Journal of Molecular Graphics and Modelling 32 (2012) 49–66

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Combinatorial support vector machines approach for virtual screening of selective multi-target serotonin reuptake inhibitors from large compound libraries

Z. Shi[a,c], X.H. Ma[a,c], C. Qin[a,c], J. Jia[a,c], Y.Y. Jiang[b,c], C.Y. Tan[b,c], Y.Z. Chen[a,c,*]

[a] Bioinformatics and Drug Design Group, Department of Pharmacy, Centre for Computational Science and Engineering, National University of Singapore, Blk S16, Level 8, 3 Science Drive 2, Singapore 117543, Singapore
[b] The Key Laboratory of Chemical Biology, Guangdong Province, The Graduate School at Shenzhen, Tsinghua University, Shenzhen, Guangdong, PR China
[c] Department of Biological Science, National University of Singapore, Blk S2, Level 5, Science Drive 4, Singapore 117543, Singapore

ARTICLE INFO

Article history:
Received 11 May 2011
Received in revised form 30 August 2011
Accepted 1 September 2011
Available online 5 October 2011

Keywords:
Multi-target
High-throughput screening
Computer aided drug design
Antidepressants
Support vector machines
Virtual screening

ABSTRACT

Selective multi-target serotonin reuptake inhibitors enhance antidepressant efficacy. Their discovery can be facilitated by multiple methods, including *in silico* ones. In this study, we developed and tested an *in silico* method, combinatorial support vector machines (COMBI-SVMs), for virtual screening (VS) multi-target serotonin reuptake inhibitors of seven target pairs (serotonin transporter paired with noradrenaline transporter, H_3 receptor, 5-HT_{1A} receptor, 5-HT_{1B} receptor, 5-HT_{2C} receptor, melanocortin 4 receptor and neurokinin 1 receptor respectively) from large compound libraries. COMBI-SVMs trained with 917–1951 individual target inhibitors correctly identified 22–83.3% (majority >31.1%) of the 6–216 dual inhibitors collected from literature as independent testing sets. COMBI-SVMs showed moderate to good target selectivity in misclassifying as dual inhibitors 2.2–29.8% (majority <15.4%) of the individual target inhibitors of the same target pair and 0.58–7.1% of the other 6 targets outside the target pair. COMBI-SVMs showed low dual inhibitor false hit rates (0.006–0.056%, 0.042–0.21%, 0.2–4%) in screening 17 million PubChem compounds, 168,000 MDDR compounds, and 7–8181 MDDR compounds similar to the dual inhibitors. Compared with similarity searching, k-NN and PNN methods, COMBI-SVM produced comparable dual inhibitor yields, similar target selectivity, and lower false hit rate in screening 168,000 MDDR compounds. The annotated classes of many COMBI-SVMs identified MDDR virtual hits correlate with the reported effects of their predicted targets. COMBI-SVM is potentially useful for searching selective multi-target agents without explicit knowledge of these agents.

1. Introduction

A primary anti-depression strategy is to inhibit monoamine reuptakes, such as serotonin reuptake, by both single target and multi target drugs [1]. Single target drugs [2,3] frequently encounter reduced efficacy and drug resistance problems caused by network robustness [2], redundancy [4], crosstalk [5], compensatory and neutralizing actions [6], anti-target and counter-target activities [7], and on-target and off-target toxicities [8]. Multi-target drugs are particularly useful for avoiding these problems.

Multi-target monoamine inhibitor drugs achieve enhanced efficacies by several mechanisms. The first one involves the inhibition of multiple monoamine reuptakes [9]. The simultaneous blockade of complementary monoamine reuptakes synergistically enhances the overall therapeutic efficacy [10]. Specific types of monoamines in CNS are reduced both by a primary monoamine transporter and by alternative transporters [11,12]. For instance, 5-HT is reduced primarily by serotonin transporter (SERT), and secondarily by noradrenaline transporter (NET) and dopamine transporter (DAT) particularly at high levels of 5-HT and/or when SERT function/expression is compromised [12]. Therefore, inhibition of one monoamine reduction route is complemented by the inhibition of the other routes to reduce their compensatory activities, leading to therapeutic synergy. This multi-target strategy is the basis for developing dual serotonin reuptake and noradrenaline reuptake inhibitors (NETSRIs) as antidepressant drugs of fast and enhanced therapeutic effects [13]. DES-VENLATAFINE and TESOFENSINE are good examples of NETSRI and dual SERT and DAT inhibitor respectively (Fig. 1).

The second mechanism involves collective monoamine reuptake inhibition and receptor antagonism. For instance, it has

* Corresponding author at: Bioinformatics and Drug Design Group, Department of Pharmacy, Centre for Computational Science and Engineering, National University of Singapore, Blk S16, Level 8, 3 Science Drive 2, Singapore 117543, Singapore.
Tel.: +65 6874 6877; fax: +65 6774 6756.
E-mail address: phacyz@nus.edu.sg (Y.Z. Chen).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.09.002

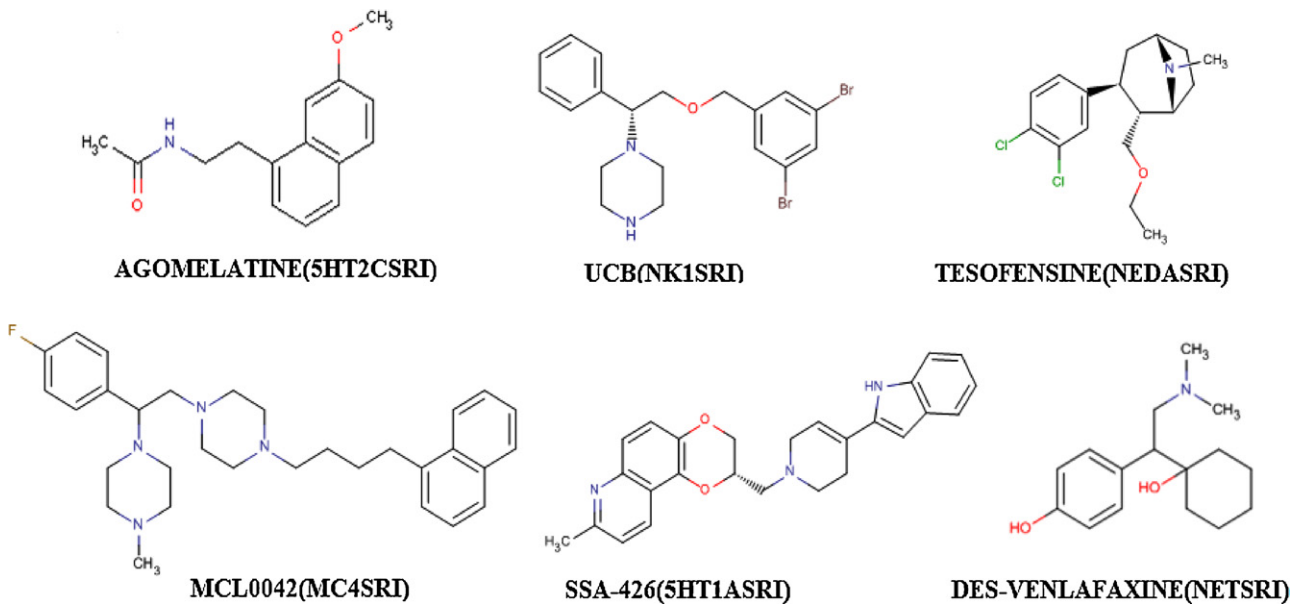

Fig. 1. Examples of multi-target serotonin reuptake inhibitors. NETSRI = dual serotonin reuptake and noradrenaline reuptake inhibitor; NEDASRI = serotonin, dopamine, and noradrenaline reuptake inhibitor; 5HT1ASRI: dual serotonin reuptake inhibitor/5-HT_{1A}; NK1SRI = dual serotonin reuptake inhibitor/neurokinin 1 receptor antagonist; MC4SRI = dual serotonin reuptake inhibitor/melanocortin 4 receptor antagonist.

been reported that increased release of 5-HT by SERT inhibition stimulates 5-HT_{1A}, 5-HT_{1B}, and 5-HT_{1C} autoreceptors, which subsequently reduces 5-HT release, thereby delaying the therapeutic effect of serotonin reuptake inhibitors until the 5-HT_{1A} and 5-$HT_{1B/1C}$ autoreceptors become desensitized [14]. This counteractive effect can be reduced by simultaneous targeting of serotonin transporters and 5-HT_{1A}, 5-HT_{1B} or 5-HT_{1C} receptors. Indeed, co-administration of a 5-HT_{1A} receptor antagonist with a selective serotonin reuptake inhibitor leads to an immediate increase in CNS 5-HT levels [15] and shortened onset of anxiolytic activity [16]. SSA-426 is an example of dual SERT and 5-HT_{1A} receptor antagonist (Fig. 1). Histamine H_3 receptor also promotes counteractive effect against serotonin reuptake inhibition by mediating the inhibition of serotonin release in the brain [17,18]. Therefore, in some circumstances, simultaneous targeting of serotonin reuptake transporter and histamine H_3 receptor achieves an improved antidepressant effect by more enhanced 5-HT release [19].

A third mechanism involves bimodal antidepressant actions. This approach aims at reducing the undesirable actions of selective serotonin reuptake inhibitors (SSRIs) through multi-targeted inhibition of other related receptors. Some of the undesirable actions of SSRIs, such as the short-term anxiety, arise from stimulation of 5-HT_{2C} receptor, and 5-HT_{2C} receptors also mediate the inhibitory effects of SSRIs on sleep, sexual function, and appetite [9]. Therefore, serotonin reuptake inhibitors with antagonist activities against 5-HT_{2C} receptor sites are expected to show a better tolerability than SSRIs [9]. AGOMELATINE is a dual serotonin reuptake inhibitor and 5-HT_{2C} receptor antagonist (5HT2cAntags) (Fig. 1) with clinically proven activity against major depression. The blockade of neurokinin 1 (NK_1) receptors by NK_1 receptor antagonists (NK1Antags) not only complement the effects of serotonin reuptake inhibition but also accelerate the long-term facilitating influence of SSRIs on serotonergic transmission [20]. Therefore, dual serotonin reuptake inhibitor and NK_1 receptor antagonist, such as UCB (Fig. 1), is expected to be more efficacious and faster in achieving therapeutic effects than SSRIs. Moreover, dual serotonin reuptake inhibitor and melanocortin 4 (MC_4) receptor antagonist (MC4Antags), such as MCL10004 (Fig. 1), has been found to interlink neuropeptide receptor antagonist activity with SRI activity to synergistically improve mood [20].

Extensive efforts have been directed at the development of multi-target serotonin reuptake inhibitors (e.g. dual serotonin reuptake and noradrenaline reuptake inhibitors (NETSRIs) [21,22], dual serotonin reuptake inhibitor and 5-HT_{1A} receptor antagonists (5HT1aSRIs) [23,24], dual serotonin reuptake inhibitor and 5-HT_{1B} receptor antagonists (5HT1bSRIs) [25], dual serotonin reuptake inhibitor and H_3 receptor antagonists (H3SRIs) [19], dual serotonin reuptake inhibitor and 5-HT_{2C} receptor antagonists (5HT2cSRIs) [26], dual serotonin reuptake inhibitor and MC_4 receptor antagonists (MC4SRIs) [27] and dual serotonin reuptake inhibitor and NK_1 receptor antagonists (NK1SRIs) [28]) based on the above mechanisms. While *in silico* methods have been extensively used for searching selective serotonin reuptake inhibitors [29,30], noradrenaline reuptake inhibitors [31,32] 5HT_{1A} receptor antagonists [33,34] and H_3 receptor antagonists [35,36], these methods have been used in a few published works for searching NETSRIs, 5HT1aSRIs, 5HT1bSRIs, H3SRIs, 5HT2cSRIs, MC4SRIs and NK1SRIs [25,37,38]. Therefore, in order to identify multi-target agents that are more sparsely distributed in the chemical space than single-target agents, there is a strong need to explore *in silico* methods more extensively, particularly those methods capable of searching large compound libraries at good yields and low false hit rates.

In this work, we used a machine learning method, support vector machines (SVM), to develop the combinatorial SVM (COMBI-SVM) virtual screening (VS) tool for searching dual-target agents NETSRIs, 5HT1aSRIs, 5HT1bSRIs, H3SRIs, 5HT2cSRIs, MC4SRIs and NK1SRIs. COMBI-SVM has recently been developed as dual kinase inhibitor VS tools with reasonably good yields, target selectivity and low false-hit rates in searching large compound libraries [39]. Hence, it is of interest to evaluate whether COMBI-SVM is equally useful for searching dual-target agents NETSRIs, H3SRIs, 5HT1aSRIs, 5HT1bSRIs, 5HT2cSRIs, MC4SRIs and NK1SRIs from large compound libraries.

COMBI-SVM VS tools are composed of single SVM VS tools constructed for each individual target in a given multi-target combination. Virtual hits simultaneously selected by all individual VS tools are considered as multi-target virtual hits [40]. The multi-target agents search capability of COMBI-SVM was rigorously tested by excluding all known multi-target inhibitors from the training datasets and only those compounds known to be active

against only one target in the target pair (these are tentatively referred to as individual-target inhibitors regardless of their possible activity against other targets outside the target pair) were used; The purpose of this exclusiveness is to test to what extent these individual-target based VS tools can identify multi-target inhibitors without explicit knowledge of known multi-target inhibitors [39]. Target selectivity of COMBI-SVM was assessed by using the known individual-target inhibitors of each target pair and those in the other six target pairs.

In order to evaluate the performance of COMBI-SVM, particularly the virtual hit rates and false-hit rates, in searching large compound libraries, the following three data sets were screened by COMBI-SVMs: 17 million compounds from PubChem database, 168,000 compounds from the MDL Drug Data Report (MDDR) database, and those MDDR compounds which are similar in structural and physicochemical properties to the collected multi-target inhibitors. MDDR contains bioactive compounds reported in the patent literature, journals, meetings and congresses. PubChem and MDDR contain high percentages of inactive or active compounds significantly different from the multi-target agents, and the easily distinguishable features may make VS enrichments artificially good [41]. Therefore, VS performance is more strictly tested by using a subset of MDDR compounds that is similar to the known multi-target agents so that enrichment is not simply a separation of trivial physicochemical features [42].

