

The development of 3D-QSAR study and recursive partitioning of heterocyclic quinone derivatives with antifungal activity

Su-Young Choi,^{a,†} Jae Hong Shin,^{b,†} Chung Kyu Ryu,^a Ky-Youb Nam,^b
Kyoung Tai No^c and Hea-Young Park Choo^{a,*}

^a*School of Pharmacy, Ewha Womans University, Seoul 120-750, Republic of Korea*

^b*Drug Discovery Division, Research Institute of Bioinformatics and Molecular Design, Seoul 120-749, Republic of Korea*

^c*Department of Biotechnology, Yonsei University, Seoul 120-749, Republic of Korea*

Received 20 July 2005; revised 5 October 2005; accepted 6 October 2005

Available online 2 November 2005

Abstract—It was reported that some 1,4-quinone derivatives such as 6-(*N*-arylamino)-7-chloro/6,7-bis[*S*-(aryl)thio]-5,8-quinolinedione and 6-arylthio-/5,6-arylamino-4,7-dioxobenzothiazoles have antifungal effects. To understand the structural basis for antifungal activity and guide in the design of more potent agents, we performed three-dimensional quantitative structure–activity relationship studies for a series of compounds using comparative molecular field analysis (CoMFA). The MIC values of 1,4-quinone derivatives on *Aspergillus niger* exhibited a strong correlation with steric and electrostatic factors of the 3D structure of molecules. The statistical results of the training set, cross-validated q^2 (0.683) and conventional r^2 (0.877) values, gave reliability to the prediction of inhibitory activity of a series of compounds. We also performed recursive partitioning (RP) analysis, used for the classification of molecules with activity using CART methods. Physicochemical, structural, and topological connectivity indices and E-state key descriptors were used for obtaining the decision tree models. The decision tree could classify the inhibitory activity of 1,4-quinone derivatives and its essential descriptors were S_{aaN}, Hbond donor, and Kappa-3.

© 2005 Elsevier Ltd. All rights reserved.

1. Introduction

Due to an increase of mycotic infections and frequent accounts of resistance there has been an interest in developing new antifungal agents with a novel mode of action. One of the potent antifungal agents with novel mode of action is UHDBT **1** (5-*n*-indecyl-6-hydroxy-4,7-dioxobenzothiazole) which blocks mitochondrial electron transport. Recently, it has been reported that a series of UHDBT **1** analogues with 1,4-quinone moiety such as 4,7-benzimidazolediones and 4,7-dioxobenzothiazoles showed potent antifungal activity against pathogenic fungi (Chart 1).^{1–5}

To understand the structural basis for the antifungal activity and to guide in the design of more potent compounds, we performed quantitative structure–activity

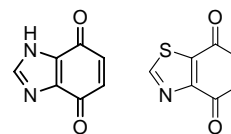


Chart 1. 1, 4,7-Dioxobenzimidazole and benzothiazole.

relationship studies for a series of compounds using Comparative molecular field analysis (CoMFA) and recursive partitioning.

CoMFA, developed by Cramer et al. is a simple extension of the standard structure–activity relationship correlating tables with mostly steric and electrostatic physicochemical parameters.⁶ The 3D QSAR methodology has been widely used in drug design.

Decision tree, a useful classification tool,^{7–9} can be used to qualitatively predict activities or activity classes in structure–activity relationship analysis or in focused library design.^{10–13} Recursive partitioning (RP) is defined as the division of compound sets into groups of higher and lower responses according to their descriptors. At

Keywords: Antifungal; CoMFA; Recursive partitioning; 4,7-Dioxobenzothiazole; 4,7-Dioxobenzimidazole.

* Corresponding author. Tel.: +82 2 3277 3042; fax: +82 2 3277 2851; e-mail: hypark@ewha.ac.kr

† These authors contributed equally to this work.

each decision point, molecules are split into two groups by a particular parameter. Each of the two subgroups split again by another parameter. This partitioning process is terminated when further splitting is impossible or a threshold value of the stopping rule is reached. The RP method overcomes the difficulties of handling non-linear relationships and strong interactions in the large SAR data by applying decision tree in which the statistically best chemical features are used to split the large data set into smaller and more homogeneous subsets. Furthermore, the tree display of the SAR results is clear, easy to understand, and often suggestive of new pharmacophore. Because of the speed and the 2D nature of this method, RP is especially good for large amounts of data that are difficult to sieve into usable divisions of classification.

We present here a report on the application of the CoMFA and the RP to analyze antifungal compounds using 3D molecular fields and 2D descriptors. To our knowledge, it is the first time RP approach is successfully compared with the CoMFA study.

