Local and Global Quantitative Structure−Activity Relationship Modeling and Prediction for the Baseline Toxicity

Hua Yuan, Yongyan Wang, and Yiyu Cheng*

Pharmaceutical Informatics Institute, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310027, China

Received July 18, 2006

The predictive accuracy of the model is of the most concern for computational chemists in quantitative structure−activity relationship (QSAR) investigations. It is hypothesized that the model based on analogical chemicals will exhibit better predictive performance than that derived from diverse compounds. This paper develops a novel scheme called "clustering first, and then modeling" to build local QSAR models for the subsets resulted from clustering of the training set according to structural similarity. For validation and prediction, the validation set and test set were first classified into the corresponding subsets just as those of the training set, and then the prediction was performed by the relevant local model for each subset. This approach was validated on two independent data sets by local modeling and prediction of the baseline toxicity for the fathead minnow. In this process, hierarchical clustering was employed for cluster analysis, *k*-nearest neighbor for classification, and partial least squares for the model generation. The statistical results indicated that the predictive performances of the local models based on the subsets were much superior to those of the global model based on the whole training set, which was consistent with the hypothesis. This approach proposed here is promising for extension to QSAR modeling for various physicochemical properties, biological activities, and toxicities.

INTRODUCTION

Quantitative structure−activity relationship (QSAR) studies have been well-developed and widely applied to the prediction of various biological activities and toxicities. The predictive accuracy and application domain of a QSAR model are the two aspects of most concern for computational chemists. In order to extend the application domain, a global QSAR model is traditionally based on the entire diverse data set to cover a wide range of chemicals. However, the predictive accuracy is not always satisfactory because such a global model is too rough and general to represent the detailed and subtle structure−activity relationship of each small subset of molecules; that is, the specific local features may be overshadowed by the more global ones. As QSAR is based on the assumption that compounds from the same chemical domain will behave in a similar manner, QSAR models built upon the analogical chemicals would have a better chance of capturing the structure−activity relationship that characterizes these molecules accurately and would give rise to the improved predictive performance compared to the QSAR models derived from the diverse data set. This hypothesis has been validated in a number of papers. For example, Bergström et al.[1] have investigated the subset-specific models for aqueous solubility by dividing 85 druglike compounds into acids, bases, and ampholytes according to the functional groups. As a result, the predictive accuracy for bases and ampholytes was improved by comparing with the global model. He and Jurs[2] have also assessed the reliability of the QSAR model's prediction and found that the activity of a query compound could be predicted more accurately by the local QSAR model generated from the compounds more similar to the query compound than that from the compounds less similar. Pan et al.[3] have attempted to construct local QSAR models for the prediction of blood−brain barrier penetration by clustering the whole data set into subsets based on 4D molecular similarity measures. The compounds in each subset were further divided into a training set and a test set. Then, a specific QSAR model was constructed on the basis of the corresponding subset of each training set. The result suggested that the compounds in the test set could be best predicted by the subset-specific model. Recently, Guha et al.[4] applied local lazy regression to predict the biological behavior of a query compound using its local neighborhood, rather than considering the whole data set. The Sheridan research group[5] developed molecular transformations (i.e., making a small change to a chemical structure) as a method of automatically organizing and displaying sets of closely related compounds to build local QSAR models. Another work[6] by them tried a set of retrospective cross validations against 20 diverse in-house activity sets in order to find a good discriminator of prediction accuracy for molecules not in the training set, and they found that the molecules with the highest similarity and the most neighbors in the training set could be best predicted. Therefore, the local QSAR modeling scheme seems promising to improve the predictive accuracy for the modeling of the data set with large diversity in both the molecular structure and the mechanisms of action, such as the data set regarding toxicity.

In recent years, toxicity has become an increasingly concerned topic of QSAR study because of its great

* Corresponding author phone: +86-571-87952509; fax: +86-571-87951138; e-mail: chengyy@zju.edu.cn.

10.1021/ci600299j CCC: $37.00
Published on Web 12/24/2006

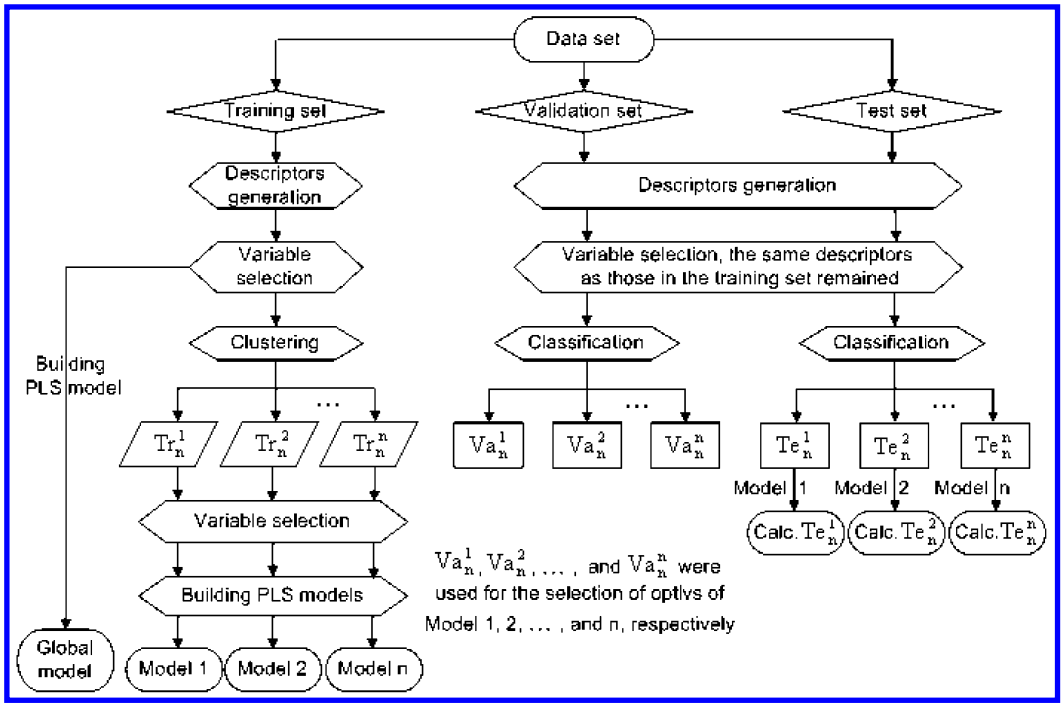

Figure 1. Scheme of QSAR modeling in this study.

importance in the virtual screening of drug candidates. There were numerous reports on the in silico prediction of various toxicity endpoints such as acute toxicity,[7,8] carcinogenicity,[9,10] mutagenicity,[11,12] genotoxicity,[13] and so on. In the field of predictive toxicology, three types of models may be summarized: the mechanistic model, mode-of-action-based model, and statistical model. Although good predictive performance could be achieved, the two former types of models were rarely developed because the mechanism and the mode of action of most toxicological endpoints were unknown. Thus, the statistical models play a dominating role in the prediction of toxicity. However, many global statistical models based on the whole data set show deficiency in the predictive performance because the toxicity data set is very complex in not only the molecular structure but also the toxicological behavior. The local QSAR modeling introduced above may be applied to such a data set, but a problem is raised in regard to how the subsets are formed. Because of the high diversity of the molecular structures and often more than one functional group involved in one molecule, the data set on toxicity cannot be easily divided into several subsets according to the functional groups as in the method employed in Bergström et al.'s work.[1] Therefore, another approach needs to be explored.

