行政院國家科學委員會專題研究計畫　期中進度報告

QSAR 與虛擬藥物篩選之研究(1/3)

計畫類別： 個別型計畫
計畫編號： NSC94-2113-M-009-014-
執行期間： 94 年 08 月 01 日至 95 年 07 月 31 日
執行單位： 國立交通大學生物科技學系(所)

計畫主持人： 楊進木

報告類型： 精簡報告
報告附件： 出席國際會議研究心得報告及發表論文
處理方式： 本計畫可公開查詢

中　華　民　國 95 年 5 月 9 日

中文摘要

在第一年的計畫中我們發展了一個結合分子鉗合與自動特徵篩選PLS的QSAR分析方法。在我們的方法中，首先利用分子鉗和預測配體與目標分子結合構型並產生其鉗和能量，稱之為 protein-ligand interaction profile，並淬取出 atom-based，group-based 與 residue-based 能量值。其主要的概念來自 COMBINE 之 QSAR 分析方法。GEMDOCK 預測分子鉗和構型，並產生蛋白質原子對配體原子的能量值，包含氫鍵，靜電力與凡得瓦力交互作用。這些蛋白質原子對配體原子的能量值稱之為 atom-based 特徵項。對每一個胺基酸的組成原子都能計算出一個原子作用能量的總和，其稱之為 residue-based 特徵項。而 group-based 特徵項則是分離自每個胺基酸的骨幹與側鏈的貢獻能量。這三種特徵淬取方式都會由 GEMPLS 建立不同的 QSAR 模型。GEMPLS 包含了特徵篩選與 QSAR 模型的建立。由演化式方法篩選可能對抑制性有貢獻的特徵，再由 PLS 建立這些特徵值與抑制性之間的回歸關係，最後建立 QSAR 分析的預測模型。

我們也同時將這套 QSAR 分析方法應用到三個不同的目標分子之 QSAR 分析上，包含神經胺酸酶，糖基磷酸酶 b 與環氧化酵素-2。我們的方法所得到的初步結果與其他知名研究者已發表的模型相較，其預測能力的表現較佳。在未來的研究中，我們將會投入更多努力提升方法的準確性與預測力，並結合虛擬藥物篩選建立一高速虛擬藥物預測平台之雛型。

英文摘要（abstract）

In first year, we developed a QSAR methodology associating molecular docking and feature selection with PLS. The feature of our model generates from the interaction energies of docked results, named as protein-ligand interaction profile and was extracted as atom-based, group-based and residue-based energetic terms. The concept of our QSAR analysis came from COMBINE analysis. First, GEMDOCK predicted the binding conformation for protein-ligand and generated atom paired energies. The atom paired energetic terms included hydrogen bonding, electrostatic and van der Waals interactions. These atom paired energetic terms would serve as atom-based features. The sum of atom paired energies of each residue was residue-based term and each residue could divide into two parts of main chain and side chain which called group-based terms. GEMPLS served as feature selection and model building in QSAR analysis. Potential features for contributing inhibition would be selected by evolutionary strategy and built regression by PLS.

We applied our QSAR methodology to build inhibitory QSAR models of neuraminidase, glycogen phosphorylase b and cyclooxygenase-2. We compared our preliminary results to published references and our performances showed more prediction power than published models. In the recently future, we will make more efforts to improving our methodology and

combine virtual screen to create a high through-put prediction environment.

關鍵詞（keywords）

QSAR, Virtual screening, GEMQSAR, COMBINE, GEMDOCK, GEMPLS, Neuraminidase, Glycogen Phosphorylase b, GPB, Cyclooxygenase-2, COX-2

前言與研究目的:

QSAR techniques are commonly regarded as a key role to computational molecular design. The major goal of QSAR is to formulate mathematical relationships between physicochemical properties of compounds and their experimentally determined *in vitro* biological activities. Thus the derived QSAR model can be subsequently used to predict the biological activities of new derivatives. A good QSAR model both enhances our understanding of the specifics of drug action and provides a theoretical foundation for lead optimization[1].

The QSAR methodology, COMparative BINding Energy (COMBINE)[2; 3], develops a linear relationship between binding free energy of the ligand-receptor complexes. In general, residue is the basic unit utilized in the feature of binding energy. But there are just a few differences between the side chains of compounds of QSAR model. For instance, the activity was changed by just one atom altered in the side chain. However, it could be difficult to detect by residue-based method. Using residue as the unit is insensitive and easy to generate outliers by comparison. In order to make QSAR model more sensitivity, we use group and atom as the new feature.