2. Methods

2.1. Data set for analysis

The MIC values (micrograms per milliliter) of 74 compounds, 5/6-aryl-4,7-dioxobenzothiazoles, 2,5-disubstituted-6-arylamino-4,7-benzimidazolidiones, 6-arylthio-/6-arylamino-4,7-dioxobenzothiazoles, 6-(*N*-arylamino)-7-chloro-5,8-quinolinediones, and 6,7-bis-[*S*-(aryl)thio]-5,8-quinolindiones of Ryu et al., were used for this analysis.^{1–5} Their structures and activities are summarized in Table 1.

2.2. CoMFA

All molecular modeling and statistical analyses were performed using SYBYL 6.9 molecular modeling software (Tripos Inc.) on Silicon Graphics Origin 300 (IRIX 6.5). The 2D structure of each compound was built using SYBYL Build program with the default SYBYL settings. The 2D structure was converted to a 3D structure using Concord 4.0 program. The structural energy minimization was performed using the SYBYL quantum mechanical charge or Gasteiger–Hückel charge, with a 0.005 kcal/mol energy gradient convergence criterion. Low energy conformation was searched from grid search. The highest active compound **4-1** is used as a template molecule for superpositioning and all other quinone derivatives were superposed with a common substructure of the 2-amino-benzo-1,4-quinone group as shown in Figure 1.

Conventional CoMFA was performed with the QSAR option of SYBYL. The steric and electrostatic energies were calculated using sp³ carbon probe atoms with +1 charge. Maximum energy cutoff for steric and electrostatic energies was 30 kcal/mol. The CoMFA grid spacing was 2.0 Å in all three dimensions within the defined region. The partial least-squares (PLS) method was used

for fitting the 3D structural features and their biological activities in terms of $-\log$ MIC values. The optimum number of components in the final PLS model was determined by the q^2 value, obtained from the leave-one-out cross-validation technique.

2.3. Recursive partitioning

A simple recursive partitioning (RP) analysis was carried out by Cerius2 version 4.9 Accelrys Inc. and the decision tree was generated in the results of RP. The biological activity of compounds was measured in terms of their MIC₅₀ values for antifungal activity. It is necessary to classify the values of the activity into two classes, active and less active. The activity classes are defined as follows by the activity of MIC₅₀ values. The class 1 contains the 54 active compounds which have the MIC₅₀ values of lower than 12.5 µg/mL and the class 2 contains the 20 less active compounds which have the MIC₅₀ values of higher than 12.5 µg/mL or equal. To generate the RP model, we calculated hundreds of descriptors, which show of 2D physicochemical, structural, spatial, and topological property of the molecules,^{14,15} using Cerius2. The calculated descriptors are summarized in Table 2.^{16–18}

We applied the CART™ (classification and regression trees) method by using Cerius2 to generate the decision tree model. The conditions of the parameters were set to default; moderate pruning, equal weight classes, and Gini impurity score for split, etc. We performed the internal validation for the decision tree model through using leave-one-out cross-validation.

3. Results and discussion


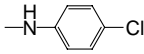

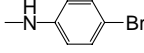

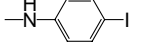

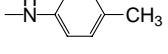

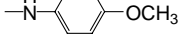

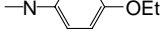

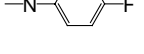

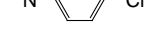

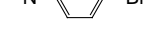

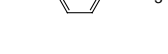

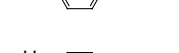

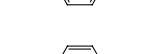
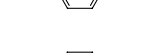
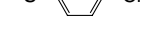
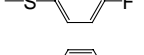
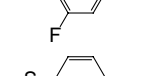
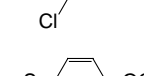
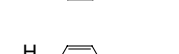
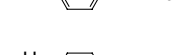
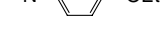
3.1. CoMFA

The statistical results of the CoMFA of training set composed of 55 compounds are summarized in Table 3. In this model, CoMFA was used as descriptors. A cross-validated value q^2 was obtained as a result of PLS analysis which served as a quantitative measure of the predictability of the CoMFA model. Good cross-validated q^2 (0.683) and conventional r^2 (0.877) values proved the correlation between the descriptors and each of their activities, and gave reliability to the prediction of the antifungal activities of the test set. The relative contributions of steric and electrostatic fields were 65.1% and 34.9%, respectively. The measured and estimated activities of the training set and test set are reported in Tables 4 and 5, respectively, and the plots of the predicted versus the actual activity values for the training set and the test set are shown in Figure 2. The great advantage of CoMFA is that the field effect on the target property can be viewed as a 3D coefficient contour plot. The coefficient contour plots are helpful to identify important regions where any change in steric and electrostatic fields may affect the biological activity and they may also help to identify the possible interaction sites. In Figures 3 and 4, the isocontour diagrams of the field contributions of different properties obtained from CoMFA are illustrated with an example ligand.

