KP211

Analysis of compounds with developmental toxicity in human and prediction of fetotoxicity using Support Vector Machines

Yusuke Kikuno, Mami Adachi, Kumiko Sakamoto, Masaru Kihara, Aiko Yamauchi*
Graduate School of Pharmaceutical Sciences, The University of Tokushima,
1-78-1 Syoumachi, Tokushima, 770-8505 JAPAN
aiko@ph.tokushima-u.ac.jp

1. Introduction

Information about drug fetotoxicity such as teratogenicity in human is very important in pharmaceutical R&D and drug therapy. Since the reproductive and developmental adverse events (fetotoxicity) cannot be tested in the human body, it is necessary to ensure the safety of drug candidates in preclinical stage. Especially, the computer-aided screening for the fetotoxicity *in silico* is of great importance before the screening assay *in vitro*. Recently we have constructed a web-based drug-teratogenicity information sharing system, anzen-drug and SimScore (http://anzen-drug.ph.tokushima-u.ac.jp/) [1].

In this study, 327 pharmaceutical compounds were analyzed based on the chemical structure and physicochemical properties of molecule, and then the fetotoxicity in human was predicted by Support Vector Machine (SVM), a statistical-learning method. Furthermore, fetotoxicity-positive compounds were grouped by the molecular type with a characteristic common substructure.

2. Method

1) Software

Creation of 39 atom-type descriptors and all SVM calculations were carried out according to the method previously reported by Lepp *et al.* [2]. The value of measured log *P* of each compound was obtained from Bio-Loom (for windows) or Clog *P* was calculated by ClogP program. Molecular weight and polar surface area (PSA) were obtained from SciFinder Scholar 2006. Cluster analysis by K-means or Ward's method and Mann-Whitney U test were performed using JMP software (version 5.1).

2) The US Food and Drug Administration (FDA) Pregnancy Category

The FDA has assigned a fetal risk category (A, B, C, D or X) to each pharmaceuticals [3]. In this study, we defined the drugs belonging to the categories D and X as fetotoxicity-positive in human, and the drugs categorized A and B were defined as fetotoxicity-negative.

3) Data Sets

Pharmaceutical compounds were divided into two groups; 149 fetotoxicity-positives and 179 fetotoxicity-negatives. They were clustered in the descriptor space by K-means clustering method. Each of clusters was roughly divided into training (60%) and test (40%) sets for SVM.

Table 1. Performance of SVM models for prediction of drug fetotoxicity in human

		Accuracy (%)		
		All	FT+	FT-
Test Sets	1	85.8 ± 4.9	82.3 ± 4.0	88.6 ± 8.6
	2	77.7 ± 1.7	75.3 ± 5.9	80.0 ± 4.2
	3	80.0 ± 1.8	73.3 ± 9.3	85.2 ± 4.9
Train Sets	1	97.8 ± 2.9	98.0 ± 1.7	97.6 ± 4.2
	2	89.3 ± 6.5	85.2 ± 9.5	91.1 ± 7.6
	3	88.6 ± 5.6	84.5 ± 6.7	92.2 ± 5.0

Mean ± S.D (n=5 runs)

