

Modeling the Ligand–Receptor Interaction for a Series of Inhibitors of the Capsid Protein of Enterovirus 71 Using Several Three-Dimensional Quantitative Structure–Activity Relationship Techniques

Yi-Yu Ke and Thy-Hou Lin*

Institute of Molecular Medicine & Department of Life Science, National Tsing Hua, University, Hsinchu, 30043, Taiwan

Received November 23, 2005

The structure of enterovirus 71 (EV71) capsid protein VP1 has been constructed by using homology modeling and molecular dynamics simulation techniques. The ligand structures were a series of EV71 VP1 inhibitors synthesized by Shia et al. in 2002 and Chern et al. in 2004. The training set was selected by the VOLSURF4.1/PCA program and the IC₅₀ values varied from 0.06 to 10.83 μM . Then, the training set was analyzed by the following three-dimensional quantitative structure–activity relationship techniques: CoMFA, CoMSIA, CATALYST4.9, and VOLSURF4.1/PCA. The model generated by a two-stage flexible docking procedure and without any structural alignment has far more significant statistics. Highly accurate activities for the test sets were then predicted by the top hypothesis of the CATALYST program and were compared with those predicted by CoMFA, CoMSIA, and VOLSURF. These studies identified some important clues for searching or making more potent inhibitors against the EV71 infection.

Introduction

There has been a serious outbreak and neurological complications due to EV71 infection in Southeast Asia, especially in Malaysia and Taiwan. A large-scale epidemic of hand-foot-and-mouth disease (HFMD) occurred in Taiwan in 1998 in which more than 80 children died of shock syndrome due to pulmonary edema/hemorrhage.¹ Since the 1998 outbreak, EV71 has been isolated throughout the whole island all year long and many fatal cases have been reported. This prompts the need to search for and develop more effective anti-EV71 agents. As a member of the family *Picornaviridae*, the genome of EV71 consists of a single stranded RNA of positive polarity encapsulated inside a capsid made by 60 copies of four coat proteins namely, VP1, VP2, VP3, and VP4. Most of the neutralization sites are densely clustered on VP1, and variations within capsid proteins VP1–VP3 are responsible for the antigenic diversity among the enteroviruses.²

The structures of all the three proteins VP1–VP3 are determined to be the same “jelly roll”-type antiparallel β -barrel structure. This structure is composed of eight β -strands arranged in two sheets, each being made with four strands: strands 1, 8, 3, and 6 form the first sheet and strands 2, 7, 4, and 5 the second one.³ Crystallographic, biochemical, and immunological data have together identified a depression within the β -barrel of VP1 that is believed to be the attachment site of piconaviruses.⁴ The structure of the attachment site is important and is regarded as the prime target for antiviral drug development that may stop the attachment of virus with the host cell. In the rhinoviruses (common cold viruses), this is particularly deep and is called a “canyon”.⁵ The canyon lies within the structure of the β -barrel. The interior pocket near the base of the canyon has been the target for several antiviral compounds developed against infection by rhinoviruses and enteroviruses.^{6,7} For example, pleconaril has been shown to bind with the pocket of capsid protein and prevent the attachment by some rhinoviruses and enteroviruses to cells and uncoating of viral RNA, thus interrupting the

infection cycle.^{8–13} This potential drug candidate is currently undergoing clinical trials, and its successful phase III results for the treatment of viral respiratory infection (VIR),⁹ often referred to as the common cold, were announced by Viropharma, Inc. However, pleconaril was found to be unable to neutralize the cytopathic effect (CPE) of cultured cells induced by EV71 isolated from the 1998 outbreak in Taiwan.¹⁴ On the basis of the skeletons of pleconaril and its related molecules, the WIN compounds,^{8,15} Shia et al.¹⁴ have synthesized a novel class of imidazolidinones with significant anti-EV71 activity, and the measured IC₅₀ of these compounds was 0.04–12.04 μM .

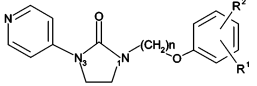
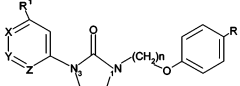
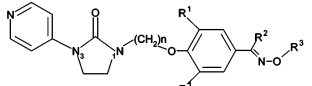







In this report, we have built a protein model of the EV71 VP1 protein by using the InsightII/Homology program¹⁶ and refined the model by some molecular dynamics (MD) simulation steps. A series of EV71 VP1 inhibitors was docked into the model protein active site. Several 3D-QSAR techniques including comparative molecular field analysis (CoMFA),¹⁷ comparative molecular similarity indices (CoMSIA),¹⁸ hypothesis generation (CATALYST),¹⁹ and VOLSURF²⁰ were then applied on a training set. The conformations of the corresponding inhibitors were generated by DOCK 4.0²¹ and refined by the CRID22A/GLUE1.0²² programs. The prediction ability of these 3D-QSAR models was tested on two test sets synthesized by Shia et al. in 2002¹⁴ and Chern et al. in 2004.²³ Mapping of the CoMFA and CoMSIA contours and the CATALYST pharmacophore features onto the structures of the two most potent inhibitors of the training and test sets was also performed and compared with the surfaces of the active site. These mapping results were consistent with each other and could be used as a clue for searching for more potent inhibitors of EV71 VP1 protein to act against EV71 infection.

Material and Methods

Homology Modeling Structure. The VP1 protein sequence of EV71(2086) genotype C 297 amino acid in length was chosen as the target sequence to be modeled.²⁴ There were three such protein structures searched and used as structure templates, namely, the poliovirus type 2 VP1 (PDB ID is 1EAH) complexed with antiviral agent Sch48973,²⁵ the poliovirus type 3 VP1 (PDB ID is 1PIV) complexed with disoxaril (WIN51711),²⁶ and the coxsackievirus

* Corresponding author. Fax: +886–3–571–5934. E-mail: thlin@life.nthu.edu.tw.

Table 1. EV71 VP1 Inhibitors Synthesized by Shia et al. in 2002 and Chern et al. in 2004

																		
compound	n	R ¹	R ²	IC ₅₀ (μM)	compound	n	X	Y	Z	R	R ¹	IC ₅₀ (μM)	compound	n	R ¹	R ²	R ³	IC ₅₀ (μM)
1 (20) ^c	5	H		0.04±	21 (42)	7	C	N	C	Cl	H	1.43±0.025	31 (8b)	5	H	H	C ₂ H ₅	0.001±0.001
2 ^d (21)	5	H		0.06±	22 (45)	6	C	N	C	Br	H	2.07±0.573	32 (8a)	5	H	H	CH ₃	0.005±0.001
3 (27)	6	H		0.30±	23 (43)	6	C	N	C	CF ₃	H	2.10±0.410	33 (11c)	6	H	H	C ₂ H ₅	0.010±0.004
4 (24)	5	H		0.33±	24 (41)	6	C	N	C	Cl	H	2.52±0.175	34 (8c)	5	H	H	n-C ₃ H ₇	0.021±0.003
5 (1)	6	H	4'-Br	0.35±	25 (40)	5	C	N	C	Cl	H	4.65±0.170	35 (11d)	7	H	H	C ₂ H ₅	0.025±0.001
6 (11)	5	H	4'-Br	0.50±	26 (44)	6	C	N	C	F	H	9.33±0.192	36 (8d)	5	H	H	n-C ₄ H ₉	0.079±0.019
7 (12)	7	H	4'-Br	0.56±	27 (54)	6	C	C	C	Br	H	10.60±	37 (11b)	4	H	H	C ₂ H ₅	0.24±0.01
8 (8)	6	H	4'-CF ₃	0.58±	28 (53)	5	C	C	C	Br	H	10.66±	38 (8i)	5	CH ₃	H	C ₂ H ₅	0.36±0.08
9 (28)	6	H		0.58±	29 (55)	6	C	C	C	Cl	H	10.83±	39 (8g)	5	H	n-C ₃ H ₇	C ₂ H ₅	0.65±0.03
10 (26)	5	H		0.58±	30 (56)	6	C	C	C	CF ₃	H	12.04±	40 (8h)	5	CH ₃	H	CH ₃	0.80±0.11
11 (10)	7	H	4'-CF ₃	0.61±									41 (8f)	5	H	C ₂ H ₅	C ₂ H ₅	1.08±0.17
12 (15)	7	H	4'-Cl	0.66±									42 (8e)	5	H	CH ₃	C ₂ H ₅	16.85±0.65
13 (13)	5	H	4'-Cl	1.35±									43 (11a)	3	H	H	C ₂ H ₅	21.35±1.57
14 (9)	5	H	4'-CF ₃	1.35±									44 (8j)	5	CH ₃	H	n-C ₃ H ₇	>25
15 (14)	6	H	4'-Cl	1.50±														
16 (25)	5	H		2.66±														
17 (16)	6	H	4'-F	3.01±														
18 (19)	6	H	4'-SCH ₃	4.96±														
19 (18)	6	H	4'-OCH ₃	5.43±														
20 (17)	6	H	4'-CH ₃	6.36±														