Table 1. The structure and activity of antifungal agents

Compound	X	R1	R2	R3	MIC (μg/mL)
1-1	S	H		H	3.2
1-2	S	H		H	3.2
1-3	S	H		H	1.6
1-4	S	H		H	12.5
1-5	S	H		H	3.2
1-6	S	H		H	3.2
1-7	S	CH ₃	H		12.5
1-8	S	CH ₃	H		3.2
1-9	S	CH ₃	H		6.3
1-10	S	CH ₃	H		12.5
1-11	S	CH ₃	H		1.6
1-12	S	CH ₃	H		6.3
1-13	S	CH ₃	H		6.3
1-14	S	CH ₃	H		12.5
1-15	S	CH ₃	H		12.5
1-16	S	CH ₃	H		12.5
1-17	S	CH ₃	H		12.5
2-1	N			H	100

Table 1 (continued)

Compound	X	R1	R2	R3	MIC (μg/mL)
2-2	N			H	3.2
2-3	N			H	1.6
2-4	N			H	100
2-5	N			H	6.3
2-6	N			H	50
2-7	N			H	50
2-8	N			Cl	25
2-9	N			Cl	6.3
2-10	N			Cl	6.3
2-11	N			Cl	25
2-12	N			Cl	25
2-13	N			Cl	25
3-1	S	CH ₃		CH ₃	25
3-2	S	CH ₃		CH ₃	100
3-3	S	CH ₃		CH ₃	12.5
3-4	S	CH ₃		CH ₃	50
3-5	S	CH ₃		CH ₃	100
3-6	S	CH ₃		OCH ₃	100
3-7	S	CH ₃		CH ₃	3.2
3-8	S	CH ₃		CH ₃	3.2

(continued on next page)

Table 1 (continued)

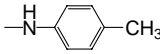
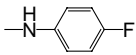
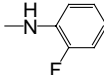
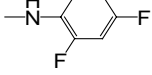
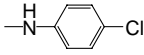
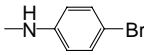
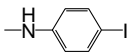
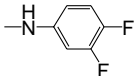
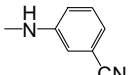
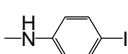
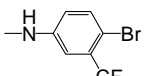
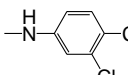
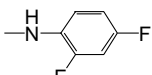
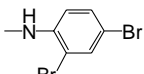
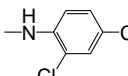
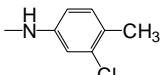
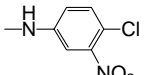
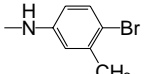
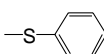
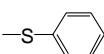
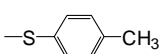
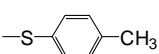
Compound	X	R1	R2	R3	MIC (μg/mL)
3-9	S	CH ₃		CH ₃	6.3
3-10	S	CH ₃		CH ₃	3.2
3-11	S	CH ₃		CH ₃	12.5
3-12	S	CH ₃		CH ₃	12.5
3-13	S	CH ₃		CH ₃	6.3
3-14	S	CH ₃		CH ₃	3.2
3-15	S	CH ₃		CH ₃	6.3
4-1	CH=CH	H		Cl	0.8
4-2	CH=CH	H		Cl	50
4-3	CH=CH	H		Cl	6.3
4-4	CH=CH	H		Cl	6.3
4-5	CH=CH	H		Cl	6.3
4-6	CH=CH	H		Cl	12.5
4-7	CH=CH	H		Cl	3.2
4-8	CH=CH	H		Cl	3.2
4-9	CH=CH	H		Cl	3.2
4-10	CH=CH	H		Cl	12.5
4-11	CH=CH	H		Cl	12.5
5-1	CH=CH	H			25
5-2	CH=CH	H			12.5

Table 1 (continued)

Compound	X	R1	R2	R3	MIC (μg/mL)
5-3	CH=CH	H			12.5
5-4	CH=CH	H			12.5
5-5	CH=CH	H			25
5-6	CH=CH	H			6.3
5-7	CH=CH	H			6.3
5-8	CH=CH	H			12.5
5-9	CH=CH	H			25
5-10	CH=CH	H			3.2
5-11	CH=CH	H			25
5-12	CH=CH	H			6.3
5-13	CH=CH	H			12.5
5-14	CH=CH	H			6.3
5-15	CH=CH	H			25
5-16		Cl			6.3
5-17		Cl			12.5
5-18					25

The electrostatic contour plots are shown in Figure 3. The blue contour defines regions where increasing positive charge will result in increasing the activity, whereas the red contour defines a region of space where increasing electron density is favorable. A predominant feature of the electrostatic plot is the presence of a blue contour locating X-position and linker of R₂ position. It could be reasonably presumed that there is a significant electrostatic interaction between the linker N–H of R₂ position instead of a sulfur atom and a possible receptor.