The partial least square (PLS) analysis[4] is able to deal with strongly collinear input data and make no restriction on the number of variables used. Unfortunately, the predictive performance of PLS model drops and the PLS model becomes complicated when the number of features increases. Several feature selection methods for PLS have been proposed, in which genetic algorithm (GA) combined with PLS approach (GAPLS) has demonstrated the improvement on the prediction and interpretation of model[5]. GEMPLS is general able to evolve the relationship between biological activities and compound features generated by COMBINE.

In the COMBINE method[2; 3], the binding energy of the receptor-ligand complex is correlated to the interaction energy components. In this study, “ELE” is electrostatic interactions; “VDW” is van der Waals interactions; “HYD” is hydrogen bond interaction. Each selected energy component u_i contributes to the binding free energy according to its weight *wi* and PLS analysis is applied to obtain the weights w_i:

$$\text{pIC50} = \sum_i w_i^{ELE} u_i^{ELE} + \sum_i w_i^{HYD} u_i^{HYD} + \sum_i w_i^{VDW} u_i^{VDW} + C \quad (1)$$

The purpose of GEMQSAR is to develop a novel *in silico* drug screening system combining 3D QSAR and virtual screening process. To archive the objective, we first developed the methodology of QSAR (GEMPLS[6]) and applied GEMPLS to evolve the QSAR models. This model was generated by the COMBINE method[2; 3] according to the three different units of interaction energy features. The 3D QSAR methodology were applied on three public data sets including 38 influenza neuraminidase inhibitor complexes[7], 76 glycogen phosphorylase b inhibitors[8] and 31 cyclooxygenase-2 inhibitors[9]. Experiments showed that the reduced units were able to improve the predictability and efficiency, and at the same time, the selected features in the yielded QSAR model were consistent with some experimental evidences.

研究方法

The final goal of GEMQSAR is to develop a novel *in silico* drug screening system. The relationships of QSAR and docking are shown in Figure 1. The molecular docking technique was well established in our past researches[10]. In the first year, we developed the core 3D QSAR methodology of GEMQSAR, named GEMPLS[11]. Figure 2 shows the main steps of applying GEMPLS[11] in the COMBINE analysis[2; 3]: 1) prepare the inhibitor set and model protein-inhibitor complexes; 2) refine protein-inhibitor complexes and calculate features (i.e., energy interactions); 3) select important features by Mahalanobis distance; 4) select features and evolve QSAR models. 5) Performance evaluation. Each step is described in the following subsections.

The COMBINE analysis is the use of structural information about ligand-receptor complexes[2; 3]. When the three-dimensional structure of macromolecule is available, ligand-receptor interaction energies could be calculated as features, which are subjected to statistical analysis in COMBINE. A subset of these features will be account for the ligand affinity. The critical interaction patterns between ligands and the receptor could be identified and be used to derive the correlation of binding affinities.