2. Methods

2.1. Compound datasets and molecular descriptors

Individual target and dual target inhibitors, each with IC50 or K_i value $\leq$10 μM, were collected from the literature [19,21,24], and the ChEMBL [43] and BindingDB [44] databases. The collected individual target inhibitors include 1125–1951 SSRIs, 1410 noradrenaline reuptake inhibitors (NRIs), 1689 H_3 receptor antagonists (H3Antags), 1144 5-HT_{1A} receptor antagonists (5HT1aAntags), 917 5-HT_{1B} receptor antagonists (5HT1bAntags), 1234 5-HT_{2C} receptor antagonists (5HT2cAntags), 1721 melanocortin 4 receptor antagonists (MC4Antags) and 1787 neurokinin 1 receptor antagonists (NK1Antags). The collected dual inhibitors include 101 dual serotonin reuptake/noradrenaline reuptake inhibitors (NETSRIs), 147 dual serotonin reuptake inhibitor/H_3 receptor antagonists (H3SRIs), 216 dual serotonin reuptake inhibitor/5-HT_{1A} receptor antagonists (5HT1aSRIs), 57 dual serotonin reuptake inhibitor/5-HT_{1B} receptor antagonists (5HT1bSRIs), 27 dual serotonin reuptake inhibitor/5-HT_{2C} receptor antagonists (5HT2cSRIs), 6 dual serotonin reuptake inhibitor/melanocortin 4 receptor antagonists (MC4SRIs) and 45 dual serotonin reuptake inhibitor/neurokinin 1 receptor antagonists (NK1SRIs), Table 1 summarises the datasets of these individual-target inhibitors, dual-inhibitors and MDDR compounds similar to at least one dual-inhibitor for each the target pair used as the training and testing sets in this work.

As few non-inhibitors have been reported, putative non-inhibitors of each target were generated by using our published method that requires no knowledge of inactive compounds or active compounds of other target classes and enables more expanded coverage of the "non-inhibitor" chemical space [45,46]. First, 17 million PubChem and 168 thousand MDDR compounds were clustered into 8993 compound families of similar molecular descriptors [47], which are consistent with the reported 12,800 compound-occupying neurons (regions of topologically close structures) for 26.4 million compounds of up to 11 atoms [48], and 2851 clusters for 171,045 natural products [49].

The putative non-inhibitors for each target were extracted from those families (5–8 per family) that contain no known individual-target inhibitors. The specific numbers of putative non-inhibitors are 60,726–62,593 from 7590 to 8018 families for SERT, 61,957 from 7937 families for NET, 61,960 from 7937 families for H_3 receptor, 62,376 from 7991 families for 5-HT_{1A} receptor, 64,790 from 8114 families for 5HT_{1B} receptor, 61,912 from 7739 families for 5-HT_{2C} receptor, 63,807 from 7976 families for MC_4 receptor and 62,733 from 7842 families for NK_1 receptor. This approach has the risk of the wrong exclusion of the compound families that contain multi-target inhibitors and undiscovered individual-target inhibitors from the non-inhibitor training dataset. The maximum possible "wrong" classification rate arising from these mistakes has been estimated at <13% even in the extreme and unlikely cases that all of the undiscovered single-target and multi-target agents are misplaced into the non-inhibitor class [46,50]. The noise level generated by up to 13% "wrong" negative compound family representation is expected to be substantially smaller than the maximum 50% false-negative noise level tolerated by SVM [51].

Molecular descriptors quantitatively represent structural and physicochemical features of molecules, and have been extensively used in deriving structure–activity relationships [52], quantitative structure–activity relationships [53] and VS tools [46,50,54] including the multi-target VS tools [39]. A set of 98 1D and 2D descriptors were selected for representing inhibitors and non-inhibitors of each target (Supplementary Table S1) [55], which include 18 descriptors in the class of simple molecular properties, 3 descriptors in the class of chemical properties, 35 descriptors in the class of molecular connectivity and shape, 42 descriptors in the class of electro-topological state. This set of 98 descriptors has been selected in our previous studies for representing diverse structural and physicochemical properties of both inhibitors of a specific target and non-inhibitors of that target distributed in large chemical space defined by 17 million Pubchem compounds. Although the structures of inhibitors of one target can be very different from those of another target, each inhibitor set plus the representatives of the non-inhibitors cover the same chemical space defined by the 17 million Pubchem compounds. Therefore, the same set of molecular descriptors was used in this work as well as our previous works. The virtual screening models of different biochemical classes (kinases, GPCR agonists/antagonists, peptidase inhibitors, DHFR inhibitors, and HDAC inhibitors) developed by this set of descriptors have shown equally good performance in screening large chemical libraries [39,40,46,50].

2.2. Support vector machines

SVM is based on the structural risk minimization principle of statistical learning theory [56]. It consistently shows outstanding classification performance; It is less penalized by sample redundancy; It has lower risk for overfitting; It is capable of accommodating large and structurally diverse training and testing datasets, and is fast in performing classification tasks [57,58]. However, like all machine learning methods, the performance of SVM is critically dependent on the diversity of training datasets. Because of the limited knowledge of known inhibitors for many targets, sufficiently good SVM VS tools may not be readily developed for these targets. Nonetheless, SVM VS tools with comparable performances or partially improved performances in certain aspects (e.g. reduced false-hit rates at comparable inhibitor yield) are useful to complement other VS tools.

In linearly separable cases, SVM constructs a hyper-plane to separate active and inactive classes of compounds with a maximum margin. A compound is represented by a vector $\mathbf{x}_i$ composed of its molecular descriptors. The hyper-plane is constructed by

Table 1
Datasets of individual-target inhibitors, dual inhibitors and MDDR compounds similar to at least one dual inhibitor used as the training and testing sets in this work.

Target pair	Inhibitors in training sets						Inhibitors and other compounds in testing set					
Target A–Target B	Training set for Target A			Training Set for Target B			Multi-target agents of Targets A and B				Inhibitors of other six targets outside target-pair	MDDR compounds similar to multi-target inhibitors of A and B
	No. of inhibitors of A that are non-inhibitor of B (no. of families)	No. of these inhibitors that are in the B inhibitor families (no. of families)	No. of these inhibitors that are in the families of multi-target agents of A and B (no. of families)	No. of inhibitors of B that are non-inhibitor of A (no. of families)	No. of these inhibitors that are in the A inhibitor families (no. of families)	No. of these inhibitors that are in the families of multi-target agents of A and B (no. of families)	No. of multi-target agents of A and B (no. of families)	No. (%) of multi-target agents in the families that contain single-target inhibitor of A in training sets	No. (%) of multi-target agents in the families that contain single-target inhibitor of B in training sets	No. (%) of multi-target agents outside the families that contain single-target inhibitor of A or B in training sets (no. of families)	No. of inhibitors	No. of compounds
SERT–NET	1125 (405)	399 (124)	113 (33)	1410 (486)	471 (124)	176 (42)	101 (73)	65 (64.3%)	46 (45.5%)	25 (24.8%)	8389	8181
SERT–H_3	1804 (604)	366 (95)	39 (16)	1689 (486)	345 (95)	124 (28)	147 (56)	97 (65.9%)	53 (36.1%)	27 (18.4%)	8191	1486
SERT–$5HT_{1A}$	1679 (590)	512 (130)	121 (26)	1144 (432)	421 (130)	151 (26)	216 (71)	130 (60.2%)	120 (55.6%)	52 (24.1%)	8354	7349
SERT–$5HT_{1B}$	1894 (631)	514 (108)	164 (22)	917 (309)	424 (108)	93 (11)	57 (35)	21 (41.2%)	42 (73.7%)	14 (24.6%)	8688	7475
SERT–$5HT_{2C}$	1924 (631)	689 (145)	28 (10)	1234 (493)	405 (145)	36 (9)	27 (23)	10 (37.0%)	13 (48.1%)	10 (37.0%)	8426	1302
SERT–MC_4	1951 (644)	175 (61)	2 (2)	1721 (248)	557 (61)	2 (2)	6 (2)	6 (100%)	6 (100%)	0	8164	7
SERT–NK_1	1910 (631)	262 (69)	39 (8)	1787 (358)	219 (69)	62 (8)	45 (23)	29 (64.4%)	9 (20%)	9 (20%)	8110	275

finding another vector $\mathbf{w}$ and a parameter b that minimizes $||\mathbf{w}||^2$ and satisfies the following conditions:

$$\mathbf{w}\cdot\mathbf{x}_i + b \geq +1, \quad \text{for} \quad y_i = +1 \quad \text{Class} \quad 1(\text{active}), \tag{1}$$

$$\mathbf{w}\cdot\mathbf{x}_i + b \leq -1, \quad \text{for} \quad y_i = -1 \quad \text{Class}\, 2(\text{inactive}), \tag{2}$$

where y_i is the class index, $\mathbf{w}$ is a vector normal to the hyperplane, $|b|/||\mathbf{w}||$ is the perpendicular distance from the hyperplane to the origin and $||\mathbf{w}||^2$ is the Euclidean norm of $\mathbf{w}$. Based on $\mathbf{w}$ and b, a given vector $\mathbf{x}$ can be classified by $f(\mathbf{x}) = \text{sign}[(\mathbf{w}\cdot\mathbf{x}) + b]$. A positive or negative $f(\mathbf{x})$ value indicates that the vector $\mathbf{x}$ belongs to the active or inactive class respectively. Linear SVM can then applied to this feature space based on the following decision function: $f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{l}\alpha_i^0 y_i K(\mathbf{x}, \mathbf{x}_i) + b\right)$, where the coefficients α_i^0 and b are determined by maximizing the following Langrangian expression: $\sum_{i=1}^{l}\alpha_i - 1/2\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$ under the conditions $\alpha_i \geq 0$ and $\sum_{i=1}^{l}\alpha_i y_i = 0$. A positive or negative $f(\mathbf{x})$ value indicates that the vector $\mathbf{x}$ belongs to the active or inactive class respectively. In nonlinearly separable cases, which frequently occur in classifying compounds of diverse structures [46,50,54], SVM maps the input vectors into a higher dimensional feature space by using a kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$. We used RBF kernel $K(\mathbf{x}_i, \mathbf{x}_j) = e^{-||\mathbf{x}_j - \mathbf{x}_i||^2/2\sigma^2}$ where σ is the kernel parameter. RBF kernel has been extensively used and consistently shown better performance than other kernel functions [46,50,54].

For a given training set of instance-label pairs (x_i, y_i), $i = 1, \ldots, l$ where $x_i \in R^n$ and $y_i \in \{1, -1\}^l$, in SVM, the task of finding the hyper-plane that separates active and inactive classes with a maximum margin, in essence, is to find the solution of the following optimization problem:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i, \\ \text{subject to} \quad & y_i(w^T\Phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0. \end{aligned}$$

$C > 0$ is the penalty parameter of the error term.

The performance of SVM in predicting non-dual inhibitors was evaluated by 5-fold cross-validation test. For each target pair, non-dual inhibitors and non-inhibitors were randomly divided into 5 groups of approximately equal size, with 4 groups used for training a SVM VS tool and 1 group used for testing it, and the test process is repeated for all 5 possible compositions to derive an average VS performance. After the 5-fold cross-validation, the σ values are chosen in the range of 0.9–5 based on the average VS performance for the model development. Table 2

shows the results of the 5-fold cross validation of SVM VS models for the target pairs SERT–NET, SERT–H_3, SERT–$5HT_{1A}$, SERT–$5HT_{1B}$, SERT–$5HT_{2C}$, SERT–MC_4 and SERT–NK_1. As for margin C, our SVM VS models were developed by using a hard margin $c = 100{,}000$. A hard margin has been proven to provide well with a more sensitive and strict classification for unbalanced datasets in which the negative data outnumbered the positive ones [36,37,43,47]. Fig. 3 illustrates the schematic diagram of COMBI-SVMs.

2.3. Tanimoto similarity searching method

Compounds similar to at least one inhibitor in a given inhibitor dataset can be identified by using the Tanimoto coefficient sim(*i*,*j*): [59].

$$\text{sim}(i,j) = \frac{\sum_{d=1}^{l} x_{di}x_{dj}}{\sum_{d=1}^{l}(x_{di})^2 + \sum_{d=1}^{l}(x_{dj})^2 - \sum_{d=1}^{l}x_{di}x_{dj}}, \tag{3}$$

where l is the number of molecular descriptors. A compound i is considered to be similar to a known inhibitor j in the inhibitor dataset if the corresponding sim(*i*,*j*) value is greater than a cut-off value. In computing sim(*i*,*j*), the molecular descriptor vectors $\mathbf{x}_i$s were scaled with respect to all of the MDDR compounds, which are representative of active compounds in the chemical spaces. The cut-off values for similarity compounds are typically in the range of 0.8–0.9 [42,60]. A stricter cut-off value of 0.9 was used in this study.

2.4. k-Nearest neighbour

k-Nearest Neighbour (k-NN) measures the Euclidean distance between a to-be-classified vector $\mathbf{x}$ and each individual vector $\mathbf{x}_i$ in the training set [2,4]. The Euclidean distances for the vector pairs are calculated using the following formula: $D = \sqrt{||\mathbf{x} - \mathbf{x}_i||^2}$. A total of k number of vectors nearest to the vector $\mathbf{x}$ are used to determine its class $\hat{f}(\mathbf{x}) \leftarrow \arg\max_{v\in V}\sum_{i=1}^{k}\delta(v, f(\mathbf{x}_i))$, where $\delta(a,b) = 1$ if $a = b$ and $\delta(a,b) = 0$ if $a \neq b$, arg max is the maximum of the function, V is a finite set of vectors $\{v_1, \ldots, v_s\}$ and $\hat{f}(\mathbf{x})$ is an estimate of $f(\mathbf{x})$. Here estimate refers to the class of the majority of the k nearest neighbours. The performance of k-NN was evaluated by 5-fold cross-validation in the same manner as in SVM and Table 3 shows the results of the 5-fold cross-validation for the target pairs SERT–NET, SERT–H_3, SERT–$5HT_{1A}$, SERT–$5HT_{1B}$, SERT–$5HT_{2C}$, SERT–MC_4 and SERT–NK_1. After the 5-fold cross-validation, the parameters of the developed k-NN models for the evaluated targets are chosen in the range of $k = 1$ or 3 according to the average performances of cross-validation.