A red contour in the R₂-substituents indicates that electronegative groups at this position could help increase activity. This is reflected in certain compounds. For example, 3-7, 3-9, and 3-11 show that a change from a methoxy to a hydrogen atom at the R₂-para-substituents decreases the potency, which may be due to a suitable increase in the electronegative groups.

The steric contribution contour maps of CoMFA are plotted in Figure 4. The yellow contour depicts steric

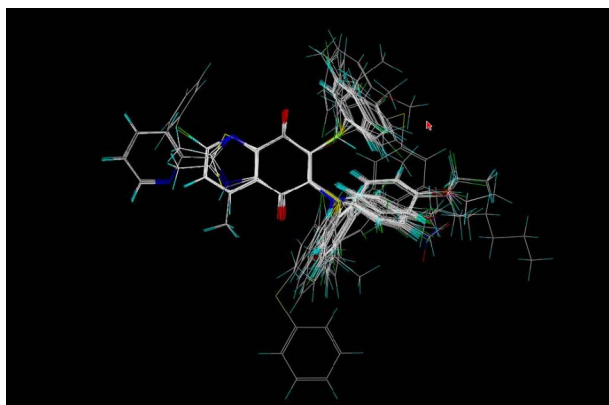


Figure 1. Superposition of the compounds in the training and test sets by using common substructure-based alignment.

limitation of the molecules at R_1 position. For instance, as the size of functional group increases at R_1 position, the antifungal activity will be diminished. Green contours obviously indicate that the presence of the R_2 -substituent on the heterocyclic quinone ring is favorable. The experimental biological activities are plotted versus the predicted activities in Figure 2.

3.2. Recursive partitioning

The statistical results are summarized in Tables 6 and 7. In class 1, 48 compounds among 54 compounds were correctly classified as class 1, and in class 2, 18 of 24 compounds were correctly classified as class 2. The enrichment factor of class 1 is 1.316 and class 2 is 2.775. The structure of decision tree consists of four terminal and three non-terminal nodes (see Fig. 5). The terminal nodes 2 and 3 are class 1 (active) and the terminal nodes 1 and 4 are class 2 (less active). The descriptors, S_{aaN} , Hbond donor, and Kappa-3, are used for construction of decision tree (Table 8). By convention, a true response to any given split follows the branch to the downside and a false response to any given split follows the branch to the upside. The first primary split is made on S_{aaN} (nitrogen atom connected with 2 aromatic bonds). The terminal node 1 contains 15 compounds with S_{aaN} values larger than 4.23104, which were as shown in class 2 (less active), and five compounds of them were misclassified. In the terminal node 1, 13 molecules can be distinguished by the existence of the pyridinyl ring at the R_1 position, but the two molecules (e.g., **5-15** and **5-18**) are deficient in it. Interestingly, the pyridine containing compounds (**2-1** to **2-13**) in the terminal node 1 are all classified as less active

Table 3. Summary of results from CoMFA

Statistical parameters	Results
n	55
q^2	0.683
r^2	0.877
Number of components	4
Standard error of estimation (SEE)	0.152
F	89.50
Steric (%)	65.1
Electrostatic (%)	34.9

compounds. In the terminal node 1, the actual activity of MIC_{50} was largely deviated from our classification. However, the CoMFA study also predicted them to be an inactive MIC range. These compounds might be in a different class of molecules or have a different mode of action, therefore requiring further study.

The second split occurred again on Hbond donor, which indicates a direct linkage of the R_2 or R_3 position of the NH group. The terminal node 2 (class 1) contains 37 compounds, the majority of 36 active compounds and only 1 compound of them (compound **4-2**) is misclassified. The molecules having one or more hydrogen bonding donors were placed on the terminal node 2. It means that hydrogen bonding donor coincides with the CoMFA contour maps as the larger one of the blue regions in Figure 3. It could be reasonably presumed with regard to electrostatic interactions between the R_2 or R_3 position of NH group and a possible receptor, which can increase the activity.

The last split is characterized by Kappa-3 and its value is by 4.5715. The Kappa-3 is one of the topological Kier's molecular shape indices (e.g., hexane is 5.333, 3-methylpentane is 3.000, and cyclohexane is 1.333). The terminal node 3 (class 1) contains 12 active compounds and 1 less active compound. The remaining 9 compounds are placed on the terminal node 4 (class 2), which contains eight less active compounds and 1 active compound. The terminal node 4 can be represented to the CoMFA contour maps of Figure 4, whose yellow surfaces describe the regions of space around the molecules. As the steric bulk depicted as yellow regions increases toward R_1 , the antifungal activity will be diminished.