^a Inhibitors with antiviral activity against EV71 (2086) genotype C and synthesized in 2002. ^b Inhibitors with antiviral activity against EV71 (4643) genotype C and synthesized in 2004. ^c The parenthetical number was the original numbering in Shia et al. and Chern et al. ^d The bold numbers signify the training set compounds for the 3D-QSAR.

A9 VP1 (PDB ID is 1D4M) complexed with disoxaril (WIN51711).¹⁵ The amino acid sequence identity between the target sequence and sequences of templates 1EAH, 1PIV, and 1D4M were 38, 37, and 36%, respectively. To construct a protein model for the target sequence, we used the InsightII/Homology programs¹⁶ implemented on a Silicon Graphics computer. Sequences of these three templates were aligned against the target sequence to find regions where structures of these proteins were most matched. The matched sequences were taken as the structures of the regions for the target sequence (see Supporting Information). Loops and missing fragments of the target sequence were generated by InsightII/Modeler. A series of energy minimization steps using the steepest descent method with the Amber force field version 2.3²⁷ was performed, and some residues were restrained at their initial positions in order to relax loops and bad contacts. To further refine the modeling structure, a ligand engulfed in the active site was conducted using SYBYL 7.0²⁸ and InsightII/Discover and Discover 3 MD simulation programs. The ligand (compound **1** in Table 1) engulfed protein model was then minimized by 20 000 steps of conjugated gradient method by SYBYL 7.0. A distance dependent dielectric constant was set to 30 and the Amber7 FF99 force field²⁹ was employed. Next, the ligand-engulfed protein model was energy minimized for 500 steps by InsightII programs with layers of waters of total thickness 5 Å for the whole protein, and the default Amber version 2.3 force field was employed. The coordinates of backbone C α atoms were constrained during the minimization. Finally, after the short energy minimization, the ligand engulfed protein model was subjected to 10 000 fs of MD simulation by releasing all the constraints imposed and soaking into several layers of waters of total thickness 10 Å.

Generation of Ligand Structures. The EV71 VP1 inhibitors used in this study were synthesized by Shia et al.¹⁴ (the 2002 set) and Chern et al.²³ (the 2004 set) (Table 1). All these compounds were generated theoretically using SYBYL 7.0 and docked into the EV71 VP1 protein model with a two-stage flexible docking procedure by DOCK 4.0²¹ and GRID22A/GLUE.²² Each com-

pound was docked rigidly into the protein model using DOCK 4.02 first and scored by the energy scoring function. Then, the docked conformation of each ligand was flexibly docked back into the protein model and scored by the chemical scoring function of DOCK 4.02. GRID22A/GLUE1.0 was used as a second filter to refine each conformation. The GLUE program fits ligands into a set of GRID maps of a target structure and uses standard probes to compute MIFs (molecular interaction fields) on protein cavity. The standard probes chosen were H, OH2, DRY, N1, N+, O, O:, and O1. A box of approximately 12 Å in dimension and centered at the ligand in the active site with a grid spacing of 1 Å was used in the calculations. The output conformations were used in the following 3D-QSAR computations.

Construction of CoMFA and CoMSIA Models. The VOL-SURF4.1/PCA²⁰ program was used to select the training set compounds out of the 30 compounds of the 2002 set.¹⁴ These compounds were clustered into five different groups and selected by the largest minimum distance (LMD) method. Twenty compounds with maximum LMD computed were selected from each group and pooled together as the training set (Figure 1). To build the first 3D-QSAR model (model 1), each ligand structure of the training set was subsequently aligned against the structure of compound **1** treated as the common structural template. The second 3D-QSAR model (model 2) was built by using the two-stage docking procedure with the same alignment rules as model 1. Model 3 used the same docking results as model 2 but no structural alignment was employed.

The CoMFA analysis¹⁷ was conducted using SYBYL 7.0 with a regularly spaced grid of 2.0 Å. The lattice was extended to 4 Å units beyond the van der Waals volume of each molecule in the X, Y, and Z directions. A sp³ carbon atom of radius 1.52 Å and charge +1.0 was used as a probe to calculate both the steric (Lennard-Jones 6–12 potential) and electrostatic (Coulombic potential) fields. The truncation for both the steric and electrostatic contributions computed was set as 15 and 30 kcal/mol. The same lattice that was used for CoMFA was used for CoMSIA. However, a different

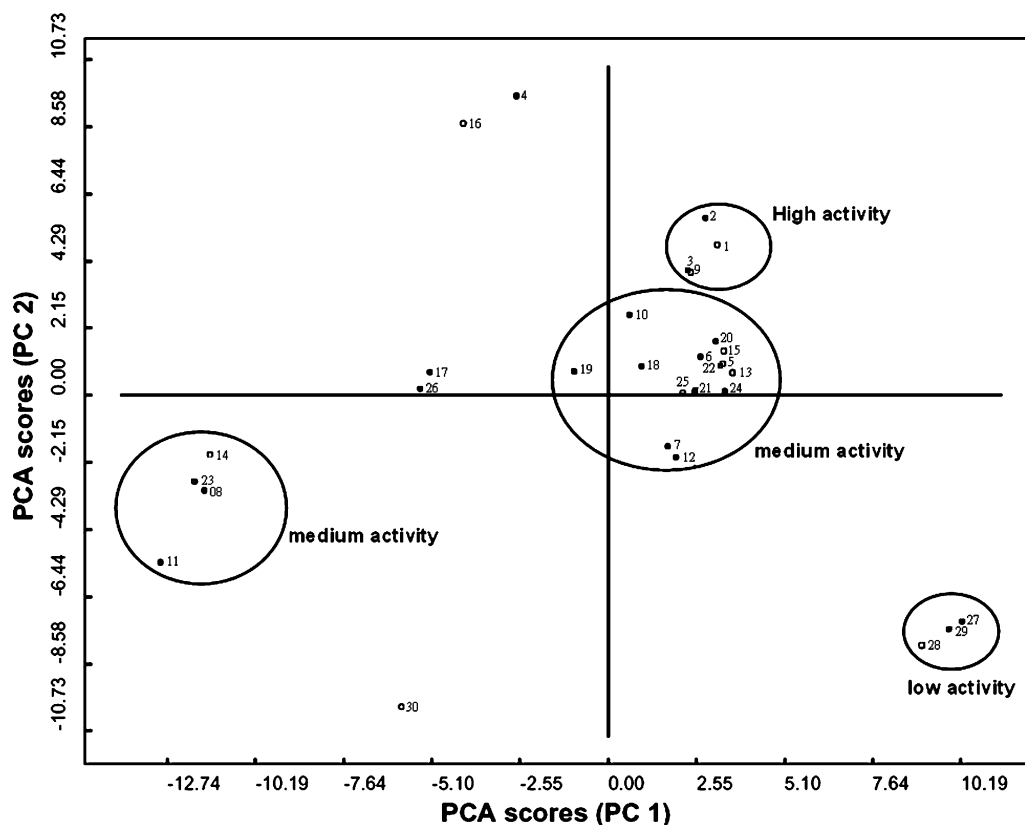


Figure 1. Results of the PCA analysis performed on the 30 EV71 VP1 inhibitors synthesized by Shia et al. in 2002. The black squares were selected as the training sets while the hollow squares were selected as the test 1 set.