2.5. Probabilistic neural network

Probabilistic neural network (PNN) is a form of neural network designed for classification through the use of Bayes' optimal decision rule [5]: $h_i c_i f_i(\mathbf{x}) > h_j c_j f_j(\mathbf{x})$, where h_i and h_j are the prior probabilities, c_i and c_j are the costs of misclassification and $f_i(x)$ and $f_j(x)$ are the probability density function for class i and j respectively. An unknown vector $\mathbf{x}$ is classified into population i if the product of all the three terms is greater for class i than for any other class j (not equal to i). In most applications, the prior probabilities and costs of misclassifications are treated as being equal. The probability density function for each class for a univariate case can be estimated by using the Parzen's nonparametric estimator,

$$g(\mathbf{x}) = \frac{1}{n\sigma}\sum_{i=1}^{n} W\left(\frac{\mathbf{x} - \mathbf{x}_i}{\sigma}\right), \tag{4}$$

where n is the sample size, σ is a scaling parameter which defines the width of the bell curve that surrounds each sample point, $W(d)$ is a weight function which has its largest value at $d = 0$ and $(\mathbf{x} - \mathbf{x}_i)$ is the distance between the unknown vector and a vector in the training set. The Parzen's nonparametric estimator was later expanded by Cacoullos for the multivariate case.

$$g(x_1, \ldots, x_p) = \frac{1}{n\sigma_1 \ldots \sigma_p}\sum_{i=1}^{n} W\left(\frac{x_1 - x_{1,i}}{\sigma_1}, \ldots, \frac{x_p - x_{p,i}}{\sigma_p}\right). \tag{5}$$

The Gaussian function is frequently used as the weight function because it is well behaved, easily calculated and satisfies the conditions required by Parzen's estimator. Thus the probability density function for the multivariate case becomes

$$g(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\exp\left(-\sum_{j=1}^{p}\left(\frac{x_j - x_{ij}}{\sigma_j}\right)^2\right). \tag{6}$$

Table 2
5-Fold cross-validation of SVM models for parameter selection and additional tests of these models for predicting dual-inhibitors and non-inhibitors, SEN sensitivity, SPE specificity, AC overall accuracy, AVE average; SD standard deviation, and SEM standard error of means.

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			NET			NETSRIs	Non-SSRIs	Non-NRIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–NET	1	84%	99.8%	99.5%	90%	99.8%	99.6%	48%	88%	81%
	2	91%	99.8%	99.7%	90%	99.7%	99.5%	48%	86%	77%
	3	89%	99.7%	99.5%	88%	99.7%	99.5%	45%	86%	81%
	4	85%	99.8%	99.5%	89%	99.7%	99.4%	43%	84%	83%
	5	87%	99.8%	99.6%	88%	99.8%	99.5%	48%	85%	82%
	AVE	87%	99.8%	99.6%	89%	100%	99%	46%	86%	81%
	S.D	0.025	0.000	0.001	0.010	0.000	0.000	0.02	0.02	0.02
	S.E.M	0.011	0.000	0.000	0.004	0.000	0.000	0.01	0.01	0.01

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			H3			H3SRIs	Non-SSRIs	Non-H3Is
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–H_3	1	93%	99.5%	99.3%	92%	99.8%	99.6%	17%	85%	100%
	2	89%	99.6%	99.3%	93%	99.7%	99.5%	31%	80%	100%
	3	88%	99.6%	99.3%	93%	99.7%	99.5%	25%	77%	100%
	4	89%	99.6%	99.3%	93%	99.7%	99.5%	24%	84%	100%
	5	86%	99.5%	99.1%	92%	99.7%	99.5%	19%	87%	100%
	AVE	89%	99.6%	99%	93%	99.7%	99.5%	23%	82%	100%
	S.D	0.025	0.001	0.001	0.006	0.000	0.000	0.06	0.04	1.00
	S.E.M	0.011	0.000	0.000	0.003	0.000	0.000	0.03	0.02	0.00

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			5HT1a			5HT1aSRIs	Non-SSRIs	Non-5HT1aIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–$5HT_{1A}$	1	99.7%	99.7%	99.4%	88%	99.8%	99.6%	48%	86%	74%
	2	99.6%	99.6%	99.3%	83%	99.8%	99.4%	45%	79%	74%
	3	99.5%	99.5%	99.2%	86%	99.7%	99.5%	44%	85%	74%
	4	99.7%	99.7%	99.3%	86%	99.8%	99.5%	45%	89%	77%
	5	99.7%	99.7%	99.4%	84%	99.8%	99.5%	45%	89%	82%
	AVE	99.7%	99.7%	99%	85%	99.8%	99.5%	45%	86%	76%
	S.D	0.001	0.001	0.001	0.021	0.000	0.001	0.01	0.04	0.04
	S.E.M	0.000	0.000	0.000	0.009	0.000	0.000	0.01	0.02	0.02

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			5HT1b			5HT1bSRIs	Non-SSRIs	Non-5HT1bIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–$5HT_{1B}$	1	84%	99.98%	99.7%	85%	99.8%	99.6%	23%	98%	99%
	2	82%	99.98%	99.7%	85%	99.8%	99.6%	19%	98%	91%
	3	78%	99.95%	99.5%	80%	99.8%	99.6%	23%	98%	96%
	4	81%	99.96%	99.6%	85%	99.8%	99.6%	23%	97%	92%
	5	80%	99.97%	99.6%	83%	99.9%	99.6%	23%	97%	93%
	AVE	81%	99.97%	99.6%	84%	99.8%	99.6%	22%	98%	94%
	S.D	0.024	0.0001	0.001	0.023	0.000	0.000	0.018	0.004	0.03
	S.E.M	0.011	0.00005	0.0003	0.010	0.000	0.000	0.008	0.002	0.02

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			5HT2c			5HT2cSRIs	Non-SSRIs	Non-5HT2cIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–$5HT_{2C}$	1	86.8%	99.5%	99.2%	91.4%	99.7%	99.3%	22%	75%	89%
	2	89.4%	99.6%	99.3%	90.6%	99.7%	99.3%	26%	84%	88%
	3	90.1%	99.6%	99.3%	90.6%	99.7%	99.4%	15%	81%	88%
	4	88.8%	99.6%	99.3%	94.1%	99.8%	99.3%	15%	82%	88%
	5	92.4%	99.6%	99.4%	98.0%	99.8%	99.4%	15%	81%	95%
	AVE	90%	99.6%	99%	92.9%	99.7%	99.3%	20%	81%	90%
	S.D	0.021	0.000	0.001	0.032	0.0006	0.0003	0.05	0.03	0.028
	S.E.M	0.009	0.000	0.000	0.014	0.0003	0.0002	0.02	0.01	0.013

Table 2 (*continued.*)

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			MC4			MC4SRIs	Non-SSRIs	Non-MC4Is
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–MC_4	1	89.3%	99.5%	99.2%	97.4%	100.0%	99.9%	83%	94%	95%
	2	92.6%	99.5%	99.3%	98.3%	99.9%	99.9%	83%	93%	91%
	3	92.1%	99.5%	99.3%	98.3%	99.9%	99.9%	67%	91%	92%
	4	92.8%	99.6%	99.4%	98.3%	99.9%	99.9%	67%	91%	94%
	5	87.7%	99.6%	99.2%	97.1%	99.9%	99.8%	67%	91%	93%
	AVE	91%	99.5%	99%	98%	99.9%	99.9%	73%	92%	93%
	S.D	0.023	0.000	0.001	0.006	0.000	0.000	0.09	0.01	0.02
	S.E.M	0.010	0.000	0.000	0.003	0.000	0.000	0.04	0.01	0.01
Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			NK1			NK1SRIs	Non-SSRIs	Non-NK1Is
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–NK_1	1	87.4%	99.5%	99.1%	93.6%	99.8%	93.6%	36%	88%	91%
	2	89.8%	99.6%	99.3%	95.5%	99.8%	95.5%	40%	88%	93%
	3	89.5%	99.5%	99.2%	96.6%	99.8%	96.6%	40%	90%	93%
	4	91.4%	99.5%	99.2%	95.0%	99.8%	95.0%	38%	86%	93%
	5	87.7%	99.6%	99.3%	95.2%	99.8%	95.2%	38%	88%	93%
	AVE	89%	99.5%	99%	95%	99.8%	95.2%	38%	88%	93%
	S.D	0.016	0.001	0.001	0.011	0.000	0.011	0.02	0.02	0.01
	S.E.M	0.007	0.000	0.000	0.005	0.000	0.005	0.01	0.01	0.00

The network architectures of PNN are determined by the number of compounds and descriptors in the training set. There are 4 layers in a PNN. The input layer provides input values to all neurons in the pattern layer and has as many neurons as the number of descriptors in the training set. The number of pattern neurons is determined by the total number of compounds in the training set. Each pattern neuron computes a distance measure between the input and the training case represented by that neuron and then subjects the distance measure to the Parzen's nonparameteric estimator. The summation layer has a neuron for each class and the neurons sum all the pattern neurons' output corresponding to members of that summation neuron's class to obtain the estimated probability density function for that class. The single neuron in the output layer then estimates the class of the unknown vector **x** by comparing all the probability density function from the summation neurons and choosing the class with the highest probability density function. The performance of PNN was validated by 5-fold cross-validation in the same manner as in SVM model development. Table 4

shows the results of the 5-fold cross-validation for the target pairs SERT–NET, SERT–H_3, SERT–$5HT_{1A}$, SERT–$5HT_{1B}$, SERT–$5HT_{2C}$, SERT–MC_4 and SERT–NK_1. After the 5-fold cross-validation, the parameters of the developed PNN models for the evaluated targets are chosen in the range of $\delta = 0.001–0.015$ based on the average performances.

3. Results and discussion

3.1. Individual target inhibitors and dual inhibitors of the studied target pairs

As shown in Table 1, high percentages of the known dual inhibitors of the seven studied target pairs are distributed in the compound families containing individual target inhibitor of at least one target in the target pair. Only 18.4–37.0% of the known dual inhibitors are not in the compound families of the known individual target inhibitors. Nonetheless, dual inhibitors have some features distinguished from those of individual target inhibitors, which are partly exhibited from the top-ranked scaffolds contained in higher percentages of dual inhibitors of the studied target pairs (Fig. 2). Table 5 gives the distribution of some of these scaffolds in the dual inhibitors of the studied target pairs and inhibitors of individual targets of these target pairs. Scaffolds A, B, C, D, E, F and G are contained in high percentages of dual inhibitors. Specifically, scaffold A is contained in 21.8% of the 101 NETSRIs, scaffold B in 17.7% of the 147 H3SRIs, scaffold C in 14.8% of the 216 5HT1aSRIs, scaffold D in 14.8% of the 27 5HT2cSRIs, scaffold E in 100% of the 6 MC4SRIs, and scaffold F and G in 44.4% and 33.3% of the 45 NK1SRIs, whereas these scaffolds are contained in single-digit percentages or less of the inhibitors of other target pairs and the individual target inhibitors of the specific target-pairs. Known 5HT1bSRIs appear to be distributed in many scaffolds each containing no more than three compounds. Nonetheless, some specific variations of side-chain groups of these and other scaffolds found in the known 5HT1bSRIs as well as known NETSRIs, H3SRIs and 5HT1aSRIs appear to be sufficient to convert individual target inhibitors into dual inhibitors. Moreover, physicochemical properties as well as structural features are also important for distinguishing individual target inhibitors and dual inhibitors.

3.2. 5-Fold cross-validation tests of SVM, k-NN and PNN models

The parameters of our SVM, k-NN and PNN models were determined by 5-fold cross-validation studies of individual target inhibitors and putative non-inhibitors of each target pair. Additionally, each 5-fold cross-validation model was tested by dual target NETSRIs, H3SRIs, 5HT1aSRIs, 5HT1bSRIs, 5HT2cSRIs, MC4SRIs and NK1SRIs and real non-inhibitors of the individual target of each target pair. Non-inhibitors of a target refer to compounds with IC50 or K_i value >20 μM. The results of these tests for SVM, k-NN and PNN are shown in Tables 2–4 respectively. The 5-fold cross-validation tests were measured by sensitivity, specificity and over all accuracy given as TP/(TP + FN), TN/(TN + FP) and TP + TN/(TP + TN + FP + FN) respectively in terms of the numbers of true positives TP (true inhibitors), true negatives TN (true non-inhibitors), false positives FP (false inhibitors), and false negatives FN (false non-inhibitors). Overall, the sensitivity of SVM, k-NN and PNN is in the range of 78.0–99.8%, 79–99.7% and 89–99.7%, the specificity in the range of 99.4–99.98%, 99–99.98%, and 95.1–99.4%, and overall accuracy in the range of 93.6–99.6%, 99.0–99.98%, and 96.5–99.3% respectively. The dual inhibitor accuracy of SVM, k-NN and PNN are in the range of 15–83%, 10–83%, and 17–58% respectively. The non-inhibitor

Table 3
5-Fold cross-validation of k-NN models for parameter selection and additional tests of these models for predicting dual-inhibitors and non-inhibitors, SEN sensitivity, SPE specificity, AC overall accuracy, AVE average; SD standard deviation, and SEM standard error of means.