It could be suggested that a modification for more electron-withdrawing group (methoxy or difluoro) of phenyl ring at R_2 and R_3 positions should lead to an increase of activity. On the other hand, the overall bulk of a molecule could influence a decrease in activity. In respect to

Table 2. Molecular descriptors used for recursive partitioning

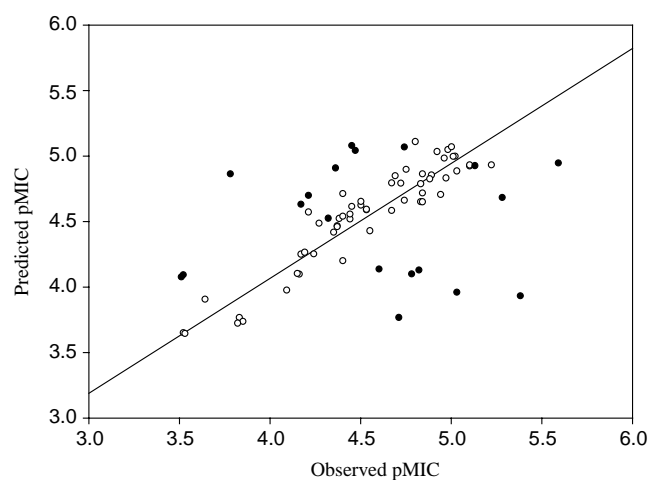
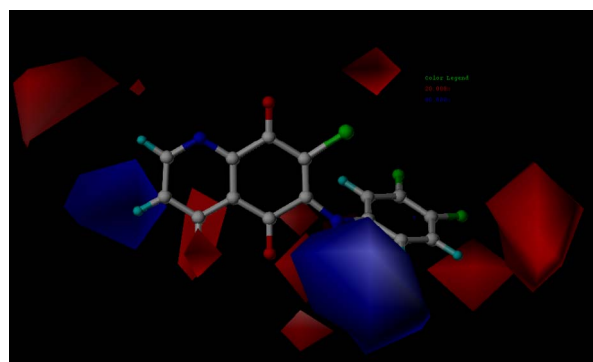
Descriptor class	Descriptors
Physicochemical	ADME_PSA_2D, ADME_Alog P98, ADME_solubility, Alog P, Alog P98
Structural	MW, Rotlbonds, Hbond acceptor, Hbond donor
Electrotopological (E-state key)	S_{sCH3} , S_{dsCH} , S_{aaCH} , S_{dssC} , S_{aasC} , S_{ssNH} , S_{aaN} , S_{dO} , S_{ssO} , S_{ssS} , S_{aaS} , $Sum_E_state_F_Cl_Br$
Topological	Kappa-1,2,3, Kappa-1,2,3-AM, Wiener, log Z, Zagreb SC-0,1,2,3_P, 3_C, CHI-0,1,2,3_P, 3_C, V-0, V-1, V-2, V-3_P
Jurs charged partial surface area	Jurs-SASA, PPSA-1,2,3, PNSA-1,2,3, DPSA-1,2,3, FPSA-1,2,3, FNSA-1,2,3, WPSA-1,2,3, WNSA-1,2,3, RPCG, RNCG, RPCS, RNCS, TPSA, TASA, RPSA, RASA

Table 4. Measured and estimated activities of 55 compounds in the training set

Compound	CoMFA	Observed (pMIC)	Predicted (pMIC)	Residual
1-01	84	4.90	4.84	0.06
1-02	84	4.93	5.02	−0.09
1-03	86	5.23	4.92	0.31
1-05	104	4.97	4.97	0.00
1-06	96	5.03	4.99	0.04
1-08	98	4.95	4.69	0.26
1-09	102	4.68	4.57	0.11
1-10	106	4.36	4.41	−0.05
1-12	106	4.73	4.78	−0.05
1-13	112	4.75	4.65	0.10
1-14	146	4.45	4.51	−0.06
1-15	100	4.39	4.51	−0.12
1-16	106	4.45	4.55	−0.09
1-17	116	4.38	4.46	−0.08
2-04	120	3.65	3.90	−0.25
2-06	126	3.84	3.76	0.09
2-07	136	3.86	3.73	0.14
2-08	126	4.17	4.09	0.08
2-11	134	4.18	4.24	−0.06
2-12	142	4.20	4.25	−0.05
2-13	134	4.16	4.09	0.07
3-01	124	4.10	3.96	0.14
3-03	114	4.41	4.19	0.22
3-04	120	3.83	3.71	0.12
3-05	118	3.53	3.64	−0.11
3-06	128	3.54	3.63	−0.09
3-07	126	4.99	5.04	−0.05
3-08	134	5.01	5.06	−0.05
3-09	122	4.68	4.78	−0.10
3-10	116	4.98	4.82	0.16
3-11	118	4.38	4.45	−0.07
3-12	122	4.41	4.53	−0.12
3-13	116	4.70	4.84	−0.14
3-14	124	5.11	4.91	0.20
3-15	120	4.76	4.89	−0.13
4-03	108	4.81	5.10	−0.29
4-04	122	4.84	4.64	0.20
4-06	100	4.41	4.70	−0.29
4-08	112	5.04	4.87	0.17
4-09	108	5.02	4.99	0.03
5-02	156	4.51	4.61	−0.10
5-03	150	4.51	4.64	−0.13
5-04	166	4.54	4.58	−0.04
5-05	144	4.25	4.24	0.01
5-06	144	4.85	4.64	0.21
5-07	150	4.85	4.85	0.00
5-09	140	4.22	4.56	−0.34
5-10	142	5.11	4.92	0.19
5-12	174	4.85	4.71	0.15
5-13	144	4.54	4.58	−0.04
5-14	164	4.84	4.78	0.06
5-15	148	4.28	4.48	−0.19
5-16	144	4.89	4.81	0.08
5-17	164	4.56	4.42	0.14
5-18	260	4.46	4.60	−0.14

