# Pharmacophore Model Generation of $P2Y_{12}$ Inhibitor

Zhen Yang[1], Yanling Zhang [1], Xing Wang [1,2], Yanjiang Qiao [1]*

1. *School of Chinese Material Medica*, Beijing University of Traditional Chinese Medicine Beijing 100102, P.R. China

2. *Henan University of Traditional Chinese Medicine, Zhengzhou 450008, P. R. China*
E-mail: yzwygb@126.com

***Abstract*—A three dimensional pharmacophore model was generated for the molecules which are responsible for anti-platelet aggregation activities targeting platelet adp receptor ($P2Y_{12}$). 24 structrurally diverse molecules were selected as training set to generate the hypothesis using Catalyst software 4.11. The best hypothesis comprises one hydrogen- bond acceptor, one aromatic ring, three hydrophobic points and one excluded volume and shows high correlation coefficient (0.999) as well as low RMS deviation (1.24). It has been further validated towards a test set and shows high correlation coefficient of test set (0.978). The values of effectively active hit A% and comprehensive evaluation index CAI are respectively 40% and 2.795. The results show that the pharmacophore we built is reliable and can be used to screen database. Furthermore, the best hypothesis was used to screen TCMD (Version 2005) database and the four hit compounds of higher predicted activity were the reported anti-platelet aggregation inhibitions, which may be useful for further study.**

***Keywords- Cardiovascular diseases; $P2Y_{12}$; Pharmacophore; Virtual screening***

## I. INTRODUCTION

According to the Chinese cardiovascular disease report 2007 [1], the number of Chinese cardiovascular patients had been up to 230 million. Two of ten adults are suffering from cardiovascular disease. Every year, three million people die of cardiovascular disease. In recent years, the incidence of cardiovascular, more important, the thromboembolic diseases have increased and become an important reason lead to die [2]. Antithrombotic drugs of clinical use are generally classified into three kinds including anticoagulant drugs, anti-platelet drugs and thrombolytic drugs. The anti-platelet drugs are the most widely used, accounting for more than 50% of antithrombotic drugs [3]. $P2Y_{12}$ is the important anti-platelet receptor [4-6].

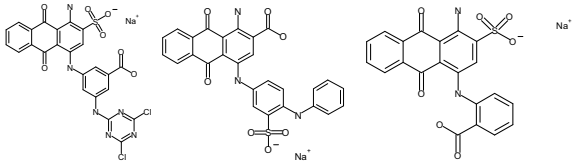
$P2Y_{12}$ is a subtype of ADP receptor which can promote platelet aggregation [7]. ADP is the most important factors related to physiology bleeding and thrombosis, so the important ways of anti-platelet drugs effects are to block the ADP receptor on platelet membrane. $P2Y_{12}$ is not only ADP effect receptor but also the target of inhibitors of ADP receptor. $P2Y_{12}$ is the main receptor in ADP induced platelet aggregation action [8]. Therefore, $P2Y_{12}$ plays a central role in platelet activation and becomes an important target of anti-platelet drugs. $P2Y_{12}$ only exists in platelet membrane. So $P2Y_{12}$ inhibitors can prevent platelet aggregation but not affect other vascular responses that mediated by ADP [9].

In this paper, we have developed a quantitative pharmacophore model whose purpose is to identify the critical pharmacophoric features necessary for potent $P2Y_{12}$ inhibitors as well as to clarify the quantitative structure-activity relationship for the known $P2Y_{12}$ inhibitors. Correlation between actual and estimated biological activities was calculated to optimize the hypothesis [10]. Further, hypothesis was evaluated by two internal databases (test database and active-MDDR database) that we have built. Then the best hypothesis was used to screen TCMD (Version 2005) database, 158 compounds were hit and the four of higher predicted activity were the reported anti-platelet aggregation inhibitions, which may be useful for further study.

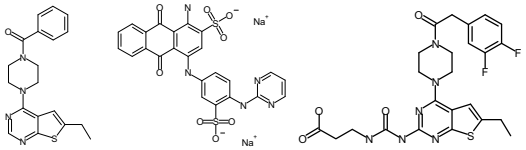
## II. MATERIALS AND METHODS

### A. Biological data

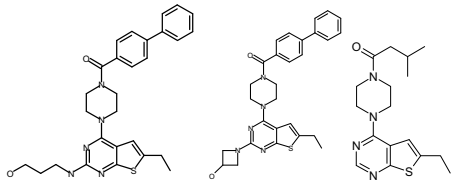
The pharmacophore modeling studies considered a total set of 32 $P2Y_{12}$ inhibitors of diverse structures derived from The Binding Database with Ki values of 0.38-25100nM (http://www.bindingdb.org/bind/index.jsp). This dataset was divided into 24 training set compounds and 8 test set compounds. The structures and biological activities of training set and test set are showed in Fig.1 and Fig.2 respectively.