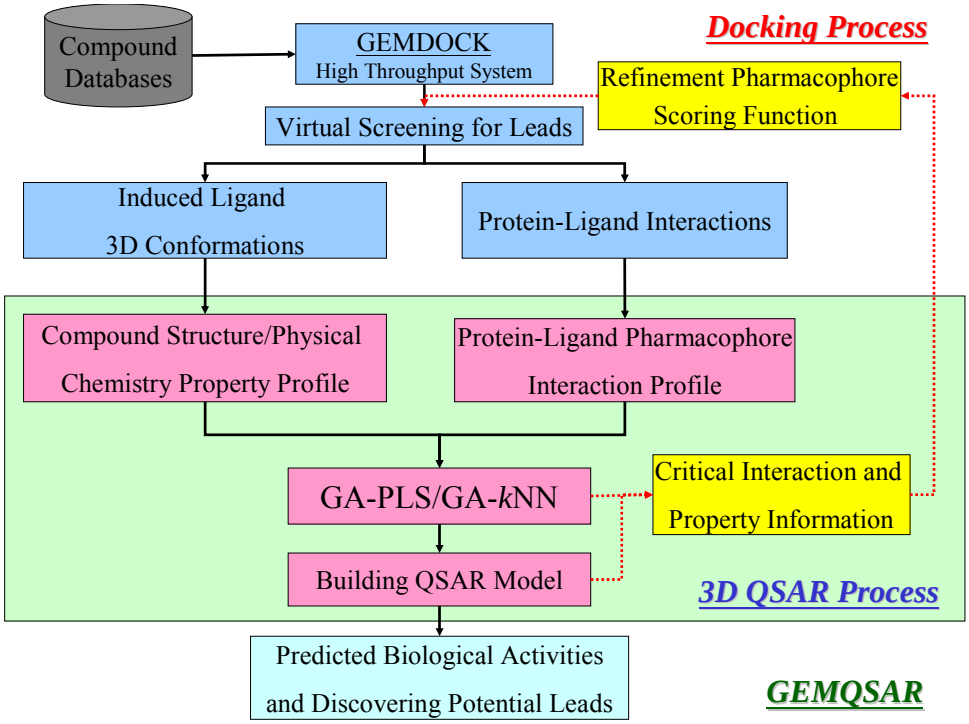

Figure 1. The relationships of 3D QSAR and docking process in GEMQSAR

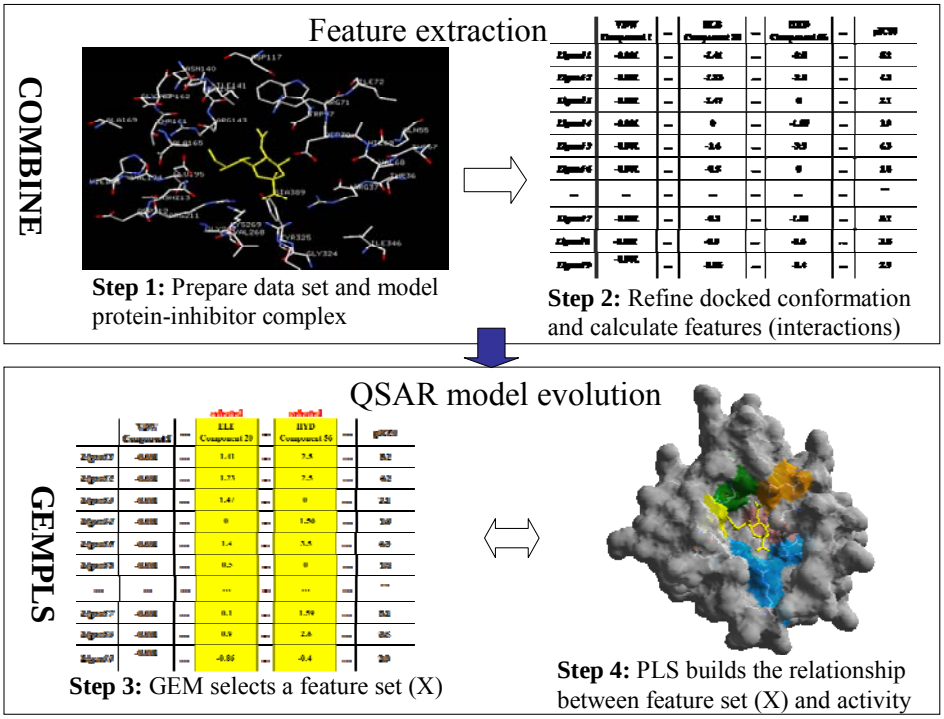

Figure 2. The framework and steps of GEMPLS applied in the COMBINE analysis

Data preparation and feature extraction

1) Prepare Data Sets and Model Protein-Inhibitor Complexes

We have used three data sets, 38 influenza neuraminidase (NA) inhibitor complexes[7], glycogen phosphorylase b (GPB)[12] and cyclooxygenase-2 (COX2)[9] as our validation and test. The inhibitory activity data and complex structures were mainly taken from the references[7; 9; 12]. Each cavity was centered the ligand in the complex with a cutoff 7.5 Å, and cavities with ligands docked by GEMDOCK. Then aligned all cavities by superposition, and the interactions of these gap residues were not considered for the COMBINE analysis.

2) Refine protein-inhibitor complexes and calculate features

The calculated ligand-receptor interaction energies were partitioned on three basis, residue, group, and atom. : 1) Residue-based method, the total binding energy of a residue is the basis unit. : 2) Group-based method, divide a residue to main chain and side chain, and the group is the basis unit. : 3) Atom-base method, the binding energy of an atom is the basis unit. The interaction profiles were outputted for QSAR model. The ligand-receptor interaction energies included van der Waals interaction (E_{vdW}) and hydrogen bonding interaction ($E_{h\text{-}bond}$) as below.