Table 5. Measured and estimated activities of 19 compounds in the test set

Compound	CoMFA	Observed (pMIC)	Predicted (pMIC)	Residual
1-04	88	4.37	4.90	−0.53
1-07	98	4.33	4.51	−0.18
1-11	116	5.29	4.67	0.62
2-01	110	3.52	4.07	−0.55
2-02	114	5.04	3.95	1.09
2-03	118	5.39	3.92	1.47
2-05	122	4.72	3.76	0.96
2-09	128	4.79	4.09	0.70
2-10	130	4.83	4.12	0.71
3-02	118	3.53	4.08	−0.55
4-01	100	5.60	4.93	0.67
4-02	108	3.79	4.85	−1.06
4-05	106	4.75	5.06	−0.31
4-07	118	5.14	4.91	0.23
4-10	114	4.46	5.07	−0.61
4-11	114	4.48	5.03	−0.55
5-01	136	4.18	4.62	−0.44
5-08	140	4.61	4.13	0.48
5-11	128	4.22	4.69	−0.47

**Figure 2.** CoMFA predicted as experimental pMIC values. Open circles represent prediction for the training set; solid circles represent predictions for the test set.**Figure 3.** Electrostatic contour map from the CoMFA model. Compound 4-1 is shown inside field. Blue contour (80% contribution) encompass regions where an increase of positive charge will enhance affinity, whereas in red contoured areas (20% contribution) more negative charges are favorable for binding properties.

the S_{aaN} and Hbond donor parameters, they seemed to be in agreement with the contour maps and the decision tree. The 19 compounds of test set reserved from training were used to validate the performance of the obtained 3D-CoMFA model. The test compounds showed moderate over-prediction shown as in Figure 2, where the experimental biological activities are plotted

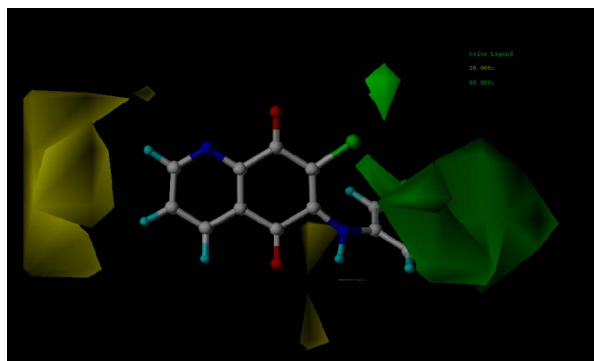


Figure 4. Steric contour map from the CoMFA model. Compound **4-1** is shown inside field. Sterically disfavored areas (contribution level of 80%) are represented by yellow surfaces. Sterically favored areas (contribution level 20%) are represented by green surfaces.

Table 6. Actual and predicted classification of the activities

Observed	Predicted	
	Class 1	Class 2
Class 1 (54)	48	6
Class 2 (20)	2	18
Total (74)	50	24

Table 7. Statistical results of recursive partitioning

Class	Number of compound	% ^a	Class % ObsCorrect ^b	Overall % PredCorrect ^c	Enrichment ^d
1	54	72	89	96	1.316
2	20	27	90	75	2.775

^a %: Percentage of compounds in each class.

^b Class % ObsCorrect is the so-called intraclass prediction. In this test, only the molecules in the corresponding class are being predicted. It provides information on false negatives as well as false positives, depending on which class you want to examine.

^c Overall % PredCorrect is the so-called overall prediction. In this test, all the molecules in the set are being predicted. This provides you with some information on the accuracy of the prediction when you predict the whole set with the model. Some classes are predicted more accurately than others. This may be a low number for actives since there could be a good proportion of false positives. This is generally acceptable as long as you get significant enrichment for the active bin.

^d The enrichment factor for a specific bin is the percentage of compounds correctly predicted to belong to that bin (Overall % PredCorrect) divided by the original percentage of compounds belonging to that bin (%).

versus the predicted activities. For this reason, the 3D-CoMFA model could tend to predict larger values of error in the limited test set. Although the 3D-CoMFA model could have a limitation of predictability, it still generated many useful interpretations.