This paper proposes a strategy called "clustering first, and then modeling", that is, to cluster the compounds in the training set into subsets according to structural similarity and then build local model for each subset. For the prediction, the test set was classified into the corresponding subsets of the training set and then predictions were made by the relevant local models. The object of this study is to evaluate the performance of local modeling by comparing with that of global modeling. The baseline toxicity (also known as narcosis, general anesthesia, or nonspecific toxicity) to the fathead minnow was used as the interested toxicological endpoint to validate this approach.

MATERIALS AND METHODS

Data Set. Two data sets with experimental values of 96 h median lethal concentration (LC_{50}, concentration in mg/L producing lethality in 50% of the test animals after 96 h of exposure) for the fathead minnow were taken from ref 14. The original source of these LC_{50} values is the EPA Fathead Minnow Acute Toxicity Database[15,16] compiled by the Mid-Continent Ecology Division of the U.S. EPA's National Health and Environmental Effects Research Laboratory (Duluth, MN). As described in ref 15, LC_{50} values were calculated using the Spearman−Karber method.[17] The geometric means of LC_{50} values were employed when more than one bioassay was conducted for the chemical. Logarithm transformation of LC_{50} was used as the dependent variable for model generation.

Data set 1 contained 310 compounds, which was a combination of the calibration data set (207 compounds after eliminating number 127, saccharin sodium salt hydrate) and the external validation set (103 compounds) of ref 14. Data set 2, including 211 compounds, was taken from the comparison data set of ref 14 by removing numbers 55 and 79, which had the highest and lowest LC_{50} values (markedly different from the others), respectively. The molecular weights of these compounds were all between 32 and 285 amu. Both data sets 1 and 2 were used to validate the local modeling scheme proposed in this paper. The 310 compounds of data set 1 were randomly divided into the training set, validation set, and test set with 155, 78, and 77 compounds in each data set, respectively. Similarly, data set 2 was also divided into the training, validation, and test sets with 106, 53, and 52 compounds, respectively. The training sets were used for the generation of QSAR models, which were validated by the validation sets. The test sets served as external samples to test the predictive ability of these models. Each data set is separately listed in Tables S1−S6 of the Supporting Information.

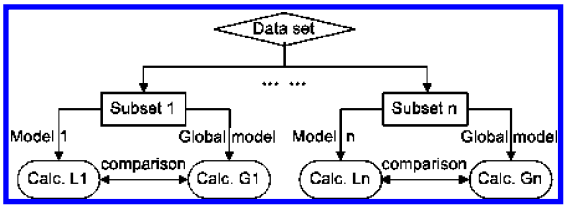

Figure 2. Scheme of results comparison based on each subset.

Generation of Molecular Descriptors. The molecular descriptors of all compounds were generated by Dragon version 5.2.[18] Except for the 3D structural descriptors, 926 0D−2D structural descriptors contained in blocks 1−10, 17−18, and 20 of the descriptor package were calculated. These descriptors consisted of constitutional descriptors, topological descriptors, walk and path counts, connectivity indices, information indices, 2D autocorrelations, edge adjacency indices, BCUT descriptors, topological charge indices, eigenvalue-based indices, functional group counts, atom-centered fragments, and molecular properties. The 2D molecular structures with hydrogens (H) saved as MDL molfiles were used as the input of Dragon for the descriptors generation.

Scheme of the QSAR Modeling. The strategy of QSAR investigation in this study is outlined as follows: (1) clustering, where the training set was clustered into *n* subsets according to structural similarity with an unsupervised pattern recognition technique; (2) classification, where the subsets of the training set were used as known chemical classes to classify the validation set and test set into subsets corresponding to those of the training set by a supervised pattern recognition method; (3) modeling, where local models for each subset of the training set and the global model for the whole training set were built and the corresponding subsets of the validation set were used to validate each model; and (4) prediction, where the baseline toxicity of the test set was predicted by both the corresponding local models and the global model and the performances of each model were evaluated. The scheme is plotted in Figure 1. In the following paragraphs, the methods and algorithms involved in each step are described in detail. All of the calculations were implemented in the Matlab environment with programs written in-house.

Variable Selection and Transformation. When the 926 calculated molecular descriptors were taken for consideration, many of them were redundant or highly correlated. In order to reduce the dimensionality of variables and eliminate the redundant features, objective variable selection was carried out according to the following two steps: (1) eliminating descriptors which have identical values for greater than 90% of the compounds by the same test and (2) removing descriptors whose pairwise correlation coefficients with other descriptors exceed *p* by a pairwise correlation test. The criterion of pairwise correlation coefficient (*p*) for variable deletion ranged mostly from 0.8000 to 0.9000 in the literature. To select an appropriate threshold for the pairwise correlation test, this paper has tried 0.8000−0.9000 in increments of every 0.0100 and found that the QSAR models derived from the descriptors left by the pairwise correlation test with *p* being 0.8600 for data set 1 and 0.8500 for data set 2 exhibited the best performances. Therefore, this paper chose 0.8600 for data set 1 and 0.8500 for data set 2 as the pairwise correlation criteria to eliminate descriptors. The descriptors for the validation set and the test set were the same as those selected for the training set.

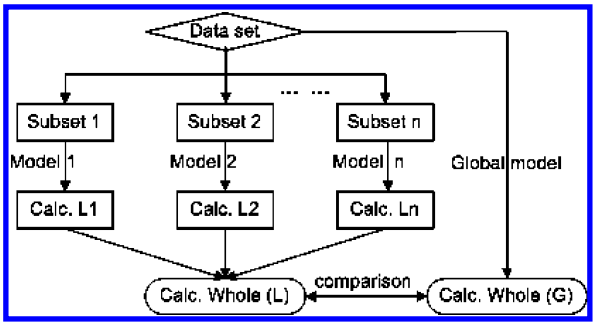

Figure 3. Scheme of results comparison based on the whole data set.

Because these descriptors characterize the molecular structural information from extensive perspectives, their magnitudes are widely different. In order to prevent the descriptors with larger ranges from outweighing those with smaller ranges, the original descriptors are all transformed into a more consistent range. There are a number of approaches for data transformation, such as mean centering, standardization, normalization, and so forth. This paper preprocessed the descriptors by standardization; that is, each descriptor was divided by its standard deviation:

$$x_{ij}^{\mathrm{pre}} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{N} (x_{ij} - \bar{x}_j)^2/(N-1)}} \qquad (1)$$

where x_{ij} is the *j*th descriptor of the *i*th compound and $\bar{x}_j$ is the mean of the *j*th descriptor of *N* compounds. The following steps of clustering, classification, and partial least-squares (PLS) modeling were all based on the transformed descriptors.

Clustering of the Training Set. Cluster analysis is one of the unsupervised pattern recognition approaches to group samples into clusters or subsets based on their similarity, such that samples within each subset are more similar to one another than they are to those in different subsets.[2,19] Similarity is usually assessed by the Euclidean distance, Manhattan distance, or Mahalanobis distance on the basis of the attribute values describing the samples. Different linkage protocols such as Ward link, single link, and complete link can be used for the measurement of the distances between subsets. Hierarchical clustering is a process of grouping samples into a tree of clusters, which is one of the most simple and easily understood clustering methods. Depending on whether the hierarchical decomposition is formed in a bottom-up or top-down fashion, the hierarchical clustering methods can be further classified into agglomerative and divisive hierarchical clustering.[20] The divisive hierarchical clustering starts from the root; that is, all the samples in the data set are in one cluster. Then, the similarities between the samples are determined by the distances between them, on the basis of which the data set is subdivided continuously until the desired number of clusters is formed. The agglomerative method works in the opposite way.

