic, hydrophobic interaction, hydrogen bond donor/acceptor and charged [54].

The SIFt-p were created by calculating the mean value for each position in three fingerprint strings obtained for a given compound in three conformations of a given receptor model. Based on the docking of all compounds and sequence alignment of 5-HT$_{1A}$ and 5-HT$_7$ receptors, a set of 34 common amino acids (sharing the same sequence position and having any contact with the docked set of ligands) was defined, and the reduced SIFt-p were created. Finally, for each compound, the reduced SIFt profiles from docking to 5-HT$_{1A}$ and 5-HT$_7$ receptors were concatenated to a single vector that handles information regarding averaged interactions between a particular ligand and both receptors.

Optimization of SVM learning parameters

The molecular fingerprints and concatenated SIFt-p were used as input to generate classification models using the SVM algorithm. To select the classification model with the best performance for a given training class, a bootstrapping procedure was used. All classes were divided into training and test sets using two ratios: 0.40 and 0.60 (i.e., 40% and 60% of all examples, respectively, were used for training, while the remaining ones constituted the testing set). For each ratio, 10-trials of resampling with replacement of the original sets was performed to optimize the kernel parameters in the SVM classification model. The models were constructed using the SVMlight library and radial base function (RBF, the Gaussian function) [55]. For each run, a grid search was performed for two parameters: a penalty of the error term $C \in \{10^{-4}, 10^{-3}, 10^{-2}, 0.5, 0.1, 1, 5, 10, 100, 500, 1000\}$ and gamma coefficient for the radial base function $\gamma \in \{10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0, 1, 5, 10, 100\}$. The example distribution of AUC and MCC values for the 10 best SVM models optimized for a training ratio of 0.40 and MACCS FP is presented in Fig 4.

CScore model generation

For each set of optimized SVM component models, the threshold (boundary value that separates positive and negative classes) for which MCC was highest was determined by sampling
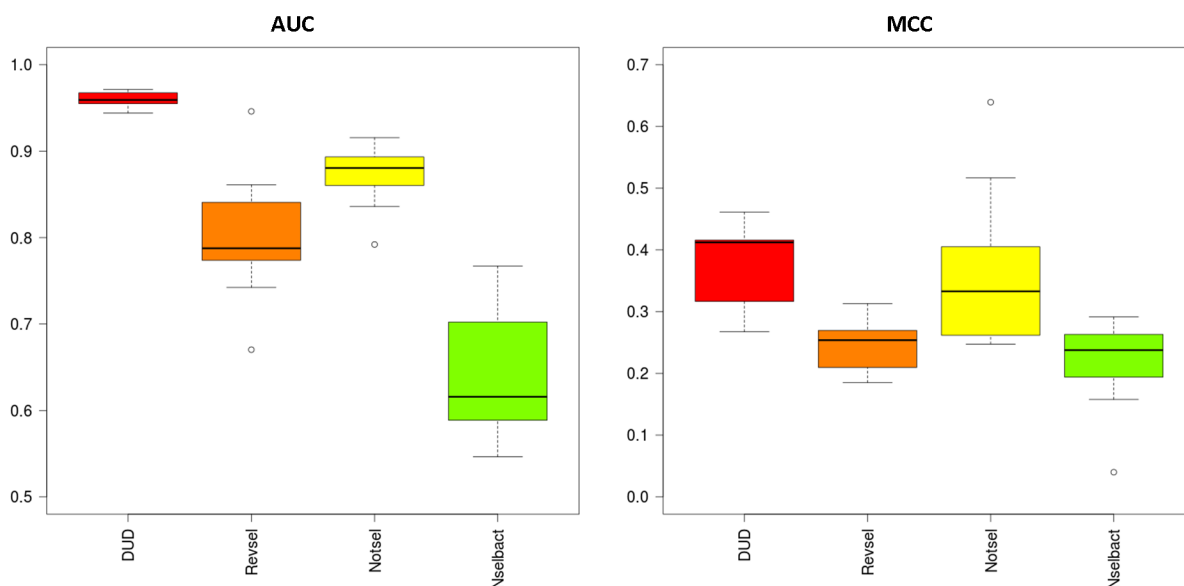

Fig 4. Box plots illustrating differences in performance (AUC and MCC) of 10 optimized SVM component models built for all training classes generated for MACCS FP and a training ratio of 0.40.