Target pair	C.V. group	5-Fold C.V. for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			NET			NETSRIs	Non-SSRIs	Non-NRIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–NET	1	85%	99.60%	99%	89%	99.60%	99.30%	40%	83%	62%
	2	84%	99.60%	99%	88%	99.40%	99.20%	38%	89%	84%
	3	88%	99.60%	99%	86%	99.60%	99.30%	37%	89%	81%
	4	89%	99.60%	99%	87%	99.60%	99.30%	36%	80%	83%
	5	87%	99.60%	99%	91%	99.50%	99.30%	39%	85%	82%
	AVE	87%	99.60%	99%	88%	99.50%	99%	38%	85%	78%
	S.D	0.023	0	0	0.018	0.001	0.001	0.02	0.04	0.09
	S.E.M	0.01	0	0	0.008	0	0	0.01	0.02	0.04

Target pair	C.V. group	5-Fold C.V. for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			H3			H3SRIs	Non-SSRIs	Non-H3Is
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–H_3	1	91%	99.50%	99.20%	93%	99.50%	99.40%	10%	85%	88%
	2	92%	99.40%	99.20%	92%	99.60%	99.40%	14%	80%	83%
	3	87%	99.30%	99.00%	92%	99.60%	99.40%	13%	77%	83%
	4	90%	99.50%	99.30%	93%	99.60%	99.50%	10%	84%	88%
	5	91%	99.40%	99.10%	91%	99.50%	99.30%	12%	87%	88%
	AVE	90%	99%	99%	92%	99.60%	99%	12%	82%	86%
	S.D	0.02	0.001	0.001	0.008	0.001	0.001	0.02	0.04	0.03
	S.E.M	0.009	0	0.001	0.003	0	0	0.01	0.02	0.01

Target pair	C.V. group	5-Fold C.V. for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			5HT1a			5HT1aSRIs	Non-SSRIs	Non-5HT1aIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–$5HT_{1A}$	1	89.30%	99.50%	99.20%	85%	99.60%	99.40%	32%	83%	79%
	2	89.30%	99.40%	99.10%	78%	99.50%	99.20%	33%	89%	81%
	3	90.80%	99.40%	99.20%	84%	99.60%	99.30%	34%	80%	82%
	4	93.80%	99.40%	99.30%	85%	99.60%	99.30%	36%	80%	78%
	5	89.60%	99.60%	99.30%	84%	99.60%	99.30%	35%	79%	82%
	AVE	91%	99%	99%	83%	99.60%	99%	34%	82%	80%
	S.D	0.019	0.001	0.001	0.029	0	0.001	0.01	0.04	0.02
	S.E.M	0.009	0	0	0.013	0	0	0.01	0.02	0.01

Target pair	C.V. group	5-Fold C.V. for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			5HT1b			5HT1bSRIs	Non-SSRIs	Non-5HT1bIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–$5HT_{1B}$	1	91%	99.50%	99.20%	84%	99.70%	99.50%	30%	82%	83%
	2	92%	99.30%	99.10%	86%	99.80%	99.60%	30%	86%	84%
	3	91%	99.40%	99.10%	86%	99.70%	99.50%	25%	82%	84%
	4	91%	99.40%	99.10%	82%	99.60%	99.40%	30%	88%	87%
	5	92%	99.30%	99.10%	83%	99.70%	99.40%	28%	88%	84%
	Average	91%	99%	99%	84%	99.70%	99%	28%	85%	84%
	S.D	0.006	0.001	0	0.019	0.001	0.001	0.02	0.03	0.01
	S.E.M	0.003	0	0	0.008	0	0	0.01	0.01	0.01

Target pair	C.V. group	5-Fold C.V. for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			5HT2c			5HT2cSRIs	Non-SSRIs	Non-5HT2cIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–$5HT_{2C}$	1	92.2%	99.4%	99.2%	86%	99%	99.2%	26%	88%	87%
	2	90.1%	99.3%	99.1%	85%	99.6%	99.3%	26%	87%	83%
	3	90.9%	99.3%	99.1%	82%	99%	99.1%	26%	86%	84%
	4	91.9%	99.3%	99.1%	79%	99.5%	99.2%	26%	88%	82%
	5	90.9%	99.4%	99.2%	81%	99.5%	99.2%	26%	87%	82%
	AVE	83%	99.5%	99%	83%	99.5%	99.2%	26%	87%	84%
	S.D	0.028	0.000	0.000	0.028	0.000	0.001	0.00	0.01	0.02
	S.E.M	0.012	0.000	0.000	0.012	0.000	0.000	0.00	0.00	0.01

Table 3 (*continued.*)

Target pair	C.V. group	5-Fold C.V. for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			MC4			MC4SRIs	Non-SSRIs	Non-MC4Is
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–MC_4	1	89.8%	99.3%	99.0%	99.7%	99.98%	99.98%	17%	84%	95%
	2	91.3%	99.3%	99.1%	98.3%	99.7%	99.7%	17%	81%	90%
	3	92.8%	99.5%	99.3%	97.7%	99.8%	99.7%	17%	86%	90%
	4	88.7%	99.4%	99.1%	98.0%	99.7%	99.7%	83%	84%	95%
	5	91.5%	99.4%	99.1%	99.1%	99.8%	99.7%	17%	83%	97%
	AVE	91%	99.4%	99%	98.5%	99.8%	99.8%	30.2%	84%	93.4%
	S.D	0.016	0.001	0.001	0.008	0.001	0.001	0.30	0.02	0.03
	S.E.M	0.007	0.000	0.000	0.004	0.0005	0.001	0.132	0.01	0.01
Target pair	C.V. group	5-Fold C.V. for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			NK1			NK1SRIs	Non-SSRIs	Non-NK1Is
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–NK_1	1	91.4%	99.3%	99.1%	95.5%	99.7%	99.6%	20%	88%	81%
	2	92.7%	99.4%	99.2%	95.0%	99.5%	99.4%	24%	89%	82%
	3	91.6%	99.3%	99.1%	95.0%	99.6%	99.4%	22%	88%	85%
	4	90.3%	99.5%	99.2%	95.8%	99.6%	99.5%	20%	87%	85%
	5	89.3%	99.4%	99.1%	95.2%	99.5%	99.4%	20%	87%	85%
	AVE	91%	99.4%	99%	95%	99.6%	99%	21%	88%	84%
	S.D	0.013	0.001	0.001	0.004	0.001	0.001	0.02	0.01	0.02
	S.E.M	0.006	0.000	0.000	0.002	0.000	0.000	0.01	0.00	0.01

prediction accuracy of SVM, k-NN and PNN are in the range of 73–100%, 62–97% and 72–89% respectively. Therefore, SVM showed comparable overall performance in these 5-fold cross-validation tests.

3.3. Virtual screening performance of combinatorial SVM in searching multi-target serotonin inhibitors from large compound libraries

The VS performance of COMBI-SVM in identifying dual inhibitors of the seven target-pairs is summarised in Table 6 together with the similarity level (sequence identity) between the drug-binding domains of each target pair. Rost has found that proteins with >40% sequence identity unambiguously distinguish similar and non-similar structures and the signal gets blurred in the twilight zone of 20–35% sequence identity [61]. Thus, target-pairs can be classified into high, intermediate, and low similarity classes with their drug-binding domains at sequence identity levels of >40%, 20–40% and <20% respectively. Based on this criterion, SERT–NET with 72.3% drug-binding domain sequence identity is of high similarity, while the other six target-pairs with 1.7–15.1% drug-binding domain sequence identities are of low sequence similarity (Table 6).

In terms of the numbers of true positives TP (true inhibitors), true negatives TN (true non-inhibitors), false positives FP (false inhibitors), and false negatives FN (false non-inhibitors), the yield and false-hit rate are given by TP/(TP + FN) and FP/(TP + FP) respectively. The dual inhibitor yields are 49.5% for NETSRIs, 25.9% for H3SRIs, 47.7% for 5HT1aSRIs, and 22.8% for 5HT1bSRIs, 22.0% for 5HT2cSRIs, 83.3% for MC4SRIs and 31.1% for NK1SRIs respectively. Therefore, COMBI-SVMs showed reasonably good capability in identifying dual inhibitors of the seven evaluated target pairs without explicit knowledge of dual inhibitors. Target selectivity was tested by using COMBI-SVM to screen the 917–1951 individual target inhibitors of each target pair, which misidentified 22.4% and 29.8% of the individual target inhibitors as dual inhibitors for the

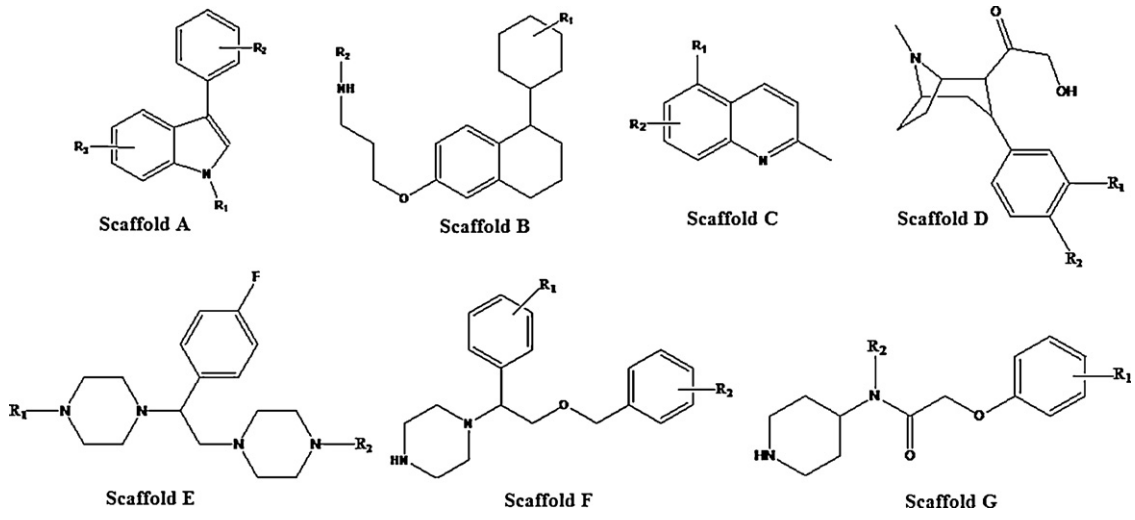

Fig. 2. Top-ranked molecular scaffolds primarily found in known multi-target serotonin reuptake inhibitors. Scaffolds A, B and C are distributed in significantly higher percentage of known multi-target NETSRIs, H3SRIs, and 5HT1aSRIs than known "individual-target" inhibitor.

Table 4
5-Fold cross-validation of PNN models for parameter selection and additional tests of these models for predicting dual-inhibitors and non-inhibitors, SEN sensitivity, SPE specificity, AC overall accuracy, AVE average; SD standard deviation, and SEM standard error of means.

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			NET			NETSRIs	Non-SSRIs	Non-NRIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–NET	1	94%	97.50%	97.50%	95%	97.00%	96.90%	58%	86%	80%
	2	94%	97.60%	97.50%	96%	96.60%	96.60%	57%	88%	79%
	3	95%	97.50%	97.50%	94%	97.00%	96.90%	50%	85%	76%
	4	94%	97.40%	97.30%	94%	96.90%	96.80%	55%	84%	72%
	5	93%	97.70%	97.60%	97%	96.90%	96.90%	55%	83%	76%
	AVE	94%	98%	98%	95%	97%	97%	55%	85%	77%
	S.D	0.005	0.001	0.001	0.013	0.002	0.002	0.03	0.02	0.03
	S.E.M	0.002	0	0	0.006	0.001	0.001	0.01	0.01	0.01

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			H3			H3SRIs	Non-SSRIs	Non-H3Is
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–H_3	1	96%	96.80%	96.80%	98%	97.50%	97.50%	40%	80%	90%
	2	96%	96.80%	96.80%	97%	97.70%	97.70%	39%	84%	84%
	3	94%	97.00%	96.90%	96%	97.80%	97.80%	41%	83%	84%
	4	96%	96.90%	96.90%	98%	97.70%	97.70%	34%	83%	84%
	5	96%	96.70%	96.60%	95%	97.90%	97.80%	37%	83%	84%
	AVE	96%	97%	97%	97%	98%	98%	38%	83%	85%
	S.D	0.011	0.001	0.001	0.014	0.002	0.001	0.03	0.01	0.03
	S.E.M	0.005	0.001	0	0.006	0.001	0.001	0.01	0.01	0.01

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			5HT1a			5HT1aSRIs	Non-SSRIs	Non-5HT1aIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–$5HT_{1A}$	1	97.30%	97.00%	97.00%	93%	97.80%	97.70%	46%	85%	72%
	2	94.90%	96.60%	96.60%	91%	97.70%	97.60%	48%	82%	73%
	3	96.40%	96.60%	96.60%	93%	97.50%	97.40%	45%	82%	71%
	4	97.30%	96.70%	96.70%	92%	97.40%	97.30%	49%	84%	72%
	5	97.30%	96.70%	96.70%	92%	97.70%	97.60%	46%	83%	73%
	AVE	97%	97%	97%	92%	98%	98%	47%	83%	72%
	S.D	0.01	0.002	0.002	0.009	0.002	0.002	0.02	0.02	0.01
	S.E.M	0.005	0.001	0.001	0.004	0.001	0.001	0.01	0.01	0

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			5HT1b			5HT1bSRIs	Non-SSRIs	Non-5HT1bIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–$5HT_{1B}$	1	98%	97.00%	97.10%	91%	98.00%	97.90%	44%	78%	83%
	2	96%	96.90%	96.90%	92%	98.40%	98.30%	35%	77%	82%
	3	96%	96.90%	96.90%	92%	98.40%	98.30%	39%	75%	82%
	4	96%	96.60%	96.50%	91%	98.20%	98.10%	47%	75%	83%
	5	97%	97.00%	97.00%	89%	98.20%	98.10%	46%	75%	81%
	AVE	97%	97%	97%	91%	98%	98%	42%	76%	82%
	S.D	0.011	0.002	0.002	0.016	0.002	0.002	0.05	0.01	0.01
	S.E.M	0.005	0.001	0.001	0.007	0.001	0.001	0.02	0.01	0

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			5HT2c			5HT2cSRIs	Non-SSRIs	Non-5HT2cIs
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–$5HT_{2C}$	1	96.1%	97.0%	97.0%	91%	97%	97.4%	33%	84%	84%
	2	98.7%	97.0%	97.0%	89%	97.2%	97.0%	33%	83%	85%
	3	95.8%	96.7%	96.7%	93%	97%	97.3%	30%	83%	84%
	4	98.2%	97.0%	97.0%	91%	97.0%	96.9%	30%	82%	87%
	5	96.4%	96.8%	96.8%	93%	97.1%	97.0%	33%	83%	84%
	AVE	91%	97.2%	97%	91%	97.2%	97.1%	32%	83%	85%
	S.D	0.016	0.002	0.001	0.016	0.002	0.002	0.02	0.01	0.01
	S.E.M	0.007	0.001	0.001	0.007	0.001	0.001	0.01	0.00	0.00

Table 4 (*continued.*)

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			MC4			MC4SRIs	Non-SSRIs	Non-MC4Is
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–MC_4	1	96.2%	96.7%	96.7%	96.5%	99.4%	99.3%	33%	82%	89%
	2	95.6%	97.0%	97.0%	98.5%	99.2%	99.2%	17%	82%	78%
	3	96.4%	97.0%	97.0%	98.5%	99.3%	99.3%	17%	81%	79%
	4	97.2%	96.4%	96.5%	98.8%	99.2%	99.1%	33%	82%	81%
	5	95.1%	97.1%	97.0%	99.7%	99.2%	99.3%	33%	82%	77%
	AVE	96%	96.9%	97%	98.4%	99.3%	99.2%	27%	82%	81%
	S.D	0.008	0.003	0.002	0.011	0.0007	0.0006	0.09	0.00	0.05
	S.E.M	0.003	0.001	0.001	0.005	0.0003	0.0002	0.039	0.00	0.02

Target pair	C.V. group	5-Fold C.V. Performance for parameter selection						5-Fold C.V. tests for dual and non-inhibitors		
		SERT			NK1			NK1SRIs	Non-SSRIs	Non-NK1Is
		SEN	SPE	AC	SEN	SPE	AC	SEN	SPE	SPE
SERT–NK_1	1	96.1%	96.8%	96.8%	97.2%	98.6%	98.6%	29%	83%	89%
	2	95.5%	97.1%	97.1%	96.9%	98.7%	98.6%	33%	83%	85%
	3	97.1%	96.7%	96.7%	98.6%	98.6%	98.6%	36%	83%	86%
	4	96.3%	96.8%	96.8%	97.8%	98.5%	98.5%	33%	82%	85%
	5	96.1%	96.9%	96.8%	98.0%	98.6%	98.6%	33%	83%	83%
	AVE	96%	96.9%	97%	98%	98.6%	99%	33%	83%	86%
	S.D	0.006	0.002	0.001	0.007	0.001	0.001	0.02	0.00	0.02
	S.E.M	0.003	0.001	0.001	0.003	0.000	0.000	0.01	0.00	0.01

SERT–NET pair, 5.4% and 8.2% for SERT–H_3, 15.4% and 19.4% for SERT–$5HT_{1A}$, 13.8% and 12.3% for SERT–$5HT_{1B}$, 14.2% and 12.4% for SERT–$5HT_{2C}$, 2.2% and 8.0% for SERT–MC_4 and 4.2% and 6.3% for SERT–NK_1 respectively. Therefore, COMBI-SVM is reasonably selective in distinguishing multi-target inhibitors from individual-target inhibitors of the same target pair.

