

Discovery of a Novel Family of SARS-CoV Protease Inhibitors by Virtual Screening and 3D-QSAR Studies

Keng-Chang Tsai,^{†,§} Shih-Yuan Chen,[†] Po-Huang Liang,[‡] I-Lin Lu,^{||} Neeraj Mahindroo,^{||} Hsing-Pang Hsieh,^{||} Yu-Sheng Chao,^{||} Lincoln Liu,[†] Donald Liu,[†] Wei Lien,[†] Thy-Hou Lin,^{*,†} and Su-Ying Wu^{*,||}

Institute of Molecular Medicine & Department of Life Science, National Tsing Hua University, Hsinchu 300, Taiwan, Republic of China, Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan, Republic of China, Genomics Research Center, Academia Sinica, Taipei 115, Taiwan, Republic of China, and Division of Biotechnology and Pharmaceutical Research, National Health Research Institutes, 35 Keyan Road, Zhunan Town, Miaoli County 350, Taiwan, Republic of China

Received August 27, 2005

The severe acute respiratory syndrome-associated coronavirus (SARS-CoV) 3C-like protease (3CL^{pro} or M^{pro}) is an attractive target for the development of anti-SARS drugs because of its crucial role in the viral life cycle. In this study, a compound database was screened by the structure-based virtual screening approach to identify initial hits as inhibitors of SARS-CoV 3CL^{pro}. Out of the 59 363 compounds docked, 93 were selected for the inhibition assay, and 21 showed inhibition against SARS-CoV 3CL^{pro} (IC₅₀ ≤ 30 μM), with three of them having common substructures. Furthermore, a search for analogues with common substructure in the Maybridge, ChemBridge, and SPECS_SC databases led to the identification of another 25 compounds that exhibited inhibition against SARS-CoV 3CL^{pro} (IC₅₀ = 3–1000 μM). These compounds, 28 in total, were subjected to 3D-QSAR studies to elucidate the pharmacophore of SARS-CoV 3CL^{pro}.

Introduction

Severe acute respiratory syndrome (SARS), an epidemic that rapidly spread in Asia, North America, and Europe resulted in severe illness and deaths in early 2003.^{1,2} This disease is mainly spread by respiratory droplets containing the virus, and many studies show that a previously unrecognized coronavirus called SARS-CoV (SARS-associated coronavirus) could be the cause of SARS.^{3–5} The genome of SARS-CoV contains 11–14 major open-reading frames and encodes several proteins, including the replicase polyproteins, S (spike protein), polymerase, M (membrane protein), N (nucleocapsid protein) and E (small envelope protein).^{6–8}

SARS-CoV 3C-like protease (SARS-CoV 3CL^{pro}), as a part of the replicase polyproteins, cleaves a functional polypeptide and, consequently, leads to the maturation of SARS-CoV. Because of its functional importance in the SARS-CoV replication cycle, SARS-CoV 3CL^{pro} is considered a potential target to develop novel anti-SARS drugs.⁹ Previous homology modeling studies for SARS-CoV 3CL^{pro}^{9–11} made it possible to design inhibitors using various computer-aided drug design methods.^{11–17} For example, an 8-mer peptide was docked into the 3D model of SARS-CoV 3CL^{pro}, and the possible binding profile between SARS-CoV 3CL^{pro} and this substrate was elucidated.¹⁸ In another case, a pharmacophore model generated from a small peptide was used as a query for screening possible SARS-CoV 3CL^{pro} inhibitors through several chemical databases.^{19,20} Furthermore, the crystal structures of the coronavirus 3CL proteases from the transmissible gastroenteritis virus (TGEV)²¹ and human coronavirus 229E (HCoV-229E)⁹ were available to facilitate inhibitors design. Recently, the crystal structure of SARS-CoV

3CL^{pro} (pdb ID: 1Q2W and 1UK4) was independently solved by two groups. Rao et al. reported the structure of SARS-CoV 3CL^{pro} in complex with a covalently attached substrate-analogue inhibitor, thus providing insights into the substrate binding site.²²

Although a number of non-peptide inhibitors of SARS-CoV 3CL^{pro}, such as bifunctional aryl boronic acids,²³ isatin derivatives,²⁴ polyphenols,²⁵ etacrynic acid analogues,²⁶ cinanserin,¹² and other chemically diverse small molecules^{27–30} have been identified, only a few of these show potent inhibitory activity. Here, we report the discovery of a novel family of potent SARS-CoV 3CL^{pro} inhibitors by virtual screening and 3D-QSAR studies. Virtual (in silico) screening by the molecular docking of chemical databases in combination with 3D-QSAR studies is one of the most powerful approaches used to discover small molecule inhibitors. The present study aimed to design novel non-peptide inhibitors against SARS-CoV 3CL^{pro} using the knowledge obtained from the 3D structure of SARS-CoV 3CL^{pro}. We conducted a virtual screening study using the DOCK4.0.2.³¹ program to identify novel small molecule inhibitors of SARS-CoV 3CL^{pro}. The lead compounds were subsequently analyzed by several 3D-QSAR techniques to fully explore the pharmacophore of SARS-CoV 3CL^{pro} and to guide further lead optimization.

Materials and Methods

Virtual Screening. The X-ray structure of the complex of SARS-CoV 3CL^{pro} (pdb code 1UK4)²² with an octapeptidyl CMK inhibitor was chosen as the template in the virtual screening. The DOCK4.0.2 program³¹ was used to screen a commercially available small molecule database, the Maybridge database, obtained from the Maybridge Chemical Company (Tintagel, Cornwall, England). The rule of five was applied as a filter to select the drug-like compounds in the virtual screening procedure. The database, containing 59 363 compounds, was screened and scored on a 64-processor Hpcserv2 Linux cluster with AMD Athlon MP 2000+ 1.7 GHz CPUs. Residues within a radius of 6 Å around the center of the CMK peptide inhibitor were defined as the active site to construct a grid for the virtual screening. The active site included residues His41, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163,

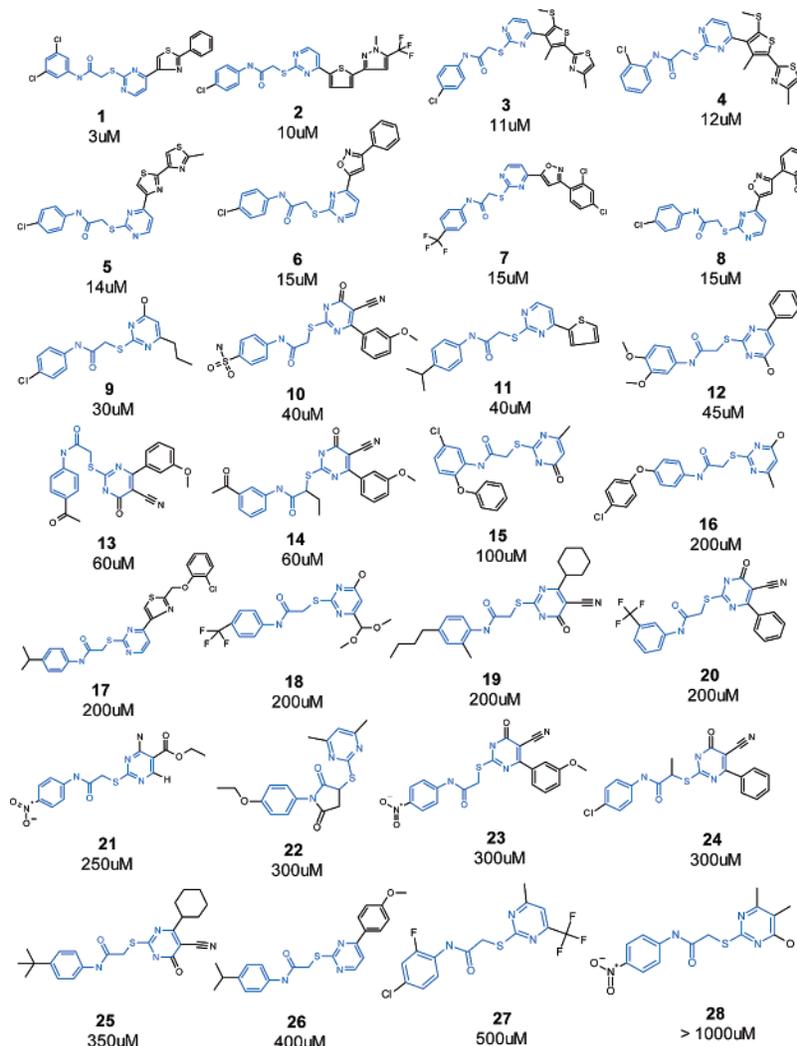
* To whom correspondence should be addressed. Tel: 886-37-246-166 ext. 35713. Fax: 886-37-586-456. E-mail: suying@nhri.org.tw (S.-Y.W.). Tel: 886-3-5742759. Fax: 886-3-571-5934. E-mail: thlin@life.nthu.edu.tw (T.-H.L.).