$$E_{vdW} = \sum_{i=1}^{lig} F_{vdW}(r_{ij}) \tag{2}$$

$$E_{h-bond} = \sum_{i=1}^{lig} F_{hb}(r_{ij}) \tag{3}$$

$$F(r_{ij}) = \begin{cases} V_6 - \dfrac{V_6 r_{ij}^{B_{ij}}}{V_1} & \text{if } r_{ij}^{B_{ij}} \le V_1 \\ \dfrac{V_5 (r_{ij}^{B_{ij}} - V_1)}{V_2 - V_1} & \text{if } V_1 < r_{ij}^{B_{ij}} \le V_2 \\ V_5 & \text{if } V_2 < r_{ij}^{B_{ij}} \le V_3 \\ V_5 - \dfrac{V_5 (r_{ij}^{B_{ij}} - V_3)}{V_4 - V_3} & \text{if } V_3 < r_{ij}^{B_{ij}} \le V_4 \\ 0 & \text{if } r_{ij}^{B_{ij}} > V_4 \end{cases} \tag{4}$$

lig presents the number of atom and r_{ij} is the distance of atom *i* and protein *j*. F_{hb} and F_{vdW} are hydrogen bonding and van der Waal interaction, respectively. $F(r_{ij})$ is the linear function of six parameters. The six parameters of V_1–V_6 are shown in figure 3.

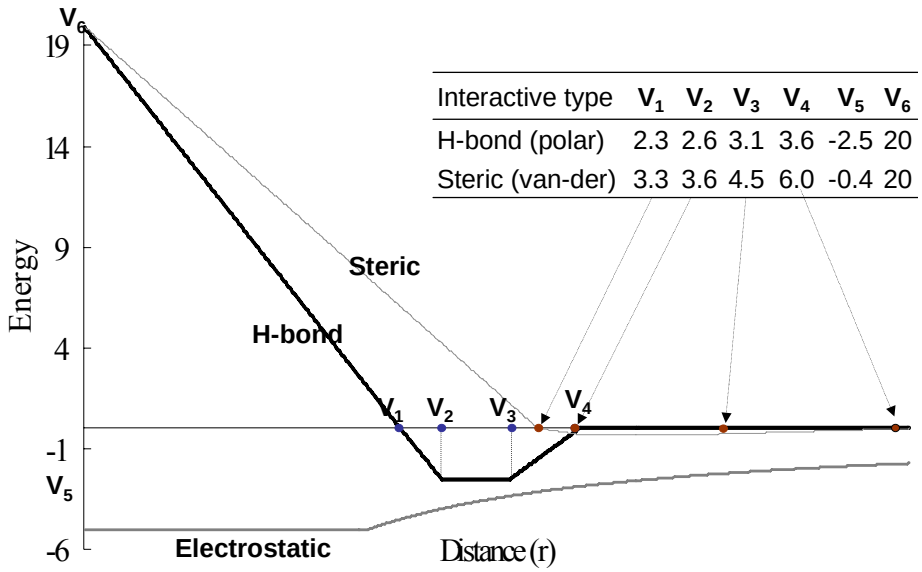

Figure 3. The linear functions and six parameters of hydrogen bonding interaction and van der Waals interaction in GEMDOCK

QSAR Model Evolution and building: GEMPLS

PLS has played a critical role in the derivation of QSAR in CoMFA or COMBINE studies. Recently, more and more people recognize the benefits of feature selection before PLS regression. GAPLS has been shown as a practical solution. But when the number of features becomes large, GAPLS still has difficulty in driving out noises. And scanning for best *lv* is too inefficient and time consuming. Here, we introduce a number of successive enhancements, which are described in the following paragraphs, to construct our model GEMPLS to overcome the drawbacks of GAPLS.

The general idea of PLS is to try to extract these latent variables, accounting for as much of the manifest feature variation as possible while modeling the inhibitory activities well. To decide both the optimum number of latent variables and prediction error of a QSAR model, we defined the weighted standard deviation error of the predictions (WSDEP) as the scoring function of our GEMPLS:

$$WSDEP = \sqrt{\frac{\sum (y_i - y_i)^2}{N - lv - 1} \left(\frac{100}{95}\right)^{lv}} \qquad (5)$$

where y_i and $y_{pred,i}$ are the observed and predicted inhibitory activities belong to inhibitor i, N is the total number of samples, and lv is the number of latent variables in the current model. In order to improve on the efficiency, we append an extra bit lv, representing the number of latent

variables, to the original chromosome and expect GEMPLS model to efficiently solve the problem of the optimum number of latent variables though evolutionary process.

3) Select Features by Mahalanobis Distance

Mahalanobis distance is able be used to measure the deviation of a sample from the mean of the distribution in multivariable calculus. Therefore, the Mahalanobis distance is adopted to identify significant features from all of those.

$$M^2 = (v-u)' \sum^{-1} (v-u) \qquad (6)$$

M is the Mahalanobis distance from the feature vector v (column vector of data matrix here) to the mean vector μ, where Σ is the covariance matrix of the features.