There are two possible reasons for the misidentification of a substantial percentage of individual target inhibitors as dual inhibitors. Firstly, SVMs were trained by using individual-target inhibitors only, which may not fully distinguish dual inhibitors from individual target inhibitors. Secondly, some of the misidentified individual target inhibitors may be true dual inhibitors not yet experimentally tested for multi-target activities. It is noted that "mistaken" selection of these individual target inhibitors is still useful for developing single-target antidepressant drug leads.

Target selectivity was further tested by using COMBI-SVM to screen the 8110–8688 (Table 1) inhibitors of the other six targets outside a given target pair with the results summarised in Table 6. We found that 2.4%, 3.5%, 7.1%, 0.95%, 4.0%, 0.58%, and 1.16% of the inhibitors of the other six targets were misclassified as NETSRIs, H3SRIs, 5HT1aSRIs, 5HT1bSRIs, 5HT2cSRIs, MC4SRIs and NK1SRIs respectively. Therefore, COMBI-SVM is fairly selective in separating multi-target inhibitors of specific target pair from antidepressant inhibitors of other targets outside the target pair.

Virtual hit rates and false hit rates of COMBI-SVM in screening compounds that resemble the structural and physicochemical

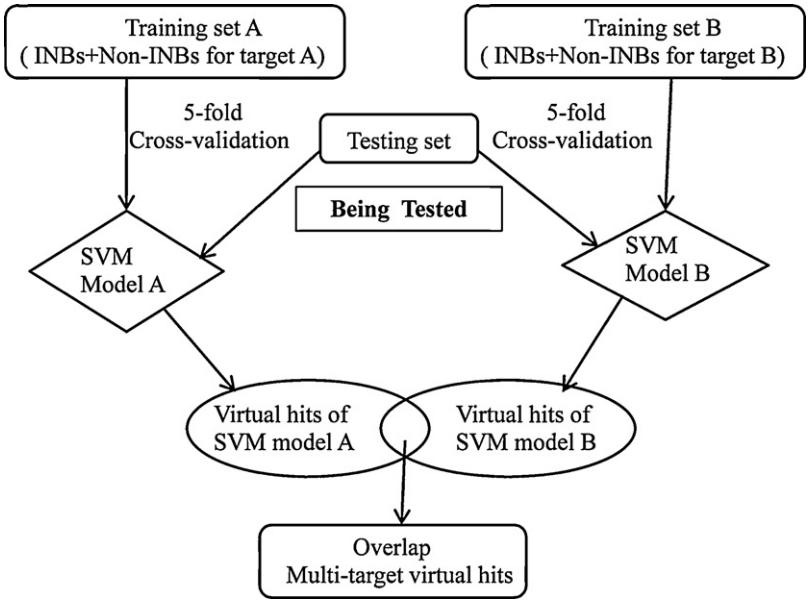

Fig. 3. The COMBI-SVMs diagram. Individual SVM models are built after 5-fold cross-validation where the training set is randomly divided into 5 sub sets and in turns 4 sets are used for training and 1 set for testing to choose the best parameters for model construction. Virtual hits simultaneously selected by all individual SVM models are considered as multi-target virtual hits. INBs = inhibitors; Non-INBs = non-inhibitors.

Table 5
Distribution of the top-ranked scaffolds in multi-target inhibitors of the seven target pairs SERT–NET, SERT–H_3, SERT–$5HT_{1A}$, SERT–$5HT_{1B}$, SERT–$5HT_{2C}$, SERT–MC_4 and SERT–NK_1.

Target pair	Datasets	Percent of inhibitors in dataset Containing scaffold						
		Scaffold A	Scaffold B	Scaffold C	Scaffold D	Scaffold E	Scaffold F	Scaffold G
SERT–NET	Multi-target NETSRIs	21.8 (22/101)	0 (0/101)	3 (3/101)	1.0 (1/101)	0 (0/101)	0 (0/101)	0 (0/101)
	NET reuptake inhibitors inactive against SERT	0.07 (1/1410)	0 (0/1410)	0 (0/1410)	3.0 (43/1410)	0.6 (8/1410)	0 (0/1410)	0 (0/1410)
	SERT reuptake inhibitors inactive against NET	2 (23/1125)	2.1 (24/1125)	1.7 (24/1125)	0.4 (5/1125)	0 (0/1125)	0.2 (2/1125)	1.3 (15/1125)
SERT–H_3	Multi-target H3SRIs	0 (0/147)	17.7 (26/147)	0 (0/147)	0 (0/147)	0 (0/147)	0 (0/147)	0 (0/147)
	H_3 receptor antagonists inactive against SERT	0 (0/1689)	0 (0/1689)	0.4 (6/1689)	0 (0/1689)	0 (0/1689)	0 (0/1689)	0 (0/1689)
	SERT reuptake inhibitors inactive against H_3 receptors	1.5 (27/1804)	0 (0/1804)	1.3 (24/1804)	1.8 (32/1804)	0 (0/1804)	1.0 (18/1804)	0.9 (16/1804)
SERT–$5HT_{1A}$	Multi-target 5HT1aSRIs	0 (0/216)	0 (0/216)	14.8 (32/216)	0 (0/216)	0 (0/216)	0 (0/216)	0 (0/216)
	$5HT_{1A}$ receptor antagonists inactive against SERT	4.8 (55/1144)	0 (0/1144)	1.4 (16/1144)	0 (0/1144)	0 (0/1144)	0 (0/1144)	0 (0/1144)
	SERT reuptake inhibitors inactive against $5HT_{1A}$ receptors	1.3 (21/1679)	1.5 (26/1678)	0.8 (13/1679)	1.8 (31/1679)	0 (0/1679)	1.1 (18/1679)	0.9 (15/1679)
SERT–$5HT_{1B}$	Multi-target 5HT1bSRIs	1.8 (1/57)	0 (0/57)	7 (4/57)	5.3 (3/57)	0 (0/57)	0 (0/57)	0 (0/57)
	$5HT_{1B}$ receptor antagonists inactive against SERT	0 (0/917)	0 (0/917)	0.3 (3/917)	0 (0/917)	0 (0/917)	0 (0/917)	0 (0/917)
	SERT reuptake inhibitors inactive against $5HT_{1B}$ receptors	1.4 (26/1894)	1.4 (26/1894)	1.1 (20/1894)	1.4 (26/1894)	0 (0/1894)	1.0 (18/1894)	0.8 (15/1894)
SERT–$5HT_{2C}$	Multi-target 5HT2cSRIs	3.7 (1/27)	0 (0/27)	3.7 (1/27)	14.8 (4/27)	0 (0/27)	0 (0/27)	0 (0/27)
	$5HT_{2C}$ receptor antagonists inactive against SERT	1.6 (20/1234)	0 (0/1234)	(3/1234)	0 (0/1234)	0 (0/1234)	0 (0/1234)	0 (0/1234)
	SERT reuptake inhibitors inactive against $5HT_{2C}$ receptors	0 (0/1924)	0 (0/1924)	1.2 (23/1924)	1.5 (29/1924)	0 (0/1924)	0.9 (18/1924)	0.7 (13/1924)
SERT–MC_4	Multi-target MC4SRIs	0 (0/6)	0 (0/6)	0 (0/6)	0 (0/6)	100% (6/6)	0 (0/6)	0 (0/6)
	MC_4 receptor antagonists inactive against SERT	0 (0/1721)	0 (0/1721)	0 (0/1721)	0 (0/1721)	2.5 (43/1721)	0 (0/1721)	0 (0/1721)
	SERT reuptake inhibitors inactive against MC_4 receptors	0 (0/1951)	0 (0/1951)	1.2 (23/1951)	1.6 (31/1951)	0 (0/1951)	0.9 (18/1951)	0.8 (15/1951)
SERT–NK_1	Multi-target NK1SRIs	0 (0/45)	0 (0/45)	0 (0/45)	0 (0/45)	0 (0/45)	44.4 (20/45)	33.3 (15/45)
	NK1 receptor antagonists inactive against SERT	0.2 (4/1787)	0 (0/1787)	0 (0/1787)	0 (0/1787)	0 (0/1787)	0.06 (1/1787)	0 (0/1787)
	SERT reuptake inhibitors inactive against NK_1 receptors	0.1 (2/1910)	0 (0/1910)	1.0 (20/1910)	1.0 (33/1910)	0 (0/1910)	0.9 (18/1910)	0.6 (11/1910)

Table 6
The virtual screening performance of combinatorial SVMs for identifying multi-target serotonin inhibitors of the seven target pairs SERT–NET, SERT–H_3, SERT–$5HT_{1A}$, SERT–$5HT_{1B}$, SERT–$5HT_{2C}$, SERT–MC_4 and SERT–NK_1. The target-pairs in this table are arranged with decreasing similarity level between their drug-binding domains. There are only 7 MDDR compounds similar to a dual-inhibitor of SERT–MC_4, the corresponding virtual hit rate was thus un-computed because the small number of compounds may not provide statistically meaningful test of the SVM performance.

Target pair	Virtual screening performance							
Target A–Target B (sequence identity between drug-binding domain)	Multi-target inhibitors		Inhibitors of individual target of the target pair inactive against another target of the target pair		Inhibitors of other three targets outside the target pair	MDDR compounds similar to multi-target inhibitors of the target pair	All 168,000 MDDR compounds	17 million PubChem compounds
	Yield	No. (%) of identified true hits outside the common training active families of both targets	False hit rate for inhibitors of Target A	False hit rate for inhibitors of Target B	False hit rate	Virtual hit rate (no. of virtual hits)	Virtual hit rate (no. of virtual hits)	Virtual hit rate (no. of virtual hits)
SERT–NET (72.3%)	49.50%	8 (7.9%)	22.40%	29.80%	2.40%	0.99% (81)	0.12% (201)	0.035% (6305)
SERT–$5HT_{1B}$ (15.1%)	22.8%	2 (3.5%)	13.80%	12.30%	0.95%	2.5% (185)	0.14% (241)	0.011% (6326)
SERT–MC_4 (11.7%)	83.33%	0	2.20%	8.02%	0.58%	–	0.042% (70)	0.007% (1252)
SERT–NK_1 (9.6%)	31.11%	13 (28.9%)	4.20%	6.30%	1.16%	0.36% (1)	0.055% (92)	0.006% (1136)
SERT–$5HT_{1A}$ (8%)	47.70%	12 (5.6%)	15.40%	19.40%	7.10%	3.5% (256)	0.28% (464)	0.054% (9603)
SERT–$5HT_{2C}$ (3.2%)	22.0%	5 (18.5%)	14.24%	12.40%	4.0%	4.0% (52)	0.21% (353)	0.042% (7574)
SERT–H_3 (1.7%)	25.90%	7 (4.8%)	5.40%	8.20%	3.50%	0.2% (3)	0.067% (112)	0.028% (4993)

properties of the training datasets were evaluated by using 7–8181 MDDR compounds (Table 1) similar to a multi-target inhibitor of each target pair. Similarity was defined by Tanimoto similarity coefficient ≥0.9 between a MDDR compound and its closest dual inhibitor [46]. As shown in Table 6, COMBI-SVM identified 81, 3, 256, 249, 66, 1 and 1 virtual-hit(s) from 8181, 1486, 7349, 7475, 1302, 7 and 275 MDDR compounds similar to NETSRI, H3SRI, 5HT1aSRI, 5HT1bSRI, 5HT2cSRI, MC4RI and NK1SRI respectively. Disregarding the target pair SERT–MC_4 with <10 MDDR compounds similar to the dual inhibitors (which is statistically less meaningful for estimating virtual hit rates), the virtual hit rates in selecting MDDR compounds similar to the dual inhibitors are in the range of 0.2–5.1%. As majority of the MDDR compounds similar to the known dual inhibitors are expected to be non-inhibitors for the target pairs, these virtual hit rates can be considered as the upper limit of the false-hit rates.

Significantly lower virtual hit rates and thus false-hit rates were found in screening large libraries of 168,000 MDDR and 17 million PubChem compounds. As shown in Table 6, the numbers of multi-target virtual hits (virtual hit rate) in screening 168,000 MDDR compounds are 201 (0.12%) for NETSRIs, 112 (0.067%) for H3SRIs, 464 (0.28%) for 5HT1aSRIs, 241 (0.14%) for 5HT1bSRIs, 353 (0.21%) for 5HT2cSRIs, 70 (0.042%) for MC4SRIs and 92 (0.055%) for NK1SRIs respectively. The numbers of multi-target virtual hits (virtual hit rate) in screening 17 million PubChem compounds are 6305 (0.035%) for NETSRIs, 4993 (0.028%) for H3SRIs, 9603 (0.054%) for 5HT1aSRIs, 6326 (0.011%) for 5HT1bSRIs, 7574 (0.042%) for 5HT2cSRIs, 1252 (0.007%) for MC4SRIs and 1136 (0.006%) for NK1SRIs respectively. Substantial percentages of the MDDR virtual hits belong to the classes of antidepressant, anxiolytic, antimigraine, and antipsychotic (Table 7, details in the next section), some of which may be true multi-target serotonin inhibitors. Therefore, the true false hits rates of our COMBI-SVM are likely smaller than the computed rates, i.e., the false hit rates of COMBI-SVM are ≤0.2–4.0%, ≤0.042–0.28% and ≤0.011–0.054% in screening MDDR similarity compounds, all MDDR compounds, and PubChem compounds respectively. These rates are similar to the false hit rates of ≤1.4–9.4%, ≤0.057–0.104%, and ≤0.013–0.036% in COMBI-SVM screening of multi-target kinase inhibitors from MDDR and PUBCHEM compounds [39]. These rates are also comparable and sometime better than the false-hit rates of 0.02–0.37% and 0.05–0.35% produced by other machine learning methods and molecular docking tools [39].