[†] National Tsing Hua University.

[‡] Institute of Biological Chemistry, Academia Sinica.

[§] Genomics Research Center, Academia Sinica.

^{||} National Health Research Institutes.

Table 1. The Chemical Structures of 28 Inhibitors and Their Corresponding IC₅₀ Values against SARS-CoV 3CL^{pro}

Met165, Glu166, and His172. Before docking, Kollmann-all atom charges³² were assigned to the protein atoms, and Gasteiger-Hückel charges^{33–35} were assigned to compound atoms using the SYBYL 6.9.1 program.³⁶ The position and conformation of each compound were optimized first by the anchor fragment orientation and then by the torsion minimization method implemented in the DOCK4.0.2 program. Fifty configurations and the maximum of 100 anchor orientations for each compound were generated, and all of the docked configurations were energy minimized by 100 iterations.

SARS-CoV 3CL^{pro} Inhibition Assay. The SARS-CoV 3CL^{pro} inhibition assay was performed on the basis of the published procedure.^{27,37} As described by Kuo et al.,³⁷ the gene encoding the SARS-CoV 3CL^{pro} was cloned from the viral whole genome by using PCR with the forward primer 5'-GGTATTGAGGGTTCG-CAGTGGTTTTAGG-3' and the reverse primer 5'-AGAGGAGAGT-TAGAGCCTTATTGGAAGGTAACACC-3' into the pET32Xa/LIC vector. The FXa cleavage site was included in these primers. The plasmid was transformed into *E. coli* BL21 to express the His-tagged protease. The purified SARS-CoV 3CL^{pro}, cleaved by FXa protease to remove the His-tag, has an authentic sequence without extra amino acids, as confirmed by *N*-terminal sequencing and mass spectroscopy. All of the kinetic measurements were performed in 20 mM bis[(2-hydroxyethyl)amino]tris(hydroxymethyl)methane (pH 7.0) at 25 °C. Enhanced fluorescence was monitored at 538 nm with excitation at 355 nm by using a fluorescence plate reader upon the cleavage of the fluorogenic substrate peptide (Dabcyl-KTSAVLQ-SGFRKME-Edans). The initial velocities of the inhibited reactions of 50 nM SARS-CoV 3CL^{pro} and 6 μM fluorogenic

substrate were plotted against the inhibitor concentrations to obtain the IC₅₀ by using the following equation

$$A[I] = A[0] \times \left\{ 1 - \left[\frac{[I]}{([I] + IC_{50})} \right] \right\}$$

where $A[I]$ is the enzyme activity with inhibitor concentration $[I]$, and $A[0]$ is the enzyme activity without the inhibitor.

CoMFA, CoMSIA, and Pharmacophore 3D-QSAR Models. Twenty-eight inhibitors were divided into training and test sets containing 18 and 10 inhibitors, respectively, for 3D-QSAR studies. The compounds of the test set were selected first on the basis of published rules,³⁸ and they were as follows: compounds **4**, **7**, **8**, **11**, **12**, **14**, **15**, **18**, **21**, and **23** (Table 1). The CoMFA steric and electrostatic potential fields were calculated using the SYBYL 6.9.1 program with a regularly spaced grid of 2.0 Å. A C.3 carbon atom with a radius of 1.52 Å and a charge of +1.0 was used as a probe to calculate the steric and electrostatic energies between the probe and the molecules using the Tripos force field.³⁹ The truncation for both the steric and electrostatic energies was set to ±30 kcal/mol. The CoMFA steric and electrostatic fields were scaled by the default value given in the program.

A C.3 atom with a radius of 1.0 Å and a charge of +1.0 was used as the probe to calculate the CoMSIA similarity indices defined by Klebe⁴⁰ with a spaced grid of 2.0 Å. The similarity indices were calculated using the Gaussian-type distance dependence between the probe and each atom of the molecules. The attenuation factor

α was set as 0.3. Both CoMFA and CoMSIA results were cross-validated using the SYBYL PLS (partial-least-squares) module. The minimum-sigma (column filtering) was set to 2.0 kcal/mol to improve the signal-to-noise ratio by omitting the grid points of energy variation less than this threshold. The CoMFA and CoMSIA descriptors were treated as independent variables, whereas the pIC_{50} values were treated as dependent variables in the PLS regression analyses to derive the 3D-QSAR models. The number of components used in the final nonvalidated model was optimized to give the highest cross-validated $r^2(q^2)$ and the lowest standard error of prediction. The noncross-validated models were assessed by the conventional correlation coefficient r^2 , standard error of prediction, and F -values.

The training set for CoMFA and CoMSIA studies was used for constructing pharmacophore models by the Catalyst 4.9 program.⁴¹ All of the parameters used were default values except for the Uncert value, which was set to 1.6 or 1.5. Each compound was energy minimized using the CHARMM-like force field⁴² within the Catalyst 4.9 program and subjected to a conformational analysis using the Poling algorithm.⁴³ The maximum number of conformers allowed for each compound was set to 250. On the basis of the CoMSIA results, four pharmacophore features, E (excluded volumes), H (hydrophobic), D (hydrogen-bond donor), and A (hydrogen-bond acceptor group), were selected to generate hypotheses. The significance of the best hypothesis generated was cross-validated using the CatScramble module of the Catalyst 4.9 program. To obtain a 95% confidence level, 19 random spreadsheets were generated and then analyzed by the HypoGen module using the same parameters as those in the initial run.

Binding Free Energy Calculation. To calculate the binding free energy, SARS-CoV 3CL^{pro} and 28 inhibitors were parametrized by the Parm99⁴⁴ and the AMBER force field (GAFF).⁴⁵ The partial charges of each compound were calculated using the HF/6-31G(d)/RESP method⁴⁶ and were fitted by the multiple RESP approach. The series of calculations were conducted by AMBER8.⁴⁷ The binding free energies for the 28 inhibitors were computed as follows

$$G = H_{\text{gas}} + G_{\text{sol}} - TS_{\text{conf}} \quad (1)$$

$$\Delta G_{\text{binding}} = \Delta G_{\text{complex}} - \Delta G_{\text{receptor}} - \Delta G_{\text{ligand}} \quad (2)$$

where H_{gas} is the gas-phase energy, G_{sol} is the solvation free energy, and S_{conf} is the sum of translational, rotational, and vibrational entropies determined by normal-mode analysis. H_{gas} included the electrostatic and van der Waals interactions between the protein and the inhibitor. A distance-dependent dielectric was used to compute electrostatic interactions, and a continuum model correction for energy was used to compute van der Waals interactions. Both interactions were calculated in AMBER8 using the default parameters. G_{sol} was computed using the GB model developed by Tsui and Case.⁴⁸ The hydrophobic contribution to the solvation free energy and the implicit solvation free energy were computed from the solvent-accessible area (SA) by Paul Beroza's Molsurf module of AMBER. The MM/GBSA approach was used to estimate the $H_{\text{gas}} + G_{\text{sol}}$ terms in eq 1. The protein complexes were solvated by the GB model, and instead of neutralizing the whole system by explicit cations, the salt contribution for calculating the generalized Born energy was set at 0.1 M. All His residues were protonated at the epsilon nitrogen (N_{ϵ}) except for His41 and His172, which were protonated at the epsilon and delta nitrogens (N_{δ}). For the rest of the acidic and basic residues in the protein, the default protonation states in AMBER8 were applied. The translational and rotational entropies were calculated as described by McQuarrie,⁴⁹ whereas the vibrational entropy was calculated by a normal-mode analysis using the Nmode module of AMBER. For the simulation of each complex, geometry optimization was performed by 200 steps of steepest descent followed by conjugated gradient minimization to converge to an energy criterion of 10^{-1} kcal·mol⁻¹·Å⁻¹.