4) Feature Selections and QSAR Models Evolution

The inhibitory activity usually correlates with few important interaction energy features, that is, most of interaction energy features are meaningless or not apparently distinct from each other. GEM was applied to find out the significant interaction energy features and PLS was used to build the QSAR models based on these selected features. *WSDEP* was used as the objective function to provide a measure of how the internal predictability with respect to the selected features. The fittest individual will have the lowest *WSDEP*.

GEM, modified and enhanced from our previous works[6], consists of five steps briefly described in the following:

(1) **Initiation and evaluation of the initial population** ($G_{t=0}$). Each chromosome is composed by an array of feature set and an *lv* value. For example, a chromosome has n+3 bits if the number of candidate feature is n and three bits for *lv* value. The initial population ($G_{t=1}$) of population size (N_p) is created by setting feature bits (0 denote the absence of corresponding feature, and 1 denote its presence) and an *lv* value (denote the number of latent variables and range in [1~5]) of each chromosome to random values and one, respectively. Then PLS is used to build a quasi-QSAR model, and evaluated by the scoring function (*WSDEP*), for each chromosome.

(2) **Selection of the reproductive population**. The chromosomes of reproductive population (P_sG_t) are selected from the population (G_t) with a fixed proportion (P_s) according to the stochastic universal sampling.

(3) **Crossover and mutate the reproductive population** (P_sG_t). The offspring population (G_{off}) is generated by uniform crossover with a probability (crossover rate: P_c) and mutation operators, including uniform and biased mutation operators, with a probability (mutation rate: P_m).

(4) **Evaluation of the offspring population** (G_{off}). PLS is then used to build a quasi-QSAR

model, evaluated by *WSDEP*, for each chromosome in the offspring population.

(5) **Reinsertion of the child population.** To form the population of the next generation (G_{next}), the chromosomes of the current population (G_t) with lower objectives in the preceding (1-P_s) proportion are protected to the next generation, while the others are replaced with better ones from the offspring population (G_{off}). Let $t = t+1$ and $G_t = G_{next}$.

(6) The cycle of above four steps (from step 2 to 5) is repeated until the number of generation reaches to the maximum number of generations (N_{max}). The values of empirical parameters are defined as follows: $N_p = 100$, $N_{max} = 200$, $P_s = 0.9$, $P == 0.6$, and $P_m = 0.05$.

Genetic operator: Biased Mutation**.** The uniform mutation may incur a risk of local convergence and slow evolution because plenty of features will raise the combinatorial complexity of feature space. To reduce the phenomena, the uniform mutation was cooperated with biased mutation to lead the evolution of GA toward significant feature set and to reduce the interference of noise features.

$$F(x_i) = MIN + (MAX - MIN) \times (\frac{N_f - X_i}{N_f - 1}) \quad (7)$$

where $F(x_i)$ is the probability of selection of feature i; x_i is the rank of feature i in the descending order of Mahalanobis distance of all features, *MIN* and *MAX* are the lower and upper bounds, respectively, of probability of biased mutation; N_f is the number of significant features. The value of $F(x_i)$ is derived from xi only when xi is ahead of N_f, otherwise $F(x_i)$ is set to *MIN*. The meaning of $F(x_i)$ is that the more significant feature, the more higher probability of selection. In this study, *MAX*=0.8, *MIN*=0.2 and N_f =39.

5) Performance Evaluation

The predictability of QSAR model was assessed by the conventional correlation coefficient (r^2), the cross-validated correlation coefficient (q^2), the cross-validated *SDEP* ($SDEP_{cv}$), and external *SDEP* ($SDEP_{ex}$):

$$q^2 = 1 - \frac{\sum (y_i - y_{pred,i})^2}{\sum (y_i - \bar{y})^2} \quad (8)$$

$$SDEP = \sqrt{\frac{\sum (y_i - y_{pred,i})^2}{N}} \quad (9)$$

where y_i and $y_{pred,i}$ are the observed and predicted activity of inhibitor i, $y_{pred,i}$, respectively, y is the average activity value of the inhibitor set, and N is the total number of inhibitors. The model with more remarkable predictability can provide the higher correlation coefficient (r^2, q^2) and

the lower *SDEP* between the observed and predicted inhibitory activities.