Table 1. Statistical Results Obtained by Local Models and the Global Model for Data Set 1 When the Number of Subsets (n) Ranges from 1 to 5

		local model				global model				
n[a]	subset[b]	R^2	RMS	RSE	optlv	R^2	RMS	RSE	optlv	N[c]
1	Tr_1^1	0.8794	0.3787	0.1828	5	0.8794	0.3787	0.1828	5	155
	Va_1^1	0.7494	0.5886	0.2844	5	0.7494	0.5886	0.2844	5	78
	Te_1^1	0.7096	0.6185	0.3137	5	0.7096	0.6185	0.3137	5	77
2	Tr_2^1	0.9459	0.2339	0.1492	5	0.8983	0.3240	0.2067	5	36
	Tr_2^2	0.8801	0.3687	0.1653	5	0.8588	0.4001	0.1794	5	119
	Va_2^1	0.5983	0.6119	0.4745	5	0.4061	0.7582	0.5879	5	22
	Va_2^2	0.7813	0.5642	0.2389	5	0.7704	0.5516	0.2335	5	56
	Te_2^1	0.7114	0.4913	0.3203	5	0.4812	0.6942	0.4526	5	19
	Te_2^2	0.7830	0.5913	0.2739	5	0.7297	0.6264	0.2901	5	58
3	Tr_3^1	0.8234	0.3158	0.1772	4	0.7747	0.3736	0.2076	5	56
	Tr_3^2	0.9804	0.1769	0.0648	9	0.8771	0.4381	0.1664	5	63
	Tr_3^3	0.9459	0.2339	0.1492	5	0.8983	0.3240	0.2067	5	36
	Va_3^1	0.6475	0.5949	0.2849	4	0.5949	0.6647	0.3122	5	30
	Va_3^2	0.9083	0.5122	0.1614	9	0.8917	0.4614	0.1616	5	26
	Va_3^3	0.5983	0.6119	0.4745	5	0.4061	0.7582	0.5879	5	22
	Te_3^1	0.8409	0.3575	0.1757	4	0.6083	0.6107	0.2941	5	29
	Te_3^2	0.8896	0.6137	0.2291	9	0.8127	0.7025	0.2873	5	29
	Te_3^3	0.7114	0.4913	0.3203	5	0.4812	0.6942	0.4526	5	19
4	Tr_4^1	0.9999	0.0126	0.0026	10	0.9188	0.3617	0.0974	5	17
	Tr_4^2	0.9805	0.1701	0.0682	8	0.8400	0.4830	0.2011	5	46
	Tr_4^3	0.8234	0.3158	0.1772	4	0.7747	0.3736	0.2076	5	56
	Tr_4^4	0.9459	0.2339	0.1492	5	0.8983	0.3240	0.2067	5	36
	Va_4^1	0.8360	*d*	0.1719	10	0.9702	0.3430	0.0846	5	9
	Va_4^2	0.8735	0.7790	0.2277	8	0.8827	0.5773	0.1949	5	17
	Va_4^3	0.6475	0.5949	0.2849	4	0.5949	0.6647	0.3122	5	30
	Va_4^4	0.5983	0.6119	0.4745	5	0.4061	0.7582	0.5879	5	22
	Te_4^1	1.0000	*d*	0.0511	10	1.0000	*d*	0.1768	5	2
	Te_4^2	0.8942	0.5692	0.2140	8	0.8049	0.7290	0.2939	5	28
	Te_4^3	0.8536	0.3540	0.1718	4	0.6372	0.6064	0.2881	5	28
	Te_4^4	0.7114	0.4913	0.3203	5	0.4812	0.6942	0.4526	5	19
5	Tr_5^1	0.9139	0.2190	0.1110	6	0.7876	0.3539	0.1831	5	30
	Tr_5^2	0.6918	0.4478	0.2591	2	0.7631	0.4358	0.2358	5	26
	Tr_5^3	0.9999	0.0126	0.0026	10	0.9188	0.3617	0.0974	5	17
	Tr_5^4	0.9805	0.1701	0.0682	8	0.8400	0.4830	0.2011	5	46
	Tr_5^5	0.9459	0.2339	0.1492	5	0.8983	0.3240	0.2067	5	36
	Va_5^1	0.6738	0.6229	0.2474	6	0.6846	0.6189	0.2529	5	23
	Va_5^2	0.8943	0.2944	0.1827	2	0.8453	1.3807	0.4284	5	6
	Va_5^3	0.8360	*d*	0.1719	10	0.9702	0.3430	0.0846	5	9
	Va_5^4	0.8508	0.7692	0.2426	8	0.7792	0.6931	0.2493	5	18
	Va_5^5	0.5983	0.6119	0.4745	5	0.4061	0.7582	0.5879	5	22
	Te_5^1	0.8141	0.4705	0.1868	6	0.4949	0.6313	0.2602	5	19
	Te_5^2	0.7049	0.4733	0.3093	2	0.7982	0.8075	0.3988	5	9
	Te_5^3	1.0000	*d*	0.0511	10	1.0000	*d*	0.1768	5	2
	Te_5^4	0.8941	0.5835	0.2151	8	0.8091	0.7302	0.2896	5	27
	Te_5^5	0.6633	0.5365	0.3388	5	0.4531	0.7076	0.4468	5	20

[a] n is the number of subsets. [b] Tr, Va, and Te are the abbreviations of the training set, validation set, and test set, respectively; for example, Tr_n^m refers to the mth subset when the whole training set is clustered into n subsets ($m \leq n$); the same applies to Va_n^m and Te_n^m. [c] N is the number of samples in each subset. [d] RMS is in the form of complex number due to $N <$ optlv; thus, the value is not shown.

In this paper, the divisive hierarchical clustering is carried out on the basis of the Euclidean distance and Ward linkage to cluster the training set into subsets. The performances of cluster analysis were evaluated by Silhouette value, which was defined as follows:

$$S(i) = \frac{\min \mathrm{av}D_b(i,k) - \mathrm{av}D_w(i)}{\max[\mathrm{av}D_w(i), \min \mathrm{av}D_b(i,k)]} \qquad (2)$$

where $\mathrm{av}D_w(i)$ is the average distance from the ith sample to the other samples within its own cluster and $\mathrm{av}D_b(i,k)$ is the average distance from the ith sample to samples in another cluster k. The Silhouette values range from -1 to $+1$, and a value closer to $+1$ indicates a higher probability of appropriate object assignment.

Classification of the Validation Set and the Test Set. Classification, often called supervised pattern recognition, is used to assign samples to a number of predefined groups or subsets. There are many classification approaches such as decision tree,[21] Bayesian classification,[20] k-nearest neighbor (kNN),[19,20] and support vector machines (SVM).[22] In this paper, when the subsets derived from the clustering of the

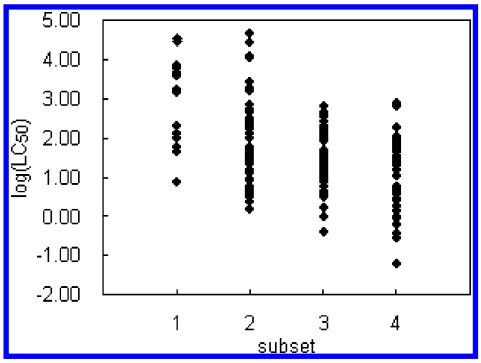

Figure 4. log(LC_{50}) values of the compounds in each subset.

training set are used as the known chemical classes, *k*NN was employed to classify the compounds in the validation set and the test set into the corresponding subsets.