Results and Discussion

Virtual Screening. The binding site of SARS-CoV 3CL^{pro}/substrate-analogue CMK complex structure (pdb code:1UK4)²²

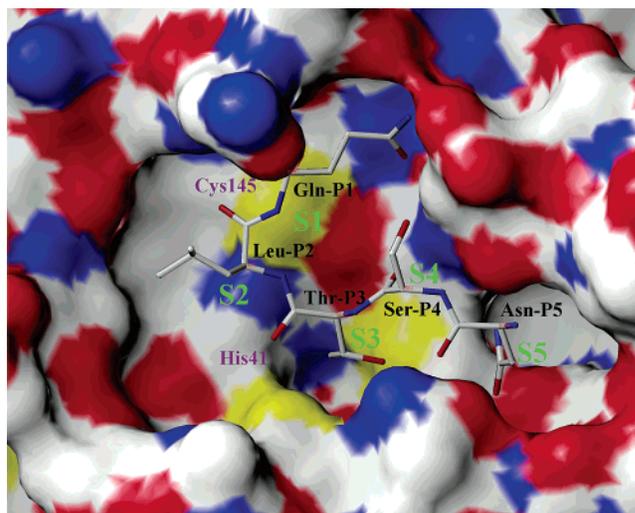


Figure 1. Active site of SARS-CoV 3CL^{pro} with the bound hexapeptidyl CMK peptide inhibitor. The subsites that complement substrate binding are designated as S1 (His163, Glu166, Cys145, Ser144, Gly143, and Phe140), S2 (Cys145 and Thr25), S3 (Met165, Met49, and His41), S4 (Glu166), and S5 (Gln189, Met165, and Glu166).

was used as the target site to perform virtual screening on the Maybridge database, a database of approximately 60 000 commercially available small molecules. The binding site includes the catalytic center (His41 and Cys145) and several subsites, designated as S1(His163, Glu166, Cys145, Ser144, Gly143, and Phe140), S2 (Cys145 and Thr25), S3 (Met165, Met49, and His41), S4 (Glu166), and S5 (Gln189, Met165, and Glu166) (Figure 1). The catalytic dyad characterized by Cys145 and His41 is located inside subsites S1, S2, and S3. The virtual screening was conducted using the DOCK4.0.2 program, and the docked molecules were ranked by the two scoring functions implemented in the program. The first one was the internal ligand–receptor binding energy, which measures the sum of the van der Waals and electrostatic energies. The second one was the electrostatic energy between the docked ligand and the protein. The top 200 compounds ranked by each scoring function were further screened by analyzing their H-bonding patterns using IDEA2.0 (<http://www.breadth.com.tw>). Because H-bonding interactions play an important role in ligand binding, as revealed by the protease–substrate complex structure, the top ranked compounds making more than two H bonds with the protease were selected for bioassay. The number of compounds preliminarily screened for the inhibition assay was 93. Of these, 21 compounds (hit rate of 22%) were found to exhibit SARS-CoV 3CL^{pro} with IC_{50} values less than 30 μ M. A careful analysis of these active compounds revealed that three of the compounds shared similar chemical structure and a core structure of *N*-phenyl-2-(2-pyrimidinylthio)acetamide (Figure 2) was identified. The core structure was then used as a query structure to search for analogues in Maybridge, ChemBridge, and SPEC-S_SC databases. This resulted in identifying 28 structural analogues, including the three from the initial round, to be evaluated for the SARS-CoV SARS 3CL^{pro} inhibitory activity (Table 1). The series of compounds exhibited inhibition with IC_{50} values in the range of 3–1000 μ M and were subjected to further 3D-QSAR studies.

CoMFA and CoMSIA Models. To proceed with the 3D-QSAR studies, compound **10** (Table 1), with the highest score ranked by program DOCK4.0.2 and DOCK5.1.1, was chosen as a template, and the rest of the compounds were aligned with it. The core structure for the 28 compounds, *N*-phenyl-2-(2-

Table 2. Summary of CoMFA and CoMSIA Results Obtained from the Training Set^a

CoMFA		CoMSIA		cross validation			noncross validation		
SARS-CoV3CL ^{pro} 18	inhibitors	SARS-CoV 3CL ^{pro} 18	inhibitors	PC	q^2_{cv}	CF	SEE	r^2	F
q^2_{loo}	0.689	S		2	0.674	2	0.216	0.923	89.446
q^2_{cv}	0.673	E		3	0.340	2	0.216	0.927	59.557
CF	2	H		4	0.613	2	0.161	0.963	83.600
r^2	0.987	D		1	-0.177	2	0.686	0.164	3.131
SEE	0.095	A		1	-0.456	2	0.637	0.279	6.180
PC	4	S + E		4	0.763	2	0.097	0.987	238.212
F	248.719	S + H		5	0.629	2	0.116	0.982	131.326
steric contribution	62.3%	S + D		4	0.611	2	0.233	0.921	38.057
electrostatic contribution	37.7%	S + A		3	0.301	2	0.225	0.921	54.390
		S + E + H		5	0.767	2	0.077	0.992	298.523
		S + E + D		5	0.705	2	0.085	0.990	245.926
		S + E + A		5	0.494	2	0.085	0.990	249.856
		S + E + D + A		5	0.401	2	0.079	0.992	289.995
		S + E + H + D		5	0.611	2	0.075	0.993	321.358
		S + E + H + A		5	0.647	2	0.070	0.993	362.985
		all fields		5	0.559	2	0.085	0.990	246.172

^a q^2_{loo} : leave-one-out; q^2_{cv} : cross-validation; CF: column filtering; r^2 : conventional; SEE: standard error of estimate; PC: principal components; F : F -values; S: steric field; E: electrostatic field; H: hydrophobic field; A: H-bond acceptor; and D: H-bond donor.

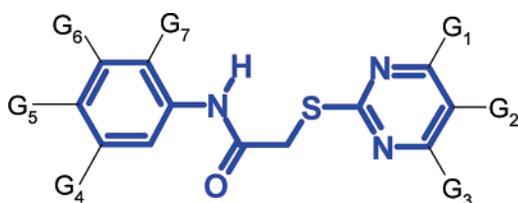


Figure 2. Core structure identified by three original hits. The core structure highlighted in blue served as the correspondence points for the structural alignment of 28 inhibitors to construct the CoMFA and CoMSIA models. G1, G2, G3, G4, G5, G6, and G7 could be any substitute or hydrogen atom.

pyrimidinylthio)acetamide, served as the correspondence points in the alignment (Figure 2). The aligned training set containing 18 inhibitors was then analyzed by the SYBYL CoMFA, CoMSIA, and PLS algorithm. The results of the CoMFA and CoMSIA models are summarized in Table 2. The best CoMFA result yielded a leave-one-out (loo) validated $r^2(q^2_{loo})$ of 0.689, q^2_{cv} of 0.673, and r^2 of 0.987. The CoMSIA model was constructed in a stepwise manner, and the results are presented in Table 2. Five different field indices (steric, denoted as S; electrostatic, denoted as E; hydrophobic, denoted as H; H-bond acceptor, denoted as A; and H-bond donor, denoted as D) were employed one by one or as a combination of various fields to perform the CoMSIA study. The S field gave a better statistical result ($q^2_{cv} = 0.674$) compared to those of other fields and, thus, was selected for further analyses by systematically adding the other four fields. A combination of the S and E fields improved the q^2_{cv} value to 0.763, which was applied to the next cycle of calculations where each of the rest of the three fields (H, D, and A) was sequentially added to the combined S and E fields. The combination of S, E, and H fields produced the best CoMSIA model with a q^2_{cv} value of 0.767 (Table 2). There was no apparent improvement by a further addition of either the D or the A fields to the combined S + E + H fields or upon the combination of all five fields. Therefore, the interactions of the inhibitors in the training set with the target protein SARS-CoV 3CL^{pro} is best described by a combination of S, E, and H fields. A decomposition of the combined S + E + H fields gave the contributions of 42.6%, 40.5%, and 17.0% from the S, E, and H fields, respectively. This result indicated that the S and E fields dominated the present CoMSIA model, although the addition of the H field could slightly improve it.

The experimental and predicted biological activities by the best CoMFA and CoMSIA models for each training set inhibitor

are listed in Table 3, whereas those for the test set are given in Table 4. The template structure, **10**, was also included in the training set. The correlation coefficients (r^2), a measure of the correlation between the predicted versus experimental activities, for the training set and test set given by the best CoMFA model were 0.987 and 0.886, respectively. The best CoMSIA model yielded the correlation coefficients (r^2) of 0.992 and 0.940 for the training set and test set, respectively. The results reveal that the activities predicted by the CoMFA and CoMSIA models are in agreement with the experiment data, demonstrating the predictive ability of both models.

Mapping the CoMFA and CoMSIA Models onto the Protein Active Site. The contour maps generated by the CoMFA and CoMSIA models were mapped on the active site of SARS-CoV 3CL^{pro} and analyzed with respect to the subsites of the protease. The most active compound, compound **1** and the template compound, compound **10**, are displayed with the contour maps to aid in visualization and discussion. Both the CoMFA and CoMSIA contour maps identify favored regions for steric interactions (displayed as green contours in both CoMFA and CoMSIA maps), which are around binding pocket S5 (Figures 1 and 3). The CoMFA and CoMSIA contour maps also show that there are disfavored regions for steric interactions (displayed as yellow contours) around the 3,5-dichloro-benzene group of **1** (Figure 3a-1 and 3b-1) and the benzene sulfonamide group of **10** (Figure 3c-1 and 3d-1), pointing toward the binding pocket S2. The favored regions for hydrophobic interactions identified by the CoMSIA model are displayed as cyan contours for both **1** and **10** (Figure 3b-1 and 3d-1). These are correctly mapped onto the hydrophobic surface of the protease (displayed as a gray surface in Figure 3b-2 and 3d-2). In comparison, the disfavored hydrophobic regions of these two compounds, shown as white contours in the CoMSIA map (Figure 3b-1 and 3d-1), indicate that the substitution with polar groups in this region could improve the binding to the protein. In the CoMFA map, the red contours representing the favorable negative charge area suggest that the substitution with electron rich groups is preferable around the pyrimidine of **1** (Figure 3a-1) and dihydro-pyrimidine of **10** (Figure 3c-1). In addition, the blue contours representing the favorable positive charge area are also correctly mapped onto the electrostatic regions of the protease (Figure 3a-2 and 3c-2).