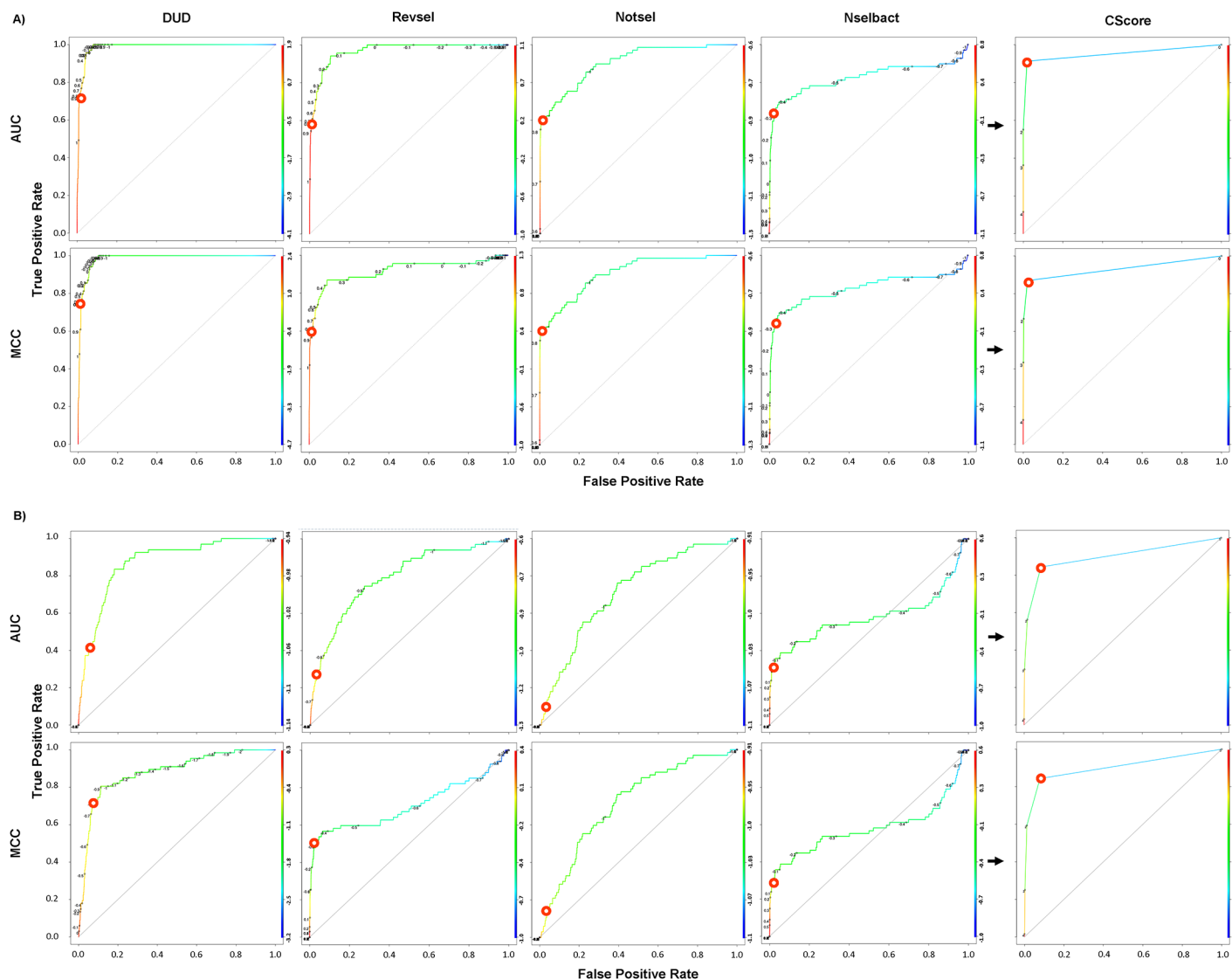doi:10.1371/journal.pone.0156986.g004

Fig 5. Example illustration of CScore model generation using component models. The ROC curves were used to show the performance of the component and CScore models and the thresholds (red circles) used to determine the classification (A—for MACCS FP and a training ratio of 0.40; B—for SIFt-p generated using the 5-HT$_{1B}$R template with loops and a training ratio of 0.40).

the range of the RBF decision function with a step of 0.1. Next, the best SVM models were identified and used to build the consensus score (CScore) classifiers (Fig 5). Two criteria were used to select the best in-class SVM model—the highest AUC or MCC values. The set of best component models and their thresholds were then used to perform the new classification. The CScore model was created by applying the SUM rule from the data fusion [56] merging the outputs of the individual component classifiers by summing the predictions that were calculated using the thresholds obtained for each component model.

The performance measures

The recall (1), precision (2), Mathews Correlation Coefficient—MCC (3), area under the receiver operator characteristic (ROC) curve (AUC) and the Boltzmann-Enhanced

Discrimination of ROC (BEDROC) metrics were used to assess the classification effectiveness of trained SVM models.

$$R = \frac{TP}{TP + FN} \tag{1}$$

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \tag{3}$$

$$BEDROC = \frac{\sum\limits_{i=1}^{n} e^{-\alpha \frac{ri}{N}}}{\frac{n}{N} \frac{(1 - e^{-\alpha})}{(e^{\alpha/N} - 1)}} \times \frac{\frac{n}{N} \sinh(\alpha/2)}{\cosh(\alpha/2) - \cosh(\alpha/3 - \alpha^n/N)} + \frac{1}{1 - e^{-\alpha(1 - \frac{n}{N})}} \tag{4}$$

where *TP* denotes the number of true positives (actives labeled as actives), *TN*–true negatives, *FP*–false positives (inactives labeled as actives), *FN*–false negatives, *n* is the number of actives among N compounds, *ri* is the rank of the *i*-th active and α is a parameter that assigns a weight towards compounds the top of the ranked list.

Recall measures the number of correctly identified positive examples, precision describes the correctness of positive predictions and MCC is a balanced measure of the binary classification effectiveness, ranging from –1 to 1, with 1 indicating a perfect prediction. The ROC presents the variation in the number of correctly classified positive examples with the number of incorrectly predicted negative examples. The BEDROC was introduced by Truchon and Bayly [57] to address the problem of "early recognition" in virtual screening. It can be interpreted as the probability that an active is ranked before a randomly selected compound that is exponentially distributed with parameter α, which controls the earliness of "early recognition" to test whether a ranking method is useful in the context of VS. The BEDROC metric ranges from 0 to 1, and it was calculated for α = 20 in the present study, as previously suggested [58].

The ROC curves and AUC values were calculated using the ROCR [59] package in R [60]. The BEDROC was also calculated in R using the enrichvs package [61].

Results

It should be emphasized that in the present analysis we focused on comparing the performance of the designed algorithm in different settings (i.e., representations, selection of the best component models) in terms of its ability to distinguish Selective from not-selective, inactive, decoy and multimodal (dual) compounds for the virtual screening of molecular databases. Moreover, the classification obtained by combining the SVM and binary molecular fingerprints cannot be interpreted at the level of chemical features that may be responsible for compound selectivity.

Initially, the performance of the CScore, the best component and Classical (trained on the Selective subset as positive class and on the sum of the Revsel, Notsel and Nselbact subsets as negative classes) were compared. The AUC, BEDROC and MCC parameters were calculated (Table 2) for ligand-based and structure-based approaches.

Interestingly, to address the "early recognition" problem (BEDROC value), the CScore models always demonstrated superior performance for recognizing selective over not-selective compounds in comparison to any single SVM-based ranking strategy (i.e., Classical and best component). However, considering the global performance (MCC, AUC), a single strategy

Table 2. Performance of CScore models obtained for the 0.40 training ratio and AUC and MCC strategies compared with the best single models trained in the classical manner and the best in-class component models. The median values for the Classical and best component strategies (ten trials were performed) are presented in parentheses.