To implement *k*NN, the Euclidean distances of an interested sample in the validation set or the test set to all the members of the training set should be calculated first. Then, these distances are ranked in order, and the *k* smallest distances are picked. The value of *k* is usually an odd number such as 3, 5, and so on. According to the subsets corresponding to these *k* members, the interested sample is assigned to the "majority vote" subset.

Partial Least-Squares (PLS) Regression. PLS regression is one of the most-used multivariate calibration methods.[23,24] It models the dependent variable (**Y**) with a small number of new, independent latent variables extracted by a linear transformation of the original independent descriptors (**X**) to a limited number of orthogonal factors. That is,

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1 + \mathbf{t}_2\mathbf{p}_2 + ... + \mathbf{t}_A\mathbf{p}_A + \mathbf{E} = \mathbf{TP} + \mathbf{E} \quad (3)$$

where **t** is the score vector, **p** is the loading vector for **X**, **q**

$$\mathbf{Y} = \mathbf{t}_1\mathbf{q}_1 + \mathbf{t}_2\mathbf{q}_2 + ... + \mathbf{t}_A\mathbf{q}_A + \mathbf{F} = \mathbf{TQ} + \mathbf{F} \quad (4)$$

$$\mathbf{t} = \mathbf{Xw} \quad (5)$$

is the loading vector for **Y**, *A* is the number of latent variables, and **E** and **F** are the residual matrices for **X** and **Y**, respectively. **w** is the loading weight vector describing the contribution of the original variables to the scores. Large loading weights indicate important variables, while small loading weights indicate less-important variables. As introduced by Wold,[24] the loading weight vector (**w1**) of the first latent variable is used to interpret the statistical model.

PLS is insensitive to the collinearity among the variables and also offers the advantage of handling data sets where the number of independent variables is much greater than the number of samples. The latent variables can represent the original information more efficiently in a reduced dimension. The optimal number of latent variables (optlv) was determined by the validation set to get the minimum relative standard error (RSE). In this paper, *A* ranging from 1 to 10 was investigated to select the optlv. All the statistical results shown in this paper were obtained by the model with optlv.

Assessment of the Statistical Results. Three statistical quantities—the square of correlation coefficient (R^2), root-mean-square error (RMS), and relative standard error (RSE)—were used for the evaluation of the performances of QSAR models. The formulations of RMS and RSE are shown as follows:

$$\mathrm{RMS} = \sqrt{\frac{\sum_{i=1}^{N}(y_{\mathrm{exp}} - y_{\mathrm{calc}})^2}{N - \mathrm{optlv}}} \quad (6)$$

In eqs 6 and 7, y_{exp} and y_{calc} are the experimental and

$$\mathrm{RSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_{\mathrm{exp}} - y_{\mathrm{calc}})^2}{\sum_{i=1}^{N} y_{\mathrm{exp}}{}^2}} \quad (7)$$

calculated log(LC_{50}) values, respectively; *N* is the number of samples in the data set of interest.

The purpose of this paper is to evaluate the performances of local modeling by comparing them with those of global modeling. We can evaluate the results on the basis of not only the subsets but also the whole data set. Both of these schemes are illustrated as Figures 2 and 3, respectively.

When the data set was classified into *n* subsets, the log(LC_{50}) values of each subset were calculated by both the corresponding local model and the global model to obtain the Calc. L1, Calc. L2, ..., Calc. L*n* and Calc. G1, Calc. G2, ..., Calc. G*n*, respectively. On the one hand, Calc. L*n* was directly compared to Calc. G*n*, which was the so-called subset-based comparison, as shown in Figure 2. On the other hand, the calculated values Calc. L1, Calc. L2, ..., and Calc. L*n* by local models for each subset were combined to a whole to get Calc. Whole (L), while the log(LC_{50}) values of the entire data set were also calculated by the global model to get Calc. Whole (G). Then, the Calc. Whole (L) was compared to the Calc. Whole (G), that is, the comparison based on the whole data set, as shown in Figure 3.

RESULTS AND DISCUSSION

Data sets 1 and 2 were both used to validate the QSAR modeling scheme of this paper. For terseness and perspicuity, data set 1 was used as an example to illustrate the QSAR modeling process and results in detail, while the results of data set 2 were presented briefly.

Results and Discussion of Data Set 1. According to the study scheme, the results of data set 1 are presented as the following subsections: Results and Toxicological Understanding of the Clustering and Classification; Results of PLS Modeling; Assessment of the Results; Detailed Description and Discussion of the Statistical Process and Results When *n* Equals 4.

Results and Toxicological Understanding of the Clustering and Classification. After the objective variable selection based on the training set, the original 926 molecular descriptors were reduced to only 169 descriptors, which also remained for both the validation set and the test set. All the descriptors were preprocessed by standardization. On the basis of these descriptors, divisive hierarchical clustering for the training set and *k*NN classification for the validation set and the test set were carried out. For *k*NN classification,

Table 2. Integrated Statistical Results of Data Set 1 with n Ranging from 1 to 5

	training set			validation set			test set		
n^a	R^2	RMS	RSE	R^2	RMS	RSE	R^2	RMS	RSE
1	0.8794	0.3787	0.1828	0.7494	0.5886	0.2844	0.7096	0.6185	0.3137
2	0.9036	0.3385	0.1634	0.7867	0.5564	0.2688	0.7829	0.5516	0.2798
3	0.9519	0.2391	0.1154	0.8041	0.5238	0.2530	0.8576	0.4426	0.2245
4	0.9552	0.2307	0.1114	0.7771	0.5610	0.2710	0.8670	0.4229	0.2145
5	0.9509	0.2417	0.1167	0.7949	0.5385	0.2602	0.8431	0.4601	0.2333

[a] n is the number of subsets.

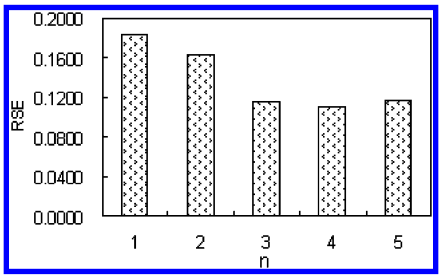

Figure 5. Plot of RSE vs number of subsets (n) for the training set of data set 1.

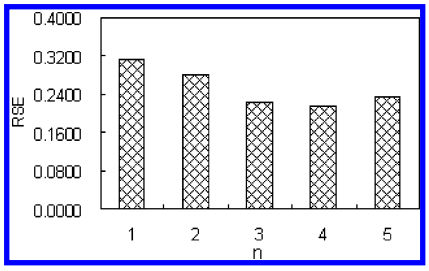

Figure 6. Plot of RSE vs number of subsets (n) for the test set of data set 1.