3.4. Evaluation of combinatorial SVM identified MDDR virtual hits

COMBI-SVM identified MDDR virtual hits were evaluated based on the known biological or therapeutic target classes specified in MDDR. Table 7 gives the MDDR classes in which higher percentage (≥5%) of COMBI-SVM identified MDDR dual inhibitor virtual hits are distributed. We found that 15–177 (21.4–38.1%), 10–76 (7.5–21.5%), and 4–53 (5.7–22.0%) of the 70–464 dual-inhibitor virtual hits of the seven target-pairs belong to the antidepressant, anxiolytic and 5HT reuptake inhibitor class respectively. It is noted that serotonin reuptake inhibitors have been used as antidepressant and anxiolytic agents [1]. Therefore, some of the COMBI-SVM virtual hits are either known SSRIs or have the same therapeutic actions of SSRIs, which were misidentified as dual inhibitors by COMBI-SVM partly because it has 2.2–22.4% false hit rates in misclassifying SSRIs as dual inhibitors of the seven target pairs (Table 6). Moreover, 20 (10.0%) of the 201 SERT–NET dual inhibitor virtual hits belong to the noradrenaline uptake inhibitor class. While some of these virtual hits might be true SERT–NET dual inhibitors, most of these individual target NET inhibitors were falsely selected as dual inhibitors by COMBI-SVM at 6.33% false-hit rate (Table 7).

We found that 118 (25.5%), 76 (31.5%), 36 (10.2%) and 14 (7.0%) MDDR virtual hits for SERT–$5HT_{1A}$, SERT–$5HT_{1B}$, SERT–$5HT_{2C}$ and SERT–NET belong to the antimigraine class respectively. Serotonin has been implicated in migraine pathophysiology with a low 5-HT state facilitating activation of the trigeminovascular nociceptive pathway [62]. Because serotonin is primarily reduced by SERT [12], serotonin reuptake inhibitors may in some circumstances have antimigraine effect in certain patients [63]. Some of the MDDR antimigraine virtual hits may be selected by COMBI-SVM partly because they are SERT inhibitors (COMBI-SVMs select individual-target inhibitors as dual-target serotonin reuptake inhibitors at

Table 7
MDDR classes in which higher percentage (≥5%) of COMBI-SVM identified MDDR multi-target virtual hits are distributed in.

Target pair (no. of COMBI-SVM virtual hits)	MDDR class that contains higher percentage of these virtual hits	Number (%) of COMBI-SVM identified multi-target virtual-hits in class	ercentage of MDDR class members as virtual hits
SERT–NET (201)	Antidepressant	56 (27.9%)	0.91%
	5-HT reuptake inhibitor	36 (17.9%)	3.68%
	Dopamine reuptake inhibitor	28 (13.9%)	14.29%
	Antipsychotic	25 (12.4%)	0.48%
	Noradrenaline uptake inhibitor	20 (10.0%)	6.33%
	Treatment of cocaine dependency	20 (10.0%)	25.97%
	Anxiolytic	15 (7.5%)	0.22%
	Calcium channel blocker	15 (7.5%)	0.88%
	Antimigraine	14 (7.0%)	0.81%
	Analgesic, non-opioid	13 (6.5%)	0.27%
SERT–H_3 (112)	Antidepressant	30 (26.8%)	0.49%
	Antipsychotic	23 (20.5%)	0.44%
	Analgesic, non-opioid	13 (11.6%)	0.27%
	5-HT reuptake inhibitor	10 (8.9%)	1.02%
	Anxiolytic	10 (8.9%)	0.15%
	Cognition disorders, agent for	10 (8.9%)	0.13%
	Antiparkinsonian	9 (8.0%)	0.48%
	Anticonvulsant	8 (7.1%)	0.26%
	Antifungal	7 (6.3%)	0.24%
	Calcium channel blocker	7 (6.3%)	0.41%
SERT–$5HT_{1A}$ (464)	Antidepressant	177 (38.1%)	11.91%
	Antimigraine	118 (25.4%)	2.27%
	Antipsychotic	113 (24.3%)	1.67%
	5-HT_{1D} receptor agonist	100 (21.6%)	10.21%
	Anxiolytic	62 (13.4%)	3.58%
	5-HT reuptake inhibitor	49 (10.6%)	8.52%
	5-HT_{1A} receptor agonist	48 (10.3%)	4.66%
	5 HT2A antagonist	47 (10.1%)	7.40%
	Dopamine (D_4) antagonist	26 (5.6%)	8.05%
	Analgesic, non-opioid	25 (5.4%)	3.63%
SERT–$5HT_{1B}$ (241)	Antidepressant	82 (34.0%)	1.33%
	Antimigraine	76 (31.5%)	4.39%
	5-HT1D receptor agonist	63 (26.1%)	9.62%
	5-HT reuptake Inhibitor	53 (22.0%)	5.41%
	Antipsychotic	47 (19.5%)	0.9%
	Anxiolytic	44 (18.32%)	0.65%
	5-HT2A receptor antagonist	25 (10.4%)	3.63%
	5-HT1Areceptor agonist	21 (8.7%)	2.0%
	Dopamine (D_4) antagonist	15 (6.2%)	2.23%
	5-HT_{1A} receptor antagonist	13 (5.4%)	2.26%
SERT–$5HT_{2C}$ (353)	Antipsychotic	126 (5.7%)	2.42%
	Antidepressant	99 (28.0%)	1.60%
	Anxiolytic	76 (21.5%)	1.12%
	5-HT_{2A} receptor antagonist	46 (13.0%)	6.68%
	Antimigraine	36 (10.2%)	2.08%
	5-HT_{1A} receptor agonist	34 (9.6%)	3.30%
	Dopamine (D_4) antagonist	32 (9.1%)	4.75%
	5-HT reuptake inhibitor	24 (6.8%)	2.45%
	Antiparkinsonian	22 (6.2%)	1.16%
	5-HT_{1D} agent agonist	22 (6.2%)	1.16%
	Antihypertensive	21 (5.9%)	0.19%
	Antiallergic/antiasthmatic	18 (5.1%)	0.17%
	Cognition disorders, agent for	18 (5.1%)	0.24%
SERT–MC_4 (70)	Antidepressant	15 (21.4%)	0.24%
	Anti-allergic/anti-asthmatic	15 (21.4%)	0.14%
	Anxiolytic	13 (18.6%)	0.19%
	Neurokinin NK_2 antagonist	9 (12.9%)	2.16%
	Neurokinin NK_3 antagonist	8 (11.4%)	4.40%
	Antipsychotic	7 (10.0%)	0.13%
	Substance P antagonist	7 (10.0%)	0.40%
	Antiviral (AIDS)	5 (7.1%)	0.11%
	Analgesic, non-opioid	4 (5.7%)	0.08%
	Cognition disorders, agent for	4 (5.7%)	0.05%
	5-HT reuptake inhibitor	4 (5.7%)	0.41%
	Anti-arthritic	4 (5.7%)	0.03%

Table 7 (*Continued*)

+ Target pair (no. of COMBI-SVM virtual hits)	MDDR class that contains higher percentage of these virtual hits	Number (%) of COMBI-SVM identified multi-target virtual-hits in class	Percentage of MDDR class members as virtual hits
SERT–NK_1 (92)	Substance P antagonist	23 (25.0%)	1.31%
	Antidepressant	18 (19.6%)	0.29%
	Anxiolytic	15 (16.3%)	0.22%
	Antipsychotic	11 (12.0%)	0.21%
	Antiallergic/antiasthmatic	11 (12.0%)	0.10%
	5-HT reuptake inhibitor	11 (12.0%)	1.12%
	Analgesic, non-opioid	10 (10.9%)	0.20%
	Neurokinin NK_2 antagonist	9 (9.8%)	2.16%
	Calcium channel blocker	9 (9.8%)	0.53%
	5-HT_{1A} receptor agonist	6 (6.5%)	0.58%
	Antihypertensive	6 (6.5%)	0.05%
	Antianginal	6 (6.5%)	0.18%
	Adrenergic (beta) blocker	6 (6.5%)	2.76%
	Antiarrhythmic	5 (5.4%)	0.19%
	Anti-inflammatory	5 (5.4%)	0.09%
	Neurokinin antagonist	5 (5.4%)	3.73%
	Cognition disorders, agent for	5 (5.4%)	0.07%
	5-HT_{1A} receptor antagonist	5 (5.4%)	0.87%
	Antiviral (AIDS)	5 (5.4%)	0.11%

2.2–29.8% false-hit rates based on the statistics in Table 6). Moreover, 25–113 (11.4–24.3%) MDDR virtual hits of six target pairs (SERT–NET, SERT–H_3, SERT–5HT_{1B}, SERT–5HT_{2C}, SERT–MC_4 and SERT–NK_1) belong to the antipsychotic class. Some antipsychotic drugs show certain level of activity against serotonin reuptakes and 5-HT receptors [64]. It is further noted that serotonin reuptake inhibitors augment and synergize with antipsychotic drugs hence serotonin reuptake inhibitors have been used in combination with antipsychotic drugs in the treatment of some psychiatric disorders [65]. Hence, some of the antipsychotic MDDR virtual hits may be selected because they have these activities.

An additional set of 87–100 (21.6–21.7%), 38–48 (9.5–10.3%) and 36–47 (9.0–10.1%) dual inhibitor virtual hits of the SERT–5HT_{1A} and SERT–5HT_{1B} target pairs belong to the 5-HT_{1D} receptor agonist, 5-HT_{1A} receptor agonist, and 5-HT_{2A} receptor antagonist classes respectively. As discussed below, some of the MDDR 5-HT_{1D} receptor agonist, 5-HT_{1A} receptor agonist, and 5-HT_{2A} receptor antagonist virtual hits were falsely selected by COMBI-SVM possibly because they have some level of structural similarity to 5-HT_{1A} receptor antagonists or 5-HT_{1B} receptor antagonists. Analogy of certain scaffolds has been found to bind to both 5-HT_{1A} and 5-HT_{1D} receptors with weak partial agonist activity in cloned receptor and antagonistic activity in *in vitro* studies [66]. Some compounds such as BMY 7378 can act as both 5-HT_{1A} agonist and antagonist depending on the location of 5-HT_{1A}. BMY 7378 shows agonist activity at 5-HT_{1A} autoreceptors in the raphe and act as antagonists or show partial agonist activity at postsynaptic 5-HT_{1A} receptors [67]. Both mixed 5-HT_{1A} and 5-HT_{2A} receptor antagonists and 5-HT_{1A} receptor agonists have been derived from the same scaffolds [68]. The human 5-HT_{1B} and 5-HT_{1D} receptors are significantly similar in sequence despite being encoded by two distinct genes, and some dual 5$HT_{1B/1D}$ receptor antagonists show substantial degree of structural similarity to dual 5$HT_{1B/1D}$ agonists [69]. Some analogs of specific scaffolds are mixed 5-HT_{1B} and 5-HT_{2A} receptor antagonists [70]. Moreover, some compounds have been reported to have dual 5-HT_{1A} receptor agonist and serotonin reuptake inhibitory activities [71]. It is possible that some of the MDDR 5-HT_{1A} receptor agonist virtual hits were selected by the COMBI-SVM of SERT–5HT_{1B} target pair because they have serotonin reuptake inhibitory activity which may be falsely recognized as multi-target 5HT1bSRIs by COMBI-SVM at 13.8% false hit rate based on the statistics in Table 6.

3.5. Comparison of the performance of combinatorial SVM with other virtual screening methods

At present, the 3D structure is unavailable for the eight targets considered in this work (serotonin transporter, noradrenaline transporter, H_3 receptor, 5-HT_{1A} receptor, 5-HT_{1B} receptor, 5-HT_{2C} receptor, NK_1 receptor and MC_4 receptor). Only some of their homologous proteins or other members from the same GPCR families, such as H1 receptor, have 3D structural information available [72,73]. While these structures give important insights into functional mechanism and allow the modelling of ligand binding to the eight evaluated targets, the modelled and homologous structures may not provide sufficiently high quality structural platforms as those of high-resolution crystal structures for fair comparison of the VS performance of COMBI-SVM with molecular docking methods. We therefore only compared the VS performance of COMBI-SVMs with three VS methods, i.e., similarity searching [59], k-NN [74], and PNN [75], by using the common testing datasets composed of 6–216 dual inhibitors of the seven evaluated target pairs, 917–1951 individual target inhibitors of the same target pairs, 8110–8688 inhibitors of the other six target pairs outside a given target pair, and 168,000 MDDR compounds respectively. Similarity searching was conducted against known dual inhibitors of each target pair. The training datasets of k-NN and PNN and the methods for estimating the yield and virtual hit rate are the same as those of SVM.

Table 8 shows the comparison of the performance of COMBI-SVM with the other three VS methods for identifying multi-target inhibitors of the seven target pairs from the four common testing datasets. Overall, the dual inhibitor yields of all VS methods are comparable, mostly in the ranges of 20–83% for the seven target-pairs with the exception of k-NN for SERT–NK_1 (7.7%) and similarity searching for SERT–5HT_{2c} (11.1%). Compared to COMBI-SVM, k-NN produced comparable false-hit rates, and similarity searching and PNN produced slightly higher false-hit rates in misidentifying individual-target inhibitors of the same target-pair and inhibitors of the other six target pairs outside a target pair as dual-inhibitors.

The false hit rates of the similarity searching method may be significantly reduced by adjusting the similarity cut-off values for individual targets, which may however lead to significantly reduced yields. The higher false hit rates likely arise in part from the difficulty in establishing optimal molecular similarity threshold values that correlate with biological activity, and in separating active and inactive close analogs of reference molecules [76]. Data

Table 8
Comparison of the performance of combinatorial SVMs with other virtual screening methods for identifying multi-target inhibitors of the four target pairs.

