Structure-Based Rational Quest for Potential Novel Inhibitors of Human HMG-CoA Reductase by Combining CoMFA 3D QSAR Modeling and Virtual Screening

Qing Y. Zhang,[†] Jian Wan,*[,†] Xin Xu,*[,‡] Guang F. Yang,[†] Yan L. Ren,[†] Jun J. Liu,[†] Hui Wang,[§] and Yu Guo[§]

Key Laboratory of Pesticide and Chemical Biology (CCNU) of Ministry of Education, College of Chemistry, Central China Normal University, Wuhan 430079, China, State Key Laboratory of Physical Chemistry of Solid Surfaces, College of Chemistry and Chemical Engineering, Center for Theoretical Chemistry, Xiamen University, Xiamen 361005, China, and Department of Pharmacology, Basic Medical School of Wuhan University, Wuhan 430071, China

Received July 19, 2006

3-Hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) catalyzes the formation of mevalonate. In many classes of organisms, this is the committed step leading to the synthesis of essential compounds, such as cholesterol. However, a high level of cholesterol is an important risk factor for coronary heart disease, for which an effective clinical treatment is to block HMGR using inhibitors like statins. Recently the structures of catalytic portion of human HMGR complexed with six different statins have been determined by a delicate crystallography study (Istvan and Deisenhofer *Science* **2001**, 292, 1160−1164), which established a solid basis of structure and mechanism for the rational design, optimization, and development of even better HMGR inhibitors. In this study, three-dimensional quantitative structure−activity relationship (3D QSAR) with comparative molecular field analysis (CoMFA) was performed on a training set of up to 35 statins and statin-like compounds. Predictive models were established by using two different ways: (1) Models-fit, obtained by SYBYL conventional fit-atom molecular alignment rule, has cross-validated coefficients (q^2) up to 0.652 and regression coefficients (r^2) up to 0.977. (2) Models-dock, obtained by FlexE by docking compounds into the HMGR active site, has cross-validated coefficients (q^2) up to 0.731 and regression coefficients (r^2) up to 0.947. These models were further validated by an external testing set of 12 statins and statin-like compounds. Integrated with CoMFA 3D QSAR predictive models, molecular surface property (electrostatic and steric) mapping and structure-based (both ligand and receptor) virtual screening have been employed to explore potential novel hits for the HMGR inhibitors. A representative set of eight new compounds of non-statin-like structures but with high pIC_{50} values were sorted out in the present study.

Introduction

Cholesterol is a blood fat needed by the body in moderate amounts. High cholesterol levels can lead to coronary artery disease, which is one of the leading causes of mortality and a significant cause of morbidity. One of the most effective classes of drugs for lowering serum low-density lipoprotein cholesterol (LDL-c) concentrations is known as statins. Statins work by blocking an enzyme, 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (HMG-CoA reductase or HMGR), which catalyzes the formation of mevalonate, the rate-limiting step in the manufacture of cholesterol.[1,2] Although current statin drugs have made a great contribution to lower LDL, there is still a demand for even higher efficiency and even less side effects.

All statins share an HMG-like moiety (Figure 1) such that statins compete with the natural substrates for the active binding sites in the enzyme, inhibiting the catalytic role of HMGR. Besides this HMG-like moiety, statins consist of various attachments, which present a hydrophobic anchor. Thus a common strategy in drug designs is to find refined artificial inhibitors with simple aromatic and heteroaromatic motifs to replace the structurally complicated decalin ring system of those naturally led compounds.

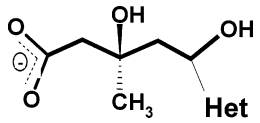


Figure 1. HMG-like moiety of the statins (in bold). Het represents heteroaromatic moiety.

Recently the structures of catalytic portion of human HMRG complexed with six different statins have been determined by a series of delicate crystallography studies,[3,4] which explained the detailed characterization of the active site and the HMGR-statin interactions and, therefore, built up a solid basis of structure and mechanism for the rational design and optimization of even better HMGR inhibitors.

