Hyper-polyhedron model applied to molecular screening of guanidines as Na/H exchange inhibitors[1]

BAO Xin-Hua[2], LU Wen-Cong, LIU Liang, CHEN Nian-Yi

Department of Chemistry, School of Sciences, Shanghai University, Shanghai 200436, China

ABSTRACT

AIM: To investigate structure-activity relationships of N-(3-Oxo-3,4-dihydro-2H-benzo[1,4]oxazine-6-carbonyl) guanidines in Na/H exchange inhibitory activities and probe into a new method of the computer-aided molecular screening. **METHODS:** The hyper-polyhedron model (HPM) was proposed in our lab. **RESULTS:** The samples with probably higher activities could be determined in such a way that their representing points should be in the hyper-polyhedron region where all known samples with high activities were distributed. And the predictive ability of different methods available was tested by the cross-validation experiment. **CONCLUTION:** The accurate rate of molecular screening of N-(3-Oxo-3,4-dihydro-2H-benzo[1,4]oxazine-6-carbonyl) guanidines by HPM was much higher than that obtained by PCA (principal component analysis) and Fisher methods for the data set available here. Therefore, HPM could be used as a powerful tool for screening new compounds with probably higher activities.

INTRODUCTION

The Na/H exchanger is a major Na$^+$ entry pathway in many types of cells and plays an important role in regulation of cell volume and ion concentration. It is rapidly activated at post-ischemic reperfusion and causes a Ca^{2+} overload, which is known to be associated with cellular dysfunction, damage, and necrosis. Therefore, Na/H exchange inhibitor is a potentially useful candidate for improvement of ischemia-reperfusion-induced injury. N-(3-Oxo-3,4-dihydro-2H-benzo[1,4] oxazine-6-carbonyl) guanidines are known to be Na/H exchange inhibitors[1].

With the increasing demand for research of structure-activity relationship (SAR), the number of multi-variate methods introduced into the literature for SAR has increased[2-7].

In recent years, we have applied some chemometric methods in drug optimal designing, material optimal design and industrial optimization[8-10]. It is also realized that a real-world data set sometimes exhibits so much noise that quantitative models such as linear or nonlinear regression, back-propagation artificial neural network (BP ANN) methods can not be used to represent them appropriately in some cases. When realistic quantitative models are not available, semi-quantitative, and qualitative pattern recognition methods are useful in classifying different kinds of samples of data set. The pattern recognition methods can be divided into 2 classes[6-7,11-12]: 1) linear methods, eg, principal component analysis (PCA), Fisher vector. 2) non-linear methods, eg, non-linear mapping (NLM), Kohonen's

self-organizing map (KSOM).

However, a real-world data set sometimes is so complicated that even qualitative methods available can not give us satisfactory solutions. The data set used in this paper is one of such examples. So it is a meaningful task to obtain a realistic mathematical model for computer prediction of new compounds with high biological activities on the basis of such data set.

In this paper, the hyper-polyhedron model (HPM) was proposed for computer-aided molecular screening of *N*-(3-Oxo-3,4-dihydro-2*H*-benzo[1,4]oxazine-6-carbonyl) guanidines as Na/H exchange inhibitors. It seemed to be a more powerful tool to classify different kinds of samples than traditional pattern recognition methods. A useful software, "Hyper Miner", has been built based on a series of new computation methods including HPM.

MATERIALS AND METHODS

HPM was a new pattern recognition method powerful for classification of samples. In this method, no projection map was made. In computation, a series of hyper-planes could be obtained to create a hyper-polyhedron that encloses all samples with known high biological activities. The characteristic parameters relating to biological activities of compounds were used to span the multi-dimensional space. The representative points of compounds were plotted into the space. If the representative points of compounds with high biological activities (class "1" samples in Fig 1) were located in a definite zone (optimal zone) in the space, HPM could be used to describe the optimal zone. HPM could be expressed by a series of inequalities describing the boundaries of two kinds of compounds. Therefore, the inequalities could be used as the criterion of compounds with high activities. Since data separation took place in the multi-dimensional space directly, the results of separation were usually very good, provided the original classification of deferent classes in the multi-dimensional space was well defined. The conceptual hyper-polyhedron in three-dimensional space was demonstrated in Fig 1.