Fingerprint	Strategy	BEDROC	MCC	AUC
CDK FP				
MCC	Best Classical	0.82 (0.72)	0.60 (0.55)	0.97 (0.92)
	Best component (notsel)	0.77 (0.74)	0.64 (0.53)	0.94 (0.91)
	CScore	0.93	0.65	0.93
AUC	Best Classical	0.84 (0.73)	0.53 (0.52)	0.96 (0.91)
	Best component (dud)	0.72 (0.68)	0.48 (0.55)	0.95 (0.94)
	CScore	0.92	0.52	0.96
MACCS FP				
MCC	Best Classical	0.81 (0.76)	0.66 (0.61)	0.93 (0.91)
	Best component (notsel)	0.77 (0.70)	0.64 (0.50)	0.90 (0.92)
	CScore	0.95	0.72	0.94
AUC	Best Classical	0.76 (0.75)	0.68 (0.66)	0.92 (0.91)
	Best component (dud)	0.82 (0.78)	0.66 (0.57)	0.98 (0.97)
	CScore	0.93	0.74	0.95
KlekR FP				
MCC	Best Classical	0.78 (0.76)	0.69 (0.61)	0.90 (0.89)
	Best component (dud)	0.77 (0.75)	0.72 (0.60)	0.93 (0.93)
	CScore	0.91	0.69	0.88
AUC	Best Classical	0.82 (0.81)	0.60 (0.61)	0.95 (0.94)
	Best component (dud)	0.86 (0.80)	0.63 (0.60)	0.97 (0.96)
	CScore	0.90	0.69	0.96
SIFt-p_D_3_loop				
MCC	Best Classical	0.12 (0.11)	0.17 (0.05)	0.61 (0.55)
	Best component (notsel)	0.45 (0.13)	0.33 (0.17)	0.60 (0.59)
	CScore	0.90	0.50	0.92
AUC	Best Classical	0.12 (0.11)	0.07 (0.05)	0.68 (0.63)
	Best component (dud)	0.24 (0.22)	0.15 (0.13)	0.71 (0.68)
	CScore	0.67	0.13	0.77
SIFt-p_D_3_nloop				
MCC	Best Classical	0.16 (0.13)	0.09 (0.05)	0.61 (0.55)
	Best component (nselbact)	0.41 (0.35)	0.43 (0.25)	0.46 (0.42)
	CScore	0.71	0.27	0.81
AUC	Best Classical	0.15 (0.13)	0.09 (0.05)	0.65 (0.60)
	Best component (revsel)	0.40 (0.39)	0.25 (0.21)	0.71 (0.64)
	CScore	0.53	0.11	0.75
SIFt-p_5-HT$_{1B}$_loop				
MCC	Best Classical	0.65 (0.22)	0.59 (0.13)	0.85 (0.77)
	Best component (revsel)	0.56 (0.16)	0.39 (0.11)	0.82 (0.68)
	CScore	0.81	0.50	0.90
AUC	Best Classical	0.63 (0.22)	0.65 (0.11)	0.82 (0.76)
	Best component (dud)	0.39 (0.37)	0.20 (0.22)	0.83 (0.80)
	CScore	0.75	0.53	0.90
SIFt-p_5-HT$_{1B}$_nloop				
MCC	Best Classical	0.12 (0.14)	0.12 (0.06)	0.65 (0.65)
	Best component (revsel)	0.49 (0.44)	0.37 (0.34)	0.51 (0.50)
	CScore	0.72	0.39	0.85

(*Continued*)

Table 2. (*Continued*)

Fingerprint	Strategy	BEDROC	MCC	AUC
AUC	Best Classical	0.23 (0.14)	0.12 (0.06)	0.69 (0.61)
	Best component (dud)	0.40 (0.37)	0.17 (0.11)	0.81 (0.77)
	CScore	0.67	0.24	0.82

sometimes provided better results than a consensus approach. In should be noted that use of MCC as the SVM model selection strategy consistently provided better CScores than AUC.

To evaluate global performance, the CScore was compared with all component models. Fig 6 shows an example panel of heat maps comparing the results obtained for ligand-based (Fig 6A) and structure-based (Fig 6B) approaches. As expected, in the majority of cases, the CScore models were better than any of the component models. The CScore models optimized for AUC and MCC ranked first in 56.7% and 40% of the analyzed cases, respectively (Fig 7).

Additionally, to study the relationships between CScore and component models, the clustering of rows (represented by vectors containing four performance parameters) using the complete linkage method with Euclidean distance measure was performed (all of the mentioned heat maps are available in S1 File). The global analysis of dendrograms revealed that the
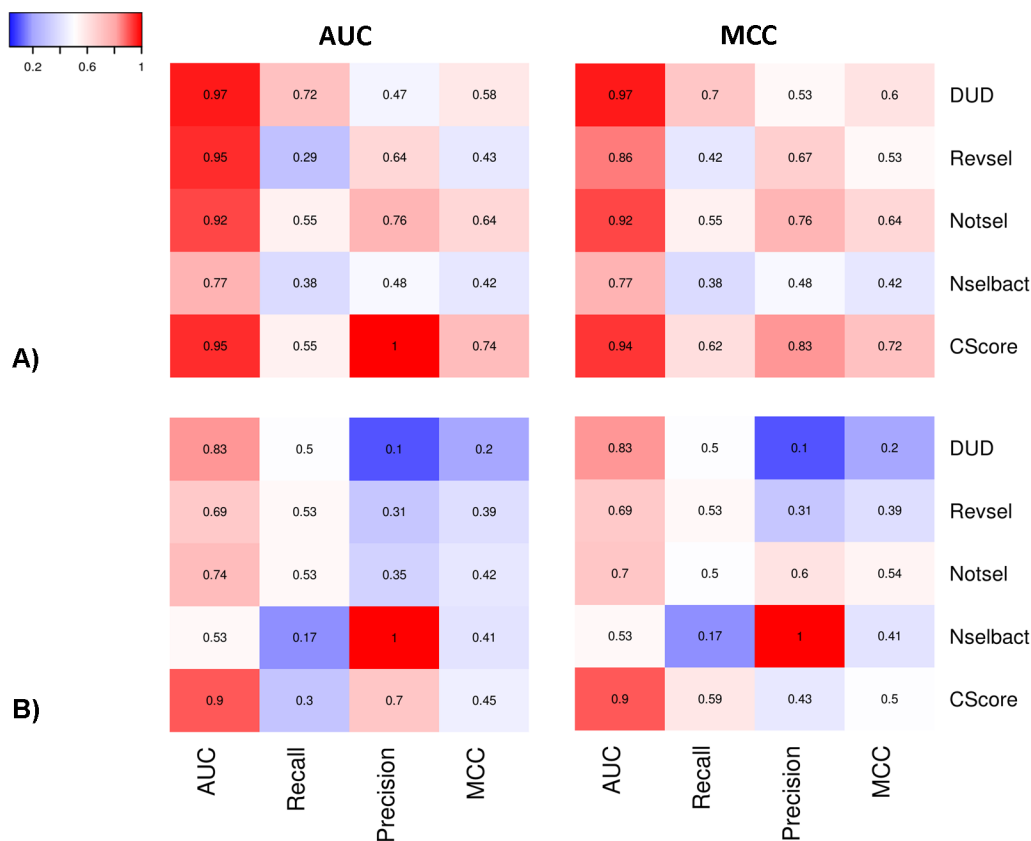