The 3D-CoMFA and the RP analysis were consistent and gave information to guide the activity both quantitatively and qualitatively. The advantage of both the methods could also give a helpful information for priority of the next development in many antifungal active chemical derivatives.

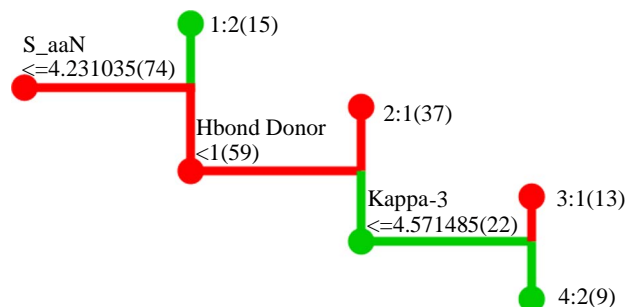


Figure 5. RP tree generated with moderate pruning; score splits using Gini impurity for prediction of antifungal activity classes. Those marked '1:2', '2:1', '3:1', and '4:2' correspond to terminal nodes 1–4 from the top to bottom and each terminal node corresponds to the value of 1 (active) or 2 (less active).

Table 8. Activity prediction by recursive partitioning

Com-pound	MIC (μg/mL)	Class	Predicted	S_aaN	H donor	Kappa-3
1-1	3.2	1	1	3.90547	1	2.65928
1-2	3.2	1	1	3.86748	1	2.88
1-3	1.6	1	1	3.85958	1	2.88
1-4	12.5	1	1	3.89962	1	2.88
1-5	3.2	1	1	3.921	1	3.5046
1-6	3.2	1	1	3.82806	1	4.34568
1-7	12.5	1	1	4.14182	1	2.88
1-8	3.2	1	1	4.09593	1	3.12245
1-9	6.3	1	1	4.13482	1	3.12245
1-10	12.5	1	1	4.15593	1	3.12245
1-11	1.6	1	1	4.16028	1	3.75309
1-12	6.3	1	1	4.03947	1	3.81944
1-13	6.3	1	1	4.05123	1	4.60118
1-14	12.5	1	1	4.20472	1	5.04167
1-15	12.5	1	1	4.13331	1	3.12245
1-16	12.5	1	1	4.15011	1	3.12245
1-17	12.5	1	1	4.17349	1	3.2
2-1	100	2	2	8.34819	2	3.98354
2-2	3.2	1	2	8.40334	2	3.98354
2-3	1.6	1	2	8.42264	2	3.98354
2-4	100	2	2	8.43158	2	3.98354
2-5	6.3	1	2	8.43328	2	3.98354
2-6	50	2	2	8.42405	2	4.22449
2-7	50	2	2	8.44044	2	4.60942
2-8	25	2	2	8.31402	2	3.93817
2-9	6.3	1	2	8.36917	2	3.93817
2-10	6.3	1	2	8.38847	2	3.93817
2-11	25	2	2	8.38988	2	4.16
2-12	25	2	2	8.40627	2	4.54179
2-13	25	2	2	8.39911	2	3.93817
3-1	25	2	2	4.19086	0	3.06238
3-2	100	2	2	4.17341	0	3.06238
3-3	12.5	1	2	4.14127	0	3.06238
3-4	50	2	2	4.08475	0	3.16535
3-5	100	2	2	4.17075	0	2.93345
3-6	100	2	2	4.15319	0	3.52
3-7	3.2	1	1	4.16689	1	3.29861
3-8	3.2	1	1	4.17628	1	3.66514
3-9	6.3	1	1	4.17234	1	3.06238
3-10	3.2	1	1	4.12276	1	3.06238
3-11	12.5	1	1	4.10423	1	2.93345
3-12	12.5	1	1	4.06624	1	3.16535
3-13	6.3	1	1	4.1549	1	3.06238
3-14	3.2	1	1	4.17135	1	3.06238

Table 8 (continued)