RESULTS

Computation of characteristic parameters of structures The theoretical parameters from computation of molecular mechanics and quantum chemistry were used to describe the molecular structures of *N*-
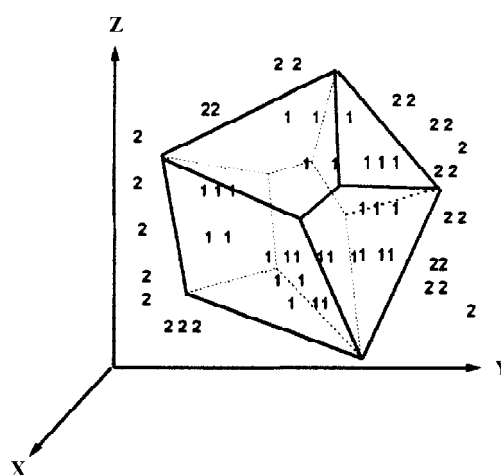

Fig 1. Conceptual hyper-polyhedron in three-dimensional space. 1: samples with high biological activities; 2: samples with low biological activities.

(3-Oxo-3,4-dihydro-2*H*-benzo[1,4]oxazine-6-carbonyl) guanidines. The molecular mechanics program (MM+) with all atom force fields was used to optimize the configurations of compounds, and the quantum chemical program (PM3) to calculate some theoretical parameters of the compounds.

Feature selection The reduced subset of parameters was also obtained by HPM. The data separability criteria had been established to select key parameters that determined the biological activities of the compounds. The rate of separability R was defined as $R=(1-N_2/N_1)$. Here N_1 was the number of the samples with high activities, and N_2 was the number of the samples with low activities remaining within the hyper-polyhedron. If R >90 %, the separability was "excellent." It was practical and reliable to build an HPM on the basis of excellent separability for all known samples with a reduced subset of parameters. After the separability analysis of sample set, 4 quantum chemical parameters were taken as the features of pattern recognition. They were HOMO (energy of highest occupied molecular orbital), LUMO (energy of lowest unoccupied molecular orbital), D (the longest diameter of the substituent R_3), Q (net charge of the oxygen atom of oxazine). It should be mentioned that the determining factors of activities of the compounds might be not merely the above four parameters, but they were enough to be used to classify the training samples. The "excellent" separability result was available by HPM for such data set with above 4 parameters. The activity (IC_{50}) was evaluated by its ability to inhibit platelet swelling

induced by sodium propionate. The training data set consisted of 18 compounds listed in Tab 1[1].

Computation of hyper-polyhedron model In the four-dimensional pattern space spanned by HOMO, LUMO, D, and Q of the samples, the two different kinds of the samples were classified according to their activities. Samples with $IC_{50} < 0.44$ mmol/L are grouped as "class 1" while the others as "class 2." The following series of inequalities used as the criterion of "class 1" samples were obtained by the hyper-polyhedron model.

$-34.7 < 5.16[HOMO] - 5.94[LUMO] + 1.00[D] - 25.6[Q] < -33.6$

$-3.31 < 0.222[HOMO] + 3.17[LUMO] - 0.052[D] - 6.16[Q] < -2.41$

$-141.2 < 17.9[HOMO] - 13.4[LUMO] - 0.080[D] - 53.9[Q] < -140.3$

Results of cross-validation In order to test the predictive ability of different methods available here, the leaving-one method was used to compare the rate of correctness predicted by HPM, PCA and Fisher methods.

The projection maps of PCA and Fisher methods by using the same features as the HPM were given in Fig 2, 3. The results of cross-validation (leaving-one method) experiment were summarized in Tab 2. The
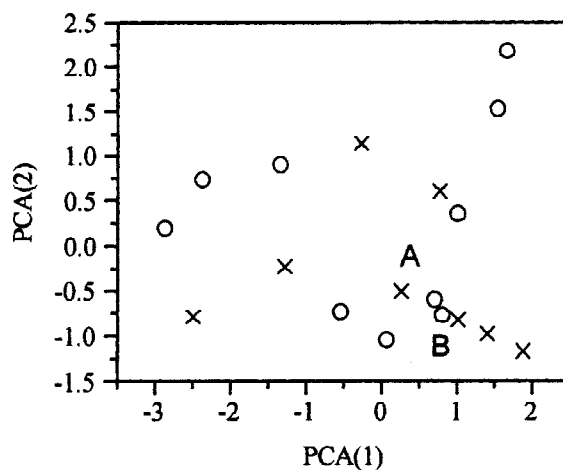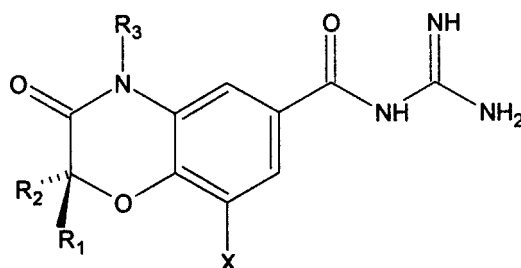

Fig 2. Pattern recognition by PCA method. ○ $IC_{50} < 0.44$ mmol/L; × $IC_{50} > 0.44$ mmol/L; A, B: Predicted samples.