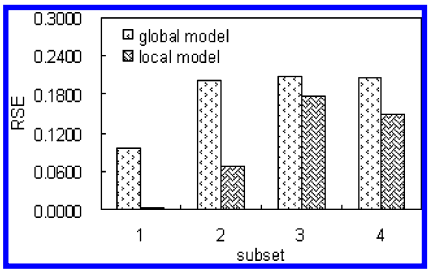

Figure 7. Plot of RSE for each subset of the training set of data set 1 when $n = 4$.

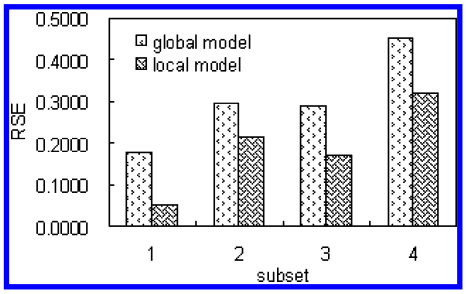

Figure 8. Plot of RSE for each subset of the test set of data set 1 when $n = 4$.

different values of k from 3 to 21 had been tried, and $k = 15$ was ultimately chosen as the best protocol judged by the subsequent statistical results. In order to investigate the performances of local modeling under different numbers of subsets (n) and to determine the optimal value of n, the training set was clustered into one to six subsets ($Tr_n^1 \sim Tr_n^n$, $n = 1 \sim 6$), and the validation set and the test set were also classified into the corresponding subsets ($Va_n^1 \sim Va_n^n$, $Te_n^1 \sim Te_n^n$, $n = 1 \sim 6$). However, when the number of subsets n was equal to 6, there was only one compound in one of the subsets of the test set, which was not fit for the statistical analysis. Therefore, only the local modeling with n ranging from 1 to 5 was investigated, and the number of samples (N) in each subset is given in Table 1.

As subsets were formed on the basis of molecular similarity, the compounds in the same subset would be analogs, while those in different subsets would be distinct from each other. Taking $n = 4$ as an example, the structural features of the compounds in four subsets derived from the training set could be summed up as follows: (1) In subset 1 (Tr_4^1), the compounds were chain molecules with the number of heavy atoms (atoms other than H) in the longest chain less or equal to four, such as methanol, 2-methyl-1-propanol, and 1,2-dichloropropane. (2) In subset 2 (Tr_4^2), the compounds were also chain molecules, but the number of heavy atoms in the longest chain was more than four, such as 4-methyl-2-pentanone, 1-hexanol, diethanolamine, and so on. (3) In subset 3 (Tr_4^3), there was an aromatic ring in each compound, and the longest chain of the substitute groups on the ring only contained one or two heavy atoms, such as acetophenon, *m*-bromobenzamide, and 2-chloro-4-nitroaniline. (4) In subset 4 (Tr_4^4), the compounds also contained an aromatic ring, but the substitute groups on the ring were much larger in length and/or bulk than those in subset 3, such as 4-hexyloxyaniline, di-*N*-butylorthophthalate, and triphenylphosphine oxide. But there were also some exceptions; for example, three compounds (pyrrole, *s*-trioxane, and pyridine) in subset 1 had ring structures; the longest chain of the substitute group in butyl phenyl ether of subset 3 had more than two heavy atoms; four compounds (*N*,*N*-bis(2,2-diethoxyethyl)methylamine, 3-hydroxy-3,7,11-trimethyl-1,6,-10-dodecatriene, dibutyl fumarate, and dibutyl adipate) in subset 4 were chain molecules without ring substructures. Although the exceptions existed, most of the compounds show chemical structural similarity in the same subset and distinctness between different subsets.

Furthermore, seen from the $\log(LC_{50})$ values of the compounds in each subset as shown in Figure 4, the $\log(LC_{50})$ values were in the range of $0.88 \sim 4.53$ for subset 1, $0.18 \sim 4.67$ for subset 2, $-0.39 \sim 2.79$ for subset 3, and $-1.21 \sim 2.80$ for subset 4. On the one hand, the upper limits of $\log(LC_{50})$ of subsets 1 and 2 were close to each other but much higher than those of subsets 3 and 4, which indicated that the chain molecules (in subsets 1 and 2) appeared to be less toxic than the molecules containing aromatic ring substructures (in subsets 3 and 4). On the other hand, the lower limit of $\log(LC_{50})$ of subset 2 was below that of subset 1, similar to the relation between subset 4 and subset 3, which

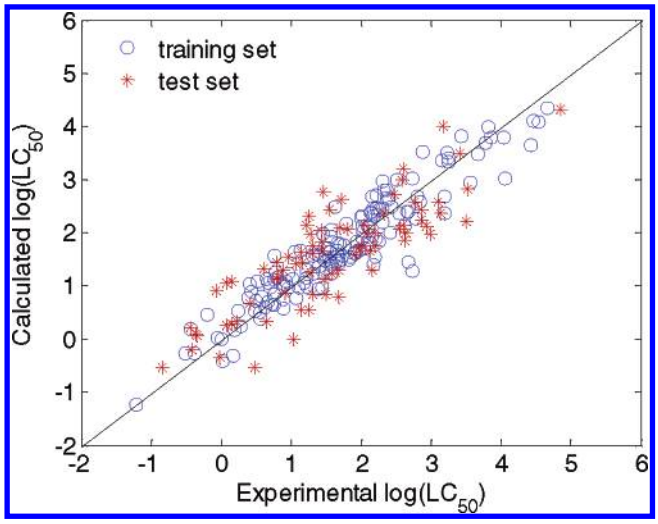

Figure 9. Plot of the calculated vs experimental log(LC_{50}) values for the training set and test set of data set 1 with the global model.

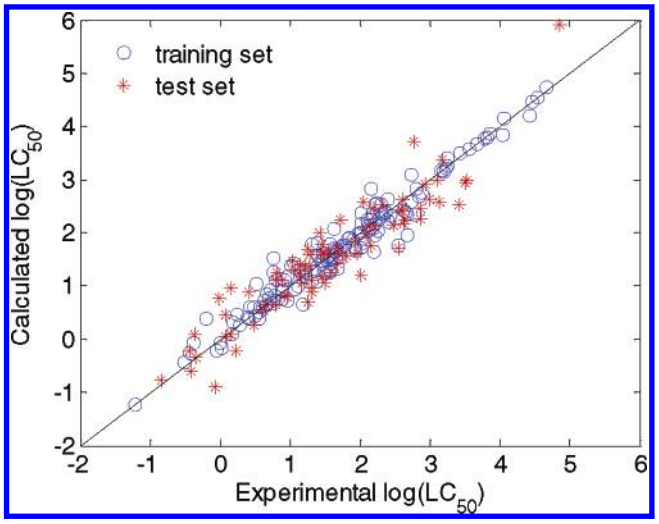

Figure 10. Plot of the calculated vs experimental log(LC_{50}) values for the training set and test set of data set 1 with local models.

implied that the larger molecular size enhanced the baseline toxicity. The above observations were very interesting, and their toxicological foundation was probed. Although a thorough mechanistic understanding of baseline toxicity (or general anesthesia) does not exist yet, a number of investigations and observations have been reported and extensively reviewed.[25] Early studies noted that there was a good correlation between the anesthetic potencies and the lipid solubility, which was formulated by Meyer[26] as the "lipid solubility theory"; that is, "narcosis commences when any chemically indifferent substance has attained a certain molar concentration in the lipids of the cell". Paton[27] and Janoff et al.[28] have found that the correlation between general anesthetic potency and lipid solubility holds for an enormous range of chemical structures and described it as "one of the most impressive correlations in biology". In addition, Mullins[29] suggested that the molecular volume was important to general anesthesia, which was confirmed by the investigations of Miller et al.[30] and Raines et al.[31] Our observation that larger molecular size enhanced the baseline toxicity might be explained in that larger molecular size increased the hydrophobicity and facilitated the compound to penetrate into the lipid bilayer of the cell membrane. The chain molecules in subsets 1 and 2 appeared to be less toxic than the molecules containing aromatic ring substructures in subsets 3 and 4, which could be inferred that different molecular shapes and functional groups might account for the specific binding sites to proteins.[32]