Virtual screening performance measure	Method	Virtual screening performance for target pair						
		SERT–NET	SERT–H_3	SERT–$5HT_{1A}$	SERT–$5HT_{1B}$	SERT–$5HT_{2C}$	SERT–MC_4	SERT–NK_1
Yield of multi-target inhibitors of target pair	SVM	49.50%	25.90%	47.70%	22.8%	22%	83.33%	31.11%
	Similarity searching	53.50%	20.40%	63.90%	40.40%	11.11%	33.3%	48.89%
	k-NN	59.40%	10.90%	34.30%	31.60%	18.52%	16.7%	6.67%
	PNN	57.40%	38.10%	45.40%	45.60%	33.33%	66.7%	20.00%
False-hit rate for "Individual-Target" inhibitors of the same target pair	SVM	22.4–29.8%	5.4–8.2%	15.4–19.4%	13.8–12.3%	14.24–12.4%	2.2–8.02%	4.2–6.3%
	Similarity searching	46.5–42.4%	35.6–21.8%	28.6–46.9%	28.6–65.4%	19.4–13.6%	8.5–26.3%	17.7–16.0%
	k-NN	19.8–25.1%	9–8.5%	16.6–24.3%	14.1–32.6%	15.0–16.3%	3.1–11.5%	4.5–4.9%
	PNN	38.4–52.3%	22.2–25.5%	34.3–38.9%	30.3–4.7%	34.8–31.8%	6.6–27.9%	11.8–9.8%
False-hit rate for inhibitors of the other six targets outside the target pair	SVM	2.40%	3.50%	7.10%	0.95%	4.0%	0.58%	1.16%
	Similarity searching	15.90%	20.50%	24.10%	11.60%	10.6%	5.0%	8.4%
	k-NN	3.00%	3.50%	9.30%	5.20%	4.2%	1.2%	2.9%
	PNN	16.90%	13.50%	24.20%	10.80%	14.2%	0.37%	8.8%
Virtual hit rate for 168,000 MDDR compounds	SVM	0.12%	0.067%	0.28%	0.14%	0.21%	0.042%	0.055%
	Similarity searching	6.80%	8.20%	7.60%	7.60%	3.81%	3.26%	3.54%
	k-NN	0.58%	0.41%	0.83%	0.75%	0.52%	0.15%	0.81%
	PNN	3.14%	2.35%	3.40%	2.83%	3.90%	0.93%	2.24%

fusion and group fusion approaches may be explored to conduct multiple similarity searches using different sets of molecular representations, similarity measure and parameters followed by the combination of the resulting search outputs to give a single fused output [77,78]. The higher false-hit rates may also arise from the bias linked to molecular complexity and size, i.e., reference molecules of increasing size generate systematically higher Tanimoto coefficient values in database searching [79]. This bias may be partly reduced by exploring bit density reduction methods [79], complexity-independent molecular representations [79] and complexity-independent similarity metrics [79].

In screening the MDDR compounds, COMBI-SVM produced slightly to substantially lower virtual hit rates (0.042–0.28%) than those of similarity searching (2.81–8.2%), k-NN (0.15–0.83%) and PNN (0.93–3.4%) in identifying the MDDR compounds as dual inhibitor virtual hits of the evaluated target pairs. The numbers of MDDR compounds in the antidepressant and 5-HT reuptake inhibitor classes are 6182 and 979 respectively. It is expected that no more than half of the MDDR antidepressant compounds are SSRIs. Therefore, the total number of labelled and unlabelled SSRIs in MDDR can be crudely estimated as ~1000–3000, most likely significantly less than 3000. Assuming that the ratio of the dual target serotonin reuptake inhibitors against SSRIs in MDDR is roughly similar to those of known dual-target serotonin reuptake inhibitors against SSRIs which are 9.0% (101 vs. 1125) for NETSRIs, 8.2% (147 vs. 1804) for H3SRIs, 12.9% (216 vs. 1679) for 5HT1aSRIs, 3.0% (57 vs. 1894) for 5HT1bSRIs, 1.4% (27 vs. 1924) for 5HT2cSRIs, 0.3% (6 vs. 1951) for MC4SRIs and 2.4% (45 vs. 1910) for NK1SRIs. Then the numbers of dual-target serotonin reuptake inhibitors in MDDR can be crudely estimated as ~3–380 ($1000 \times 0.3\%$ to $3000 \times 12.9\%$), most likely significantly less than 380. Therefore the numbers of COMBI-SVM identified MDDR dual inhibitor virtual hits of the evaluated target pairs (70–464) are consistent to the crudely estimated numbers of dual inhibitors in MDDR than the identified numbers from the other three methods (971–12,698).

4. Conclusion

In silico methods have been increasingly explored for facilitating multi-target drug discovery, and shown promising potential in identifying selective multi-target agents. This study further suggested that combinatorial SVM VS tools developed from individual target inhibitors are capable of identifying dual target serotonin reuptake inhibitors at comparably good yields and low false-hit rates, and in some cases substantially lower false-hit rates than some of the other VS tools in screening large chemical libraries. COMBI-SVMs, in combination with other methods, may be useful for facilitating the search of novel multi-target antidepressants by screening larger chemical libraries. With increasing knowledge of newly discovered selective multi-target agents from the current and future drug discovery efforts [80,81], and further improvement of the algorithms and parameters of VS methods [50,82–87], the capability and application ranges of COMBI-SVMs and other *in silico* methods may be further enhanced, particularly in facilitating multi-target drug discovery. The introduction of more comprehensive elements of distinguished structural and physicochemical features of selective multi-target agents and multi-target activity and binding site profiles enable the development of more effective and relevant tools for the identification of selective multi-target agents against selected targets.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.09.002.

References

[1] S.K. Kulkarniand, A. Dhir, Current investigational drugs for major depression, Expert Opin. Investig. Drugs 18 (2009) 767–788.
[2] K.S. Smalley, N.K. Haass, P.A. Brafford, M. Lioni, K.T. Flaherty, M. Herlyn, Multiple signaling pathways must be targeted to overcome drug resistance in cell lines derived from melanoma metastases, Mol. Cancer Ther. 5 (2006) 1136–1144.
[3] C.T. Keith, A.A. Borisy, B.R. Stockwell, Multicomponent therapeutics for networked systems, Nat. Rev. Drug Discov. 4 (2005) 71–78.
[4] Y. Pilpel, P. Sudarsanam, G.M. Church, Identifying regulatory networks by combinatorial analysis of promoter elements, Nat. Genet. 29 (2001) 153–159.
[5] R. Muller, Crosstalk of oncogenic and prostanoid signaling pathways, J. Cancer Res. Clin. Oncol. 130 (2004) 429–444.
[6] N.V. Sergina, M. Rausch, D. Wang, J. Blair, B. Hann, K.M. Shokat, M.M. Moasser, Escape from HER-family tyrosine kinase inhibitor therapy by the kinase-inactive HER3, Nature 445 (2007) 437–441.
[7] M. Christopher, Overall, O. Kleifeld, Validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy, Nat. Rev. Cancer 6 (2006) 227–239.
[8] T. Force, D.S. Krause, R.A. Van Etten, Molecular mechanisms of cardiotoxicity of tyrosine kinase inhibition, Nat. Rev. Cancer 7 (2007) 332–344.
[9] M.J. Millan, Multi-target strategies for the improved treatment of depressive states: conceptual foundations and neuronal substrates, drug discovery and therapeutic application, Pharmacol. Ther. 110 (2006) 135–370.
[10] X.H. Ma, C.J. Zheng, L.Y. Han, B. Xie, J. Jia, Z.W. Cao, Y.X. Li, Y.Z. Chen, Synergistic therapeutic actions of herbal ingredients and their mechanisms from molecular interaction and network perspectives, Drug Discovery Today 14 (2009) 579–588.

[11] L.D. Jayanthiand, S. Ramamoorthy, Regulation of monoamine transporters: influence of psychostimulants and therapeutic antidepressants, AAPS J. 7 (2005) E728–E738.
[12] L.C. Daws, Unfaithful neurotransmitter transporters: focus on serotonin uptake and implications for antidepressant efficacy, Pharmacol. Ther. 121 (2009) 89–99.
[13] G.A. Whitlock, M.D. Andrews, A.D. Brown, P.V. Fish, A. Stobie, F. Wakenhut, Design of monoamine reuptake inhibitors: SSRIs, SNRIs and NRIs, Transporters Targets Drugs: Top. Med. Chem. 4 (2009) 42–53.
[14] C. Davidsonand, J.A. Stamford, Evidence that 5-hydroxytryptamine release in rat dorsal raphe nucleus is controlled by 5-HT1A, 5-HT1B and 5-HT1D autoreceptors, Br. J. Pharmacol. 114 (1995) 1107–1109.
[15] L. Romero, I. Hervas, F. Artigas, The 5-HT1A antagonist WAY-100635 selectively potentiates the presynaptic effects of serotonergic antidepressants in rat brain, Neurosci. Lett. 219 (1996) 123–126.
[16] F. Artigas, L. Romero, C. de Montigny, P. Blier, Acceleration of the effect of selected antidepressant drugs in major depression by 5-HT1A antagonists, Trends Neurosci. 19 (1996) 378–383.
[17] E. Schlicker, R. Betz, M. Gothert, Histamine H3 receptor-mediated inhibition of serotonin release in the rat brain cortex, Naunyn. Schmiedebergs Arch. Pharmacol. 337 (1988) 588–590.
[18] S. Threlfell, S.J. Cragg, I. Kallo, G.F. Turi, C.W. Coen, S.A. Greenfield, Histamine H3 receptors inhibit serotonin release in substantia nigra pars reticulata, J. Neurosci. 24 (2004) 8704–8710.
[19] K.S. Ly, M.A. Letavic, J.M. Keith, J.M. Miller, E.M. Stocking, A.J. Barbier, P. Bonaventure, B. Lord, X. Jiang, J.D. Boggs, L. Dvorak, K.L. Miller, D. Nepomuceno, S.J. Wilson, N.I. Carruthers, Synthesis and biological activity of piperazine and diazepane amides that are histamine H3 antagonists and serotonin reuptake inhibitors, Bioorg. Med. Chem. Lett. 18 (2008) 39–43.
[20] M.J. Millan, Dual- and triple-acting agents for treating core and co-morbid symptoms of major depression: novel concepts, new drugs, Neurotherapeutics 6 (2009) 53–77.
[21] G.A. Whitlock, J. Blagg, V.P. Fish, 1-(2-Phenoxyphenyl)methanamines: SAR for dual serotonin/noradrenaline reuptake inhibition, metabolic stability and hERG affinity, Bioorg. Med. Chem. Lett. 18 (2008) 596–599.
[22] M.J. Fray, P.V. Fish, G.A. Allan, G. Bish, N. Clarke, R. Eccles, A.C. Harrison, J.L. Le Net, S.C. Phillips, N. Regan, C. Sobry, A. Stobie, F. Wakenhut, D. Westbrook, S.L. Westbrook, G.A. Whitlock, Second generation N-(1,2-diphenylethyl)piperazines as dual serotonin and noradrenaline reuptake inhibitors: improving metabolic stability and reducing ion channel activity, Bioorg. Med. Chem. Lett. 20 (2010) 3788–3792.
[23] K. Takeuchi, T.J. Kohn, N.A. Honigschmidt, V.P. Rocco, P.G. Spinazze, S.K. Hemrick-Luecke, L.K. Thompson, D.C. Evans, K. Rasmussen, D. Koger, D. Lodge, L.J. Martin, J. Shaw, P.G. Threlkeld, D.T. Wong, Advances toward new antidepressants beyond SSRIs: 1-aryloxy-3-piperidinylpropan-2-ols with dual 5-HT1A receptor antagonism/SSRI activities. Part 5, Bioorg. Med. Chem. Lett. 16 (2006) 2347–2351.
[24] Z. Shen, P. Siva Ramamoorthy, N.T. Hatzenbuhler, D.A. Evrard, W. Childers, B.L. Harrison, M. Chlenov, G. Hornby, D.L. Smith, K.M. Sullivan, L.E. Schechter, T.H. Andree, Synthesis and structure–activity relationship of novel lactam-fused chroman derivatives having dual affinity at the 5-HT(1A) receptor and the serotonin transporter, Bioorg. Med. Chem. Lett. 20 (2010) 222–227.
[25] L. Matzen, C. van Amsterdam, W. Rautenberg, H.E. Greiner, J. Harting, C.A. Seyfried, H. Bottcher, 5-HT reuptake inhibitors with 5-HT(1B/1D) antagonistic activity: a new approach toward efficient antidepressants, J. Med. Chem. 43 (2000) 1149–1157.
[26] M.J. Millan, M. Brocco, A. Gobert, A. Dekeyne, Anxiolytic properties of agomelatine, an antidepressant with melatoninergic and serotonergic properties: role of 5-HT2C receptor blockade, Psychopharmacology (Berl.) 177 (2005) 448–458.
[27] S. Chaki, Y. Oshida, S. Ogawa, T. Funakoshi, T. Shimazaki, T. Okubo, A. Nakazato, S. Okuyama, MCL0042: a nonpeptidic MC4 receptor antagonist and serotonin reuptake inhibitor with anxiolytic- and antidepressant-like activity, Pharmacol. Biochem. Behav. 82 (2005) 621–626.
[28] T. Ryckmans, L. Balancon, O. Berton, C. Genicot, Y. Lamberty, B. Lallemand, P. Pasau, N. Pirlot, L. Quere, P. Talaga, First dual NK(1) antagonists-serotonin reuptake inhibitors: synthesis and SAR of a new class of potential antidepressants, Bioorg. Med. Chem. Lett. 12 (2002) 261–264.
[29] A. Rupp, K.A. Kovar, G. Beuerle, C. Ruf, G. Folkers, A new pharmophoric model for 5-HT reuptake-inhibitors: differentiation of amphetamine analogues, Pharm. Acta Helv. 68 (1994) 235–244.
[30] R. Bureau, C. Daveu, J.C. Lancelot, S. Rault, Molecular design based on 3D-pharmacophore. Application to 5-HT subtypes receptors, J. Chem. Inf. Comput. Sci. 42 (2002) 429–436.
[31] C.Y. Kim, P.E. Mahaney, O. McConnell, Y. Zhang, E. Manas, D.M. Ho, D.C. Deecher, E.J. Trybulski, Discovery of a new series of monoamine reuptake inhibitors, the 1-amino-3-(1H-indol-1-yl)-3-phenylpropan-2-ols, Bioorg. Med. Chem. Lett. 19 (2009) 5029–5032.
[32] D.J. O'Neill, A. Adedoyin, P.D. Alfinito, J.A. Bray, S. Cosmi, D.C. Deecher, A. Fensome, J. Harrison, L. Leventhal, C. Mann, C.C. McComas, N.R. Sullivan, T.B. Spangler, A.J. Uveges, E.J. Trybulski, G.T. Whiteside, P. Zhang, Discovery of novel selective norepinephrine reuptake inhibitors: 4-[3-aryl-2, 2-dioxido-2,1,3-benzothiadiazol-1(3H)-yl]-1-(methylamino)butan-2-ols (WYE-103231), J. Med. Chem. 53 (2010) 4511–4521.
[33] A.J. Bojarski, Pharmacophore models for metabotropic 5-HT receptor ligands, Curr. Top. Med. Chem. 6 (2006) 2005–2026.
[34] K.C. Weber, L.B. Salum, K.M. Honorio, A.D. Andricopulo, A.B. da Silva, Pharmacophore-based 3D QSAR studies on a series of high affinity 5-HT1A receptor ligands, Eur. J. Med. Chem. 45 (2010) 1508–1514.
[35] S. Lorenzi, M. Mor, F. Bordi, S. Rivara, M. Rivara, G. Morini, S. Bertoni, V. Ballabeni, E. Barocelli, P.V. Plazzi, Validation of a histamine H3 receptor model through structure–activity relationships for classical H3 antagonists, Bioorg. Med. Chem. 13 (2005) 5647–5657.
[36] B. Schlegel, C. Laggner, R. Meier, T. Langer, D. Schnell, R. Seifert, H. Stark, H.D. Holtje, W. Sippl, Generation of a homology model of the human histamine H(3) receptor for ligand docking and pharmacophore-based screening, J. Comput. Aided Mol. Des. 21 (2007) 437–453.
[37] N. Dessalew, QSAR study on dual SET and NET reuptake inhibitors: an insight into the structural requirement for antidepressant activity, J. Enzyme Inhib. Med. Chem. 24 (2009) 262–271.
[38] F. Micheli, P. Cavanni, D. Andreotti, R. Arban, R. Benedetti, B. Bertani, M. Bettati, L. Bettelini, G. Bonanomi, S. Braggio, R. Carletti, A. Checchia, M. Corsi, E. Fazzolari, S. Fontana, C. Marchioro, E. Merlo-Pich, M. Negri, B. Oliosi, E. Ratti, K.D. Read, M. Roscic, I. Sartori, S. Spada, G. Tedesco, L. Tarsi, S. Terreni, F. Visentini, A. Zocchi, L. Zonzini, R. Di Fabio, 6-(3,4-Dichlorophenyl)-1-[(methyloxy)methyl]-3-azabicyclo[4.1.0]heptane: a new potent and selective triple reuptake inhibitor, J. Med. Chem. 53 (2010) 4989–5001.
[39] X.H. Ma, R. Wang, C.Y. Tan, Y.Y. Jiang, T. Lu, H.B. Rao, X.Y. Li, M.L. Go, B.C. Low, Y.Z. Chen, Virtual screening of selective multitarget kinase inhibitors by combinatorial support vector machines, Mol. Pharmacol. 7 (2010) 1545–1560.
[40] X.H. Ma, Z. Shi, C. Tan, Y. Jiang, M.L. Go, B.C. Low, Y.Z. Chen, In silico approaches to multi-target drug discovery: computer aided multi-target drug design, multi-target virtual screening, Pharm. Res. 27 (2010) 739–749.
[41] M.L. Verdonk, V. Berdini, M.J. Hartshorn, W.T. Mooij, C.W. Murray, R.D. Taylor, P. Watson, Virtual screening using protein-ligand docking: avoiding artificial enrichment, J. Chem. Inf. Comput. Sci. 44 (2004) 793–806.
[42] N. Huang, B.K. Shoichet, J.J. Irwin, Benchmarking sets for molecular docking, J. Med. Chem. 49 (2006) 6789–6801.
[43] A. Bender, Databases: compound bioactivities go public, Nat. Chem. Biol. 6 (2010) 309.
[44] T. Liu, Y. Lin, X. Wen, R.N. Jorissen, M.K. Gilson, BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities, Nucleic Acids Res. 35 (2007) D198–D201.
[45] S. Yamane, S. Ishida, Y. Hanamoto, K. Kumagai, R. Masuda, K. Tanaka, N. Shiobara, N. Yamane, T. Mori, T. Juji, N. Fukui, T. Itoh, T. Ochi, R. Suzuki, Proinflammatory role of amphiregulin, an epidermal growth factor family member whose expression is augmented in rheumatoid arthritis patients, J. Inflamm. (Lond.) 27 (2008), 5:5.
[46] X.H. Ma, R. Wang, S.Y. Yang, Z.R. Li, Y. Xue, Y.C. Wei, B.C. Low, Y.Z. Chen, Evaluation of virtual screening performance of support vector machines trained by sparsely distributed active compounds, J. Chem. Inf. Model. 48 (2008) 1227–1237.
[47] T.I. Opreaand, J. Gottfries, Chemography: the art of navigating in chemical space, J. Comb. Chem. 3 (2001) 157–166.
[48] T.F.a.J.-L. Reymond, Virtual exploration of the chemical universe up to 11 atoms of C, N, O, F: assembly of 26.4 million structures (110.9 million stereoisomers) and analysis for new ring systems, stereochemistry, physicochemical properties, compound classes, and drug discovery, J. Chem. Inf. Model. 47 (2007) 342–353.
[49] M.A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl, H. Waldmann, Charting biologically relevant chemical space: a structural classification of natural products (SCONP), Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 17272–17277.
[50] L.Y. Han, X.H. Ma, H.H. Lin, J. Jia, F. Zhu, Y. Xue, Z.R. Li, Z.W. Cao, Z.L. Ji, Y.Z. Chen, A support vector machines approach for virtual screening of active compounds of single and multiple mechanisms from large libraries at an improved hit-rate and enrichment factor, J. Mol. Graphics Modell. 26 (2008) 1276–1286.
[51] M. Glick, J.L. Jenkins, J.H. Nettles, H. Hitchings, J.W. Davies, Enrichment of high-throughput screening data with increasing levels of noise using support vector machines, recursive partitioning, and Laplacian-modified naive Bayesian classifiers, J. Chem. Inf. Model. 46 (2006) 193–200.
[52] W. Tong, Q. Xie, H. Hong, L. Shi, H. Fang, R. Perkins, Assessment of prediction confidence and domain extrapolation of two structure–activity relationship models for predicting estrogen receptor binding activity, Environ. Health Perspect. 112 (2004) 1249–1254.
[53] I. Ijjaali, F. Petitet, E. Dubus, O. Barberan, A. Michel, Assessing potency of c-Jun N-terminal kinase 3 (JNK3) inhibitors using 2D molecular descriptors and binary QSAR methodology, Bioorg. Med. Chem. 15 (2007) 4256–4264.
[54] H. Briemand, J. Gunther, Classifying "kinase inhibitor-likeness" by using machine-learning methods, Chembiochem 6 (2005) 558–566.
[55] Y. Xue, C.W. Yap, L.Z. Sun, Z.W. Cao, J.F. Wang, Y.Z. Chen, Prediction of P-glycoprotein substrates by a support vector machine approach, J. Chem. Inf. Comput. Sci. 44 (2004) 1497–1505.
[56] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.
[57] N. Pochet, F. De Smet, J.A. Suykens, B.L. De Moor, Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction, Bioinformatics 20 (2004) 3185–3195.
[58] F. Liand, Y. Yang, Analysis of recursive gene selection approaches from microarray data, Bioinformatics 21 (2005) 3741–3747.
[59] P. Willett, Chemical Similarity Searching, J. Chem. Inf. Comput. Sci. 38 (1998) 983–996.