Fig 6. Heat maps comparing CScore with their component models for two strategies of selecting the best component models (highest AUC and MCC) and input data representations (A—for MACCS FP and a training ratio of 0.40; B—for SIFt-p generated using a 5-HT$_{1B}$R template with loops and a training ratio of 0.40).
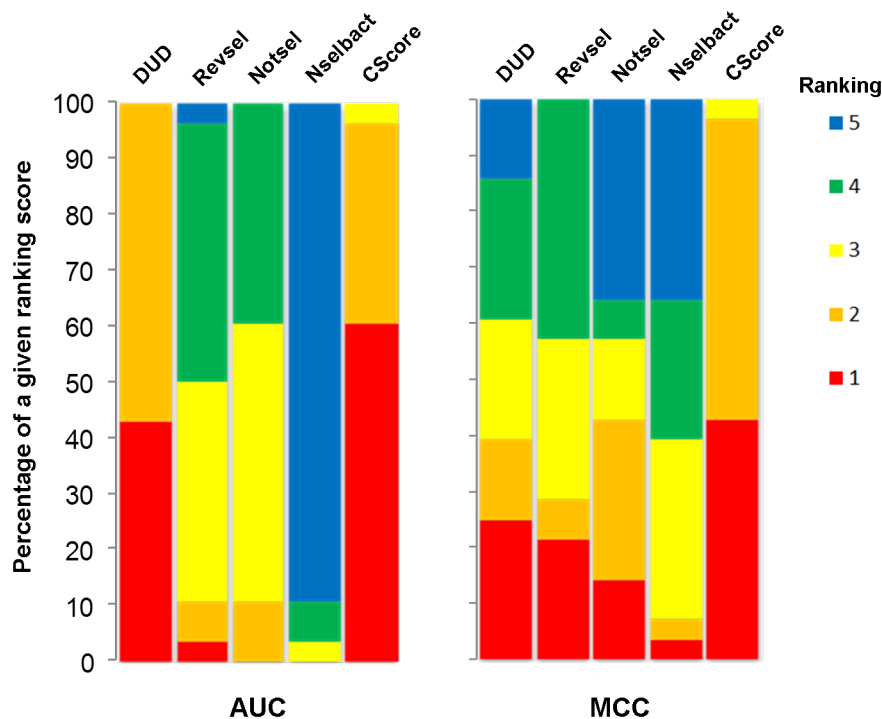
Fig 7. The ranking percentage (from 1 –best to 5 –worst) of a given SVM component and CScore model counted for AUC and MCC strategy.

performance of the CScore models was coupled with component models on different levels of performance similarity. For example, considering the highest level of similarity (i.e., the shortest Euclidean distance between two models), the performance of the CScore model was most similar to DUD, Revsel, Notsel and Nselbact in eight, six, eight and one cases, respectively. In the five remaining cases, CScore was coupled on the second level (i.e., with a cluster formed by two component models). Interestingly, eight cases had singleton component models (Nselbact and Notsel in six and two, respectively) and generally displayed poor performance.

Comparison of the approaches used to generate the input data (Fig 8) revealed that significantly better CScore models were obtained for ligand-based (average BEDROC = 0.95, MCC = 0.67 and AUC = 0.96) compared with structure-based (average BEDROC = 0.79, MCC = 0.38, and AUC = 0.86) approaches. Among all the molecular fingerprints that were utilized, MACCS FP (BEDROC = 0.94, MCC = 0.70 and AUC = 0.95) and KlekFP (BEDROC = 0.95, MCC = 0.69 and AUC = 0.95) performed at a comparable level that was greater than that of CDKFP (average BEDROC = 0.94, MCC = 0.61 and AUC = 0.95).

Regarding the homology modeling approach, the results showed that for both 5-HT$_{1B}$R and D$_3$R templates, slightly better CScore models were obtained when homology models with extracellular loops were used for SIFt-p generation (MCC for both models = 0.50, Fig 8). Additionally, the template also appears to influence the performance of the CScore models—better results were obtained for more homologous 5-HT$_{1B}$R template.

The CScore models obtained for component models selected using MCC criteria were slightly more efficient than those constructed using component models with the best AUC values. Because we optimized MCC to identify the threshold enabling the best classification effectiveness for each component model, the approach based on CScore model generation for MCC provides, globally, SVM models with the highest performance on classification tasks.
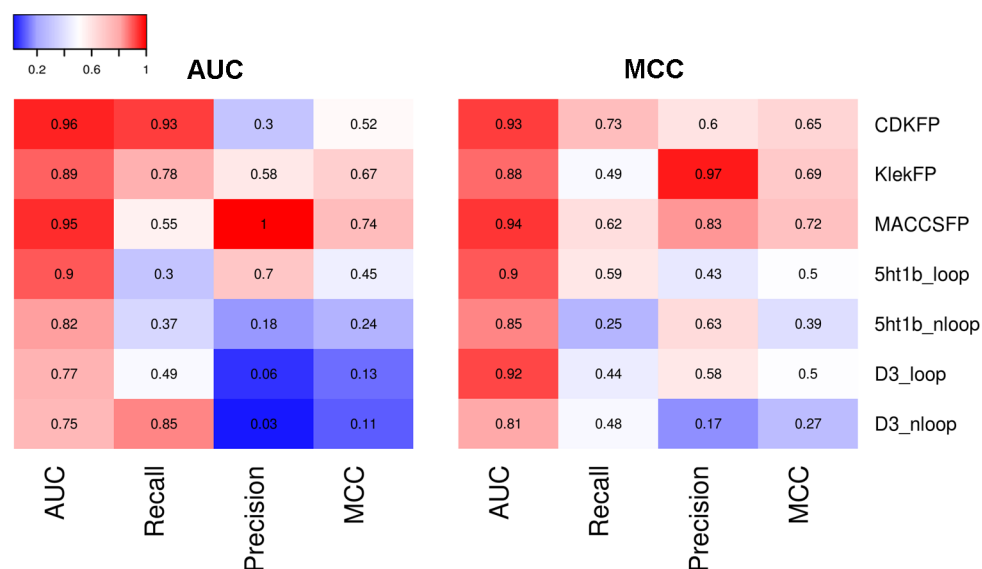
Fig 8. Heat maps comparing the performance of CScore models obtained for all the studied cases, i.e., representation of the data (three molecular fingerprints and SIFt-p for models with and without loops) and strategy for component models selection (AUC, MCC).