sp^3 carbon atom of radius 1.0 Å and charge +1.0 was used as the probe to compute the CoMSIA similarity index. The attenuation factor R was set as 0.3. To validate the 3D-QSAR model, a partial least-squares (PLS)³⁰ linear regression was used to correlate the activities (pIC_{50}) with CoMFA and CoMSIA values. The optimum number (N) of PLS components corresponding to the smallest standard error estimate (SEE) of prediction was determined by the leave-one-out (LOO) cross-validation procedure. To speed up the analysis and reduce noise, column filtering was set as 2.0 kcal/mol so that only those steric and electrostatic energies with values greater than 2.0 kcal/mol were considered in the PLS analysis.³¹ The optimal N obtained was used in a non-cross-validation PLS analysis to get the model parameters such as correlation coefficient (r^2), SEE, and F value. Except the LOO cross-validation, a second cross-validation (leave-five-out) in which 80% of the compounds were randomly selected to build a model for predicting the activities for the 20% compounds left was also conducted. This analysis was repeated 100 times and the mean and standard deviation of q_{cv}^2 values were recorded.³² To further assess the statistical confidence limits of the analyses, a PLS analysis using 100 bootstrap groups with the optimum number of components chosen was performed. Moreover, the extent of chance correlation was tested by randomly scrambling the activities of the training set compounds. The predictive ability of each model was tested using two test sets namely, test 1 and test 2, which were a set of 10 compounds extracted from the 2002 set and the total 14 compounds of the 2004 set,^{14,23} respectively.

Construction of Pharmacophores. The same training set used in CoMFA and CoMSIA was utilized for constructing some pharmacophore models by CATALYST4.9.¹⁹ The hypothesis generation was performed by the HypoGen module automatically with an activity-based alignment derived from a collection of model 3 conformations. In parameter setting, the Uncert value was usually defaulted to 3. However, a value of 1.5 was chosen in our case because of the narrow range of activities (IC_{50} values of 0.04–12.04 μ M). The parameter of IdealHBondGeomOnly was set as 1, since only three samples were used for the hydrogen torsion angle

(one position on OH and two lone pair positions on carbonyl O). Two chemical functions predefined in the CATALYST Feature Dictionary, namely, hydrogen bond acceptor (HBA) and hydrophobic, were employed and HypoRefine was also used to automatically generate the exclusion volumes in the simulated annealing optimization process. The quality of the pharmacophore hypothesis constructed was evaluated by the cost functions calculated,³³ which contain the total cost, fixed cost, null cost, configuration cost, and error cost. As suggested by the CATALYST program, the difference between total and null cost of the generated hypothesis should be as large as possible. A difference of 40–60 bits may indicate that it has a 75–90% chance of representing a true correlation in the data set used. The total cost of any hypothesis should be toward the value of fixed cost. The CATALYST/CatScramble module was performed to further assess the statistical significance of the pharmacophore hypothesis generated. This validation technique was based on Fischer's randomization test, and the activities of the training set molecules were randomly reassigned as some new spreadsheets. The number of spreadsheets needed to achieve a 98% confidence level is 49. In our studies, 19 spreadsheets were created to achieve a confidence level of 95%.

Modeling with VOLSURF. The same training set analyzed by CoMFA, CoMSIA, and CATALYST was also subjected to the PLS analysis via the VOLSURF4.1 program.²⁰ The VOLSURF descriptors, namely, hydrophilic, hydrophobic, carbonyl oxygen atom, and amide NH group, were chosen only for model 3 conformations. The QSAR models constructed by the program were cross-validated with 0.5 Å grid resolution by both the leave-one-out and leave-five-out procedures to obtain the cross-validated r^2 , or q_{LOO}^2 and q_{cv}^2 values, respectively.

Results and Discussion

The EV71 VP1 protein model was built using a homology modeling technique for the EV71 (2086) genotype C sequence. The template structures searched from the NCBI database were 1EAH, 1PIV, and 1D4M and each of these carries an antiviral

agent in the crystallized structure. This would be advantageous to build the protein model, since the possible conflict caused by docking molecules into the narrow hydrophobic canyon may be avoided. The first validation for the protein model built by the InsightII/Modeler program was conducted by feeding the predicted coordinates into the ERRAT³⁴ protein verification server. Regions for residues between 43–55, 156–162, and 251–280 were found to be most erroneous and were recognized as the loop regions by a subsequent secondary structure prediction. The protein model was refined by several steps of MD simulation, and the refined structure was fed into the server again. The overall quality factor estimated for the primary refinement on the protein model was 83.07. To proceed with further refinement, the structure of the most active ligand, compound **1** of the 2002 set (Table 1), was docked into the protein model and refined together by the InsightII simulation programs. All the ligand structures including both the 2002 and 2004 sets (Table 1) were docked rigidly and then softly into the active site of the protein model by the DOCK 4.02 program. Furthermore, the conformation of each docked ligand structure was individually refined by GRID22A/GLUE1.0. The effectiveness of these refinements for all the ligand structures was assessed by a linear regression of the ranks given by the chemical scoring system of DOCK 4.02 or the MIF system of GLUE onto the corresponding pIC₅₀ values of each ligand (see the Supporting Information). The regression coefficients r^2 of the docked score given by DOCK 4.02 for the 2002 and 2004 set were 0.61 and 0.66, while those of the score docked and refined by the GRID22A/GLUE1.0 program were 0.83 and 0.94, respectively. This shows that the docked conformation of each ligand from DOCK 4.02 was successfully refined by GRID22A/GLUE1.0.

To proceed with 3D-QSAR analyses, VOLSURF4.1/PCA was used to select 20 compounds (**2–4**, **6–8**, **10–12**, **17–24**, **26**, **27**, and **29**) out of the 30 compounds from the 2002 set as the training set (Table 1). This analysis generated a plot where PC-2 (the second principal components) were plotted against PC-1 (the first principal components), and the compounds were clustered into five different groups in the plot (Figure 1). By using the LMD analysis on each cluster, the maximum values for some compounds in each cluster were then pooled together as the training set (highlighted in Table 1). The compounds left from the 2002 set was treated as the test 1 set, while those of the 2004 set were entirely treated as the test 2 set. Note that the IC₅₀ values of the 2004 set were measured against VP1 of EV71 strain 4643 and those of 2002 set were measured against VP1 of EV71 strain 2086. No difference in protein sequence of VP1 between the two strains was detected, and the measuring methods for IC₅₀ were the same for each.^{14,23}

There were three 3D-QSAR models, namely, model 1, model 2, and model 3, constructed as the training set for the CoMFA and CoMSIA programs. The superposition of all the conformations for each model is presented in the Supporting Information. Apparently, structures of model 1 were better superposed on each other than those of model 2 or model 3. To select the best CoMFA and CoMSIA results, the SEE and F values were used as criteria for accompanying the q_{cv}^2 value computed. In general, larger F -value means fewer explanatory variables and more target properties obtained for a model, which implies that the model is more statistically significant. Therefore, a good model was considered as having larger q_{cv}^2 and F and smaller SEE values. As demonstrated by the better q_{cv}^2 , F , and SEE values obtained, the CoMFA result from model 3 appears to be superior to that from either model 1 or model 2 (Table 2). The CoMSIA