FT+, fetotoxicity positive; FT-, fetotoxicity negative

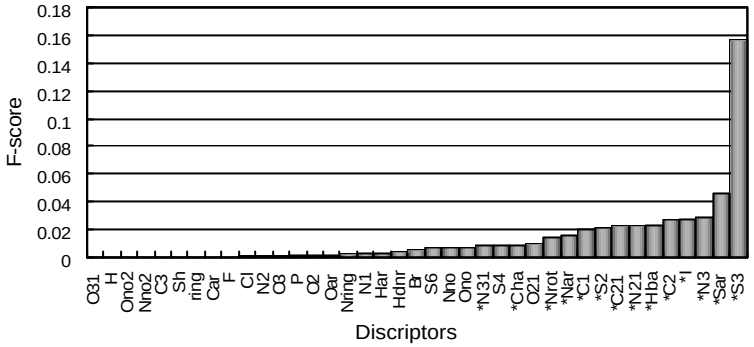

Fig. 1. F-score of every feature
* $p<0.05$, FT+ vs. FT-

Table 2. Significant atom-type descriptors

name	description
C1	sp1 carbon
C2	sp2 carbon
C21	carboxylic type group R-CX1-X2-Y X1=O,N,S X2=O,N,S Y=any
Chan	No, of atoms inside chains
Hbac	No, of H-bond acceptor
I	iodine
N21	sp2 N with adjacent pi orb. (C=NH, R-C=C=N-R,etc.)(X1 in case of C21 type)
N3	sp3 N
N31	sp3 N in ester or conj. Ether(X=C-O-R)(X2 in case of C21 type)
Nar	N in aromatic ring
Nrot	No, of rotatable bonds
S2	sp2 S(C=S)
S3	sp3 S in thioether
Sar	S in aromatic ring

3. Results and Discussion

The SVM model created by 39 atom-type descriptors (Set 1) showed excellent performance (Table 1). The feature selection based on F-score (Fig. 1) was unable to show the significant descriptors. So, we determined the difference of distribution pattern in each descriptor between fetotoxicity-positive and negative groups using the Mann-Whitney U test. Then 14 atom-type descriptors (Table 2), logP and PSA was found to be significant ($p<0.05$). PSA defined as the sum of surface of polar atoms in a molecule, has shown to correlate well passive molecular transport through membranes [4]. We defined those 14 descriptors as Set 2. The set added logP and PSA to Set 2 was defined as Set 3. Table 1 shows the performance of three SVM models. Although the accuracy of SVM in Set 2 and Set 3 were lower than that in Set 1, those 2 models are more efficiency about cost-cutting of calculation than Set 1 model.

The high accuracy SVM model created by atom-type descriptors indicates the strong relationship between fetotoxicity and chemical structure of compound. Thus, we tried to divide fetotoxicity-positive compounds into some groups based on the similarity of chemical structure, using K-means clustering followed by Ward's clustering method. Consequently, 119 out of 149 compounds were categorized into 18 groups, and 18 unique substructures were determined. Some examples were shown in Fig. 2.

While the mechanism of drug developmental toxicity in the human body is complex and unknown, nevertheless this study suggests the possibility of fetotoxicity prediction.

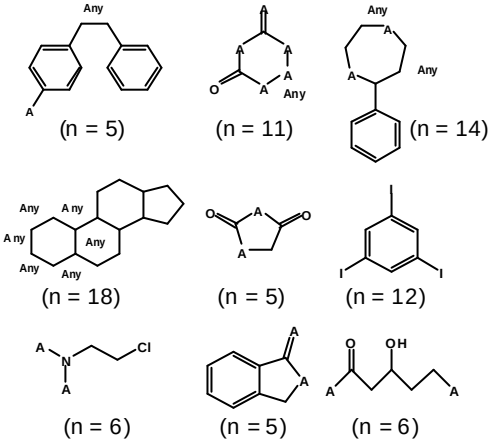

Fig. 2. Examples of the unique structure contained in fetotoxicity positive molecules
A, any atom Any, any hybridization
n, number of compounds

Acknowledgements

We wish to thank Dr. Lepp and Prof. Chuman for providing the atcls_2 and SVM programs.

References

1. Sakamoto K, Yamauchi A, Sasaki M, Chuman H. *J Comput. Chem. Jpn.*, **2007**, *6*, 117-122.
2. Lepp Z, Kinoshita T, Chuman H. *J.Chem.inf.Model.* **2006**, *46*, 158-167.
3. Briggs GG, Freeman RK, Yaffe S, eds., *Drug in pregnancy and lactation,7th ed.*,Williams & Wilkins, Philadelphia (**2005**).
4. Ertl P, Rohde B, Selzer P, *J.Med. Chem.* **2000**, *43*, 3714-3717.