Pharmacophore Generation with the Catalyst Program. The structural features, including hydrophobic, hydrogen-bond

Table 3. Measured and Predicted Activities of the Training Set Obtained from CoMFA, CoMSIA, and Hypo1

SARS-CoV 3CL ^{pro} inhibitors	act pIC ₅₀	CoMFA	CoMSIA	act IC ₅₀ (μ M)	pred IC ₅₀ (μ M)	catalyst pharmacophore hypothesis Hypo1					supplier
		pred pIC ₅₀	pred pIC ₅₀			error	act activity scale ^a	pred activity scale ^a	uncert	principal	
1	5.52	5.45	5.52	3	6.3	+2.1	+++	+++	1.6	2	Maybridge
2	5.00	5.20	5.04	10	15	+1.5	+++	+++	1.6	0	Maybridge
3	4.95	4.94	4.95	11	21	+1.9	+++	+++	1.6	0	Maybridge
5	4.85	4.91	4.90	14	8.3	-1.7	+++	+++	1.6	0	Maybridge
6	4.82	4.74	4.79	15	9.6	-1.6	+++	+++	1.6	0	Maybridge
9	4.52	4.31	4.33	30	41	+1.4	+++	+++	1.6	0	ChemBridge
10	4.39	4.37	4.38	40	29	-1.4	+++	+++	1.6	2	ChemBridge
13	4.22	4.17	4.18	60	55	-1.1	++	++	1.6	0	ChemBridge
16	3.69	3.69	3.79	200	310	+1.6	++	++	1.6	0	ChemBridge
17	3.69	3.66	3.67	200	130	-1.6	++	++	1.6	0	Maybridge
19	3.69	3.73	3.66	200	82	-2.4	++	++	1.6	0	Maybridge
20	3.69	3.71	3.65	200	270	+1.4	++	++	1.6	0	SPECS SC
22	3.52	3.60	3.59	300	290	-1.0	++	++	1.6	0	ChemBridge
24	3.52	3.52	3.60	300	450	+1.5	++	++	1.6	0	ChemBridge
25	3.45	3.46	3.48	350	260	-1.3	++	++	1.6	0	ChemBridge
26	3.39	3.37	3.36	400	340	-1.2	++	++	1.6	0	Maybridge
27	3.30	3.24	3.26	500	450	-1.1	+	+	1.5	1	ChemBridge
28	<3.00	3.06	2.98	>1000	1000	+1.0	+	+	1.5	1	ChemBridge

^a Activity scale: highly active (<50 μ M, +++), moderately active (50–450 μ M, ++), and inactive (>450 μ M, +).

Table 4. Measured and Predicted Activities of the Test Set Obtained from CoMFA, CoMSIA, and Hypo1

SARS-CoV 3CL ^{pro} inhibitors	act pIC ₅₀	CoMFA	CoMSIA	act IC ₅₀ (μ M)	pred IC ₅₀ (μ M)	catalyst pharmacophore hypothesis Hypo1			supplier
		pred pIC ₅₀	pred pIC ₅₀			error	act activity scale ^a	pred activity scale ^a	
4	4.92	5.00	4.84	12	11	-1.1	+++	+++	Maybridge
7	4.82	4.75	4.90	15	15	-1.0	+++	+++	Maybridge
8	4.82	4.77	4.84	15	6.3	-2.4	+++	+++	Maybridge
11	4.39	4.13	4.02	40	350	+8.8	+++	++	Maybridge
12	4.34	4.17	4.05	45	42	-1.1	+++	+++	ChemBridge
14	4.22	4.21	3.85	60	130	+2.1	++	++	ChemBridge
15	4.00	3.79	4.04	100	440	+4.4	++	++	SPECS SC
18	3.69	3.30	3.44	200	230	+1.2	++	++	ChemBridge
21	3.60	3.63	3.30	250	290	+1.2	++	++	Maybridge
23	3.52	3.78	3.31	300	760	+2.5	++	+	ChemBridge

^a Activity scale: highly active (<50 μ M, +++), moderately active (50–450 μ M, ++), and inactive (>450 μ M, +).

acceptor, hydrogen-bond donor, and excluded volume, were selected on the basis of the best CoMSIA results (Table 2) to generate the pharmacophore hypotheses for SARS-CoV 3CL^{pro} inhibitors using the Catalyst 4.9 program. Among the 10 hypotheses generated by Catalyst 4.9, the Hypo1 hypothesis was the best pharmacophore hypothesis as characterized by the lowest error cost, highest cost difference (Δ cost), lowest root-mean-square deviation, and the best correlation coefficient (Table 5). The configuration cost of Hypo1 hypothesis was only 10.4 bits, which was also smaller than the criterion of 17 bits for a good hypothesis. The Hypo1 hypothesis was then validated using the CatScramble module of the Catalyst 4.9 program. The validation proceeded with a random reassignment of activity values, that is, the generation of random spreadsheets among the molecules of the training set. To achieve a confidence level of 95% (significance = $1 - ((1+0)/(19+1)) \times 100\% = 95\%$), 19 random spreadsheets (random hypotheses) were generated, and the corresponding statistics are listed in Table 6. The validation clearly shows that the Hypo1 hypothesis is not generated by chance because its statistics are far more superior to those of the 19 random hypotheses generated (Table 6).

The Hypo 1 hypothesis was also evaluated for its capability to predict compound activities using the same training and test sets as those in the CoMFA and CoMSIA studies. The HypoRefine module of the Catalyst 4.9 program was used to further refine the Hypo1 hypothesis. The activities predicted by the refined Hypo1 hypothesis (Tables 3 and 4) were labeled according to the activity scales defined as +++ for highly active

Table 5. Information of Statistical Significance and Predictive Power Presented in Cost Values for the Top 10 Hypotheses Generated^a

hypothesis no.	total cost	Δ cost	rms deviation	correlation (<i>r</i>)
1	64.320	95.564	0.919	0.966
2	77.009	82.875	1.506	0.906
3	77.873	82.011	1.537	0.902
4	78.190	81.694	1.547	0.901
5	81.873	78.011	1.654	0.886
6	83.518	76.366	1.726	0.875
7	87.336	72.548	1.845	0.856
8	87.549	72.335	1.852	0.855
9	93.743	66.141	2.013	0.826
10	94.749	65.135	2.059	0.817

^a Null cost of top-ten score hypotheses is 159.884 bits. Fixed cost is 56.547 bits. Configuration cost is 10.456 bits.

(IC₅₀ < 50 μ M), ++ for moderately active (IC₅₀ = 50–450 μ M), and + for inactive (IC₅₀ > 450 μ M). The accuracy of the hypothesis can be judged by the agreement of the predicted activities with the measured ones (Tables 3 and 4). The prediction accuracy of the Hypo1 hypothesis for the training set is 100% because all of the activity scales of the predicted ones were consistent with the measured ones (Table 3). However, the prediction accuracy for the test set is 80% because the activity scales of two compounds, **11** and **23**, were in conflict with the measured ones (Table 4). Linear regression of the predicted versus measured activities for the training and test sets yielded correlation coefficients of 0.966 and 0.875, respectively.