Compound	MIC ($\mu\text{g/mL}$)	Class	Predicted	S_aaN	H donor	Kappa-3
3-15	6.3	1	1	4.16614	1	3.06238
4-1	0.8	1	1	3.84196	1	3.29861
4-2	50	2	1	3.91016	1	3.44047
4-3	6.3	1	1	3.93644	1	3.2
4-4	6.3	1	1	3.80805	1	3.98354
4-5	6.3	1	1	3.91299	1	3.29861
4-6	12.5	1	1	3.83133	1	3.29861
4-7	3.2	1	1	3.93953	1	3.29861
4-8	3.2	1	1	3.91148	1	3.29861
4-9	3.2	1	1	3.93044	1	3.29861
4-10	12.5	1	1	3.86799	1	3.75888
4-11	12.5	1	1	3.94535	1	3.29861
5-1	25	2	2	4.15692	0	4.3795
5-2	12.5	1	1	4.18263	0	4.85493
5-3	12.5	1	1	4.18859	0	4.85493
5-4	12.5	1	1	4.2143	0	5.01018
5-5	25	2	2	4.13785	0	4.54179
5-6	6.3	1	1	4.14141	0	4.85493
5-7	6.3	1	1	4.14407	0	4.85493
5-8	12.5	1	1	4.12857	0	5.01018
5-9	25	2	2	4.02907	0	4.54179
5-10	3.2	1	1	4.05451	0	4.85493
5-11	25	2	1	4.07304	0	4.85493
5-12	6.3	1	1	3.9452	0	5.01018
5-13	12.5	1	1	4.20697	0	5.33333
5-14	6.3	1	1	4.17044	0	5.33333
5-15	25	2	2	4.24777	0	5.08636
5-16	6.3	1	1	4.16478	0	5.16797
5-17	12.5	1	1	4.20334	0	5.16797
5-18	25	2	2	4.77635	0	9.522

4. Conclusion

The CoMFA method has been successfully applied to a series of 1,4-quinone derivatives with antifungal activity. The resulting QSAR model provided a significant correlation between steric and electrostatic fields with MIC values. The RP analysis showed that the number of nitrogen atoms at aromatic ring, hydrogen bonding donors, and Kappa-3 were essential features to classify the antifungal activity. The study of 3D-QSAR is expected to give a better understanding of relationships between the chemical structure and biological activity. Although the 3D-QSAR provides a possibility to intuitively understand the relationship, the derivation of 3D-QSAR has proven to be much more difficult and time consuming than 2D-QSAR. The major difficulties in deriving the 3D-QSAR model are to determine the active conformations and to find the correct binding modes of the molecules in a practical data set, which are almost unknown. From the results of this work, decision tree model and CoMFA results were proven to be coincided by the reasonable SARs, as the descriptors of RP could be well represented in the contour maps of CoMFA. We can use the RP analysis to classify the given activity classes by grouping the active compounds. Also, the RP model is probably another useful approach to select

the structural and chemical features, which can be designed to directly affect the active class (e.g., class 1) by understanding the descriptors. The RP method overcomes the difficulties of handling non-linear relationships and strong interactions in the large SAR data by applying a decision tree, where the statistically best chemical features are selected and used to classify the larger data set into smaller and more homogeneous subsets. Consequently, the decision tree and 3D-QSAR results introduced an easy way to understand the structure–activity relationship and successfully suggested new pharmacophores.

Acknowledgments

This work was supported by the grant from Research Institute of Pharmaceutical Science in Ewha Womans University and Vision 21 Program from Korea Institute of Science and Technology.

References and notes

- Ryu, C. K.; Kang, H. Y.; Yi, Y. J.; Shin, K. H.; Lee, B. H. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1589.
- Ryu, C. K.; Song, E. H.; Shim, J. Y.; You, H. J.; Choi, K. U.; Choi, I. K.; Lee, E. Y.; Chae, M. J. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 17.
- Ryu, C. K.; Choi, K. U.; Shim, J. Y.; You, H. J.; Choi, I. H.; Chae, M. J. *Bioorg. Med. Chem.* **2003**, *11*, 4003.
- Ryu, C. K.; Kim, H. J. *Arch. Pharm. Res.* **1994**, *17*, 139.
- Ryu, C. K.; Sun, Y. J.; Shim, J. Y.; You, H. J.; Choi, K. U.; Lee, H. *Arch. Pharm. Res.* **2002**, *25*, 795.
- Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
- Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. In *Classification and Regression Trees*; Wadsworth International Group: Belmont, CA, 1984.
- Chou, P. A. *IEEE Trans. Pattern Anal. Mach. Learn.* **1991**, *13*, 340.
- Bose, S. *Comput. Stat. Data Anal.* **2003**, *42*, 685.
- Hawkins, D. M.; Young, S. S.; Rusinko, A., III. *Quant. Struct. Act. Relat.* **1997**, *16*, 296.
- Chen, X.; Rusinko, A., III; Young, S. S. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1054.
- Michiel van Rhee, A.; Stocker, J.; Printzenhoff, D.; Creech, C.; Wagoner, P. K.; Spear, K. L. *J. Comb. Chem.* **2001**, *3*, 267.
- Rao, S. N.; Stockfisch, T. P. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1614.
- Cerius2. Descriptors + version 4.6, Accelrys, Inc., San Diego, CA, 92121, 2001.
- Brown, R. D.; Hassan, M.; Waldman, M. J. *Mol. Graph. Model.* **2000**, *18*, 427.
- Stanton, D. T.; Jurs, P. C. *Anal. Chem.* **1990**, *62*, 2323.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Wiley: New York, 1986.