Tab 1. Features and Na/H exchange inhibitory activities of 18 compounds.

| No	R_1	R_2	R_3	X	HOMO/eV	LUMO/eV	D/Å	Q	$IC_{50}/\mu mol \cdot L^{-1}$
1	H	H	H	H	-8.93	-0.628	0.99	-0.200	10.0
2	H	H	Et	H	-8.84	-0.575	3.42	-0.197	0.33
3	H	H	Iso-Pr	H	-8.82	-0.516	3.44	-0.198	0.25
4	H	Et	Iso-Pr	H	-8.78	-0.478	3.44	-0.189	0.17
5	H	Ph	Iso-Pr	H	-8.79	-0.500	3.44	-0.188	1.00
6	Me	Me	Iso-Pr	H	-8.78	-0.467	3.44	-0.196	0.12
7	Me	Me	H	H	-8.87	-0.558	0.99	-0.196	7.40
8	Me	Me	Me	H	-8.77	-0.508	2.12	-0.196	0.81
9	Me	Me	Et	H	-8.77	-0.498	3.42	-0.195	0.38
10	Me	Me	Pr	H	-8.77	-0.499	4.65	-0.195	0.74
11	Me	Me	Butyl	H	-8.77	-0.498	5.96	-0.195	1.30
12	Me	Me	Hexyl	H	-8.78	-0.487	7.63	-0.194	3.70
13	H	H	Me	Cl	-8.95	-0.728	2.12	-0.194	0.33
14	H	H	Et	Cl	-8.96	-0.726	3.42	-0.190	0.22
15	H	H	Iso-Pr	Cl	-8.90	-0.659	3.44	-0.188	0.16
16	H	H	Me	OMe	-8.80	-0.591	2.12	-0.187	0.50
17	H	H	Et	OMe	-8.72	-0.511	3.45	-0.179	0.27
18	Me	Me	Iso-Pr	OMe	-8.75	-0.473	3.44	-0.182	0.37
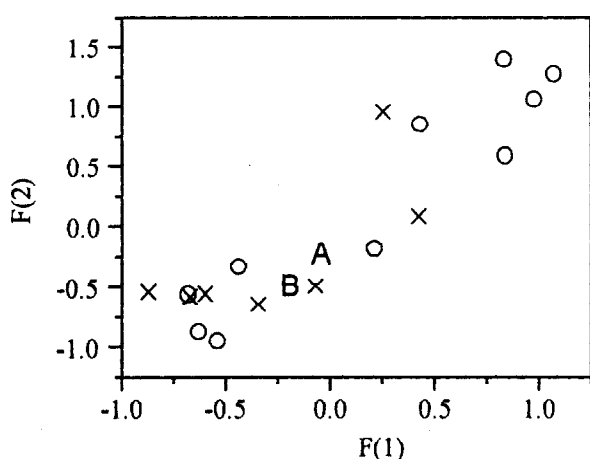
Fig 3. Pattern recognition by Fisher method. ○ $IC_{50}<0.44$ mmol/L; × $IC_{50}>0.44$ mmol/L; A, B: Predicted samples.

Tab 2. Results of cross-validation experiment by the leaving-one method.

Methodology	PCA	Fisher	HPM
Accuracy/%	70	65	85

predictive ability of HPM outperformed that of PCA and Fisher for the data.

Application of HPM According to the series of inequalities obtained by HPM, we could predict the activity of any new compound of *N*-(3-Oxo-3,4-dihydro-2*H*-benzo[1,4]oxazine-6-carbonyl) guanidines. Here were two examples to demonstrate the application of HPM. The new sample labeled A (Fig 2, 3) was found in the hyper-polyhedron expressed by above inequalities describing the boundaries of two different kinds of samples, while the other one labeled B (Fig 2, 3) was found out of the hyper-polyhedron. So we could predict that sample A would be a compound with high activity and sample B with low activity. These results were in agreement with the experimental facts listed in

Tab 3.

DISCUSSION

Generally speaking, molecular structures can be described by various parameters, including electronic parameters, geometric parameters and hydrophobic parameters. Some parameters come from theoretical methods by using computational results of molecular mechanics and quantum chemistry, others are obtained from experimental results. Because of the large number of parameters, it is very difficult to select and use the available parameters to describe the molecular structure. The main difficulty is how to choose the right set, or a reduced subset, of parameters that correlate well the biological activity with the structure of drug molecule. The benefit of feature selection by using HPM is that the reduced subset of parameters represents directly the activity response of the compounds. So the satisfactory results of modeling can be obtained on the basis of the feature reduction.