Table 2. Summary of the CoMFA and Stepwise CoMSIA Statistics for the Training Set of EV71 VP1 Inhibitors

model	parameters	CoMSIA				
		CoMFA: S + E ^a	S	S + E	S + E + A	S + E + H + A
1	q_{cv}^{2b}	0.521	0.369	0.504	0.435	0.359
	SEE	0.346	0.349	0.251	0.301	0.276
	F -values	21.225	14.210	32.556	20.943	25.829
2	q_{cv}^2	0.465	0.536	0.359	0.339	0.291
	SEE	0.172	0.227	0.320	0.288	0.337
	F -values	56.617	20.924	26.331	23.328	22.940
3	q_{L00}^2	0.813	0.896	0.859	0.852	0.607
	components	5	5	6	6	2
	q_{cv}^2	0.721	0.876	0.775	0.709	0.530
	r_{conv}^{2c}	0.991	0.969	0.990	0.989	0.843
	SEE	0.072	0.126	0.072	0.077	0.257
	F -values	269.717	87.232	225.495	197.534	45.695
	q_{rs}^{2d}	0.996	0.984	0.995	0.994	0.872
SD _{bs} ^e	0.004	0.083	0.048	0.054	0.218	

^a S, steric; E, electrostatic; H, hydrophobic; A, H-bond acceptor. ^b Cross-validation by leave-five-out and optimum number of components chosen and averaged from 100 runs. ^c Conventional r^2 values. ^d Results from 100 runs of bootstrapped analyses. ^e Standard error of estimate from 100 runs of bootstrapped analyses.

was conducted in a stepwise manner by choosing field indexes one-by-one. There were four different field indexes (steric, denoted as S; electrostatic, denoted as E; hydrophobic, denoted as H; and H-bond acceptor, denoted as A) being chosen for the stepwise CoMSIA and the best results of model 3 were presented in Table 2. With larger q_{cv}^2 and F and smaller SEE and bootstrapped values analyzed, the CoMSIA results of stepwise S, S + E, and S + E + A fields of model 3 were obviously more statistically significant than those of any other stepwise fields of either model 1 or model 2 (Table 2). However, we consider the best CoMSIA result as that obtained for the stepwise S + E fields of model 3 ($q_{cv}^2 = 0.775$, $F = 225.495$, SEE = 0.072), because the corresponding F and SEE values computed were among the largest and smallest of all the three aforementioned stepwise fields used (Table 2). Therefore, the CoMFA and stepwise CoMSIA results of model 3 not only agreed with each other statistically but also with respect to the components of interaction fields (S + E) employed.

The contour of the best CoMFA model was mapped onto the structure of compound **2**, the most potent compound of the training set, as presented in Figure 2. As expected, the favorable regions for electrostatic interaction represented by red (favor negative charge) and blue (favor positive charge) contours were mapped around the pyridine–imidazolidine part, while those for steric interaction represented with yellow (disfavor bulk group) and green (favor bulk group) contours were mapped around the biphenyl portion of compound **2**. The CoMFA contours were projected further into the active site of the EV71 VP1 protein model, where red and blue surfaces represent regions of positive and negative electrostatic potential, respectively. Apparently, the red CoMFA contours around the ligand pyridine group were mapped well with the deep brown (positive electrostatic) surface of the active site where LYS274 was located and highlighted. Moreover, the long hydrocarbon chain and biphenyl parts of the ligand were correctly positioned into a hydrophobic pocket highlighted by residues ILE111, ILE113, MET230, ALA133, PHE155, PRO177, VAL192, and MET195 of the active site which was in parallel with those observed experimentally.^{5–7,35} The complementarity of the CoMSIA hydrogen-bond-donor and -acceptor contours with the hydrogen-bond potential map of the active site surface is even more dramatic. As depicted in Figure 3a, the mapping identified a feature of hydrogen-bond acceptor expressed as magenta

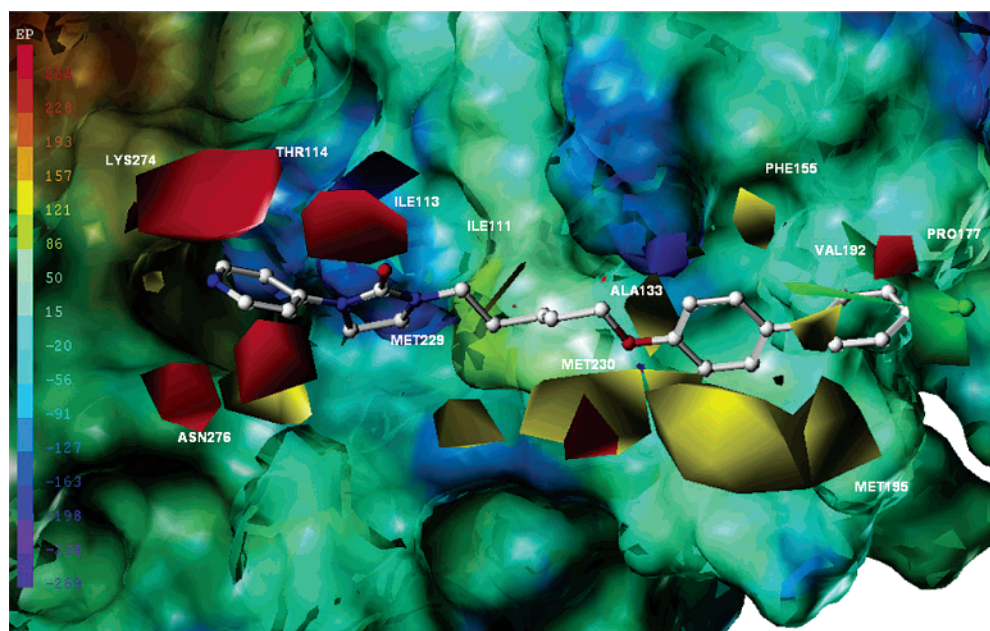


Figure 2. Transparent view of projection of the CoMFA stdev*coeff contours from model 3 plus compound 2 onto the electrostatic potential surface (blue, negative potential; red/brown, positive potential) of the EV71 VP1 active site. Green contours represent regions where bulky groups are favored, whereas yellow contours represent regions where bulky groups are disfavored. Blue contours represent regions where positive charges are favored, and red contours represent regions where negative charges are favored.

contours around the pyridine group, which was unrecognized by the CoMFA results. This feature of hydrogen-bond acceptor was correctly projected onto the hydrogen-bond-donor potential surface expressed in deep red color in the active site. As Figure 3b indicates, the stepwise CoMSIA results were projected onto the potential surfaces of the active site, which confirmed that the long hydrocarbon chain and biphenyl portion of the ligand were correctly docked into a hydrophobic pocket. These hydrophobic contours were correctly mapped with the brown (high lipophilicity) or blue (high hydrophilicity) surfaces of the hydrophobic pocket in the active site.