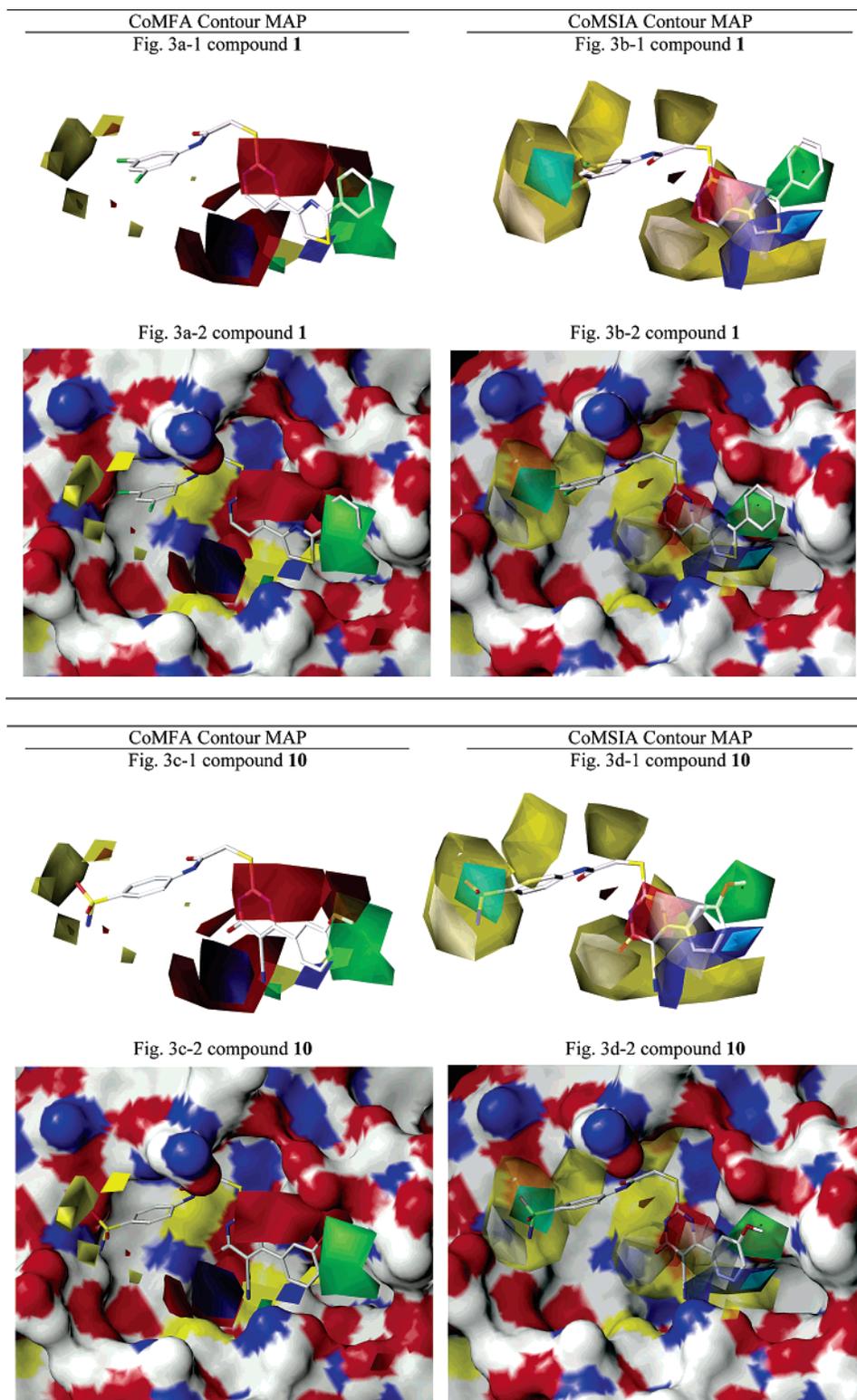


Figure 3. (a) CoMFA contour map displayed with **1** and superimposed on the SARS-CoV 3CL^{pro} active site. (b) CoMSIA contour map displayed with **1** and superimposed on the SARS-CoV 3CL^{pro} active site. (c) CoMFA contour map displayed with **10** and superimposed on the SARS-CoV 3CL^{pro} active site. (d) CoMSIA contour map displayed with **10** and superimposed on the SARS-CoV 3CL^{pro} active site. The red contours represent favored regions for negative charge, blue contours represent favored regions for positive charge, yellow contours represent disfavored regions for steric interaction, green contours represent favored regions for steric interaction, cyan contours represent favored regions for hydrophobic interaction, and white contours represent disfavored regions for hydrophobic interaction.

Mapping the Hypo1 Hypothesis onto the Protein Active Site. The Hypo1 hypothesis displayed with template structure **10** was superimposed on the active site of SARS-CoV 3CL^{pro} (Figure 4). As shown in Figure 4, the hydrophobic feature represented by blue spheres was correctly mapped onto subsite

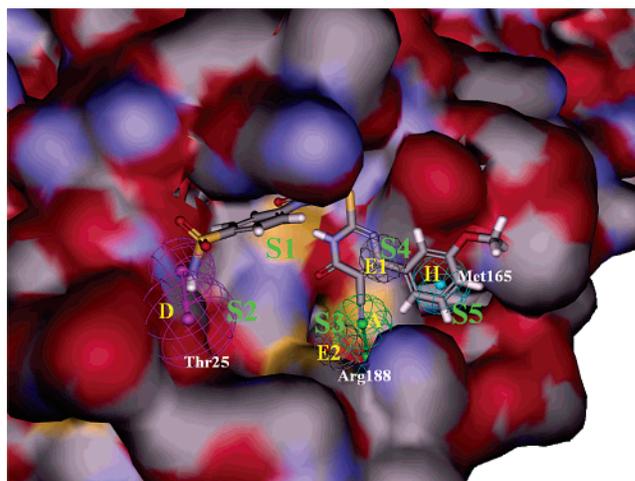
S5, where hydrophobic interactions are preferable as represented by gray surfaces. The hydrogen-bond donor features, represented by magenta spheres, were also correctly mapped onto subsite S2, near Thr25 (Figure 4). Furthermore, the hydrogen-bond acceptor features shown by green spheres of the Hypo1

Table 6. Validation of the Hypo1 Hypothesis using the CatScramble Program Implemented in the Catalyst Package^a

validation no.	total cost	fixed cost	rms deviation	correlation (r)	configuration cost
results for unscrambled					
	64.320	56.547	0.919	0.966	10.456
results for scrambled					
trial 01	77.177	57.932	1.434	0.913	11.840
trial 02	110.535	52.911	2.524	0.695	6.820
trial 03	106.805	55.589	2.373	0.741	9.497
trial 04	106.482	55.880	2.369	0.743	9.789
trial 05	134.756	53.770	2.997	0.562	7.679
trial 06	122.493	57.562	2.680	0.650	11.471
trial 07	107.593	56.547	2.355	0.744	10.456
trial 08	110.834	55.880	2.467	0.705	9.789
trial 09	123.938	52.911	2.803	0.595	6.820
trial 10	154.702	44.966	3.491	0.000	0.000
trial 11	100.979	58.007	2.174	0.786	11.916
trial 12	90.083	56.547	1.777	0.864	10.456
trial 13	65.001	57.932	0.835	0.971	11.840
trial 14	155.464	44.966	3.503	0.000	0.000
trial 15	66.481	55.376	1.110	0.948	9.285
trial 16	81.002	52.911	1.766	0.863	6.820
trial 17	93.446	55.880	2.036	0.813	9.789
trial 18	90.438	55.880	1.924	0.834	9.789
trial 19	156.513	44.966	3.520	0.000	0.000

^a Null cost = 159.884.

hypothesis were observed around subsite S4 near Arg188. To further explore the nature of the pharmacophore of SARS-CoV 3CL^{pro}, the Hypo1 hypothesis was mapped onto the structures of eight highly active SARS-CoV 3CL^{pro} inhibitors, where **1**, **2**, **3**, **5**, and **6** were selected from the training set (Table 3 and Figure 5a–e), and **4**, **7**, and **8** were selected from the test set (Table 4 and Figure 5f–h). All of these inhibitors matched well with three features, including hydrogen-bond donor, hydrogen-bond acceptor, and hydrophobic features. The hydrogen-bond donor features were mapped onto the N atom of the amide

**Figure 4.** Features of the Hypo1 hypothesis displayed with **10** and superimposed on the SARS-CoV 3CL^{pro} active site. The pharmacophore features of the Hypo1 hypothesis are color coded as follows: the black spheres represent two excluded volumes (E1 and E2), the blue spheres represent the hydrophobic (H) feature, the magenta spheres represent the hydrogen-bond donor (D), and the green spheres represent the hydrogen-bond acceptor (A).

group, the hydrogen-bond donor for all these inhibitors. The hydrogen-bond acceptor features were mapped onto the S atom of the thiazole group of **1**, the N atom of the methyl-pyrazole group of **2**, the S atom of the thiophene group of **3**, the S atom of thiazole group of **5**, the N atom of the isoxazole group of **6**, the S atom of the thiophene group of **4**, the N atom of the isoxazole group of **7**, and the N atom of the isoxazole group of **8** (Figure 5). The hydrophobic features are mapped onto various substituted aryl or heteroaryl groups, such as chlorobenzene, dichlorobenzene, and thiazole. Finally, the features of the

Table 7. Measured Activities and Calculated Binding Free Energy of the SARS-CoV 3CL^{pro} Inhibitors