In order to investigate the SAR of the compounds, a multi-dimensional space with 4 effective descriptors (HOMO, LUMO, D, and Q) was spanned to represent the different kinds of samples. Because HOMO and LUMO were the important factors affecting the activity of a compound, they were represented in the series of inequalities of HPM. So far as the geometric factors, 3 effective diameters of the substituents R_1, R_2, and R_3 were calculated, respectively before application of HPM to the sample set. However, the result of HPM indicated that R_3 represented by descriptor D was related inhibitory activity, while the substituents R_1, R_2 at 2-position of 2*H*-benzo[1,4]oxazine ring were not significant. At last, the inhibitory activity was affected by descriptor Q, which was the integrated embodiment induced, resonance, and hydrogen bond effect as well.

The cross-validation experiment comparing the performance of HPM with 2 other pattern recognition methods, PCA and Fisher methods. PCA and Fisher

Tab 3. Predicted results by hyper-polyhedron model.

No	R_1	R_2	R_3	X	HOMO/eV	LUMO/eV	D/Å	Q	IC_{50} /μmol·L^{-1}	Predicted Class
A	H	Me	Et	H	-8.80	-0.524	3.42	-0.192	0.29	1
B	Me	Me	$(CH_2)_2OEt$	H	-8.82	-0.536	6.44	-0.195	5.60	2

methods are 2 traditional pattern recognition methods widely used in structure-activity relationship (SAR) of drugs. Although they are very useful tools and successful in many cases of SAR research, PCA and Fisher methods could not build a satisfactory model for the data set used here (Fig 2, 3). As Tab 2 shown, the predictive ability of HPM outperformed that of PCA and Fisher for the data set used here. Of course, it should be noted that no single method or paradigm is uniformly superior. It is possible to use HPM in the area of SAR research.

In conclusion, HPM could be used to screen new compounds with probably higher activities. The new compounds with probably higher activities were determined in such way that their representing points should be in the hyper-polyhedron region formed by HPM, where all known samples with high activities were distributed. Also, HPM is a useful tool for feature selection. The computer prediction of *N*-(3-Oxo-3,4-dihydro-2*H*-benzo[1,4]oxazine-6-carbonyl) guanidines with high activities is feasible by using HPM while PCA and Fisher models could not provide us satisfactory results for the data set used here.

REFERENCES

1   Yamamoto T, Hori M, Watanabe I, Tsutsui H, Harada K, Ikeda S, *et al*. Synthesis and quantitative structure-activity relationships of *N*-(3-Oxo-3,4-dihydro-2*H*-benzo[1,4]oxazine-6-carbonyl) guanidines as Na/H exchange inhibitors. Chem Pharm Bull 1998; 46: 1716-23.

2   Huang XQ, Jiang HL, Luo XM, Rong SB, Gu JD, Tan XJ, *et al*. Molecular modeling on solvent effect and interaction mechanism of fentanyl analogs to μ-opioid receptor. Acta Pharmacol Sin 2000; 21: 46-54.

3   Song HF, Tang ZM, Yuan SJ, Zhu BZ. Application of secondary structure prediction in antisense drug design targeting protein kinase C-α mRNA and QSAR analysis. Acta Pharmacol Sin 2000; 21: 80-6.

4   Yu HX, Lin ZF, Feng JF, Xu TL, Wang LS. Development of quantitative structure-activity relationships in toxicity prediction of complex mixtures. Acta Pharmacol Sin 2001; 22: 45-9.

5   Luo XM, Shen JK, Chen KX, Ji RY, Cao Y, Xue H, *et al*. A 3D-structural model of memapsin 2 protease generated from theoretical study. Acta Pharmacol Sin 2001; 22: 50-6.

6   Domine D, Devillers J, Chastrette M, Karcher W. Non-linear mapping for structure-activity and structure-property modeling. J Chemom 1993; 7: 227-42.

7   Fukunaga K. Introduction to statistical pattern recognition. Academic. New York; 1972.

8   Chen NY, Lu WC, Chen RL, Li CH, Qin P. Chemometric methods applied to industrial optimization and materials optimal design. Chemometr Intell Lab Sys Lab Inform Manage 1999; 45: 329-33.

9   Liu HL, Chen NY, Lu WC, Zhu XW. Multi-target classification pattern recognition applied to computer-aided materials design. Anal Lett 1994; 27: 2195-203.

10 Lu WC, Yan LC, Liu HL, Huang SP, Chen NY, Li JL, *et al*. Studies on structure-activity relationship of ethofenprox analogous of pesticide by pattern recognition method. Chin Sci Bull 1993; 38: 1534-7.

11 Lu WC, Yan LC, Zhu YC, Chen KX, Chen NY. Studies on structure-activity relationship of 3-methylfentanyl derivatives by EHMO and pattern recognition method. J Mol Sci (in China) 1993; 9: 123-8.

12 Lu WC, Yan LC, Zhu YC, Chen KX, Chen NY. T. Kohonen self-organization ANNs applied to structure-activity relationship of fentanyl derivatives. J Shanghai Univ 1996; 2: 301-7.