1 (Ki 3760nM) 2(Ki 21nM) 3 (Ki 25100nM)

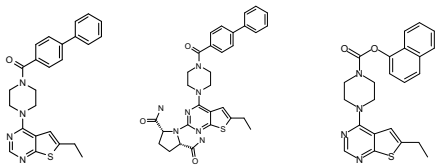


4 (Ki 4400nM) 5 (Ki 50.7nM) 6 (Ki 140nM)

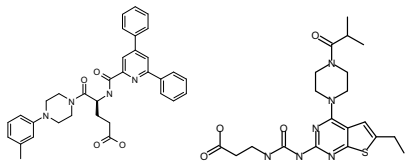


7 (Ki 78nM) 8 (Ki 27nM) 9 (Ki 2100nM)

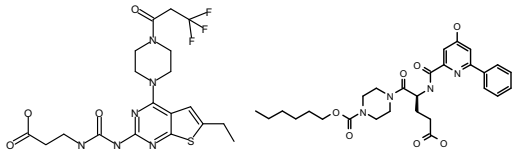
978-0-7695-4706-0/12 $26.00 © 2012 IEEE
DOI 10.1109/iCBEB.2012.293



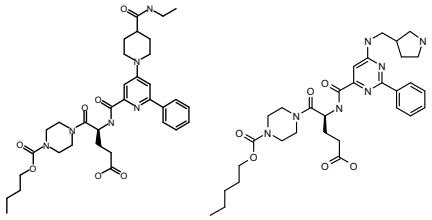
10 (Ki 29nM) 11 (Ki 21nM) 12 (Ki 130nM)



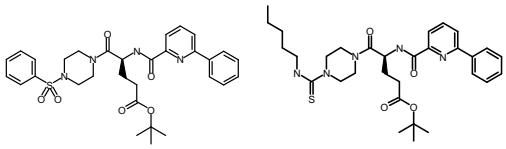
13 (Ki 150nM) 14 (Ki 36nM)



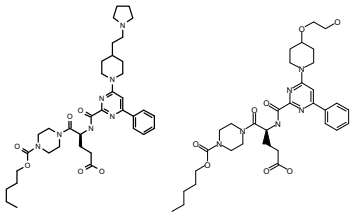
15 (Ki 177nM) 16(Ki 20nM)



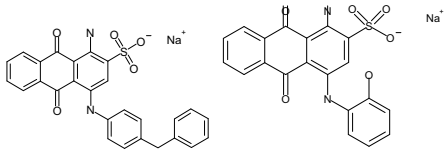
17 (Ki 0.38nM) 18 (Ki 4.1nM)



19 (Ki 842nM) 20(Ki 6nM)

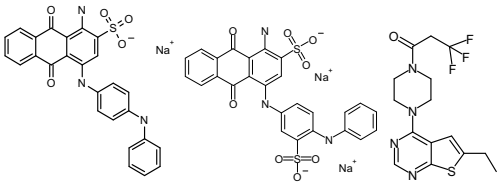


21 (Ki 4.2nM) 22(Ki 10nM)

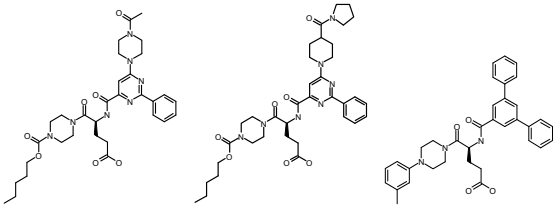


23(Ki 614nM) 24 (Ki 10000nM)

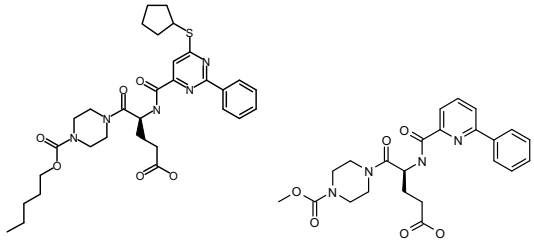
Figure.1 Training set compounds



25 (Ki 1850nM) 26 (Ki 24.9nM) 27 (Ki 2990nM)



28 (Ki 3.7nM) 29 (Ki 0.46nM) 30(Ki 1430nM)



31 (Ki 17nM) 32 (Ki 2440nM )

Figure2. Test set compounds

### B. Modeling tool

All the pharmacophore modeling calculations were carried out by using the HypoGen module implemented in Catalyst 4.11 software package (Accelrys) [11] on IBM workstation.