Finally, increasing the size of the training set (from 40% to 60% actives) improved the performance of both component and CScore models in the majority of cases (Fig 9), which is consistent with our previous findings [62,63].

Discussion

Selectivity threshold

As demonstrated in the present analysis, machine learning classification models trained on a set of ligands with different selectivity and activity profiles can provide a consensus model, the performance of which is significantly better than the component models. It must be stressed that the ligand was regarded as selective for 5-HT$_7$R as long as the ratio of K_i(5-HT$_{1A}$)/K_i(5-HT$_7$) was greater than 5. The rationale for this criterion was based on a close investigation of the data retrieved from ChEMBL database v17, which showed that there were 69 compounds for selectivity threshold $\geq$ 5, whereas there were only 34 compounds for threshold $\geq$ 20 that could be used for training and testing of the SVM models (S2 Fig). It should be noted that different thresholds for the selectivity index and rationale for their assessment have been used in similar studies, however, a quantitative definition is lacking. For example, Ma et al. used a selectivity index $\geq$ 10, which was selected based on the findings for the selective CB$_1$/CB$_2$ cannabinoid ligand by J.W. Huffman [10]. However, Wang et al. used a threshold for the selectivity index $\geq$ 3 for the kNN QSAR Classification model for 5-HT$_{1E}$/5-HT$_{1F}$ receptor selectivity [5]. To minimize errors resulting from, e.g., uncertainty regarding K_i values, they tested four selectivity thresholds and demonstrated that a threshold higher than 3 led to unacceptable models, generally due to a number of selective ligands that was too small. However, a higher selectivity index threshold was used by Wassermann [9] and Ning (50-fold) [8].

Performance of the CScore

To test general performance, as well as the "early recognition" preferences of the proposed algorithm, different performance metrics were applied. Although widely used, AUC is not a
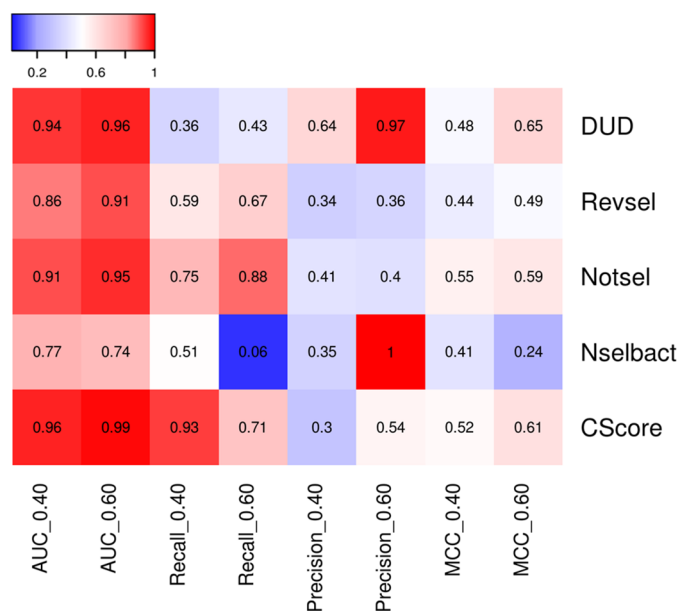
Fig 9. Comparison of the influence of training set size on the performance of separate models for MACCS FP.

doi:10.1371/journal.pone.0156986.g009

sufficient metric to address the "early recognition" problem specific to VS [57,58]. Additionally, the application of AUC to rank any database necessitates the selection of a decision threshold that enables binary classification (e.g., active/inactive or selective/not-selective). Consequently, different metrics can potentially be used to identify the optimal threshold. For example, Alvarsson et al. used net reclassification improvement (NRI) [41]. Because MCC is a more balanced summary statistic of the confusion matrix when unbalanced classes (see Table 1) are used [57], we decided to apply it in our algorithm.

Our analyses revealed that significantly better CScore models were obtained when the component models with the best MCC compared with the best AUC value were selected. This observation is explained in Fig 5, in which red circles on ROC curves depict the decision thresholds that were determined by maximizing MCC. These circles are localized in the area of the curve in which the number of true positives is ranked before false positives, which in some cases corresponds to the "early recognition" of the BEDROC.

Fingerprint influence

The influence of diverse parameters on the performance of the proposed algorithm was tested. Interestingly, CScore models based on the molecular fingerprints showed better performance than averaged interaction fingerprints (SIFt-p). All of the used fingerprints have different lengths, ranging from 166 (MACCS FP) to 4860 bits (Klekota-Roth FP), whereas concatenated SIFt-p had lengths of 1494 and 1458 bits for receptors with and without EL, respectively. There was no correlation between the length of the representation and the performance of the CScore model. The superior performance of ML models based on molecular rather than on interaction fingerprints in retrieving selectivity patterns may be due to uncertainty in predicting the correct binding mode by docking. It should be noted that because models obtained for receptors with EL showed better performance than those without loops, these additional four amino acids belonging to EL could play a role in the recognition of selective ligands.

The method presented herein could be especially useful for the virtual screening of chemical databases and for assessing combinatorial libraries to prioritize compounds for synthesis. It also offers more control capabilities in virtual screening searches for selective ligands because it enables the construction of a CScore model using different classification thresholds and performance parameters, e.g., one can generate a CScore model to optimize its performance for recall or precision. Additionally, the proposed algorithm is flexible, and after redefining the training classes, it can be used to, e.g., predict multimodal ligands.