SARS-CoV 3CL ^{pro} inhibitor	energy components calculated for binding ^a						exp
	ΔH_{vdw}	ΔH_{elec}	ΔG_{GB}	ΔG_{np}	$-T\Delta S_{\text{conf}}$	$\Delta G_{\text{binding}}$	IC ₅₀ ^b
1	-50.40	-29.75	40.17	-6.22	23.03	-23.17	3
2	-40.74	-20.61	27.48	-5.97	17.16	-22.68	10
3	-54.02	-16.31	28.01	-6.35	24.17	-24.50	11
4	-36.36	-20.81	28.82	-4.22	9.27	-23.30	12
5	-47.97	-27.08	35.10	-5.38	21.57	-23.76	14
6	-49.38	-25.46	30.05	-6.00	29.27	-21.52	15
7	-45.96	-29.07	33.28	-6.19	26.86	-21.08	15
8	-39.58	-30.68	37.54	-5.41	16.34	-21.79	15
9	-37.96	-43.40	46.54	-4.96	17.91	-21.87	30
10	-45.31	-50.22	60.04	-4.48	24.91	-15.06	40
11	-38.80	-15.00	27.73	-4.97	11.96	-19.08	40
12	-38.02	-31.16	36.97	-3.86	12.86	-17.32	45
13	-42.11	-37.27	45.11	-5.62	23.15	-16.74	60
14	-40.03	-52.47	58.68	-5.84	22.21	-17.45	60
15	-37.31	-34.14	36.34	-5.14	20.31	-19.94	100
16	-36.17	-33.27	28.74	-3.56	22.28	-21.98	200
17	-42.37	-36.72	37.85	-3.99	24.58	-20.65	200
18	-42.18	-36.39	37.81	-5.14	23.50	-22.40	200
19	-47.39	-29.60	36.67	-6.32	29.62	-17.02	200
20	-38.08	-28.02	40.05	-5.11	14.75	-16.41	200
21	-38.90	-31.77	36.29	-5.41	25.09	-14.70	250
22	-39.35	-21.99	32.60	-3.40	17.15	-14.99	300
23	-44.11	-36.49	45.95	-5.66	25.83	-14.48	300
24	-31.76	-48.47	53.06	-4.44	18.58	-13.02	300
25	-30.07	-43.11	45.69	-3.61	18.06	-13.04	350
26	-46.91	-23.44	34.00	-5.75	29.18	-12.92	400
27	-35.30	-13.75	22.93	-5.46	19.30	-12.28	500
28	-35.86	-27.12	29.23	-4.76	30.91	-7.60	>1000

^a All values are in kcal/mol at 300 K. ΔH_{vdw} , van der Waals energy; ΔH_{elec} , Coulombic energy; ΔG_{GB} , polar solvation free energy; ΔG_{np} , nonpolar solvation free energy; $\Delta H_{\text{gas}} = \Delta H_{\text{vdw}} + \Delta H_{\text{elec}}$ and $\Delta G_{\text{solv}} = \Delta G_{\text{GB}} + \Delta G_{\text{np}}$; $T\Delta S_{\text{conf}}$, total entropy contribution; $\Delta G_{\text{binding}}$ (binding free energy) = $\Delta H_{\text{gas}} + \Delta G_{\text{solv}} - T\Delta S_{\text{conf}}$. ^b Measured IC₅₀ values are in μM .

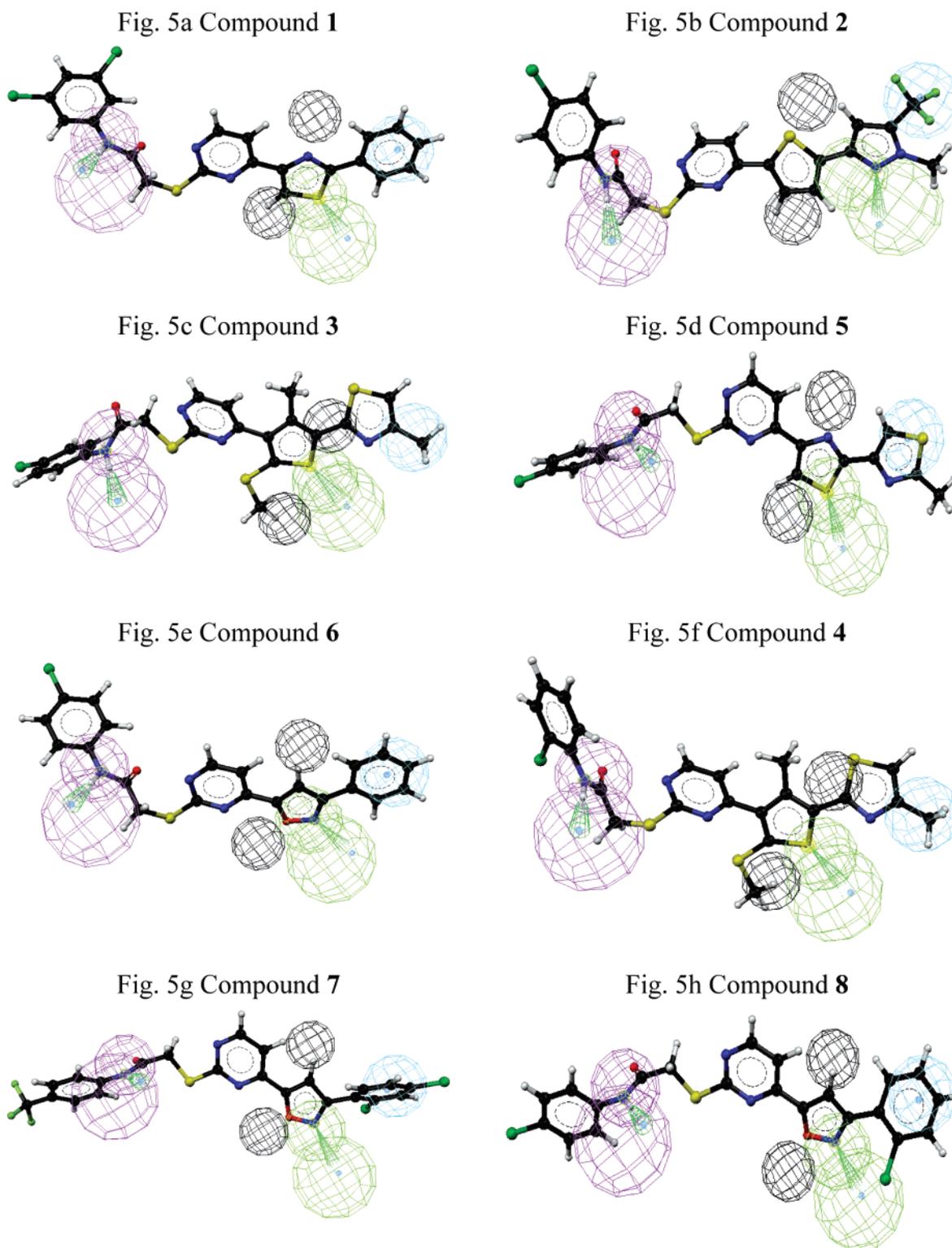


Figure 5. Features of the Hypo1 hypothesis mapped onto the structures of eight highly active SARS-CoV 3CL^{pro} inhibitors 1, 2, 3, 5, 6, 4, 7, and 8. The pharmacophore features are color coded as follows: the black spheres represent two excluded volumes, the blue spheres represent the hydrophobic feature, the violet spheres represent the hydrogen-bond donor, and the green spheres represent the hydrogen-bond acceptor.

excluded volume represented by black spheres were correctly located because no bulky groups were found around those spheres (Figure 5).