### C. Selection of training and test set

In the selection of the training set, some basic requirements namely the activity range in the total set (4-5 orders of magnitude) [12-13] and the presence of maximum structural information were considered. From the methods used for the classification of the dataset into training and test sets, the Kennard-Stone (KS) method was chosen so as to ensure the compounds of training set and test set can be evenly distributed in a space distance and can make sure that the training set has a good representative [14].

### D. Conformational analysis

All molecular structures in the training set and test set were built in 2D/3D Visualizer within Catalyst and minimized to the closest local minimum. Diverse conformational models for each compound were generated using an energy range of 20 Kcal/mol by the BEST flexible conformation generation option available in Catalyst. Maximum number of conformers was specified to 250 for each molecule to ensure maximum exploration of the conformational space.

### E. Generation of pharmacophore hypotheses

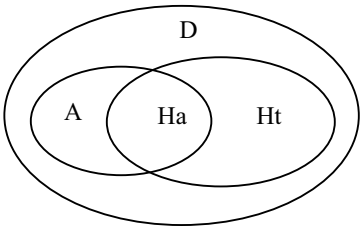
Pharmacophore models were generated from the training set using the HypoGen module in Catalyst which can be used

The Specialized Research Found for the Doctoral Program of Higher Education of China (20092213120006)

to correlate the observed biological activities for a series of compounds with their chemical structures. The quality of the generated pharmacophore hypothesis was evaluated by the cost functions calculated using the Catalyst/HypoGen module during hypothesis generation. The overall cost contains the correlation coefficient (correl), root mean square deviation (RMS), total cost, fixed cost, and null cost. A good pharmacophore model should have a high correlation coefficient, low total cost and RMS values. The total cost should be close to the fixed cost and away from the null cost. The difference between the cost of the generated hypothesis and the cost of the null hypothesis (△Cost) signifies the reliability of a pharmacophore model. A meaningful pharmacophore hypothesis may result when the value of △Cost is more than at least 20 bits. A value of 40-60 bits of △Cost for a pharmacophore hypothesis may indicate that it has 75-90% probability of correlating the data [15].

For evaluation purposes, two internal databases were built within Catalyst using the compounds of test set (test database) and 35 compounds with known $P2Y_{12}$ inhibitory activity in MDDR (MDL Drug Database Report: Version 2007.2) (active-MDDR database).

Besides the cost functions calculated in the Catalyst/HypoGen module, the screening database method was also selected to evaluate the pharmacophore model. Fig.3 is the diagram of the results of the database screening. D is for the total number of compounds in database and A represents the number of active compounds. Ht is for the total number of hit compounds from database D and Ha represents the number of hit compounds from database A. The pharmacophore was used to screen database and evaluated by a series of indices. Indices include effectively active hit A% (A higher value of A% is for a stronger ability to identify active compounds), identify effective index N (A higher value of N is for a stronger ability to distinguish between active compounds and inactive compounds). In order to evaluate N and A% comprehensively, our research group puts forward the comprehensive evaluation index CAI (A higher value of CAI is for a better pharmacophore model)




$A\%=(Ha/A)\times 100$
$N= (Ha/ Ht) \div (A/D)$
$CAI=N\times A\%$

Figure3. Results of the database screening

### F. Database screening

Chemical feature-based 3D pharmacophore models built within the Catalyst software can be used as queries for 3D database screening. Virtual screening of such databases can serve two main purposes: first, validating the quality of the generated pharmacophore models by selective detection of compounds with known $P2Y_{12}$ inhibitory activity, and second, finding novel, potential leads suitable for further development.

In Catalyst, two algorithms for database screening can be chosen: the Fast Flexible Search and the Best Flexible Search. In our study, all screening experiments were performed by using the Best Flexible Search algorithms.