In the present study, an attempt of structure-based rational

* Address whom correspondence to either author. E-mails: jianwan@mail.ccnu.edu.cn (J.W.); xinxu@xmu.edu.cn (X.X.).
[†] Central China Normal University.
[‡] Xiamen University.
[§] Basic Medical School of Wuhan University.

10.1021/cc060101e CCC: $37.00
Published on Web 11/21/2006

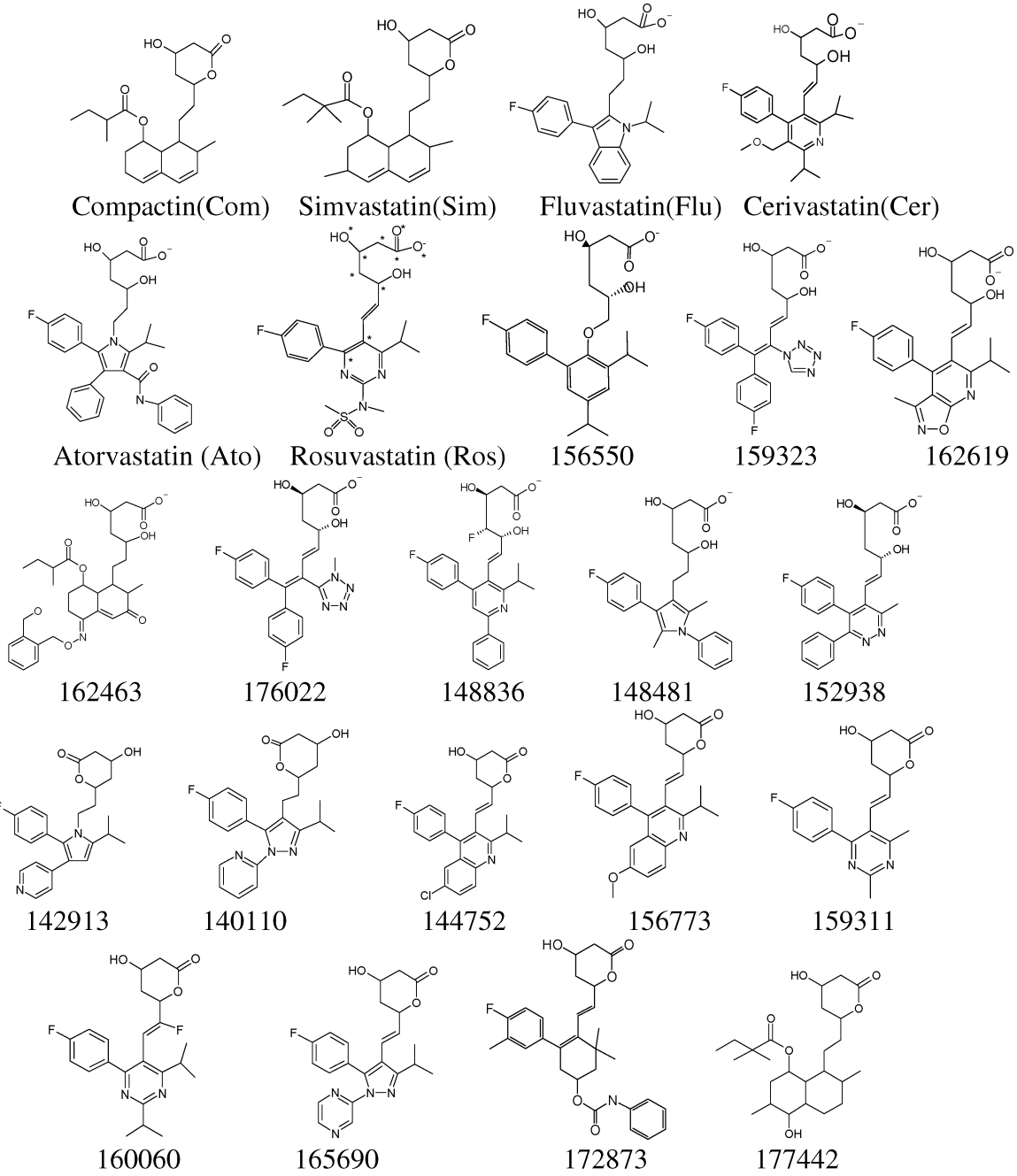
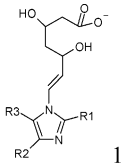


