In Silico Functional Profiling of Small Molecules for Assessing Their Physiological Effects

Tomohiro Sato¹,²
tomohiro@gsc.riken.jp

Yo Matsuo²,³
matsuo@gsc.riken.jp

Shigeyuki Yokoyama¹,²,⁴
yokoyama@biochem.s.u-tokyo.ac.jp

¹ Department of Biophysics and Biochemistry, Graduate school of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
² RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Japan
³ Department of Supramolecular Biology, International Graduate School of Arts and Sciences, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Japan
⁴ RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo-gun, Hyogo 679-5148, Japan

Keywords: drug function prediction, functional profile, cluster analysis, support vector machine

1 Introduction

Predicting physiological effects of compounds from their chemical structure is useful for drug discovery. Functional prediction of compounds in drug discovery often focuses on a single important physiological effect like affinity for a target protein. Nevertheless, drugs usually have several physiological effects which are useful, harmless or adverse. To assess effects of compounds comprehensively, it is necessary to predict many functions simultaneously and analyze the pattern of predicted functions. In this study, functional profile of a compound was defined as a vector consisting of the results of predictions about physiological functions. Functional profiles of 871 known drugs were calculated and analyzed by clustering.

2 Method and Results

2.1 Calculation of Functional Profile

Functional profile of a compound was defined as a vector as (x₁, x₂, …, xₙ), where xᵢ represented the prediction about the function of number “i”. In this study, 125 physiological effects recorded in MDL Drug Data Report were subject to prediction. To predict each function, the method we reported in GIW2005 using Support Vector Machine (SVM) [1] was used. Each compound was represented by a 173 dimensional vector consisting of similarity among the compound and 173 reference compounds covering KEGG COMPOUND database. SVM models for the 125 physiological effects were built by leaning the vectors representing known drugs in MDDR. SVM implementation used here was SVMlight [2]. If a compound was predicted to have the function of number “i”, xᵢ of the functional profile was 1, otherwise xᵢ was 0.

2.2 Clustering Functional Profiles of Known Drugs

Functional profiles of 871 known launched drugs in MDDR were calculated. The numbers of functions which each compound was predicted to have ranged from 0 to 27, the average value was 8.06 and the standard deviation was 4.59. The average of the numbers of functions annotated for each of the 871 drugs in MDDR was 1.37. The method predicted about six times as many functions as annotations in MDDR.

Functional profiles of 871 drugs were hierarchically clustered by Ward’s method (Figure 1). Compounds were predicted to have similar functions especially in clusters named “A”, “B”, and “C”. Cluster A included compounds predicted to have similar functions.
interact with cerebral receptors, cluster B included compounds predicted to have antiinflammatory activities and cluster C included compounds predicted to affect circulation (Figure 2).

Three subclusters C1, C2 and C3 constructing C were characterized by Ca²⁺ channel blocker, ACE inhibitor and other agents relating with circulation respectively. In practice, 14 out of the 18 calcium channel blockers in 871 compounds belonged to C1 and all 9 compounds in C3 were annotated as ACE inhibitors in MDDR. A difference between subclusters A1 and A2 constructing cluster A was observed in the prediction of NMDA receptor inhibitor. While no compound in A1 was predicted as NMDA receptor inhibitor, 61.5% of the compounds in A2 were predicted as NMDA receptor inhibitor. Defining two subclusters of A2 as A2,1 and A2,2 (Figure 1) in which 100% and 85.1% of the compounds were predicted as NMDA receptor inhibitor, functional differences between A2,1 and A2,2 were obtained.

NMDA receptor is known as a target protein of antiischemic agents [3]. The ratios of the compounds in A2,1 and A2,2 predicted as antiischemic agents were 36.4% and 80.9% respectively. Other difference was obtained in prediction of gastric antisecretary agents. The ratios were 72.7% for A2,1 and only 0.043% for A2,2. Blockage of gastric histamine H2 receptor was known to cause antisecretary action. Histamine release from nerve endings is enhanced in ischemia and contributes neuroprotection and blockage of cerebral histamine H2 receptors aggravates ischemic injury [4]. These results suggested compounds in A2,1 were not predicted to have antiischemic effects because of blockage of histamine H2 receptors and affinities to histamine receptors should be avoided in order to design antiischemic agents.

3 Discussion

Functional profiles would be useful in order to obtain detailed mechanism of actions. For example, the cluster analysis indicated the pattern of affinities of antiischemic agents to not only NMDA receptor which is a known target protein, but also histamine H2 receptor. In this study, functional profiles were analyzed by cluster analysis. Using other various methods like analysis of gene expression profiles would be beneficial for further data mining from the functional profiles.

References