## III. Results and Discussion

### A. Generation of pharmacophore hypotheses

An initial analysis of training set revealed that five chemical feature types such as hydrogen-bond acceptor (A), hydrogen-bond donor (D), hydrophobic (H), negative ionizable (N) and ring aromatic (R) effectively map almost of the compounds in the training set. These features were selected and set excluded volume value of 1 to build a series of pharmacophore models using a default uncertainty value of 3. Other parameters were kept at their default values. In this part, ten top ranked pharmacophore models were generated and were used to screen the test database and calculated the linear correlation coefficient (Correltest) by using the actual activity value and estimated value of the hit compounds. The results are showed in Table 1.

TABLE1. The top ten ranked pharmacophore and the results of database screen

| Hypo | Feature | Max Fit | Total Cost | RMS | Correl | Null Cost | △ Cost | Correltest |
|---|---|---|---|---|---|---|---|---|
| 1 | AHHRV | 10.07 | 116.75 | 1.04 | 0.91 | 152.72 | 36.0 | 0.05 |
| 2 | DHHRV | 7.81 | 116.96 | 1.11 | 0.89 | 152.72 | 35.8 | 0.04 |
| 3 | AAHH | 8.70 | 119.13 | 1.18 | 0.88 | 152.72 | 33.6 | 0.49 |
| 4 | AHHHV | 8.73 | 119.33 | 1.19 | 0.87 | 152.72 | 33.4 | 0.69 |
| 5 | AHHHRV | 11.04 | 119.55 | 1.19 | 0.87 | 152.72 | 33.2 | 0.11 |
| 6 | HHHHRV | 10.51 | 119.95 | 1.21 | 0.87 | 152.72 | 32.8 | only hit 1 |
| 7 | AAHH | 9.76 | 120.05 | 1.18 | 0.88 | 152.72 | 32.7 | 0.52 |
| 8 | AHHHRV | 9.76 | 120.72 | 1.24 | 0.86 | 152.72 | 32.0 | 0.69 |
| 9 | ADDH | 9.18 | 120.82 | 1.23 | 0.87 | 152.72 | 31.9 | no hit |
| 10 | ADHR | 10.23 | 121.21 | 1.20 | 0.87 | 152.72 | 31.5 | 0.10 |

Table1 showed that Hypo4 and Hypo8 pharmacophores have higher values of correlation coefficient of training set and test set but were not perfect, so they were selected and optimized by changing the allowed space tolerance (Tolerance, T) (The value of T is represent for the radius of sphere of the feature and the size of sphere is for the accuracy of the position.) between each feature to be the 0.9 times, 0.8times and 0.7 times of the default values. Then we use the produced pharmacophore models to screen the test database, active-MDDR database and MDDR (MDL Drug Data Report: Version 2007.2) and calculate values of Correltest, A%, N and CAI (D=177953, A=35). The results are showed in Table 2. (4t0.9-4t0.7 are the pharmacophore models that change the values of T to be 0.9 times, 0.8 times and 0.7 times of the default values of Hypo4 pharmacophore and 8t0.9-8t0.7 are the pharmacophore models that change the values of T to be 0.9 times, 0.8 times and 0.7 times of the default values of Hypo8 pharmacophore)

TABLE2. Optimization of Hypo4 and Hypo8

| Hypo | Feature | Ha | Ht | A% | N | CAI | Correltest | Correltrain |
|---|---|---|---|---|---|---|---|---|
| 4t0.9 | AHHHV | 22 | 42046 | 63% | 2.660 | 1.672 | 0.004 | 0.694 |
| 4t0.8 | AHHHV | 21 | 33715 | 60% | 3.167 | 1.900 | 0.002 | 0.266 |
| 4t0.7 | AHHHV | 16 | 25735 | 46% | 3.161 | 1.445 | 0.955 | 0.999 |
| 8t0.9 | AHHHRV | 17 | 32837 | 49% | 2.632 | 1.279 | 0.003 | 0.994 |
| 8t0.8 | AHHHRV | 17 | 19515 | 49% | 4.429 | 2.151 | 0.95 | 0.996 |
| 8t0.7 | AHHHRV | 14 | 10188 | 40% | 6.987 | 2.795 | 0.978 | 0.999 |

Through the analysis of the results, we select 8t0.7 to be the optimal pharmacophore model and the model is showed in Fig.4.