Results of PLS Modeling. In this section, two kinds of models, that is, the global model based on the whole training set and the local models based on each subset of the training set, were built. Prior to local modeling, variable selection and transformation were performed again on the basis of the 169 descriptors of each subset because the redundant information and pairwise correlation may change with a different composition of the subset. The optimal number of latent variables of each model was determined by the corresponding subset of the validation set. For example, the optlv of the global model was determined by the whole validation set, and the optlv of the local model for Tr_n^m was determined by Va_n^m. The log(LC_{50}) values of each subset of the training set, validation set, and test set were all calculated by the global model and the relevant local models. The statistical results with n ranging from 1 to 5 were shown in the columns titled "global model" and "local model" in Table 1. According to eq 6 for the calculation of RMS, a complex number would occur if N was less than optlv, as could be seen for the subsets Va_4^1, Te_4^1, Va_5^3, and Te_5^3.

Assessment of the Results. When n equaled 1, it meant that no clustering and classification were conducted; the statistical results listed in the columns "global model" and "local model" in Table 1 were identical because they were both generated by the global model. When n ranged from 2 to 5, it was obvious that the performances of the local models were superior to those of the global model in that the RMS and RSE of each subset of the test set obtained by local models were all smaller than those obtained by the global model, while R^2 showed the reverse trend.

Besides the comparison based on each subset, the holistic assessment of the results is more important in some sense because it can indicate the final statistical performances of local modeling. The optimal number of subsets that the data set should be divided into would be determined by the best holistic statistical results of the test set. Therefore, the calculated log(LC_{50}) values by each local model under a certain value of n were combined to a whole for the training set, validation set, and test set, respectively. The integrated statistical results for each data set with n varying from 1 to 5 are shown in Table 2. The values of RSE for the training set and the test set decreased with the augment of n from 1 to 4 but increased when $n = 5$, which could also be seen intuitively from the plot of RSE versus n for the training set and the test set in Figures 5 and 6, respectively. Therefore, it is not the case that the more subsets the training set is clustered into, the better predictive results will be achieved. The reason may be illustrated from two aspects: first, the number of samples in the subsets of the training set reduces with the increase of the number of subsets, which impairs the correctness of the classification model for the validation set and the test set; second, the performances of local QSAR models deteriorated as a result of the limited size of the training subsets. As indicated in Table 2, the best prediction results (the highest R^2 and lowest RSE for the test set) were obtained when the training set of data set 1 was clustered

Table 3. Most Important Molecular Descriptors for the Local and Global Models of Data Set 1 when $n = 4$

model	descriptor	meaning	block
global model	Me	mean atomic Sanderson electronegativity (scales on carbon atom)	constitutional descriptors
	BELm1	lowest eigenvalue *n*. 1 of Burden matrix/ weighted by atomic masses	Burden eigenvalues
	BEHm1	highest eigenvalue *n*. 1 of Burden matrix/ weighted by atomic masses	Burden eigenvalues
	PW2	path/walk 2—Randic shape index	topological descriptors
	HNar	Narumi harmonic topological index	topological descriptors
local model 1	IC1	information content index (neighborhood symmetry of first-order)	information indices
	AAC	mean information index on atomic composition	information indices
	Ms	mean electrotopological state	constitutional descriptors
	SIC1	structural information content (neighborhood symmetry of first-order)	information indices
	SIC0	structural information content (neighborhood symmetry of zero-order)	information indices
local model 2	Me	mean atomic Sanderson electronegativity (scales on carbon atom)	constitutional descriptors
	BELm1	lowest eigenvalue *n*. 1 of Burden matrix/ weighted by atomic masses	Burden eigenvalues
	ATS4m	Broto−Moreau autocorrelation of a topological structure − lag 4/ weighted by atomic masses	2D autocorrelations
	ATS4e	Broto−Moreau autocorrelation of a topological structure − lag 4/ weighted by atomic Sanderson electronegativities	2D autocorrelations
	HNar	Narumi harmonic topological index	topological descriptors
local model 3	BELm1	lowest eigenvalue *n*. 1 of Burden matrix/ weighted by atomic masses	Burden eigenvalues
	PW2	path/walk 2—Randic shape index	topological descriptors
	ATS4e	Broto−Moreau autocorrelation of a topological structure − lag 4/weighted by atomic Sanderson electronegativities	2D autocorrelations
	X3A	average connectivity index chi-3	connectivity indices
	HNar	Narumi harmonic topological index	topological descriptors
local model 4	Me	mean atomic Sanderson electronegativity (scales on carbon atom)	constitutional descriptors
	BELm1	lowest eigenvalue *n*. 1 of Burden matrix/ weighted by atomic masses	Burden eigenvalues
	PW2	path/walk 2—Randic shape index	topological descriptors
	ATS4 m	Broto-Moreau autocorrelation of a topological structure − lag 4/weighted by atomic masses	2D autocorrelations
	BELv2	lowest eigenvalue *n*. 2 of Burden matrix/ weighted by atomic van der Waals volumes	Burden eigenvalues

into four subsets; that is, the optimal number of subsets was determined as four.

Detailed Description and Discussion of the Statistical Process and Results When n Equals 4. The training set was first clustered into four subsets ($Tr_4^1 \sim Tr_4^4$) by hierarchical clustering, and each subset contained 17, 46, 56, and 36 compounds, respectively. According to the subsets of the training set, both the validation set and the test set were classified into four subsets, that is, $Va_4^1 \sim Va_4^4$ and $Te_4^1 \sim Te_4^4$, respectively. Then, four local models were built with partial least-squares based on subsets $Tr_4^1 \sim Tr_4^4$; the optlv of each model was determined by the minimal RSE of the corresponding validation subset ($Va_4^1 \sim Va_4^4$). As such, the optlvs of the local models for subsets $Tr_4^1 \sim Tr_4^4$ were 10, 8, 4, and 5, respectively. In addition, the global model based on the entire training set was also set up with the optlv of 5. The $\log(LC_{50})$ values of each subset of the training set, validation set, and test set were calculated by not only the corresponding local model but also the global model and are listed in Tables S1−S3 of the Supporting Information. The statistical quantities R^2, RMS, and RSE based on the calculated $\log(LC_{50})$ values by both the local models and the global model were all computed and were given in Table 1. For subset Va_4^1, there were only nine compounds, while the optlv of this local model was 10, so the RMS was in the form of a complex number and not given in Table 1. The RSEs for each subset

Table 4. Integrated Statistical Results for Each Subset of Data Set 2 with *n* Ranging from 1 to 5

	training set			validation set			test set		
n^a	R^2	RMS	RSE	R^2	RMS	RSE	R^2	RMS	RSE
1	0.8641	0.4409	0.2646	0.6217	0.8479	0.4940	0.2981	1.0499	0.5962
2	0.9331	0.3092	0.1856	0.5963	0.8871	0.5169	0.4141	0.9974	0.5664
3	0.9490	0.2701	0.1621	0.5970	0.8776	0.5113	0.4619	0.9035	0.5130
4	0.9552	0.2532	0.1519	0.6371	0.8225	0.4792	0.4196	0.9375	0.5323
5	0.9480	0.2727	0.1637	0.6005	0.8933	0.5205	0.3513	1.0190	0.5786

[a] *n* is the number of subsets.