In addition to proving the 3D-QSAR studies, the pharmacophore features of the docked ligand structures were also explored by the CATALYST programs. The conformations of model 3 were directly employed to generate the hypothesis. The results of the top hypothesis were as follows: total cost, 65.32; fixed cost, 56.47; null cost, 162.65; error cost, 55.44; RMS, 0.90; correlation, 0.96; and configuration cost, 8.01 (Table 3). The top hypothesis was deemed to be 90% chance in statistical significance, since the cost differences between null and fixed or null and total cost were both greater than 60. Moreover, the configuration cost of the top hypothesis generated was smaller than 17, which indicated that the hypothesis was not generated by chance and was unlikely to correlate with others. To further assess the statistical significance of the pharmacophore hypothesis, the top hypothesis was validated by using the CATALYST/CatScramble module. As shown in Table 3, none of these randomly generated hypotheses was better in statistics (correlation > 0.96, error cost < 55.44, and RMS < 0.90) than the top one. As characterized by features of hydrogen-bond acceptor (displayed as green spheres) and hydrophobic (displayed as blue spheres) and excluded volume (displayed as black spheres), the top hypothesis generated was also projected onto the surfaces of the active site (Figure 4). It is interesting to note that the feature of hydrogen-bond acceptor (green spheres in Figure 4) is also positioned around the ligand pyridine group and projected onto the surface of LYS274, which is similar to the CoMFA (Figure 2) and stepwise CoMSIA (Figure 3a) results. The

Table 3. Validation Results for the Top Hypothesis Using the CatScramble Module^a

validation	total cost	fixed cost	error cost	RMS	correlation	configuration cost
Without CatScramble						
original	65.32	56.47	55.44	0.90	0.96	8.01
With CatScramble						
trial_1	92.89	55.46	84.72	1.93	0.82	7.00
trial_2	129.57	55.78	120.12	2.69	0.61	7.32
trial_3	112.69	55.40	104.19	2.38	0.71	6.94
trial_4	127.23	53.27	121.30	2.72	0.60	4.81
trial_5	86.84	56.47	76.89	1.72	0.86	8.01
trial_6	127.08	56.29	117.36	2.65	0.63	7.83
trial_7	164.91	53.54	148.26	3.17	0.36	5.09
trial_8	124.26	55.46	113.14	2.56	0.66	7.00
trial_9	115.89	54.48	107.70	2.46	0.69	6.02
trial_10	121.76	54.48	113.26	2.56	0.65	6.02
trial_11	129.44	53.27	123.28	2.76	0.58	4.81
trial_12	103.57	55.85	94.74	2.17	0.76	7.39
trial_13	117.44	55.40	108.89	2.48	0.68	6.94
trial_14	118.78	55.46	110.52	2.51	0.67	7.00
trial_15	125.11	55.85	113.16	2.58	0.66	7.39
trial_16	103.78	56.47	93.33	2.14	0.78	8.01
trial_17	114.55	55.85	105.25	2.41	0.71	7.39
trial_18	123.58	55.40	112.85	2.56	0.66	6.94
trial_19	113.07	55.46	104.88	2.39	0.71	7.00

^a Null cost = 162.65. All costs are in unit of bits.

hydrophobic feature (blue spheres in Figure 4) of the top hypothesis projected onto the hydrophobic pocket also coincides with those identified by the stepwise CoMSIA contours (Figure 3b). Besides, the feature of excluded volume (black spheres) of the top hypothesis was projected onto the surface of ASN276 where the yellow contours of stepwise CoMFA result were also projected (Figure 2).

The prediction ability of CoMFA, CoMSIA, and CATALYST results from model 3 were evaluated on both test 1 and test 2 sets (Table 4). The goodness of prediction was judged by the linear regression coefficient r_{pred}^2 computed from regression of the predicted onto the actual pIC₅₀ values. The best prediction (r_{pred}^2) of the CoMFA result on both test 1 and test 2 sets were

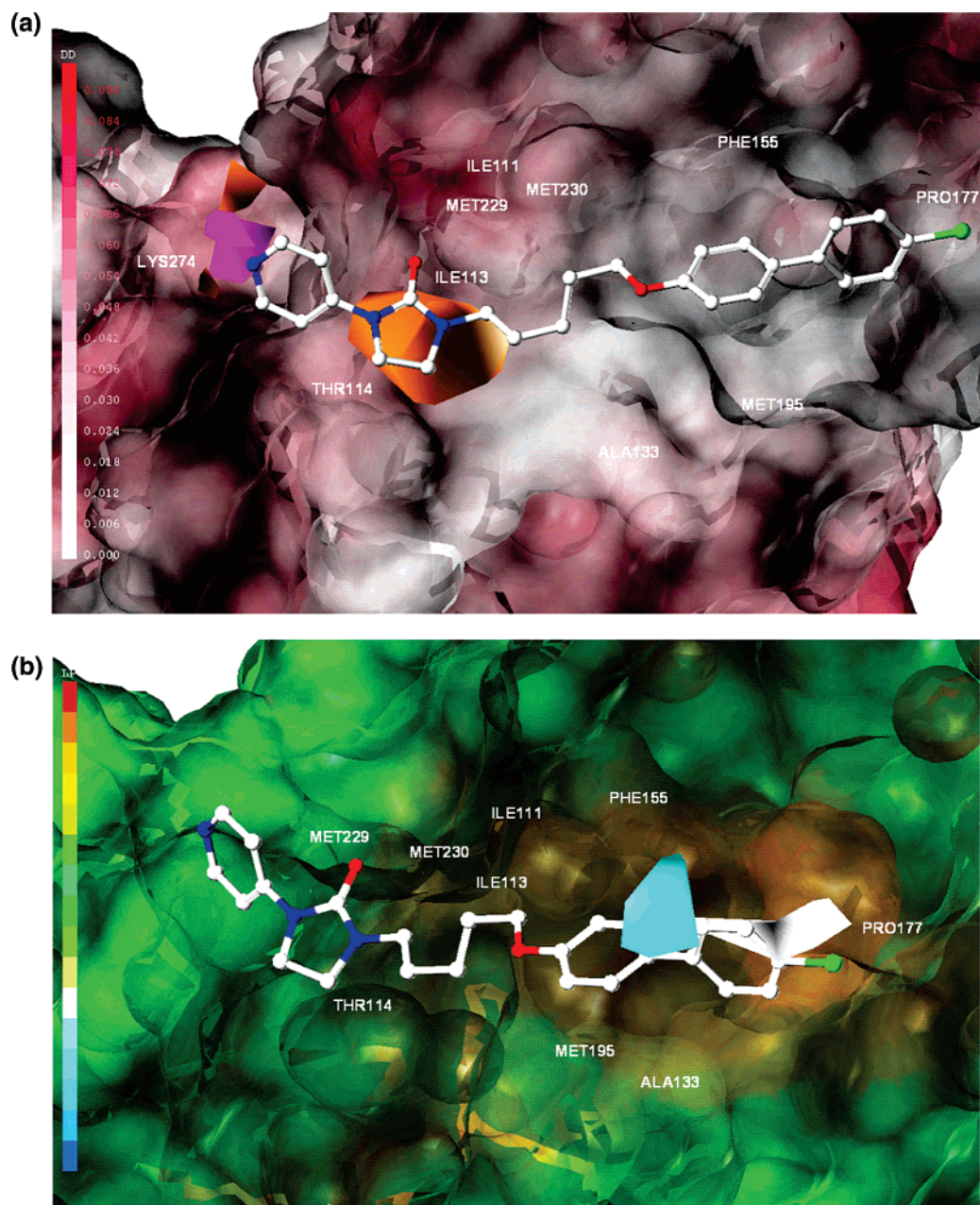


Figure 3. (a) Transparent view of projection of CoMSIA hydrogen-bond-acceptor contours from model 3 plus compound 2 onto the hydrogen-bond-donor potential surface (red, donor favor regions; white, donor disfavor regions) of the EV71 VP1 active site. Magenta contours represent regions with a hydrogen-bond-acceptor character, whereas orange contours represent regions where hydrogen-bond-acceptor character is disfavored. (b) Transparent view of projection of CoMSIA hydrophobic contours from model 3 plus compound 2 onto the lipophilic potential surface (brown, hydrophobic; blue, hydrophilic) of the EV71 VP1 active site. White contours represent regions with hydrophobic favor, and cyan contours represent regions with hydrophobic disfavor.