Conclusions

In this study, a new algorithm is presented to identify new target-selective ligands and is evaluated based on its selectivity prediction for $5\text{-}HT_7$ receptor ligands over the $5\text{-}HT_{1A}$ subtype. We adopted data fusion and SVM component models (class-specific) that were trained on four datasets, i.e., selective toward $5\text{-}HT_7R$ (Selective) or $5\text{-}HT_{1A}R$ (Revsel), not-selective (Notsel) and not-selective but active (Nselbact), to construct the consensus classifier—CScore. The primary objective of this study was to obtain a virtual screening algorithm, which was evaluated in terms of its "early recognition" performance using the BEDROC metric. The analyses showed that the CScore was a significantly better scoring strategy than the best single models trained in a classical manner and the best in-class component models. The selection of component models to construct the consensus classifier is crucial and is significantly influenced by the molecular representation and performance parameter applied. In all studied cases, selection of the component models with the best MCC versus AUC value improved "early recognition" (measured by BEDROC).

Considering the successful implementation of the proposed algorithm, it will be incorporated into our screening protocol [64] and applied to analyze combinatorial libraries to prioritize the synthesis of selective $5\text{-}HT_7R$ ligands. Further improvements in the functionality of the algorithm will be conducted to improve its utility for other research groups (S2 File).

Supporting Information

S1 Fig. Heat map showing all average pairwise intra- and inter-class similarities calculated using the Tanimoto metric and CDK FP.
(TIF)

S2 Fig. Histogram of the compound selectivity index.
(TIF)

S1 File. Heat maps with row clustering comparing the CScore and component models developed for all ligand- and structure-based approaches to the data representation.
(PDF)

S2 File. A zip file containing scripts, datasets and optimized SVM models used in this study.
(ZIP)

Acknowledgments

Author Contributions

Conceived and designed the experiments: RK AJB. Performed the experiments: RK. Analyzed the data: RK VC PZ AJB. Contributed reagents/materials/analysis tools: RK VC PZ AJB. Wrote the paper: RK VC PZ AJB.

References

1. Stockwell BR, Exploring biology with small organic molecules. Nature. 2004; 432(7019): 846–854. PMID: 15602550

2. Fang L, Zhang H, Cui W, Ji M. Studies of the mechanism of selectivity of protein tyrosine phosphatase 1B (PTP1B) bidentate inhibitors using molecular dynamics simulations and free energy calculations. J Chem Inf Model. 2008; 48: 2030–2041. doi: 10.1021/ci800104s PMID: 18831546

3. Sabbah DA, Vennerstrom JL, Zhong HA. Binding selectivity studies of phosphoinositide 3-kinases using free energy calculations. J Chem Inf Model. 2012; 52: 3213–3224. doi: 10.1021/ci3003057 PMID: 23157418

4. Huang YY, Li Z, Cai YH, Feng LJ, Wu Y, Li X, et al. The molecular basis for the selectivity of tadalafil toward phosphodiesterase 5 and 6: A modeling study. J Chem Inf Model. 2013; 53: 3044–3053. doi: 10.1021/ci400458z PMID: 24180640

5. Wang XS, Tang H, Golbraikh A, Tropsha A. Combinatorial QSAR modeling of specificity and subtype selectivity of ligands binding to serotonin receptors 5HT1E and 5HT1F. J Chem Inf Model. 2008; 48: 997–1013. doi: 10.1021/ci700404c PMID: 18470978

6. Wang Q, Mach RH, Luedtke RR, Reichert DE. Subtype selectivity of dopamine receptor ligands: Insights from structure and ligand-based methods. J Chem Inf Model. 2010; 50: 1970–1985. doi: 10.1021/ci1002747 PMID: 20936866

7. Sciabola S, Stanton RV, Wittkopp S, Wildman S, Moshinsky D, Potluri S, et al. Predicting kinase selectivity profiles using Free-Wilson QSAR analysis. J Chem Inf Model. 2008; 48: 1851–1867. doi: 10.1021/ci800138n PMID: 18717582

8. Ning X, Walters M, Karypisxy G. Improved machine learning models for predicting selective compounds. J Chem Inf Model. 2012; 52: 38–50. doi: 10.1021/ci200346b PMID: 22107358

9. Wassermann AM, Gepper 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: 582–592. doi: 10.1021/ci800441c PMID: 19249858

10. Ma C, Wang L, Yang P, Myint KZ, Xie XQ. LiCABEDS II. Modeling of ligand selectivity for G-Protein-Coupled Cannabinoid Receptors. J Chem Inf Model. 2013; 53: 11–26. doi: 10.1021/ci3003914 PMID: 23278450

11. Nikiforuk A. Targeting the serotonin 5-HT7 receptor in the search for treatments for CNS disorders: rationale and progress to date. CNS Drugs. 2015; 29: 265–275. doi: 10.1007/s40263-015-0236-0 PMID: 25721336

12. Abbas AI, Hedlund PB, Huang XP, Tran TB, Meltzer HY, Roth BL. Amisulpride is a potent 5-HT7 antagonist: relevance for antidepressant actions in vivo. Psychopharmacology. 2009; 205: 119–128. doi: 10.1007/s00213-009-1521-8 PMID: 19337725

13. Ishibashi T, Horisawa T, Tokuda K, Ishiyama T, Ogasa M, Tagashira R, et al. Pharmacological profile of lurasidone, a novel antipsychotic agent with potent 5-hydroxytryptamine 7 (5-HT7) and 5-HT1A receptor activity. J Pharmacol Exp Ther. 2010; 334: 171–181. doi: 10.1124/jpet.110.167346 PMID: 20404009

14. Bonaventure P, Aluisio L, Shoblock J, Boggs JD, Fraser IC, Lord B, et al. Pharmacological blockade of serotonin 5-HT7 receptor reverses working memory deficits in rats by normalizing cortical glutamate neurotransmission. PloS One. 2011; 6: e20210. doi: 10.1371/journal.pone.0020210 PMID: 21701689

15. Zajdel P, Canale V, Partyka A, Marciniec K, Satała G, Kurczab R, et al. Arylsulfonamide derivatives of (aryloxy)ethyl-piperidines as 5-HT7 antagonists and their antidepressant and pro-cognitive properties. Med Chem Comm. 2015; 6: 1272–1277.