結果與討論

The ligands were divided to training set and testing set according to the reference[7; 9; 12]. The three methods of feature exaction (residue-based, group-based, and atom-based) were used to train three QSAR models by Leave-One-Out method to optimize WSDEP. Our results of neuraminidase were shown in Table 1. The $q^2$of the reference in training set was better than the results of the three GEMQSAR models, but the q^2 of the GEMQSAR models in testing set were better the q^2 than the reference. Our model although had sight lower training correlation to Wade et al but we showed more superior prediction power than Wade et al. The energy basis unit was reduced (residue->group->atom), and the q^2 in training set improved with feature unit reduced. However the q^2 in testing set didn't become better like in training set.

Table 1. Performance comparison of GEMPLS and the reference on neuraminidase

Model	Original features[a]	Selected features[b]	*Lv* [c]	Train r^2 [d]	Train q^2 [e]	Test r^2	Test q^2
Residue-based	153	13	3	0.753	0.611	0.85	0.754
Group-based	306	14	3	0.737	0.621	0.876	0.798
Atom-based	1233	33	3	0.794	0.688	0.830	0.769
Wang et al[7]	770	330	3	0.877	0.875	0.582	0.566

[a] The number of features extracted from original data by different methods

[b] The number of features selected by GEMPLS

[c] Latent variable

[d] The conventional correlation coefficient.

[e] The cross-validated correlation coefficient

Generally, PLS easily trended to over-fit when the number of features increasing. Therefore, a well-developed strategy of feature extraction should consider the balance of feature number and predicting ability. We develop three QSAR model regarding three strategies of feature extraction and analysis the relationship of performance and feature numbers. Comparing with group-based and residue-based models, the residue-based model had better training quality but lower predicting power. The higher predicting ability showed that group-based method extracted more accuracy information in our QSAR analysis. In other words, one residue-based unit might contain the noises and meaningful features. After dividing feature unit form residue to backbone

and side-chain, the meaningful feature was correctly recovered in GEMPLS (Table 2.). Sum up the above, the advantages of group-based were: 1) Include residue-based information and delete noise of feature by feature selection. : 2) Select fewer features than atom-based and uneasy over fitting. According to present results, it is better to use group-based unit to build a QSAR model.

Table 2. The important groups and their GEMPLS coefficient.

Residue	**Group type**	**Energy type**	**GEMPLS coefficient**
ARG118	side	VDW	-0.192
ARG118	side	ELE	-0.12
GLU119	side	ELE	-0.1
ARG152	main	VDW	-0.396
TRP178	side	HYD	-0.359
SER179	main	VDW	-0.135
SER179	side	VDW	-0.124
ILE222	side	VDW	-0.236
GLU227	main	HYD	-0.313
ALA246	main	VDW	-0.133
ALA250	side	VDW	-0.674
ALA250	side	ELE	-0.169
VAL275	side	ELE	-0.227
ARG292	main	VDW	-0.232
VAL349	side	VDW	-0.207

The predicted pIC50 values were plotted against experimental pIC50 values for the group model with three latent variables in Figure 4. From Figure 4, the group GEMQSAR model has a good prediction. The results of GEMPLS were shown in Figure 5 (a), which showed the coefficients of the electrostatic, hydrogen bond and van der Waals. The negative coefficients contributed to the activity, and the important groups were listed in table 2. The important structural features for a strong inhibitor and corresponding groups were shown in Figure 5 (b). Some interactions played critical role to contribute higher activity. : 1) The electrostatic interactions in orange groups. : 2) The hydrogen bond interactions in green groups. : 3) The van der Waals interactions in gray groups. On the basis of the above COMBINE analysis, we could

develop new inhibitors which have high activity according to the suggested properties of the GEMQSAR model.