0.64 and 0.91, respectively. Furthermore, the corresponding r_{pred}^2 values computed by the CoMSIA combination fields (S + E + A) were 0.74 and 0.90 for test 1 and test 2 set, respectively. As judged by r_{pred}^2 values, the prediction accuracy of test 1 set by model 3 appears to be the CATALYST > CoMSIA > CoMFA (Table 4). However, this order was slightly reversed on test 2, where the best prediction ($r_{\text{pred}}^2 = 0.91$) was given by the CoMFA result. Although both CoMFA and CoMSIA gave better regression coefficients r_{pred}^2 computed for test 2 set than that calculated by CATALYST, the predicted activities given by the former were smaller than that given by the latter (Table 4). Mapping of the CoMFA contours and the pharmacophore features of the top hypothesis on the most potent compound (31) of test 2 set is presented in Figure 5. As shown in Figure 2, the bulky group favored regions of CoMFA contours showed

that the 2002 compound set may be extended further beyond the R² group (phenylchloro). This was correctly reflected by the mapping of CoMFA contours on compound 31 (top of Figure 5), where some yellow contours around the phenyl ring represent that a bulky group is disallowed in the region while green, red, and blue contours around the ethyl-oxime group represent that groups of any feature may be added in that region. On the other hand, the top hypothesis with good mapping (bottom of Figure 5) and closely predicted pIC₅₀ for compound 31 (Table 4) also confirmed that the hydrogen bond on the pyridine part or the hydrophobic effect on the long hydrocarbon chain of the ligand may be influential for the compound's activity. Note that no compound of the 2004 set was included in the training process (Table 1). These results implied that a

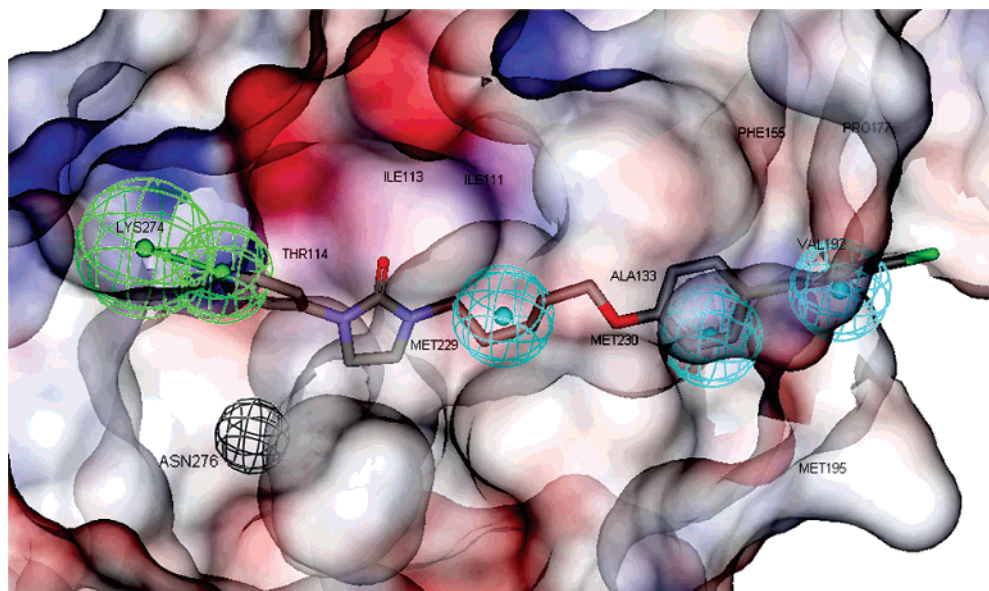


Figure 4. Transparent view of projection of the top hypothesis plus compound **2** onto the electrostatic potential surface (blue, positive potential; red, negative potential) of the EV71 VP1 active site. The top hypothesis contains five features: three hydrophobic characters (cyan spheres), one hydrogen-bond acceptor (green spheres), and one excluded volume character (black sphere).

Table 4. Predicted Activities of the Test Sets of Model 3 Obtained from the CoMFA, CoMSIA (S + E + A), CATALYST, and VOLSURF Programs

	compd	actual PIC ₅₀	estimated PIC ₅₀			
			CoMFA	CoMSIA	CATALYST	VOLSURF
<i>r</i> _{pred} ² test 1	1	7.40	7.04	7.29	7.34	6.41
	5	6.46	5.68	5.77	6.82	6.22
	9	6.24	6.35	6.74	6.40	6.11
	13	5.87	6.16	6.17	6.46	5.93
	14	5.87	6.38	6.52	5.62	6.10
	15	5.82	5.65	5.71	5.64	5.79
	16	5.58	5.66	5.49	6.16	5.48
	25	5.33	5.76	5.43	5.62	5.67
	28	4.97	5.69	5.23	5.31	4.74
	30	4.92	5.26	5.18	5.46	5.55
			0.64	0.74	0.91	0.70
<i>r</i> _{pred} ² test 2	31	9.00	7.05	7.43	8.18	4.91
	32	8.30	6.85	7.25	7.70	4.99
	33	8.00	6.92	7.16	8.23	5.43
	34	7.68	7.05	7.53	6.47	4.86
	35	7.60	6.48	6.79	7.66	4.70
	36	7.10	6.52	6.72	6.38	4.86
	37	6.62	6.41	6.38	6.10	4.74
	38	6.44	6.32	6.43	5.82	4.82
	39	6.19	6.39	6.59	6.64	4.67
	40	6.10	6.19	6.30	5.80	4.32
	41	5.97	6.18	6.34	6.74	4.15
	42	4.77	5.41	5.28	5.25	4.35
	43	4.67	5.42	5.20	5.04	4.60
	44	4.60	5.34	4.92	5.38	4.32
			0.91	0.90	0.89	0.54

model built from previous compounds (2002 set) could be used in predicting the later compounds (2004 set).

The VOLSURF method was designed to take into account the 3D structures of molecules without requiring a structural alignment that is unlike the methods of CoMFA and CATALYST. The interpretation of the VOLSURF model can be done both visually and in terms of the descriptors used. The q_{cv}^2 (leave-five-out) and q_{loo}^2 (leave-one-out) values evaluated by the model on the training set were 0.89 and 0.90, respectively. This suggests a good correlation between the VOLSURF descriptors chosen and activities of the compounds studied. This is also consistent with the PLS coefficient plots made for compounds **2** and **29**, the most and least potent compounds of

the training set studied and presented in the top part of Figure 6, respectively. There were significant coefficient (positive) values for the hydrophilic, hydrophobic, and amide NH probes chosen for compound **2** but not for compound **29** (top of Figure 6). A hydrophilic contour made at -3.0 kcal/mol intervals around the pyridine group was also very visible for compound **2** (middle of Figure 6), in accord with the CATALYST feature of hydrogen-bond acceptor (green spheres) shown in Figures 4 and 5. Further, the hydrophobic effect was manifested by the hydrophobic contours made at -1.0 kcal/mol intervals around both compounds (bottom of Figure 6). Being much more potent in activity and with a biphenyl ring attached, compound **2** was subjected to a stronger hydrophobic effect than compound **29**, as evidenced by larger hydrophobic contours made around it than those around the latter. Therefore, the VOLSURF analyses were not only consistent with the CoMFA and stepwise CoMSIA results but also confirmed that the pharmacophore features selected by the CATALYST analyses were adequate. However, with no structural alignment employed, the prediction by the VOLSURF model on both test 1 ($r_{pred}^2 = 0.70$) and test 2 ($r_{pred}^2 = 0.54$) sets were somewhat worse than those by both CoMFA and stepwise CoMSIA results (Table 4).