16. Neill JC, Barnes S, Cook S, Grayson B, Idris NF, McLean SL, et al. Animal models of cognitive dysfunction and negative symptoms of schizophrenia: focus on NMDA receptor antagonism. Pharmacol Ther. 2010; 128: 419–432. doi: 10.1016/j.pharmthera.2010.07.004 PMID: 20705091

17. Nikiforuk A, Kos T, Fijal K, Holuj M, Rafa D, Popik P. Effects of the Selective 5-HT7 Receptor Antagonist SB-269970 and Amisulpride on Ketamine-Induced Schizophrenia-like Deficits in Rats. PLoS ONE. 2013; 8: e66695. doi: 10.1371/journal.pone.0066695 PMID: 23776692

18. Perez-García GS, Meneses A. Effects of the potential 5-HT7 receptor agonist AS 19 in an autoshaping learning task. Behav Brain Res. 2005; 163: 136–140. PMID: 15936093

19. Viquier F, Michot B, Hamon M, Bourgoin S. Multiple roles of serotonin in pain control mechanisms-implications of 5-HT$_7$ and other 5-HT receptor types. Eur J Pharmacol. 2013; 716: 8–16. doi: 10.1016/j.ejphar.2013.01.074 PMID: 23500207

20. Di Pilato P, Niso M, Adriani W, Romano E, Travaglini D, Berardi F, et al. Selective agonists for serotonin 7 (5-HT7) receptor and their applications in preclinical models: an overview. Rev Neurosci. 2014; 25: 401–415. doi: 10.1515/revneuro-2014-0009 PMID: 24622785

21. Ruocco LA, Treno C, Gironi Carnevale UA, Arra C, Boatto G, Nieddu M, et al. Prepuberal stimulation of 5-HT7-R by LP-211 in a rat model of hyper-activity and attention-deficit: permanent effects on attention, brain amino acids and synaptic markers in the fronto-striatal interface. PLoS One. 2014; 9: e83003. doi: 10.1371/journal.pone.0083003 PMID: 24709857

22. Costa L, Sardone LM, Lacivita E, Leopoldo M, Ciranna L. Novel agonists for serotonin 5-HT7 receptors reverse metabotropic glutamate receptor-mediated long-term depression in the hippocampus of wild-type and Fmr1 KO mice, a model of Fragile X Syndrome. Front Behav Neurosci. 2015; 9: 65. doi: 10.3389/fnbeh.2015.00065 PMID: 25814945

23. Naumenko VS, Popova NK, Lacivita E, Leopoldo M, Ponimaskin EG. Interplay between serotonin 5-HT1A and 5-HT7 receptors in depressive disorders. CNS Neurosci The. 2014; 20: 582–590.

24. Medina RA, Sallander J, Benhamu B, Porras E, Campillo M, Pardo L, et al. Synthesis of New Serotonin 5-HT Receptor Ligands. Determinants of 5-HT7/5-HT1A Receptor Selectivity. J Med Chem. 2009; 52: 2384–2392. doi: 10.1021/jm8014553 PMID: 19326916

25. Bonaventure P, Nepomuceno D, Kwok A, Chai W, Langlois X, Hen R, et al. Reconsideration of 5-hydroxytryptamine (5-HT)7 receptor distribution using [3H]5-carboxamidotryptamine and [3H]8-hydroxy-2-(di-n-propylamino)tetraline: analysis in brain of 5-HT1A knockout and 5-HT1A/1B double-knockout mice. J Pharmacol Exp Ther. 2002; 302: 240–248. PMID: 12065723

26. Salerno L, Pittalà V, Modica MN, Siracusa MA, Intagliata S, Cagnotto A, et al. Structure-activity relationships and molecular modeling studies of novel arylpiperazinylalkyl 2-benzoxazolones and 2-benzothiazolones as 5-HT7 and 5-HT1A receptor ligands. Eur J Med Chem. 2014; 85: 716–726. doi: 10.1016/j.ejmech.2014.08.023 PMID: 25128671

27. Leopoldo M, Lacivita E, Berardi F, Perrone R. 5-HT(7) receptor modulators: a medicinal chemistry survey of recent patent literature (2004–2009). Expert Opin Therp Patent. 2010; 20: 739–754.

28. Canale V, Guzik P, Kurczab R, Verdie P, Satała G, Kubica B, et al. Solid-supported synthesis, molecular modeling, and biological activity of long-chain arylpiperazine derivatives with cyclic amino acid amide fragments as 5-HT7 and 5-HT1A receptor ligands. Eur J Med Chem. 2014; 78: 10–22. doi: 10.1016/j.ejmech.2014.03.005 PMID: 24675176

29. Renner U, Zeug A, Woehler A, Niebert M, Dityatev A, Dityateva G, et al. Heterodimerization of serotonin receptors 5-HT1A and 5-HT7 differentially regulates receptor signalling and trafficking. J Cell Sci. 2012; 25: 2486–2499.

30. Lovell PJ, Bromidge SM; Dabbs S, Duckworth DM, Forbes IT, Jennings AJ, et al. A novel, potent, and selective 5-HT(7) antagonist: (R)-3-(2-(2-(4-methylpiperidin-1-yl)ethyl)pyrrolidine-1-sulfonyl) phenol (SB-269970). J Med Chem. 2000; 43: 342–345. PMID: 10669560

31. Leopoldo M, Lacivita E, Contino M, Colabufo NA, Berardi F, Perrone R. Structure-activity relationship study on N-(1,2,3,4-tetrahydronaphthalen-1-yl)-4-aryl-1-piperazinehexanamides, a class of 5-HT7 receptor agents. 2. J Med Chem. 2007; 50: 4214–4221. PMID: 17649988

32. Volk B, Barkóczy J, Hegedus E, Udvari S, Gacsályi I, Mezei T, et al. (Phenylpiperazinyl-butyl)oxindoles as selective 5-HT7 receptor antagonists. J Med Chem. 2008; 51: 2522–2532. doi: 10.1021/jm070279v PMID: 18361484

33. Medina RA, Vázquez-Villa H, Gómez-Tamayo JC, Benhamú B, Martín-Fontecha M, de la Fuente T, et al. The extracellular entrance provides selectivity to serotonin 5-HT7 receptor antagonists with anti-depressant-like behavior in vivo. J Med Chem. 2014; 57: 6879–6884. doi: 10.1021/jm500880c PMID: 25073094