Figure 2. Structures of statins[3,4] and statin-like inhibitors of human HMGR from MDDR database.[8,9] Rosuvastatin is chosen as the template molecule in Models-fit. The atoms used for fitting are marked with asterisks.

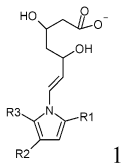
searching of novel active pharmacophore as potentially potent HMGR inhibitors was made. First, comparative molecular field analysis (CoMFA)[5] three-dimensional quantitative structure−activity relationship (3D QSAR) studies were performed on a training set consisting of statins and statin-like compounds (Figures 2 and 3) using two different molecular alignment strategies. Second, the as-built CoMFA models were validated by an external testing set. Meanwhile, the distributions of the electrostatic and steric fields created by the CoMFA studies were further checked by comparing and contrasting them to the MOLCAD-generated electrostatic and steric potential surface maps based upon the crystal structure of the HMGR active site. Finally, a new definition of active pharmacophore was generated by the fingerprints hypothesis technology of Tuplets,[6] which was then used in the virtual screening of the LeadQuest database.[7] In addition to the conventional statin-like compounds with HMG-like moiety, some compounds with totally different pharmacophore structure have also been screened out as well. Eight compounds with high screening scores obtained by jointly using a series of virtual screening methods were further supported by the CoMFA models with high pIC_{50} values, lending credit to our current attempts in the quest for potential HMGR inhibitors with new active pharmacophore.

Materials and Methods

Data Sets for CoMFA Modeling. A set of HMGR inhibitors was generated carefully. This data set contains the six statins (Figure 2) used in the crystal structure determination of the enzyme−statin complexes by Deisenhofer and



1-16



17-24

Compound	R_1	R_2	R_3
1	CF_3	4-F-C_6H_4	4-F-C_6H_4
2	CH_3	4-F-C_6H_4	4-F-C_6H_4
3	Me_2N	4-F-C_6H_4	4-F-C_6H_4
4	4-F-C_6H_4	4-F-C_6H_4	Me_2CH
5	Me_2CH	4-F-C_6H_4	3-Cl-C_6H_4
6	Me_2CH	4-F-C_6H_4	3,5-Di-Cl-C_6H_3
7	Me_2CH	4-F-C_6H_4	3,5-Di-Me-C_6H_3
8	Me_2CH	4-F-C_6H_4	2-Me-4-F-C_6H_3
9	Me_2CH	4-F-C_6H_4	3,5-Di-Me-4-F-C_6H_2
10	Me_2CH	3,5-Di-Me-4-Cl-C_6H_2	4-F-C_6H_4
11	Me_2CH	3-Pyridyl	4-F-C_6H_4
12	Me_2CH	3-SO_2Me-C_6H_4	4-F-C_6H_4
13	Me_2CH	3-NHMe-C_6H_4	4-F-C_6H_4
14*	Me_2CH	4-F-C_6H_4	4-F-C_6H_4
15*	Me_2CH	4-F-C_6H_4	4-F-C_6H_4
16	Me_2CH	4-F-C_6H_4	4-F-C_6H_4
17	Me_2CH	4-F-C_6H_4	4-F-C_6H_4
18	Me_2CH	C_6H_5	4-F-C_6H_4
19	Me_2CH	3-Cl-C_6H_4	4-F-C_6H_4
20	Me_2CH	3-Br-C_6H_4	4-F-C_6H_4
21	Me_2CH	4-Pyridyl	4-F-C_6H_4
22	Me_2CH	2-Pyridyl	4-F-C_6H_4
23*	Me_2CH	4-F-C_6H_4	4-F-C_6H_4
24	Me_2CH	4-F-C_6H_4	C_6H_5