Conclusion

It is well-known that the accuracy of a protein model built by the homology modeling technique will strongly depend on the degree of the template sequences used. Here we have shown that while this degree of homology may be important, a judicious choice of strategies for refining the structures constructed is even more important. To assess the goodness of the structural refinements for the ligands, we have created two control models, namely, ligand conformations literally generated with only point-to-point alignment (model 1) and ligand conformations generated by a two-stage flexible docking procedure with point-to-point alignment (model 2). The two-stage docking refinement method was effective since the regression of the ranks of the GLUE score over the ranks of activity were much better than that of the DOCK score for all the 2002 and 2004 compounds studied. The results of both CoMFA and stepwise CoMSIA obtained from model 3 (docking results without alignment) were also much superior to those obtained either from model 1 or model

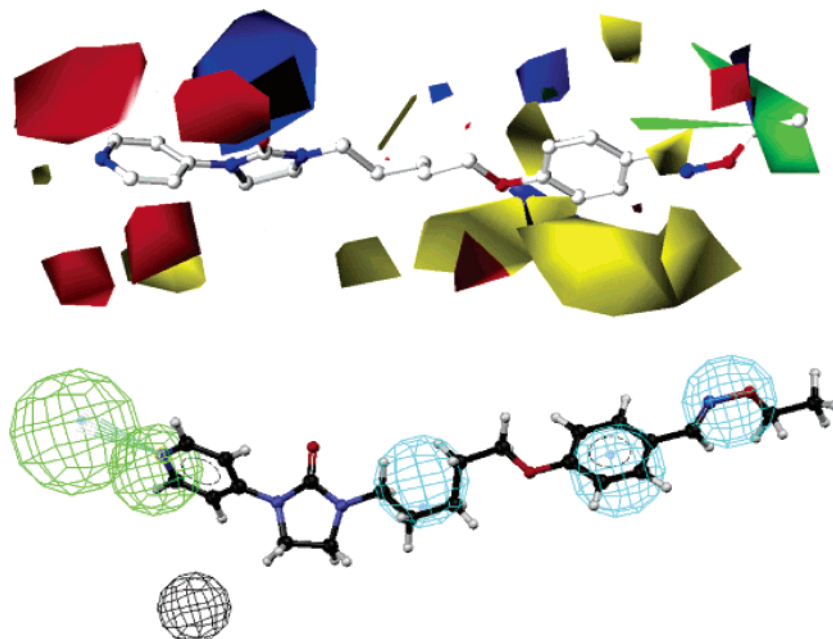


Figure 5. (Top) Mapping of the CoMFA stdev*coeff contours from model 3 onto the structure of the most potent inhibitor **31** of test 2 set. (Bottom) Mapping of the top hypothesis onto the structure of the most potent inhibitor **31** of test 2 set.

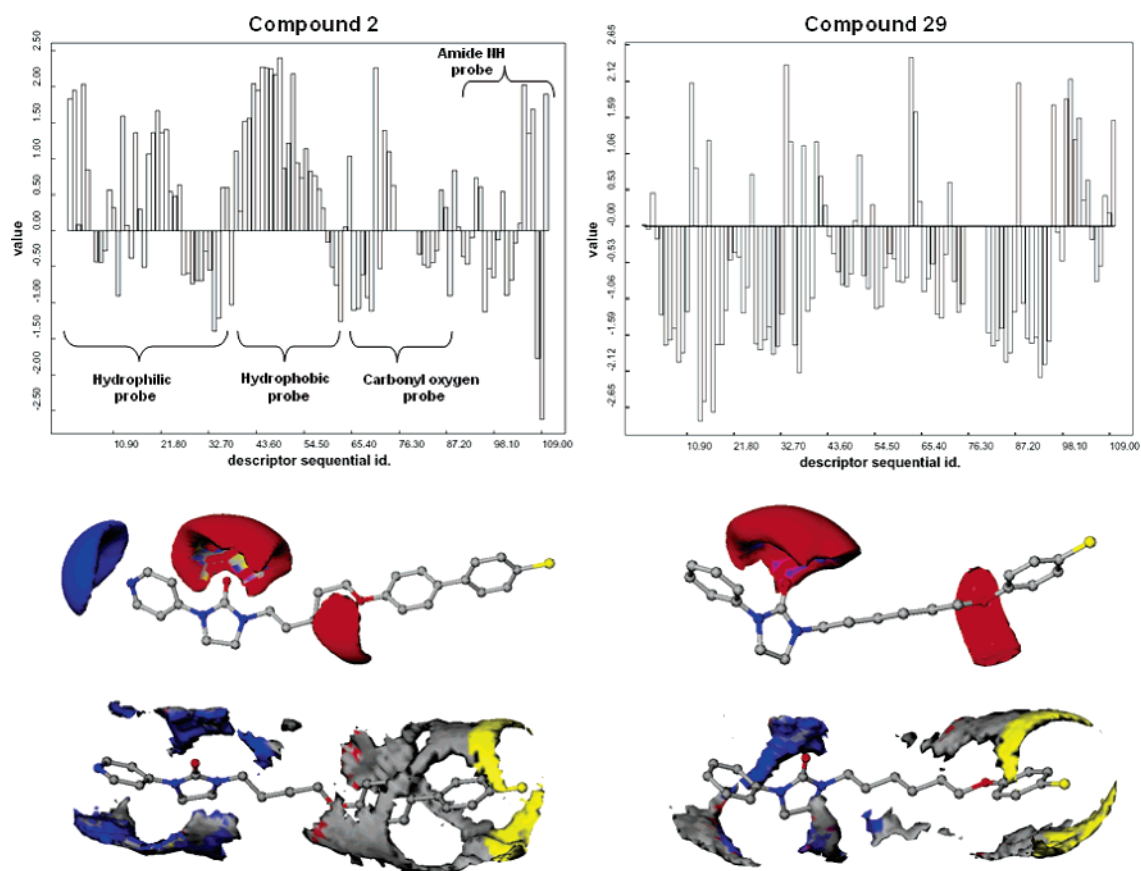


Figure 6. A comparison for the VOLSURF analysis results for the most (compound **2**) and least (compound **29**) potent inhibitor of the training set. (Top) The profiles of the VOLSURF descriptors used in the analyses that are directly (positive values) or inversely (negative values) correlated to the activity (consensus Y). (Middle) The VOLSURF hydrophilic fields contoured at -3.0 kcal/mol intervals around the two compounds. (Bottom) The VOLSURF hydrophobic fields contoured at -1.0 kcal/mol intervals around the two compounds.

2. For the test sets, the predicted activities calculated by both CoMFA and stepwise CoMSIA results from model 3 were more significant in statistics than control models. We conducted an independent VOLSURF4.1 analysis to confirm that the 3D-QSAR analyses of CoMFA, CoMSIA, and pharmacophore hypothesis were consistent with each other. The pharmacophore

features determined by these 3D-QSAR models were all projected back and matched well with the surface characteristics of the active site of the EV71 VP1 protein model. In conclusion, our study showed that most regions of the EV71 VP1 active site were rather hydrophobic or exclusive for the bulky groups. While the feature of hydrogen-bond acceptor on one side may

be influential, more versatile features on the other side of the active site were found and may be worthy of further exploration in the search for more potent inhibitors against the EV71 infection.

Acknowledgment. This work is supported in part by a grant (NSC94-2313-B007-001) from the National Science Council, Taiwan. The SYBYL, CATALYST, InsightII, and VOLSURF computations were conducted using the facilities provided by the National Center for High Performance Computing, Taiwan. We also thank Molecular Discovery Ltd for providing the GRID22A program as freeware.

