View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Tokushima University Institutional Repository

PROCEEDING

Virtual screening models for finding novel antidepressants

Zsolt Lepp, Takashi Kinoshita, and Hiroshi Chuman

Department of Molecular Analytical Chemistry, Institute of Health Biosciences, The University of Tokushima Graduate School, Tokushima, Japan

Abstract : Virtual screening was carried out against various biological targets related to depression by support vector machine classification using the atom-type descriptors. The models were effective as over 75 and 95% of the molecules in external test datasets could be correctly classified, depending on target. Antidepressant compounds had predicted activity against 2.3 targets, on average. An introduction is given to virtual screening and the results of classification experiments are presented J. Med. Invest. 52 Suppl. : 297-299, November, 2005

Keywords **:** *database, lead searching, virtual screening, antidepressant, support vector machine, atom-type descriptors, diversity, multiple activities*

INTRODUCTION

The rapid development of computational tools made it possible to effectively predict various properties, such as activity, toxicity, absorption, etc. of biologically relevant compounds. These virtual methods that can aid in finding compounds with desirable properties, without the time and effort of synthesizing and testing them, are becoming increasingly popular. The method when active compounds are identified from a database of molecules by computer is called virtual screening.

The aim of this work was to develop virtual screening method(s) for a number of targets related to depression, and use models to identify natural compounds with possible positive effects.

DISCUSSION AND RESULTS

To perform predictions some information is necessary about the 2 or 3 dimensional structure(s) of target and/or some active molecules.

When there is no 3D structural information available about the potential target proteins, like in case of many depression-related targets, the approach should be to examine known active molecules, and to extract characteristic structural information that is suitable to characterize novel compounds. Two approaches are possible ; one is to find similar structures to the already known active ones using various similarity measures ; the other is to utilize a method that can discriminate between positive and negative examples. In the current study the second approach was applied (1-5, 11-13).

A computational learning tool, called support vector machines (SVM) were used to make differentiation between the sets of active and non-active compounds (Figure 1). Computational learning methods are aimed at embedding human knowledge into workable approaches, and these methods are broadly used in chemoinformatics (6) .

It is necessary to translate the structural information of molecules into a set of numbers, so called descriptors, which can be interpreted by the computational algorithm. There are many different way to do this, in this study atom type descriptors were used that were generated by counting the number of different kinds of atoms in the molecules. The advantages of characterizing a molecule this way are various. These are mainly

Received for publication September 9, 2005 ; accepted September 13, 2005.

Address correspondence and reprint requests to Zsolt Lepp, Ph. D., Department of Molecular Analytical Chemistry, Institute of Health Biosciences, The University of Tokushima Graduate School, 1-78, Shomachi, Tokushima 770-8505, Japan and Fax : +81-88-633-9508.

the accuracy (theoretically hundred percent) , the ease and speed of generating these features. Furthermore the method can be used for any kind of compounds with large diversity (14-17).

By using SVM the sets of positive and negative examples were separated by making a non-linear transformation of the original descriptors into a high dimensional feature space, where an optimal separating hyperplane can be found. Any novel structure can be classified by calculating its position relative to this hypersurface .

SVM models were prepared for a number of targets relevant to antidepressants (7-10) . These models bore very good prediction power with accuracies of prediction over 75% and 95% for positive and negative examples, respectively (Table 1) . The method was effectively used to filter large compounds libraries with high speed. It was also found that antidepressants are usually active against multiple targets (Figure 2).

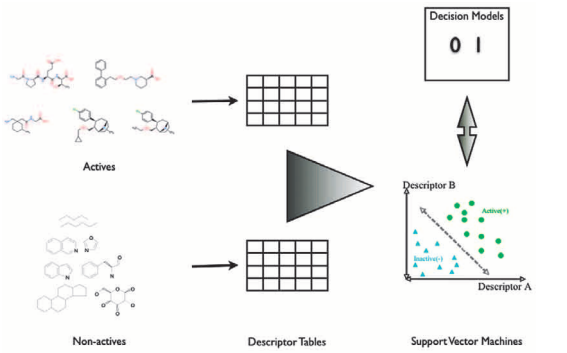

Figure 1. General scheme of computational learning method used.

Table 1. Success rate of prediction for the external test sets of screening models of depression related targets.