14*=3S,5R; 15*= 3R,5R; 23* = Bromine substitution at C_4 of pyrrole ring.

Figure 3. Structures of 24 pyrimidine-/pyrrole-substituted 3,5-dihydroxyheptenoates.[10,11]

Table 1. Summary of Results for the CoMFA Binding Models with Two Types of Molecular Alignments[a]

		q^2	n	r^2	F	SE	E %	S %
Models-fit	STD	0.361	6	0.974	173.548	0.150	25.7	74.3
	RF	0.652	6	0.977	200.120	0.140	25.0	75.0
Models-dock	STD	0.562	6	0.935	66.910	0.237	41.9	58.1
	RF	0.731	6	0.947	83.064	0.214	42.0	58.0

[a] STD, standard scaling. RF, region focusing (StDev × coefficient). n = the optimal number of components to be used in the final analysis. F = the ratio of r^2 to 1.0 − r^2 (explained to unexplained), weighted so that the fewer the explanatory properties and the more the values of the target property, the higher the F ratio. SE, standard error. E %, contribution of electrostatic field. S %, contribution of steric field.

co-workers.[3,4] The rest of the 17 compounds with ID numbers shown in Figure 2 were taken from the MACCS-II Drug Data Report (MDDR)[8,9] database. Additionally, 24 pyridine-/pyrrole-substituted 3,5-dihydroxyheptenoates (Figure 3) were reported[10,11] to be potent inhibitors of HMGR and, therefore, were selected into the data set as well.

The training set used for developing CoMFA binding models was composed of 35 inhibitors with the pIC_{50} (−log IC_{50}: the term IC_{50} represents the concentration of an inhibitor that is required for 50% inhibition of an enzyme in vitro) values in a range from 5.7 to 9.52. These inhibitors were basically randomly chosen from the data set generated above but were aimed to cover the pIC_{50} values as wide as possible to be more representative. The testing set comprised the remaining 12 compounds of the data set (see Table 2 for details). The three-dimensional structures of all compounds were obtained by using the SYBYL7.0 program package.[12] Partial atomic charges were calculated by the Gasteiger−Huckel method,[13] and energy minimizations were performed using the Tripos force field[14] and the Powell conjugate gradient algorithm[15] with a convergence criterion of 0.05 kcal/(mol·Å). All calculations in the present study were performed on a SGI origin 300 server.

Molecular Alignment Rules for CoMFA Modeling. Two different alignment rules were adopted in the CoMFA modeling: (1) SYBYL conventional fit-atom molecular alignment rule was applied by using the module of SYBYL/Analyze/Fit-atom. Fit-atom module adjusted the geometry

of the molecule such that its steric and electrostatic fields matched the template molecule. The template molecule chosen in this work is rosuvastatin, which is the most potent statin currently available. The atoms used for fitting are marked with asterisks in the template molecule shown in Figure 2. (2) Molecular alignment using molecular docking active conformation was achieved by FlexE,[16] docking the compounds into the active site of HMGR. FlexE improves over FlexX[17] of SYBYL[18] by not only taking into account the flexible ligand structures as it is in FlexX but also adding the ability to account for the protein structural variability. The efficiency of FlexE has been well-documented in the literature.[19]