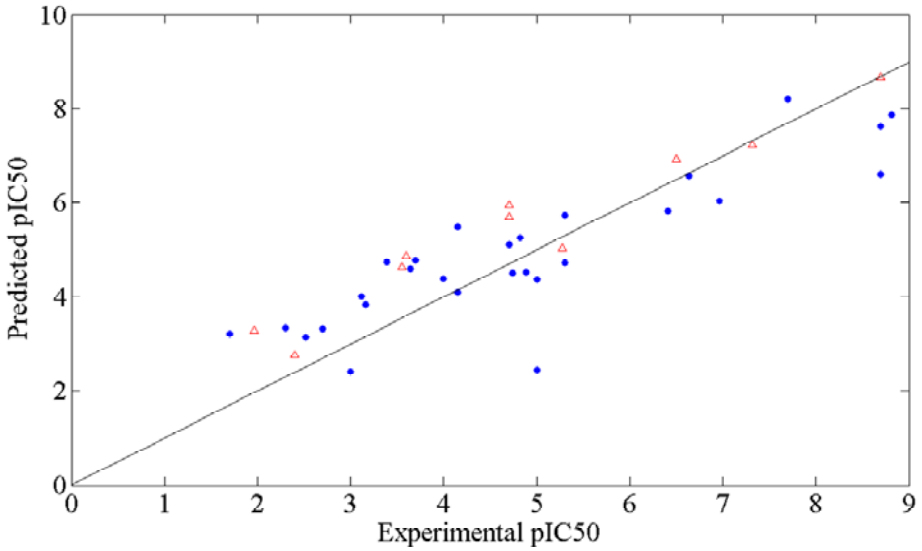

Figure 4. Experimental pIC50 values versus predicted pIC50 values for the group model derived from 38 complexes: ●, predicted values for training set from leave-one-out cross-validation at three latent variables; △, predicted values for testing set.

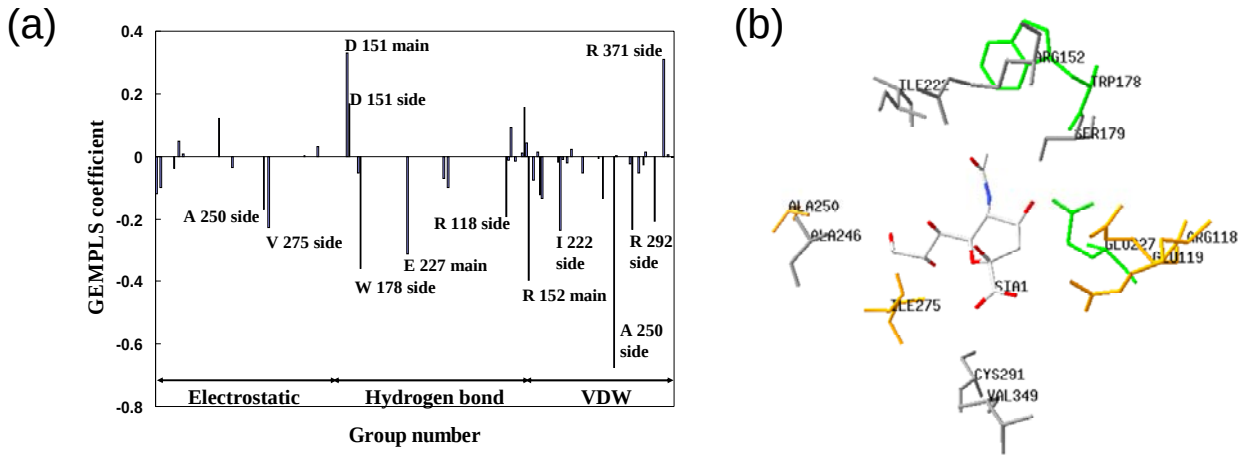

Figure 5. Selected features of neuraminidase inhibitor model (a) The coefficients of the electrostatic, hydrogen bond and van der Waals. (b) The important structural features of binding site.

Table 3. Performance comparison of GEMPLS and the reference for glycogen phosphorylase b

Model	Original features[a]	Selected features[b]	*Lv* [c]	Train r^2 [d]	Train q^2 [e]	Test r^2	Test q^2
Residue-based	120	16	3	0.481	0.337	0.115	0.002
Group-based	240	18	3	0.666	0.546	0.286	0.271
Atom-based	918	32	3	0.699	0.584	0.28	0.26
Silber et al[12]	na [f]	na [f]	na [f]	0.92	na [f]	0.59	na [f]

[a] The number of features extracted from original data by different methods

[b] The number of features selected by GEMPLS

[c] Latent variable

[d] The conventional correlation coefficient

[e] The cross-validated correlation coefficient

[f] Data not available

Table 4. Performance comparison of GEMPLS and the reference for cyclooxygenase-2

Model	Original features[a]	Selected features[b]	*Lv* [c]	Train r^2 [d]	Train q^2 [e]	Test r^2	Test q^2
Residue-based	135	21	3	0.504	0.362	0.066	-2.131
Group-based	270	19	3	0.527	0.385	0.026	-1.736
Atom-based	1230	62	3	0.688	0.566	0.066	-1.203
Pei et al[9]	na [f]	na [f]	na [f]	0.61	na [f]	0.34	na [f]