34. Zajdel P, Subra G, Verdie P, Gabzdyl E, Bojarski AJ, Duszyńska B, et al. Sulfonamides with the N-alkyl-N'-dialkylguanidine moiety as 5-HT7 receptor ligands. Bioorg Med Chem Lett. 2009; 19: 4827–4831. doi: 10.1016/j.bmcl.2009.06.038 PMID: 19560916

35. Zajdel P, Marciniec K, Maslankiewicz A, Paluchowska MH, Satala G, Partyka A, et al. Arene- and quinoline-sulfonamides as novel 5-HT7 receptor ligands. Bioorg Med Chem. 2011; 19: 6750–6759. doi: 10.1016/j.bmc.2011.09.044 PMID: 22001327

36. Zajdel P, Kurczab R, Grychowska K, Satała G, Pawłowski M, Bojarski AJ. The multiobjective based design, synthesis and evaluation of the arylsulfonamide/amide derivatives of aryloxyethyl- and

arylthioethyl piperidines and pyrrolidines as a novel class of potent 5-HT7 receptor antagonists. Eur J Med Chem. 2012; 56: 348–360. doi: 10.1016/j.ejmech.2012.07.043 PMID: 22926225

37. Zajdel P, Canale V, Partyka A, Marciniec K, Satała G, Kurczab R, et al. Arylsulfonamide derivatives of (aryloxy)ethylpiperidines as selective 5-HT7 receptor antagonists and their psychotropic properties. Med Chem Comm. 2015; 6: 1272–1277.

38. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995; 20: 273–297.

39. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 2012; 40: 1100–1107.

40. Kalliokoski T, Kramer C, Vulpetti A, Gedeck P. Comparability of mixed IC50 data—a statistical analysis. PLoS One. 2013; 8: e61007. doi: 10.1371/journal.pone.0061007 PMID: 23613770

41. Alvarsson J, Eklund M, Engkvist O, Spjuth O, Carlsson L, Wikberg JES, et al. Ligand-based target prediction with signature fingerprints. J Chem Inf Model. 2014; 54: 2647–2653. doi: 10.1021/ci500361u PMID: 25230336

42. Warszycki D, Mordalski S, Kristiansen K, Kafel R, Sylte I, Chilmonczyk Z, et al. A linear combination of pharmacophore hypotheses as a new tool in search of new active compounds—An application for 5-HT1A receptor ligands. PLoS ONE. 2013; 8(12): e84510. doi: 10.1371/journal.pone.0084510 PMID: 24367669

43. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem. 2012; 55: 6582–6594. doi: 10.1021/jm300687e PMID: 22716043

44. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E. The Chemistry Development Kit (CDK): An open-source Java library for chemo- and bioinformatics. J Chem Inf Comput Sci. 2003; 43: 493–500. PMID: 12653513

45. Klekota J, Roth FP. Chemical substructures that enrich for biological activity. Bioinformatics. 2008; 24: 2518–2525. doi: 10.1093/bioinformatics/btn479 PMID: 18784118

46. San Diego, CA, USA: MACCS Structural keys, Accelrys; [www.accelrys.com].

47. Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem. 2011; 32: 1466–1474. doi: 10.1002/jcc.21707 PMID: 21425294

48. Chuaqui C, Deng Z, Singh J. Interaction profiles of protein kinase-inhibitor complexes and their application to virtual screening. J Med Chem. 2005; 48 (1): 121–133. PMID: 15634006

49. Witek J, Smusz S, Rataj K, Mordalski S, Bojarski AJ. An application of machine learning methods to structural interaction fingerprints—a case study of kinase inhibitors. Bioorg Med Chem Lett. 2014; 24: 580–585. doi: 10.1016/j.bmcl.2013.12.017 PMID: 24374279

50. Canale V, Kurczab R, Partyka A, Satała G, Witek J, Jastrzębska-Więsek M, et al. Towards novel 5-HT7 versus 5-HT1A receptor ligands among LCAPs with cyclic amino acid amide fragments: Design, synthesis, and antidepressant properties. Part II. Eur J Med Chem. 2015; 92: 202–211. doi: 10.1016/j.ejmech.2014.12.041 PMID: 25555143

51. Schrödinger Release 2015–4: LigPrep, version 3.6, Schrödinger, LLC, New York, NY, 2015.

52. Schrödinger Release 2015–4: Epik, version 3.4, Schrödinger, LLC, New York, NY, 2015.

53. Small-Molecule Drug Discovery Suite 2015–4: Glide, version 6.9, Schrödinger, LLC, New York, NY, 2015

54. Mordalski S, Kosciolek T, Kristiansen K, Sylte I, Bojarski AJ. Protein binding site analysis by means of structural interaction fingerprint patterns. Bioorg Med Chem Lett. 2011; 21: 6816–6819. doi: 10.1016/j.bmcl.2011.09.027 PMID: 21974955

55. Joachims T. http://svmlight.joachims.org, 2002.

56. Willett P. Enhancing the effectiveness of ligand-based virtual screening using data fusion. QSAR Comb Sci. 2006; 25: 1143–1152.

57. Truchon J-F, Bayly CI. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. J Chem Inf Model. American Chemical Society; 2007; 47: 488–508. PMID: 17288412

58. Baldi P, Brunak S, Chauvin Y, Andersen C, Nielsen H. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics. 2000; 16: 412–424. PMID: 10871264

59. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005; 21: 3940–3941. PMID: 16096348

60. R Development Core Team. R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria, 2008; ISBN 3-900051-07-0.

61. Hiroaki Y. Package "enrichvs" 2015,

62. Kurczab R, Smusz S, Bojarski AJ. The influence of negative training set size on machine learning-based virtual screening. J Cheminformatics. 2014; 6: 32.

63. Smusz S, Kurczab R, Bojarski AJ. The influence of the inactives subset generation on the performance of machine learning methods. J Cheminformatics. 2013; 5: 17.

64. Kurczab R, Nowak M, Chilmonczyk Z, Sylte I, Bojarski AJ. The Development and Validation of a Novel Virtual Screening Cascade Protocol to Identify Potential Serotonin 5-HT7R Antagonists. Bioorg Med Chem Lett. 2010; 20: 2465–2468, doi: 10.1016/j.bmcl.2010.03.012 PMID: 20346662