Table 5. Statistical Results Obtained by Local Models and the Global Model for Data Set 2 when the Number of Subsets (*n*) Equals 3

	local model				global model				
subset[a]	R^2	RMS	RSE	optlv	R^2	RMS	RSE	optlv	N^b
Tr_3^1	0.9717	0.2660	0.1265	8	0.8543	0.6038	0.2872	8	29
Tr_3^2	0.9784	0.1496	0.1186	6	0.7876	0.5170	0.3794	8	20
Tr_3^3	0.9202	0.3272	0.1846	7	0.8780	0.4087	0.2283	8	57
Va_3^1	0.5851	1.2160	0.5881	8	0.6906	1.1485	0.5554	8	16
Va_3^2	0.5838	*c*	0.5998	6	0.6113	*c*	0.6395	8	4
Va_3^3	0.6086	0.9111	0.4769	7	0.6068	0.9028	0.4634	8	33
Te_3^1	0.5978	0.8401	0.4818	8	0.4678	0.9148	0.5246	8	18
Te_3^2	0.4678	*c*	0.7227	6	0.2355	*c*	0.8711	8	5
Te_3^3	0.4409	1.0097	0.4891	7	0.3093	1.2109	0.5731	8	29

[a] Tr, Va, and Te are the abbreviations of the training set, validation set, and test set, respectively; for example, Tr_3^m refers to the *m*th subset when the whole training set is clustered into three subsets ($m \leq 3$); the same applies to Va_3^m and Te_3^m. [b] *N* is the number of samples in each subset. [c] RMS is in the form of complex number $N <$ optlv; thus, the value is not shown.

of the training set and the test set were plotted in Figures 7 and 8, respectively. It was observed that the RSE of the local model was much lower than that of the global model for all of the subsets. To represent the superiority of local modeling to global modeling more intuitively, the calculated log(LC_{50}) values of the training set and test set were plotted against the experimental values in Figures 9 and 10 for the local models and the global model, respectively. It was obvious that the calculated log(LC_{50}) values obtained by local models were in much better agreement with the experimental ones than those by the global model for both the training set and the test set.

On the basis of the loading weight vector of the first latent variable of each local and global model, the most important structural descriptors were picked out and are shown in Table 3. These descriptors were mainly from five blocks such as topological descriptors, information indices, 2D autocorrelations, Burden eigenvalues, and constitutional descriptors. Baseline toxicity was generally understood as a disruption of the functions of biological membranes, although the detailed mechanism remained unclear. Thus, the structural features related to the penetration of the membrane and the binding to a specific protein would affect the baseline toxicity. The molecular descriptors employed in these statistical models may give some insights into the toxicological behavior. For example, the constitutional descriptors Me and Ms mainly reflect the electrostatic information, which is closely related to the molecular polarity and may be responsible for the penetration of the molecule into the lipid bilayer of the cell membrane. BELm1, BEHm1, and BELv2 are the descriptors characterizing molecular size, which may be relevant to the hydrophobicity and also may have a significant effect on the penetration behavior of the molecule into the cell membrane. The descriptors describing molecular shape such as PW2, IC1, SIC0, and SIC1 probably relate to the interactions of chemicals to the specific protein target. Furthermore, by comparison of the descriptors for each model, four of the five most important descriptors of the global model came from those of local models 1−4, which could be well-interpreted because the data set for the generation of the global model was the combination of four subsets for the derivation of models 1−4. For each local model, the important descriptors were different from each other because of the specific structural features of each subset.

Results and Discussion of Data Set 2. Data set 2 was used as a comparison data set to validate the local QSAR modeling scheme. A total of 926 molecular descriptors were calculated, and variable selection was performed as data set 1. On the basis of the molecular structural similarity, 106 compounds of the training set were clustered into one to six subsets; the validation set and the test set were also classified to the corresponding subsets. Local models were built when the number of subsets (*n*) ranged from 1 to 5. No statistical analysis was performed when $n = 6$ because no compounds were classified to one of the subsets of the test set. The integrated statistical results with *n* from 1 to 5 are listed in Table 4.

As shown in Table 4, the integrated statistical results of local modeling ($n \geq 2$) were better than those of global modeling ($n = 1$) in that the correlation coefficients (R^2) were higher and the RMS and RSE were lower for both the training set and the test set. With the increase of *n* from 1 to 3, the statistical results were improved gradually but deteriorated when *n* increased from 3 to 5. The performances of local modeling exhibited a parabola trend with the increase of *n*. For this data set, the lowest predictive RMS and RSE were achieved when the data set was clustered into three subsets, so the optimal number of subsets was determined as three. Taking $n = 3$ as example, the statistical results of

local and global models for each subset of the training, validation, and test set were illustrated in detail and shown in Table 5.

A total of 106 compounds of the training set were clustered into three subsets (Tr_3^1, Tr_3^2, and Tr_3^3) with 29, 20, and 57 compounds, respectively. Local models were built for each subset with an optlv of 8, 6, and 7, respectively, which were determined by the corresponding subsets of the validation set (Va_3^1, Va_3^2, and Va_3^3, including 16, 4, and 33 compounds in each of them, respectively). The global model was also constructed on the basis of the whole training set with an optlv of 8. The log(LC_{50}) values of each subset of the training, validation, and test sets were all calculated by both the global and local models and were listed in Tables S4−S6 of the Supporting Information. As indicated by the statistical results (shown in Table 5) of each subset of the training set and test set, the performances of local models were confirmed to be superior to those of the global model.

Discussion on the Scheme of this Study. The scheme of this paper was an integration of unsupervised and supervised pattern recognition and multivariate calibration. There are many other alternate algorithms to be explored in each step, although only one of them has been tried in this study. For example, in the cluster analysis, *k*-means and neural networks can be tried besides the hierarchical clustering. Discriminant analysis, SVM, and probabilistic neural networks other than *k*NN may be used in the classification. For the QSAR modeling, many other calibration approaches such as radial basis function neural networks and SVM can also be explored. Besides the methods, different parameters in each algorithm may be tried and discussed. Therefore, it is an extensive topic of data mining for further study. The objective of this paper is only to present the rough outline of this scheme. More detailed and thorough studies need to be further investigated.

CONCLUSIONS

Motivated by the hypothesis that the QSAR models built upon the analogical chemicals will exhibit better performances than those derived from the diverse data set, this paper developed a scheme of local modeling and prediction based on the subsets of the data set rather than the entire data set for the prediction of baseline toxicity. The subsets were formed by clustering for the training set and by classification for the validation set and test set. Two data sets were employed to validate this scheme and found that the statistical results obtained by local models were much superior to those obtained by the global model, which confirmed the above hypothesis. This scheme is promising in the extension to QSAR modeling for various physicochemical properties, biological activities, and toxicities because of its improvement of the predictive accuracy. However, it is not the case that the larger the number of subsets is, the better the performances of the local models will be. For a given data set, the optimal number of subsets can be determined with the objective of achieving the best predictive performances.