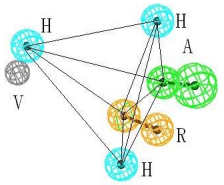



Figure4. The optimal pharmacophore model (8t0.7)

## B. *Database screening*

The best model, 8t0.7, was used as a 3D query to screen TCMD (Version 2005). TCMD screening yielded a hit list of 158 compounds and the four of higher predicted activity (Acutifolin palmitata, Dauricine, Fangchinoline and Nobiletin) were the reported anti-platelet aggregation inhibitions. The best model mapped to the four hit compounds are showed in Fig5, Fig6, Fig7 and Fig8 respectively.

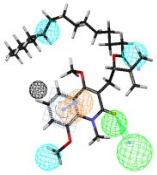


Figure5. 8t0.7 mapped with Acutifolin palmitata

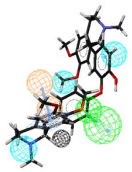


Figure6. 8t0.7 mapped with Dauricin

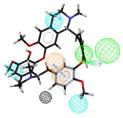


Figure7. 8t0.7 mapped with Fangchinoline

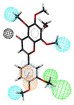


Figure8. 8t0.7 mapped with Nobiletin

# IV. CONCLUSION

In this study, we have built pharmacophore models of $P2Y_{12}$ inhibitors using Catalyst software. The optimal pharmacophore model contains six features: one hydrogen-bond acceptor (A), one ring aromatic (R), three hydrophobic points (H) and one excluded volume (V). The values of linear correlation of actual activity and estimated activity of compounds in training set and the hit compounds in test set are respectively 0.999 and 0.978. And the values of effectively active hit A% and comprehensive evaluation index CAI are respectively 40% and 2.795. The results of database screen also showed the pharmacophore model is reliable and can be used for database screening and lead compounds found, etc.

The authors wish to express sincere appreciation to the “Innovation Team of TCM Information Engineering (2011-CXTD-11)”, “the Fund from The Specialized Research Found for the Doctoral Program of Higher Education of China (20092213120006)” in Beijing University of Chinese Medicine.

# REFERENCES

[1] Chinese cardiovascular disease report 2007[R]. Beijing: Encyclopedia of China Publishing House, 2009.7.

[2] Wang E H,Liu Y. Clinical study overview of platelet P2Y12 receptor antagonist [J]. Medical Innovation of China, 2010,7(29):194.

[3] Ben C, ChristineH. Antithrombotic drugmarket [J].Na2 true Reviews Drug Discovery, 2003,2,11.

[4] Pan X, Li L L, Liu H F, etc. Progress in the development of ADP receptor agonist [J]. Chemical Industry Times, 2009, 23 (11): 51.

[5] Han C B, Jia N, Wang H L. The Clinical Effects of New P2Y12 Inhibitors In Coronary Artery Disease [J]. Medical Recapitulate, Jun, 2011, 17 (11):1705.

[6] Huang Z H. New platelet P2Y12 receptor antagonists [J]. Chin J New Drugs Clin Rem, 2011, 30 (8):579-583.

[7] Bom GV. Aggregation of blood platelets by adenosine diphosphate and its reversal [J].Nature, 1962,194,927-929.

[8] Gachet C,Léon C,Hechler B. The platelet P2 receptors in arterial thrombosis[J].Blood Cells Mol Dis,2006,(36):223-226.

[9] Chen, H. New progess of anti-platelet drugs [J]. Progess of angiocardiology, 2009, 30(1):10.

[10] Hyong-Ha Kim, Yongseong Kim, and Keun Woo Lee. Pharmacophore Design for Anti-inflammatory Agent Targeting Interleukin-2 Inducible Tyrosine Kinase [J]. Bull Korean Chem Soc, 2010, (31), 11: 3333.

[11] Yan H,Jiang F C. Acta Phys [J].Chim. Sin,2006,22(3):359.

[12] Bao H J, Zhang Y L, Qiao Y J. Pharmacophore Model Generation of HMG-CoA Reductase Inhibitors [J]. Acta Phys Chim Sin, 2008,24(2):302.

[13] F. Liu, Q.D. You, and Y.D. Chen, Pharmacophore indentification of KSP inhibitors, Bioorg. Med. Chem[J]. Lett. 17(2007), pp. 722-726.

[14] Galvao, R.; Araujo, M.; José,G. et al. A method for calibration and validation subset partitioning [J].Talanta,2005,67(4):736-740.

[15] P. Aparoy, K Kumar Reddy and Suresh K., etc. Pharmacophore modeling and virtual screening for designing potential 5-Lipoxygenase inhibitors[J]. Bioorganic& Medicinal Chemistry Letters,2010:1013-1018.