The whole docking operation of FlexE was done as follows. The first step was the preparation of the ligands, which is similar to those for FlexX. The second step was the preparation of the ensemble structures of proteins. For each structure, the description of an ensemble contained the definition of the protein atoms, the resolution of ambiguities in the PDB file, the location of hydrogen atoms at the hetero atoms, and the definition of the active site atoms. Torsion angles and the optimal tautomeric histidine states were selected by visual inspection of the protein. The side chains of the lysine and arginine residues were protonated, and the carboxylate groups of aspartic and glutamic acids were ionized. Water molecules contained in the PDB file were removed. In order to define the active site of the protein, all members of an ensemble were superimposed together with their reference ligand structure. Six members[3] (1HW8, 1HW9, 1HWI, 1WHJ, 1HWK, and the reference structure of 1HWL) were taken from the Brookhaven Protein Database (PDB, http://www.rcsb.org/pdb). All atoms located within the range of 6.5 Å distance from any atom of ligands of the ensemble were selected into the active site. Then, a residue was included into the active site if at least one of its atoms was picked out. The superimposed protein structures and the reference ligand positions were stored for later cross docking experiments with FlexX using the same definitions of the active sites. The third step was to dock all ligands into the united protein structure with FlexX. All active conformations were selected from the binding orientation at the active site of HMGR, jointly evaluated by the consideration of FlexE Energy Scores (total score) and molecular geometry.

CoMFA 3D QSAR Modeling. CoMFA steric and electrostatic interaction fields were calculated at each lattice intersection on a regularly spaced grid of 3 Å. The grid pattern was generated automatically by the SYBYL/CoMFA routine. An sp^3 carbon atom with a van der Waals radius of 1.52 Å and a +1.0 charge was used as the probe to calculate the steric (Lennard−Jones 6-12 potential) field and the electrostatic (Coulombic potential) field with a constant dielectric constant at each lattice point. The electrostatic energy at the point where the steric energy exceeded the steric cutoff for any molecule in the analysis was set to the mean value of the non-excluded electrostatic field. Values of the steric and electrostatic fields were truncated at 30.0 kcal/mol. The CoMFA steric and electrostatic fields were scaled by the CoMFA-STD[20] method in SYBYL.

A partial least-squares (PLS) approach,[21−23] which is an extension of multiple regression analysis, was used to derive the 3D QSAR models in which the CoMFA descriptors were used as independent variables, and the experimental pIC_{50} values were used as dependent variables. The cross-validation with Leave-One-Out (LOO) option and the SAMPLS program[24] were applied to obtain the optimal number of components to be used in the final analysis. After the optimal number of components was determined, a non-cross-validated analysis was performed without column filtering. Two best CoMFA models were obtained finally by using the Region focusing[25] with the weights of StDev × coefficient.

Molecular Surface Physicochemical Properties. Surface physicochemical property (surface) maps (i.e., electrostatic potential), hydrophobicity (lipophilicity) potential, and hy-

Table 2. Experimental, Predicted Activities and Residual Values by the CoMFA-fit-RF and CoMFA-dock-RF Models