Supporting Information Available: The sequence alignment of EV 71 VP1 with templates, the superposition for all the conformations of each model, and the activity ranks, compared with the ranks of DOCK4.02 and GLUE1.0 scores. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Ho, M.; Chen, E. R.; Hsu, K. H.; Twu, S. J.; Chen, K. T.; Tsai, S. F.; Wang, J. R.; Shih, S. R. (Taiwan Enterovirus Epidemic Working Group) An epidemic of enterovirus 71 infection in Taiwan. *N. Engl. J. Med.* **1999**, *341*, 929–935.
- Rueckert, R. R. Picornaviruses and their replication. *Virology*; Lippincott-Raven: New York, 1990; p 507.
- Dimmock, N. J.; Easton, A. J.; Leppard, K. N. The structure of virus. *Introduction to Modern Virology*; 5th ed.; Blackwell Science: Oxford, 2001; pp 34–36.
- Hogle, J. M.; Chow, M.; Filman, D. J. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* **1985**, *229*, 1358–1365.
- Reisdorph, N.; Thomas, J. J.; Katpally, U.; Chase, E.; Harris, K.; Siuzdak, G.; Smith, T. J. Human rhinovirus capsid dynamics is controlled by canyon flexibility. *Virology* **2003**, *314*, 34–44.
- Filman, D. J.; Syed, R.; Chow, M.; Macadam, A. J.; Minor, P. D.; Hogle, J. M. Structural factors that control conformational transitions and serotype specificity in type 3 poliovirus. *EMBO J.* **1989**, *8*, 1567–1579.
- Lewis, J. K.; Bothner, B.; Smith, T. J.; Siuzdak, G. Antiviral agent blocks breathing of the common cold virus. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6774–6778.
- Smith, T. J.; Kremer, M. J.; Luo, M.; Vriend, G.; Arnold, E.; Kamer, G.; Rossmann, M. G.; McKinlay, M. A.; Diana, G. D.; Otto, M. J. The site of attachment in human rhinovirus 14 for antiviral agents that inhibit uncoating. *Science* **1986**, *233*, 1286–1293.
- Pevear, D. C.; Tull, T. M.; Seipel, M. E.; Groarke, J. M. Activity of pleconaril against enteroviruses. *Antimicrob. Agents Chemother.* **1999**, *43*, 2109–2115.
- Kaiser, L.; Crump, C. E.; Hayden, F. G. In vitro activity of pleconaril and AG7088 against selected serotypes and clinical isolates of human rhinoviruses. *Antiviral Res.* **2000**, *47*, 215–220.
- Patick, A. K.; Binford, S. L.; Brothers, M. A.; Jackson, R. L.; Ford, C. E.; Diem, M. D.; Maldonado, F.; Dragovich, P. S.; Zhou, R.; Prins, T. J.; Fuhrman, S. A.; Meador, J. W.; Zalman, L. S.; Matthews, D. A.; Worland, S. T. In vitro antiviral activity of AG7088, a potent inhibitor of human rhinovirus 3C protease. *Antimicrob. Agents Chemother.* **1999**, *43*, 2444–2450.
- Romero, J. R. Pleconaril: A novel antipicornaviral drug. *Expert Opin. Invest. Drugs* **2001**, *10*, 369–379.
- McKinlay, M. A.; Pevear, D. C.; Rossmann, M. G. Treatment of the picornavirus common cold by inhibitors of viral uncoating and attachment. *Annu. Rev. Microbiol.* **1992**, *46*, 635–654.
- Shia, K. S.; Li, W. T.; Chang, C. M.; Hsu, M. C.; Chern, J. H.; Leong, M. K.; Tseng, S. N.; Lee, C. C.; Lee, Y. C.; Chen, S. J.; Peng, K. C.; Tseng, H. Y.; Chang, Y. L.; Tai, C. L.; Shih, S. R. Design, synthesis, and structure–activity relationship of pyridyl imidazolidinones: A novel class of potent and selective human enterovirus 71 inhibitors. *J. Med. Chem.* **2002**, *45*, 1644–1655.
- Hendry, E.; Hatanaka, H.; Fry, E.; Smyth, M.; Tate, J.; Stanway, G.; Santti, J.; Maaronen, M.; Hyypia, T.; Stuart, D. The crystal structure of coxsackievirus A9: New insights into the uncoating mechanisms of enteroviruses. *Structure Fold Des.* **1999**, *7*, 1527–1538.
- InsightII Homology Model*; Version 2000 ed.; Accelrys Inc.: San Diego, CA.
- Cramer, R. D., 3rd; Patterson, D. E.; Bunce, J. D. Recent advances in comparative molecular field analysis (CoMFA). *Prog. Clin. Biol. Res.* **1989**, *291*, 161–165.
- Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.
- CATALYST, version 4.9; Accelrys Inc. (previously known as Molecular Simulations, Inc.): San Diego, CA.
- Cruciani, G.; Pastor, M.; Guba, W. VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000**, *11 Suppl 2*, S29–39.
- Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.* **2001**, *15*, 411–428.
- GLUE, version 1.0; Molecular Discovery Ltd: Perugia, Italy.
- Chern, J. H.; Lee, C. C.; Chang, C. S.; Lee, Y. C.; Tai, C. L.; Lin, Y. T.; Shia, K. S.; Lee, C. Y.; Shih, S. R. Synthesis and antienteroviral activity of a series of novel, oxime ether-containing pyridyl imidazolidinones. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5051–5056.
- Shih, S. R.; Ho, M. S.; Lin, K. H.; Wu, S. L.; Chen, Y. T.; Wu, C. N.; Lin, T. Y.; Chang, L. Y.; Tsao, K. C.; Ning, H. C.; Chang, P. Y.; Jung, S. M.; Hsueh, C.; Chang, K. S. Genetic analysis of enterovirus 71 isolated from fatal and non-fatal cases of hand, foot and mouth disease during an epidemic in Taiwan, 1998. *Virus Res.* **2000**, *68*, 127–136.
- Lentz, K. N.; Smith, A. D.; Geisler, S. C.; Cox, S.; Buontempo, P.; Skelton, A.; DeMartino, J.; Rozhon, E.; Schwartz, J.; Girijavallabhan, V.; O'Connell, J.; Arnold, E. Structure of poliovirus type 2 Lansing complexed with antiviral agent SCH48973: Comparison of the structural and biological properties of three poliovirus serotypes. *Structure* **1997**, *5*, 961–978.
- Hiremth, C. N. Binding of the antiviral drug WIN51711 to the sabin strain of type 3 poliovirus: Structural comparison with drug binding in rhinovirus 14. *Acta Crystallogr. D Biol. Crystallogr.* **1995**, *51*, 473–489.
- Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **1986**, *7*, 230–252.
- SYBYL; Version 6.9.1 ed.; Tripos Associates, Inc.: St. Louis, MO.
- Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- Dunn, W. J.; Wold, S.; Edelund, U.; Helberg, S. Multivariate Structure–Activity Relationships between Data from a Battery of Biological Tests and an Ensemble of Structure Descriptors: The PLS Methodology. *Quant. Struct.-Act. Relat.* **1984**, *3*, 131–137.
- Xue, C. X.; Cui, S. Y.; Liu, M. C.; Hu, Z. D.; Fan, B. T. 3D QSAR studies on antimalarial alkoxyated and hydroxylated chalcones by CoMFA and CoMSIA. *Eur. J. Med. Chem.* **2004**, *39*, 745–753.
- Pastor, M.; Cruciani, G.; Clementi, S. Smart region definition: A new way to improve the predictive ability and interpretability of three-dimensional quantitative structure–activity relationships. *J. Med. Chem.* **1997**, *40*, 1455–1464.
- Sutter, J.; Guner, O. F.; Hoffman, R.; Li, H.; Waldman, M. *Effect of Variable Weight and Tolerances on Predictive Model Generation. In Pharmacophore Perception, Development, and Use in Drug Design*; Guner, O. F., Ed.; International University Line: La Jolla, CA, 1999; pp 501–511.
- Colovos, C.; Yeates, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* **1993**, *2*, 1511–1519.
- Shih, S. R.; Tsai, M. C.; Tseng, S. N.; Won, K. F.; Shia, K. S.; Li, W. T.; Chern, J. H.; Chen, G. W.; Lee, C. C.; Lee, Y. C.; Peng, K. C.; Chao, Y. S. Mutation in enterovirus 71 capsid protein VP1 confers resistance to the inhibitory effects of pyridyl imidazolidinone. *Antimicrob. Agents Chemother* **2004**, *48*, 3523–3529.