Correlation between Binding Free Energy and Inhibitory Activity. The MM/GBSA (molecular mechanics-generalized Born surface area) simulation method was applied to calculate the binding free energies of 28 inhibitors. The solvent contribution and entropy penalty upon binding to the protein were

included in the calculation. The calculated binding free energies ($\Delta G_{\text{binding}}$) of 28 inhibitors and their corresponding experimental activities (IC_{50}) are shown in Table 7. The correlation between calculated binding energies and experimental activities gave a correlation coefficient (r^2) of 0.667. To further understand major determinants for the binding of inhibitors to the protein, the binding free energy of each inhibitor was decomposed to different contributions (i.e., van der Waals energy, Coulombic

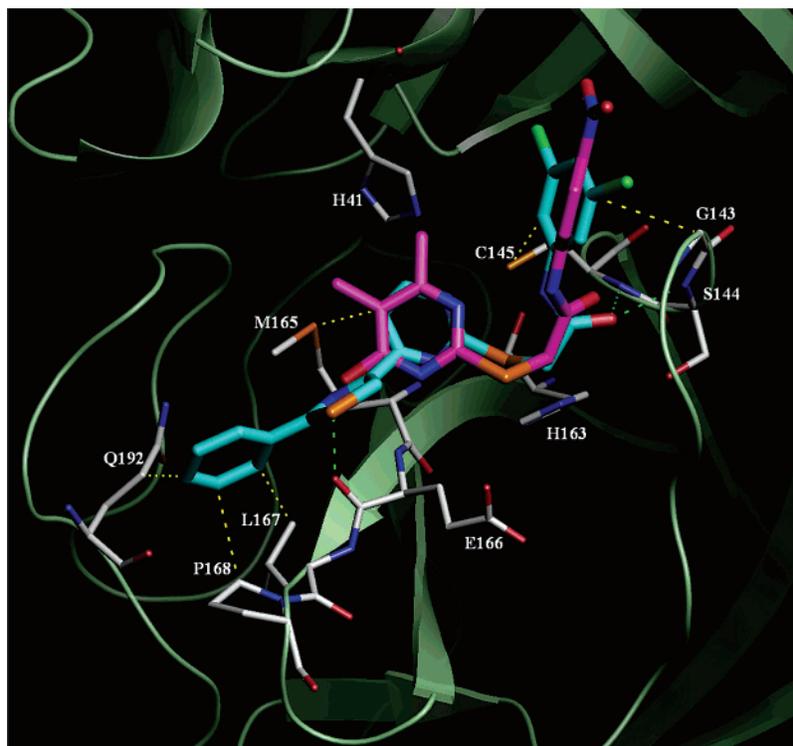


Figure 6. Superposition of the best inhibitor **1** (cyan) with the worst inhibitor **28** (magenta) in the active site of SARS-CoV 3CL^{pro}. The secondary structure elements are shown as a ribbon drawing, and the important residues involved in inhibitor binding are labeled. The green and yellow dotted lines represent H bonding and hydrophobic interactions of **1** with SARS-CoV 3CL^{pro}, respectively.

energy, solvation free energy, and entropy). As shown in Table 7, both van der Waals and electrostatic interactions with the protein dominate the binding process. Particularly, the van der Waals interactions are a major contributor to the binding of inhibitors. Furthermore, superimposition of the predicted complex structure of the best inhibitor (**1**) with that of the worst one (**28**) provides insights into their difference in binding (Figure 6). The core structure of **1** superimposes well with that of **28** except for the slight movement of the benzene group to have a better fit with the surrounding residues including Gly143 and Cys145. However, the two additional aromatic rings of **1**, thiazole and benzene groups, form strong interactions with Glu166, Leu167, Pro168, and Gln192, leading to its increased potency over **28** (Figure 6). All of these results demonstrate that the MM/GBSA analysis could be used to predict the binding free energy that correlates with experimental activities and, consequently, could be combined with current 3D-QSAR studies to design the next generation of drug leads with more potency against SARS-CoV.

Conclusion

In this article, we present a successful example of employing structure-based virtual screening in combination with an analogue search to discover a novel family of SARS-CoV 3CL^{pro} inhibitors. Twenty-eight compounds in the family with IC₅₀ values that ranged from 3 to 1000 μ M were selected for further 3D-QSAR studies. Three QSAR models, including CoMFA, CoMSIA, and pharmacophore hypothesis, have been constructed and showed consistency among these three models. The chemical interpretation of the contour maps generated by CoMFA, CoMSIA, and the pharmacophore hypothesis could be mapped well into the binding site of the SARS-CoV 3CL^{pro} and reveal the important sites where steric, hydrophobic, and electrostatic interactions could significantly contribute to the

inhibition of the target protein. These 3D-QSAR models could, therefore, guide the direction for compounds modification and facilitate further lead optimization. Moreover, the compound activities predicted by all of these 3D-QSAR models are in good agreement with experimental data, demonstrating their predictive ability and indicating that they could be used to estimate the activities of new inhibitors. Finally, the binding free energy of each inhibitor is calculated to gain insights into their binding to the protein and assist in the explanation of the structure–activity relationship obtained.

In conclusion, our study proves that the combination of structure-based virtual screening and 3D-QSAR study could be a useful approach to efficiently identify novel inhibitors from a large chemical database and provide rationales for further lead optimization. Our 3D-QSAR models could also be employed to give reasonable estimations of the activity of newly designed inhibitors before biological testing so that inhibitors predicted to have strong affinity could be prioritized for chemical synthesis.

Acknowledgment. We thank Professor Lindsay Sawyer (Institute of Cell & Molecular Biology, University of Edinburgh, U.K.), Professor Malcolm D. Walkinshaw (Institute of Cell & Molecular Biology, University of Edinburgh, U.K.), Dr. Min-Yong Li (Department of Chemistry, Georgia State University, Atlanta, GA), and Professor Kung-Chia Young, (Department of Medical Laboratory Science and Biotechnology, National Cheng Kung University, Tainan, Taiwan) for fruitful discussions. We also thank Ms. Hsiao-Wen Edith Chu, Mr. Tai-Tsung Chen, and Miss Hsiu-Hsiu Huang for administrative support. Financial support was provided by National Health Research Institutes and National Science Council of the Republic of China Grant Nos. NSC 92-2751-B-400-001-Y, NSC93-2811-B-400-001, NSC92-2313-B-007-002, and NSC93-2313-B-007-002. The CoMFA, CoMSIA, PLS, and Catalyst 4.9 studies were

conducted at the National Center for High Performance Computing, Taiwan. IDEA2.0 package was provided by BREADTH (<http://www.breadth.com.tw>).