[a] The number of features extracted from original data by different methods

[b] The number of features selected by GEMPLS

[c] Latent variable

[d] The conventional correlation coefficient

[e] The cross-validated correlation coefficient

[f] Data not available

We applied the same strategy to the other data sets: GPB[12] and COX-2[9], and the results of GPB and COX-2 were shown in Table 3 and Table 4, respectively. The low test q^2 in both data

sets mean that the descriptors of interaction profiles didn't have a correlation with activities. There might be some reasons to explain that. :1) Some atom types weren't defined clearly in GEMDOCK scoring function. For example, the atom type F, was regarded as like C. :2) The hydrogen-bond interaction wasn't sensitive in GEMDOCK scoring function. In QSAR model, there was little difference in the side chains of ligands, and it was necessary to generate the accurate descriptors of interactions to determine the difference. We would correct the two problems to improve the QSAR model in the recently future.

In summary, we apply GEMQSAR to influenza neuraminidase inhibitor complexes and compare three methods of feature exaction (residue-based, group-based, and atom-based).The results show that the group-based method has better prediction. The important interactions are found in this model, and some suggestions are given to design new inhibitors. In residue-based method, a residue may contain useful information and noise. It will reduce the accuracy of prediction. In atom-based method, there are too many features in training, and it is easily over fitting. Therefore, group-based method is useful to COMBINE model. But we just divide a residue to main chain and side chain as a group-based method. In future, it is necessary to research how to define a "group". A good definition of group-based method will improve the COMBINE model.

計畫成果自評

We developed a QSAR methodology associating molecular docking and feature selection with PLS. The feature of our model generates from the interaction energies of docked results, named as protein-ligand interaction profile and is extracted as atom-based, group-based and residue-based terms. We applied our QSAR methodology to build inhibitory models of neuraminidase, glycogen phosphorylase b and cyclooxygenase-2. Our results also compare to published references and our performances show more prediction power than other models. In the recently future, we will make more efforts to improving our methodology and combine virtual screen to create high through-put prediction environment.

參考文獻

1. **Kubinyi, H. (1997). QSAR and 3-D QSAR in drug design. 2. Applications and problems. *Drug Discovery Today* 2, 538-546.**
2. **Ortiz, A. R.;Pisabarro, M. T.;Gago, F. &;Wade, R. C. (1995). Prediction of drug binding affinities by comparative binding energy analysis. *Journal of Medicinal Chemistry* 38, 2681-2691.**
3. **Perez, C.;Pastor, M.;Ortiz, A. R. &;Gago, F. (1998). Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. *Journal of Medicinal Chemistry* 41, 836-852.**

4. **Wold, S.;Johansson, E. &;Cocchi, M. (1993). *PLS - Partial least-squares projections to latent structures. In H.Kubinyi Ed., 3D QSAR in Drug Design; Theory, Methods and Applications.*, ESCOM Science Publishers, Leiden Holland.**
5. **Wanchana, S.;Yamashita, F. &;Hashida, M. (2003). QSAR analysis of the inhibition of recombinant CYP 3A4 activity by structurally diverse compounds using a genetic algorithmcombined partial least squares method. *Pharmaceutical Research* 20, 1401-1408.**
6. **Yang, J. M. &;Kao, C. Y. (2001). *IEEE/OSA Journal of Lightwave Technology*.**
7. **Wang, T. &;Wade, R. C. (2001). Comparative Binding Energy (COMBINE) analysis of influenza neuraminidase-Inhibitor complexes. *Journal of Medicinal Chemistry* 44, 961-971.**
8. **Sutherland, J. J.;O'Brien, L. A. &;Weaver, D. F. (2004). A comparison of methods for modeling quantitative structure-activity relationships. *Journal of Medicinal Chemistry* 47, 5541-5554.**
9. **Pei, J.;Chen, H.;Liu, Z.;Han, X.;Wang, Q.;Shen, B.;Zhou, J. &;Lai, L. (2005). Improving the quality of 3D-QSAR by using flexible-ligand receptor models. *Journal of Chemical and Information Modeling* 45, 1920-1933.**
10. **Yang, J. M. &;Chen, C. C. (2004). GEMDOCK: a generic evolutionary method for molecular docking. *Proteins: Structure, Function, and Bioinformatics* 55, 288-304.**
11. **Chen, Y. C.;Yang, J. M.;Tsai, C. H. &;Kao, C. Y. (2005). *Lecture Notes in Computer Science*.**
12. **Silber, K.;Heidler, P.;Kurz, T. &;Klebe, G. (2005). AFMoC enhances predictivity of 3D QSAR: a case study with DOXP-reductoisomerase. *Journal of Medicinal Chemistry* 48, 3547 -3563.**