[60] J. Bostrom, A. Hogner, S. Schmitt, Do structurally similar ligands bind in a similar fashion? J. Med. Chem. 49 (2006) 6716–6725.
[61] B. Rost, Twilight zone of protein sequence alignments, Protein Eng. 12 (1999) 85–94.
[62] E. Hamel, Serotonin and migraine: biology and clinical implications, Cephalalgia 27 (2007) 1293–1300.
[63] T.A. Smitherman, A.B. Walters, M. Maizels, D.B. Penzien, The use of antidepressants for headache prophylaxis, CNS Neurosci. Ther. (2010).
[64] S.M. Stahland, D.K. Shayegan, The psychopharmacology of ziprasidone: receptor-binding properties and real-world psychiatric practice, J. Clin. Psychiatry 64 (Suppl. 19) (2003) 6–12.
[65] Y. Chertkow, O. Weinreb, M.B. Youdim, H. Silver, Molecular mechanisms underlying synergistic effects of SSRI-antipsychotic augmentation in treatment of negative symptoms in schizophrenia, J. Neural Transm. 116 (2009) 1529–1541.
[66] T.D. Heightman, L.M. Gaster, S.L. Pardoe, J.P. Pilleux, M.S. Hadley, D.N. Middlemiss, G.W. Price, C. Roberts, C.M. Scott, J.M. Watson, L.J. Gordon, V.A. Holland, J. Powles, G.J. Riley, T.O. Stean, B.K. Trail, N. Upton, N.E. Austin, A.D. Ayrton, T. Coleman, L. Cutler, 8-Piperazinyl-2,3-dihydropyrrolo[3,2-g]isoquinolines: potent, selective, orally bioavailable 5-HT1 receptor ligands, Bioorg. Med. Chem. Lett. 15 (2005) 4370–4374.
[67] M. Bykuand, R.L. Gannon, Effects of the 5HT1A agonist/antagonist BMY 7378 on light-induced phase advances in hamster circadian activity rhythms during aging, J. Biol. Rhythms 15 (2000) 300–305.
[68] P. Zajdel, G. Subra, A.J. Bojarski, B. Duszynska, E. Tatarczynska, A. Nikiforuk, E. Chojnacka-Wojcik, M. Pawlowski, J. Martinez, Novel class of arylpiperazines containing N-acylated amino acids: their synthesis, 5-HT1A, 5-HT2A receptor affinity, and in vivo pharmacological evaluation, Bioorg. Med. Chem. 15 (2007) 2907–2919.
[69] A. Slassi, Recent advances in 5-HT1B/1D receptor antagonists and agonists and their potential therapeutic applications, Curr. Top. Med. Chem. 2 (2002) 559–574.
[70] G. McCort, C. Hoornaert, M. Aletru, C. Denys, O. Duclos, C. Cadilhac, E. Guilpain, G. Dellac, P. Janiak, A.M. Galzin, M. Delahaye, F. Guilbert, S. O'Connor, Synthesis and SAR of 3- and 4-substituted quinolin-2-ones: discovery of mixed 5-HT(1B)/5-HT(2A) receptor antagonists, Bioorg. Med. Chem. 9 (2001) 2129–2137.
[71] T. Heinrich, H. Bottcher, K. Schiemann, G. Holzemann, M. Schwarz, G.D. Bartoszyk, C. van Amsterdam, H.E. Greiner, C.A. Seyfried, Dual 5-HT1A agonists and 5-HT re-uptake inhibitors by combination of indole-butyl-amine and chromenonyl-piperazine structural elements in a single molecular entity, Bioorg. Med. Chem. 12 (2004) 4843–4852.
[72] R.E. Hubbard, Structure-based drug discovery and protein targets in the CNS, Neuropharmacology 60 (2011) 7–23.
[73] T. Shimamura, M. Shiroishi, S. Weyand, H. Tsujimoto, G. Winter, V. Katritch, R. Abagyan, V. Cherezov, W. Liu, G.W. Han, T. Kobayashi, R.C. Stevens, S. Iwata, Structure of the human histamine H1 receptor complex with doxepin, Nature 475 (2011) 65–70.
[74] J. Liand, P. Gramatica, Classification and Virtual Screening of Androgen Receptor Antagonists, J. Chem. Inf. Model. 50 (5) (2010) 861–874.
[75] S. Derksen, O. Rau, P. Schneider, M. Schubert-Zsilavecz, G. Schneider, Virtual screening for PPAR modulators using a probabilistic neural network, ChemMedChem 1 (2006) 1346–1350.
[76] Y. Wangand, J. Bajorath, Advanced fingerprint methods for similarity searching: balancing molecular complexity effects, Comb. Chem. High Throughput Screen. 13 (2010) 220–228.
[77] J. Hert, P. Willett, D.J. Wilton, P. Acklin, K. Azzaoui, E. Jacoby, A. Schuffenhauer, New methods for ligand-based virtual screening: use of data fusion and machine learning to enhance the effectiveness of similarity searching, J. Chem. Inf. Model. 46 (2006) 462–470.
[78] P. Willett, Similarity searching using 2D structural fingerprints, Methods Mol. Biol. 672 (2011) 133–158.
[79] D.R. Flower, On the properties of bit string-based measures of chemical similarity, J. Chem. Inf. Comput. Sci. 38 (3) (1998) 379–386.
[80] M. Krugand, A. Hilgeroth, Recent advances in the development of multi-kinase inhibitors, Mini Rev. Med. Chem. 8 (2008) 1312–1327.
[81] A.L. Gill, M. Verdonk, R.G. Boyle, R. Taylor, A comparison of physicochemical property profiles of marketed oral drugs and orally bioavailable anti-cancer protein kinase inhibitors in clinical development, Curr. Top. Med. Chem. 7 (2007) 1408–1422.
[82] A. Bender, J.L. Jenkins, M. Glick, Z. Deng, J.H. Nettles, J.W. Davies, Bayes affinity fingerprints" improve retrieval rates in virtual screening and define orthogonal bioactivity space: when are multitarget drugs a feasible concept? J. Chem. Inf. Model. 46 (2006) 2445–2456.
[83] A. Givehchi, A. Bender, R.C. Glen, Analysis of activity space by fragment fingerprints, 2D descriptors, and multitarget dependent transformation of 2D descriptors, J. Chem. Inf. Model. 46 (2006) 1078–1083.
[84] S. Renner, S. Derksen, S. Radestock, F. Morchen, Maximum common binding modes (MCBM): consensus docking scoring using multiple ligand information and interaction fingerprints, J. Chem. Inf. Model. 48 (2008) 319–332.
[85] D. Erhan, J.P. L'Heureux, S.Y. Yue, Y. Bengio, Collaborative filtering on a family of biological targets, J. Chem. Inf. Model. 46 (2006) 626–635.
[86] H. Dragos, M. Gilles, V. Alexandre, Predicting the predictability: a unified approach to the applicability domain problem of QSAR models, J. Chem. Inf. Model. 49 (2009) 1762–1776.
[87] X.H. Liu, X.H. Ma, C.Y. Tan, Y.Y. Jiang, M.L. Go, B.C. Low, Y.Z. Chen, Virtual screening of Abl inhibitors from large compound libraries by support vector machines, J. Chem. Inf. Model. 49 (2009) 2101–2110.
[88] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.
[89] K.J. Miller, Additive methods in molecular polarizability, J. Am. Chem. Soc. 112 (23) (1990) 8533–8542.
[90] H.P. Schultz, Topological organic chemistry. 1. Graph theory and topological indices of alkanes, J. Chem. Inf. Comput. Sci. 29 (1989) 227–228.
[91] L.B.K. Lowell, H. Hall, Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information, J. Chem. Inf. Comput. Sci. 35 (6) (1995) 1039–1045.