References

- Lee, N.; Hui, D.; Wu, A.; Chan, P.; Cameron, P.; Joynt, G. M.; Ahuja, A.; Yung, M. Y.; Leung, C. B.; To, K. F.; Lui, S. F.; Szeto, C. C.; Chung, S.; Sung, J. J. A major outbreak of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* **2003**, *348*, 1986–94.
- Poutanen, S. M.; Low, D. E.; Henry, B.; Finkelstein, S.; Rose, D.; Green, K.; Tellier, R.; Draker, R.; Adachi, D.; Ayers, M.; Chan, A. K.; Skowronski, D. M.; Salit, I.; Simor, A. E.; Slutsky, A. S.; Doyle, P. W.; Krajden, M.; Petric, M.; Brunham, R. C.; McGeer, A. J. Identification of severe acute respiratory syndrome in Canada. *N. Engl. J. Med.* **2003**, *348*, 1995–2005.
- Holmes, K. V. SARS-associated coronavirus. *N. Engl. J. Med.* **2003**, *348*, 1948–1951.
- Leng, Q.; Bentwich, Z. A novel coronavirus and SARS. *N. Engl. J. Med.* **2003**, *349*, 709.
- Rota, P. A.; Oberste, M. S.; Monroe, S. S.; Nix, W. A.; Campagnoli, R.; Icenogle, J. P.; Penaranda, S.; Bankamp, B.; Maher, K.; Chen, M. H.; Tong, S.; Tamin, A.; Lowe, L.; Frace, M.; DeRisi, J. L.; Chen, Q.; Wang, D.; Erdman, D. D.; Peret, T. C.; Burns, C.; Ksiazek, T. G.; Rollin, P. E.; Sanchez, A.; Liffick, S.; Holloway, B.; Limor, J.; McCaustland, K.; Olsen-Rasmussen, M.; Fouchier, R.; Gunther, S.; Osterhaus, A. D.; Drosten, C.; Pallansch, M. A.; Anderson, L. J.; Bellini, W. J. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **2003**, *300*, 1394–1399.
- Marra, M. A.; Jones, S. J.; Astell, C. R.; Holt, R. A.; Brooks-Wilson, A.; Butterfield, Y. S.; Khattri, J.; Asano, J. K.; Barber, S. A.; Chan, S. Y.; Cloutier, A.; Coughlin, S. M.; Freeman, D.; Girn, N.; Griffith, O. L.; Leach, S. R.; Mayo, M.; McDonald, H.; Montgomery, S. B.; Pandoh, P. K.; Petrescu, A. S.; Robertson, A. G.; Schein, J. E.; Siddiqui, A.; Smailus, D. E.; Stott, J. M.; Yang, G. S.; Plummer, F.; Andonov, A.; Artsob, H.; Bastien, N.; Bernard, K.; Booth, T. F.; Bowness, D.; Czub, M.; Drebot, M.; Fernando, L.; Flick, R.; Garbutt, M.; Gray, M.; Grolla, A.; Jones, S.; Feldmann, H.; Meyers, A.; Kabani, A.; Li, Y.; Normand, S.; Stroher, U.; Tipples, G. A.; Tyler, S.; Vogrig, R.; Ward, D.; Watson, B.; Brunham, R. C.; Krajden, M.; Petric, M.; Skowronski, D. M.; Upton, C.; Roper, R. L. The Genome sequence of the SARS-associated coronavirus. *Science* **2003**, *300*, 1399–1404.
- Snijder, E. J.; Bredenbeek, P. J.; Dobbe, J. C.; Thiel, V.; Ziebuhr, J.; Poon, L. L.; Guan, Y.; Rozanov, M.; Spaan, W. J.; Gorbalenya, A. E. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* **2003**, *331*, 991–1004.
- Thiel, V.; Ivanov, K. A.; Putics, A.; Hertzog, T.; Schelle, B.; Bayer, S.; Weissbrich, B.; Snijder, E. J.; Rabenau, H.; Doerr, H. W.; Gorbalenya, A. E.; Ziebuhr, J. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **2003**, *84*, 2305–2315.
- Anand, K.; Ziebuhr, J.; Wadhvani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus main proteinase (3CL^{pro}) structure: basis for design of anti-SARS drugs. *Science* **2003**, *300*, 1763–1767.
- Liu, S.; Pei, J.; Chen, H.; Zhu, X.; Liu, Z.; Ma, W.; He, F.; Lai, L. Modeling of the SARS coronavirus main proteinase and conformational flexibility of the active site. *Beijing Daxue Xuebao* **2003**, *35*, 62–65.
- Xiong, B.; Gui, C. S.; Xu, X. Y.; Luo, C.; Chen, J.; Luo, H. B.; Chen, L. L.; Li, G. W.; Sun, T.; Yu, C. Y.; Yue, L. D.; Duan, W. H.; Shen, J. K.; Qin, L.; Shi, T. L.; Li, Y. X.; Chen, K. X.; Luo, X. M.; Shen, X.; Shen, J. H.; Jiang, H. L. A 3D model of SARS_CoV 3CL proteinase and its inhibitors design by virtual screening. *Acta Pharm. Sin.* **2003**, *24*, 497–504.
- Chen, L.; Gui, C.; Luo, X.; Yang, Q.; Gunther, S.; Scandella, E.; Drosten, C.; Bai, D.; He, X.; Ludewig, B.; Chen, J.; Luo, H.; Yang, Y.; Yang, Y.; Zou, J.; Thiel, V.; Chen, K.; Shen, J.; Shen, X.; Jiang, H. Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication in vitro. *J. Virol.* **2005**, *79*, 7095–7103.
- Chen, S.; Chen, L.; Tan, J.; Chen, J.; Du, L.; Sun, T.; Shen, J.; Chen, K.; Jiang, H.; Shen, X. Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization. Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. *J. Biol. Chem.* **2005**, *280*, 164–173.
- Jenwitheesuk, E.; Samudrala, R. Identifying inhibitors of the SARS coronavirus proteinase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3989–3992.
- Liu, B.; Zhou, J. SARS-CoV protease inhibitors design using virtual screening method from natural products libraries. *J. Comput. Chem.* **2005**, *26*, 484–490.
- Liu, Z.; Huang, C.; Fan, K.; Wei, P.; Chen, H.; Liu, S.; Pei, J.; Shi, L.; Li, B.; Yang, K.; Liu, Y.; Lai, L. Virtual screening of novel noncovalent inhibitors for SARS-CoV 3C-like proteinase. *J. Chem. Inf. Model.* **2005**, *45*, 10–17.
- Toney, J. H.; Navas-Martin, S.; Weiss, S. R.; Koeller, A. Sabadine: a potential non-peptide anti-severe acute-respiratory-syndrome agent identified using structure-aided design. *J. Med. Chem.* **2004**, *47*, 1079–1080.
- Chou, K. C.; Wei, D. Q.; Zhong, W. Z. Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 148–151.
- Sirois, S.; Wei, D. Q.; Du, Q.; Chou, K. C. Virtual screening for SARS-CoV protease based on KZ7088 pharmacophore points. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1111–1122.
- Zhang, X. W.; Yap, Y. L.; Altmeyer, R. M. Generation of predictive pharmacophore model for SARS-coronavirus main proteinase. *Eur. J. Med. Chem.* **2005**, *40*, 57–62.
- Anand, K.; Palm, G. J.; Mesters, J. R.; Siddell, S. G.; Ziebuhr, J.; Hilgenfeld, R. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* **2002**, *21*, 3213–3224.
- Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13190–13195.
- Bacha, U.; Barrila, J.; Velazquez-Campoy, A.; Leavitt, S. A.; Freire, E. Identification of novel inhibitors of the SARS coronavirus main protease 3CL^{pro}. *Biochemistry* **2004**, *43*, 4906–4912.
- Chen, L. R.; Wang, Y. C.; Lin, Y. W.; Chou, S. Y.; Chen, S. F.; Liu, L. T.; Wu, Y. T.; Kuo, C. J.; Chen, T. S.; Juang, S. H. Synthesis and evaluation of isatin derivatives as effective SARS coronavirus 3CL protease inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3058–3062.
- Chen, C. N.; Lin, C. P.; Huang, K. K.; Chen, W. C.; Hsieh, H. P.; Liang, P. H.; Hsu, J. T. Inhibition of SARS-CoV 3C-like protease activity by theaflavin-3,3'-digallate (TF3). *Evidence-Based Complement. Alternat. Med.* **2005**, *2*, 209–215.
- Kaeppeler, U.; Stiefl, N.; Schiller, M.; Vicik, R.; Breuning, A.; Schmitz, W.; Rupprecht, D.; Schmuck, C.; Baumann, K.; Ziebuhr, J.; Schirmeister, T. A new lead for nonpeptidic active-site-directed inhibitors of the severe acute respiratory syndrome coronavirus main protease discovered by a combination of screening and docking methods. *J. Med. Chem.* **2005**, *48*, 6832–6842.
- Wu, C. Y.; Jan, J. T.; Ma, S. H.; Kuo, C. J.; Juan, H. F.; Cheng, Y. S.; Hsu, H. H.; Huang, H. C.; Wu, D.; Brik, A.; Liang, F. S.; Liu, R. S.; Fang, J. M.; Chen, S. T.; Liang, P. H.; Wong, C. H. Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10012–10017.
- Dooley, A. J.; Shindo, N.; Taggart, B.; Park, J. G.; Pang, Y. P. From genome to drug lead: identification of a small-molecule inhibitor of the SARS virus. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 830–833.
- Blanchard, J. E.; Elowe, N. H.; Huitema, C.; Fortin, P. D.; Cecchetto, J. D.; Eltis, L. D.; Brown, E. D. High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase. *Chem. Biol.* **2004**, *11*, 1445–1453.
- Kao, R. Y.; Tsui, W. H.; Lee, T. S.; Tanner, J. A.; Watt, R. M.; Huang, J. D.; Hu, L.; Chen, G.; Chen, Z.; Zhang, L.; He, T.; Chan, K. H.; Tse, H.; To, A. P.; Ng, L. W.; Wong, B. C.; Tsoi, H. W.; Yang, D.; Ho, D. D.; Yuen, K. Y. Identification of novel small-molecule inhibitors of severe acute respiratory syndrome-associated coronavirus by chemical genetics. *Chem. Biol.* **2004**, *11*, 1293–1299.
- Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18*, 1175–1189.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.
- Marsili, M.; Gasteiger, J. Pi-charge distributions from molecular topology and pi-orbital electronegativity. *Croat. Chem. Acta* **1980**, *53*, 601–614.

- (35) Purcell, W. P.; Singer, J. A. Brief review and table of semiempirical parameters used in the Hückel molecular orbital method. *J. Chem. Eng. Data* **1967**, *12*, 235–246.
- (36) SYBYL 6.9.1; The Tripos Associates: St. Louis, MO.
- (37) Kuo, C. J.; Chi, Y. H.; Hsu, J. T.; Liang, P. H. Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. *Biochem. Biophys. Res. Commun.* **2004**, *318*, 862–867.
- (38) Sutter, J. G. N. O.; Hoddmann, R.; Li, H.; Waldman, M. Pharmaco-phore, perception, development and use in drug design; International University Line: La Jolla, CA 2000; pp 504–506.
- (39) M. Clark, R. D. C., III.; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (40) Richard, D. C., III; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (41) Catalyst, v. s. p.; Accelrys, Inc.: San Diego, CA 2003. (This was previously known as Molecular Simulations, Inc.; <http://www.accelrys.com>.)
- (42) Brooks, B. R. B. R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (43) Smellie, A. T. S. L.; Tobwin, P. Poling: Promoting conformational variation. *J. Comput. Chem.* **1995**, *16*, 171–187.
- (44) 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.
- (45) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (46) Bayly, C. L. C. P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (47) Case, D. A. P. D. A.; Caldwell, J. W.; Cheatham, T. E., III.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER*, version 8; University of California: San Francisco, CA, 2004.
- (48) Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* **2000**, *56*, 275–291.
- (49) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row. *New York* **1976**.

JM050852F