ACKNOWLEDGMENT

This work was financially supported by the key Program in Major Research Plan of National Natural Science Foundation of China (No. 90209005) and National Basic Research Program of China (No. 2005CB523402). The authors thank M.Sc Zhong-Ying Lin (School of Pharmacy, Second Military Medical University, Shanghai, China) for his kind help in writing program for data analysis.

Supporting Information Available: Tables S1−S6 listed the CAS numbers, SMILES strings, experimental and calculated log(LC_{50}) values, and the assignment of subsets for the training, validation, and test sets of data sets 1 and 2, respectively. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Bergström, C. A. S.; Wassvik, C. M.; Norinder, U.; Luthman, K.; Artursson, P. Global and Local Computational Models for Aqueous Solubility Prediction of Drug-Like Molecules. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1477−1488.
(2) He, L.; Jurs, P. C. Assessing the Reliability of a QSAR Model's Predictions. *J. Mol. Graphics Modell.* **2005**, *23*, 503−523.
(3) Pan, D.; Iyer, M.; Liu, J.; Li, Y.; Hopfinger, A. J. Constructing Optimum Blood Brain Barrier QSAR Models Using a Combination of 4D-Molecular Similarity Measures and Cluster Analysis. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2083−2098.
(4) Guha, R.; Dutta, D.; Jurs, P. C.; Chen, T. Local Lazy Regression: Making Use of the Neighborhood to Improve QSAR Predictions. *J. Chem. Inf. Model.* **2006**, *46*, 1836−1847.
(5) Sheridan, R. P.; Hunt, P.; Culberson, J. C. Molecular Transformations as a Way of Finding and Exploiting Consistent Local QSAR. *J. Chem. Inf. Model.* **2006**, *46*, 180−192.
(6) Sheridan, R. P.; Feuston, B. P.; Maiorov, V. N.; Kearsley, S. K. Similarity to Molecules in the Training Set Is a Good Discriminator for Prediction Accuracy in QSAR. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1912−1928.
(7) Serra, J. R.; Jurs, P. C.; Kaiser, K. L. E. Linear Regression and Computational Neural Network Prediction of Tetrahymena Acute Toxicity for Aromatic Compounds from Molecular Structure. *Chem. Res. Toxicol.* **2001**, *14*, 1535−1545.
(8) Smieško, M.; Benfenati, E. Thermodynamic Descriptors Derived from Density Functional Theory Calculations in Prediction of Aquatic Toxicity. *J. Chem. Inf. Model.* **2005**, *45*, 379−385.
(9) Bahler, D.; Stone, B.; Wellington, C.; Bristol, D. W. Symbolic, Neural, and Bayesian Machine Learning Models for Predicting Carcinogenicity of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 906−914.
(10) Hemmateenejad, B.; Safarpour, M. A.; Miri, R.; Nesari, N. Toward an Optimal Procedure for PC-ANN Model Building: Prediction of the Carcinogenic Activity of a Large Set of Drugs. *J. Chem. Inf. Model.* **2005**, *45*, 190−199.
(11) Helma, C.; Cramer, T.; Kramer, S.; Raedt, L. D. Data Mining and Machine Learning Techniques for the Identification of Mutagenicity Inducing Substructures and Structure Activity Relationships of Noncongeneric Compounds. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1402−1411.
(12) Kazius, J.; McGuire, R.; Bursi, R. Derivation and Validation of Toxicophores for Mutagenicity Prediction. *J. Med. Chem.* **2005**, *48*, 312−320.
(13) Li, H.; Ung, C. Y.; Yap, C. W.; Xue, Y.; Li, Z. R.; Cao, Z. W.; Chen, Y. Z. Prediction of Genotoxicity of Chemical Compounds by Statistical Learning Methods. *Chem. Res. Toxicol.* **2005**, *18*, 1071−1080.
(14) Öberg, T. A QSAR for Baseline Toxicity: Validation, Domain of Application, and Prediction. *Chem. Res. Toxicol.* **2004**, *17*, 1630−1637.
(15) EPA Fathead Minnow Acute Toxicity Database (EPAFHM). http://www.epa.gov/ncct/dsstox/index.html (accessed Nov 2006).
(16) Russom, C. L.; Bradbury, S. P.; Broderius, S. J.; Hammermeister, D. E.; Drummond, R. A. Predicting Modes of Toxic Action from Chemical Structure: Acute Toxicity in the Fathead Minnow (*Pimephales promelas*). *Environ. Toxicol. Chem.* **1997**, *16*, 948−967.
(17) Hamilton, M. A.; Russo, R. C.; Thurston, R. V. Trimmed Spearman−Karber Method for Estimating Median Lethal Concentrations in Toxicity Bioassays. *Environ. Sci. Technol.* **1977**, *11*, 714−719.
(18) Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. *Dragon 5.2*; Milano Chemometrics and QSAR Research Group: University of Milano-Bicocca, Milan, Italy, 2004.
(19) Brereton, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; John Wiley & Sons, Ltd.: West Sussex, U. K., 2003; pp 224−251.
(20) Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Morgan Kaufmann: San Francisco, CA, 2001; pp 279−363.

(21) Lee, J. Y.; Olafsson, S. In *Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques*; Triantaphyllou, E., Felici, G., Eds.; Springer: New York, 2006; Vol. 6, Chapter 10, pp 327−358.
(22) Schölkopf, B.; Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, U. K., 2002; pp 189−213.
(23) Lindström, A.; Pettersson, F.; Almqvist, F.; Berglund, A.; Kihlberg, J.; Linusson, A. Hierarchical PLS Modeling for Predicting the Binding of a Comprehensive Set of Structurally Diverse Protein−Ligand Complexes. *J. Chem. Inf. Model.* **2006**, *46*, 1154−1167.
(24) Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109−130.
(25) Little, H. J. How Has Molecular Pharmacology Contributed to Our Understanding of the Mechanism(s) of General Anesthesia? *Pharmacol. Ther.* **1996**, *69*, 37−58.
(26) Meyer, K. H. Contributions to the Theory of Narcosis. *Trans. Faraday Soc.* **1937**, *33*, 1062−1068.
(27) Paton, W. D. M. Unconventional Anesthetic Molecules. In *Molecular Mechanisms in General Anesthesia*; Halsey, M. J., Millar, R. A., Sutton, J. A., Eds.; Churchill Livingston: Edinburgh, Scotland, 1974; pp 48−64.
(28) Janoff, A. S.; Pringle, M. J.; Miller, K. W. Correlation of General Anesthetic Potency with Solubility in Membranes. *Biochim. Biophys. Acta* **1981**, *649*, 125−128.
(29) Mullins, M. J. Some Physical Mechanisms in Narcosis. *Chem. Rev.* **1954**, *54*, 289−323.
(30) Miller, K. W.; Paton, W. D. M.; Smith, R. A.; Smith, E. B. The Pressure Reversal of General Anesthesia and the Critical Volume Hypothesis. *Mol. Pharmacol.* **1973**, *9*, 131−149.
(31) Raines, D. E.; Korten, S. E.; Hill, W. A. G.; Miller, K. W. Anesthetic Cut-Off in Cycloalkanemethanols. *Anesthesiology* **1993**, *78*, 918−927.
(32) Franks, N. P.; Lieb, W. R. Mechanisms of General Anesthesia. *Environ. Health Perspect.* **1990**, *87*, 199−205.

CI600299J