ID	Target Name	Recall (%) Pos [a]	Neg [b]	E [c]
	Serotonin Receptor agonists			
M 1	HT 1 A	79.7	95.2	13.2
M 2	HT 1 C	80.3	99.6	144.7
M 3	HT 1 D	88.9	97.9	29.3
	Serotonin receptor antagonists			
M 4	HT 3	84.5	96.9	19.6
M 5	HT 2 A	80.8	96.2	17.3
M 6	HT 2 B	73.2	99.1	67.3
M 7	HT 1 A	67.4	98.9	39.0
M 8	Serotonin reuptake inhibits	70.9	97.5	21.0
	Adenosine agonists			
M 8	A 1	67.7	99.5	83.7
M 9	A 2	71.9	97.9	29.4
	Dopamine Antagonists			
M 10	D 2	70.8	97.5	23.4
M 11	Norepinephrine uptake i.	80.8	99.2	63.3
	Monoamino-oxidase inhibitors			
M 12	CRF antagonists	88.2	99.2	60.0
M 13	Substance P antagonists	82.0	96.6	13.5

[a] positive examples, [b] negative examples, [c] enrichment

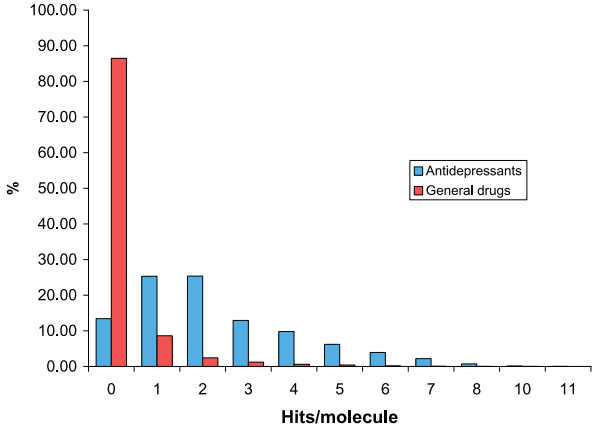

Figure 2. Percents of antidepressants predicted active by the given number of filters.

REFERENCES

1. Adell A, Castro E, Celada P, Bortolozzi, A, Pazos A, Artigas F : Strategies for producing faster acting antidepressants. Drug Discovery Today 10(8) : 578-585, 2005
2. Niwa T : Prediction of Biological Targets Using Probabilistic Neural Networks and Atom-Type Descriptors. J Med Chem 47(10):2645-2650, 2004
3. Vapnik VN : Statistical Learning Theory, John Wiley & Sons, New York, 1998
4. Chen N, Lu W, Li J Y G : Support Vector Machine In Chemistry, World Scientific Pub Co Inc, SINGAPORE, 2004
5. Jorissen RN, Gilson MK : Virtual Screening of Molecular Databases Using a Support Vector Machine. J Chem Inf Model 45 (3) : 549-561, 2005
6. Gasteiger J : Handbook of Chemoinformatics : From Data to Knowledge. Wiley, Welnheim, 2003
7. MDL Drug Data Report 2004 and Screening Compounds Directory Elsevier MDL, Hayward, CA, USA, 2005
8. Chang C-C, Lin C-J, LIBSVM:a library for support vector machines, 2001. Software available at http : //www.csie.ntu.edu.tw/～cjlin/lib
9. JChem 3.0 Chem Axon Ltd,Budapest, Hungary
10. XLSTATPro 7.5, Addinsoft, Brooklyn, NY, USA
11. Kecman V : Learning and Soft Computing, The MIT Press, Cambridge, MA, 2001
12. Kearns MS, Solla SA, Cohn DA : Advances in

Neural Information Processing Systems, 10& 11, MIT Press, Cambridge, MA, 1998
13. Kelley LA, Gardner SP, Sutcliffe MJ:An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies. Protein Eng 9 : 1063-1065, 1996
14. Martin YC, Kofron JL, Traphagen LM:Do Structurally Similar Molecules Have Similar Biological Activity? J Med Chem 45(19): 4350-4358, 2002
15. Thimm M, Goede A, Hougardy S, Preissner R : Comparison of 2D Similarity and 3D Superposition. Application to Searching a Conformational Drug Database J Chem Inf Comput Sci 44 (5) : 1816-1822, 2004
16. Willett P : Searching Techniques for Databases of Two-and Three-Dimensional Chemical Structures. J Med Chem 48 (13) : 4183-4199,2005
17. Klebe G : Virtual Screening : An Alternative or Complement to High Throughput Screening? Kluwer Academic Publishers, Dororecht, 2000