	pIC_{50}					
	model-fit-RF			model-dock-RF		
compd	actual	predicted	residues	predicted	residues	sets[a]
13	8.52	8.669	−0.15	8.706	−0.19	TR
152938	9.0	9.113	−0.11	8.785	0.22	TR
156773	7.89	7.873	0.02	7.858	0.03	TR
159311	6.89	6.877	0.01	6.995	−0.1	TR
159323	6.92	6.974	−0.05	6.912	0.01	TR
14	6.15	6.155	−0.01	6.009	0.14	TR
160060	8.74	8.777	−0.04	8.780	−0.04	TR
162463	7.74	7.713	0.03	7.597	0.14	TR
162619	8.05	7.948	0.10	8.007	0.04	TR
165690	8.18	8.155	0.03	8.263	−0.08	TR
15	5.7	5.807	−0.11	5.675	0.02	TR
176022	7.72	7.855	−0.13	7.702	0.02	TR
177442	7.31	7.375	−0.06	7.218	0.09	TR
16	9.0	8.735	0.26	9.285	−0.28	TR
17	8.7	8.938	−0.24	8.648	0.05	TR
2	7.0	6.944	0.06	7.237	−0.24	TR
18	8.7	8.800	−0.10	8.514	0.19	TR
19	8.7	8.374	0.33	8.759	−0.06	TR
20	9.3	9.306	−0.01	8.726	0.57	TR
21	8.3	8.222	0.08	8.397	−0.1	TR
22	8.7	8.713	−0.01	8.690	0.01	TR
23	9.52	9.498	0.02	9.079	0.44	TR
24	7.6	7.914	−0.31	8.004	−0.4	TR
3	8.04	8.175	−0.14	8.012	0.03	TR
4	9.0	8.935	0.06	9.213	−0.21	TR
5	8.1	8.043	0.06	7.877	0.22	TR
6	7.66	7.534	0.13	7.770	−0.11	TR
7	7.25	7.276	−0.03	7.476	−0.23	TR
8	8.4	8.424	−0.02	8.389	0.01	TR
Ato	8.1	8.098	0.00	8.210	−0.11	TR
Cer	8	7.828	0.17	8.014	−0.01	TR
Com	7.64	7.632	0.01	7.537	0.10	TR
Flu	7.55	7.427	0.12	7.546	0.00	TR
Ros	8.3	8.171	0.13	8.555	−0.26	TR
Sim	7.96	8.049	−0.09	7.883	0.08	TR
1	7.89	8.430	−0.54	7.824	0.07	TS
9	7.52	7.680	−0.16	8.136	−0.62	TS
10	8.7	8.481	0.22	9.027	−0.33	TS
11	9.0	8.564	0.44	8.921	0.08	TS
12	8.7	8.808	−0.11	8.835	−0.14	TS
140110	7.41	8.499	−1.09	7.349	0.06	TS
142913	7.89	7.396	0.49	8.013	−0.12	TS
144752	7.49	7.893	−0.40	7.714	−0.22	TS
148481	8.06	7.962	0.10	8.190	−0.13	TS
148836	8.07	7.756	0.31	7.660	0.41	TS
156550	8.64	7.311	1.33	8.537	0.10	TS
172873	8.74	8.862	−0.12	8.278	0.46	TS

[a] TS = testing set; TR = training set.

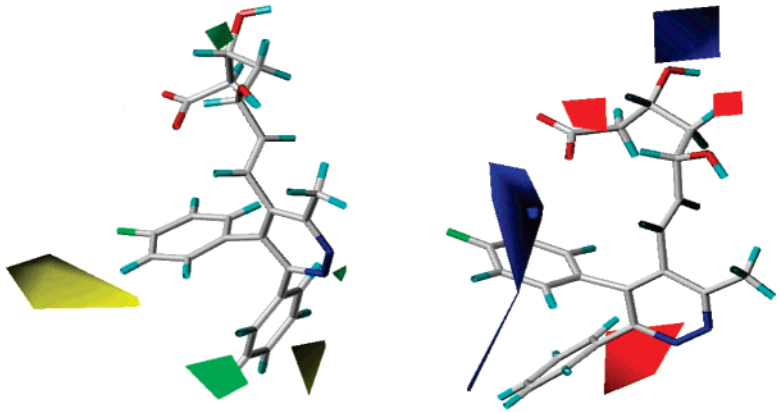


Figure 4. Steric and electrostatic maps from the CoMFA model using fit-atom molecular alignment. Compound 152938 is shown inside the field. Sterically favored areas (contribution level of 80%) are represented by green polyhedra. Sterically disfavored (contribution level of 20%) areas are represented by yellow polyhedra. Blue contours (80% contribution) encompass regions where an increase of positive charge will enhance affinity, whereas in red contoured areas (contribution level of 20%) more negative charges are favorable for binding properties.

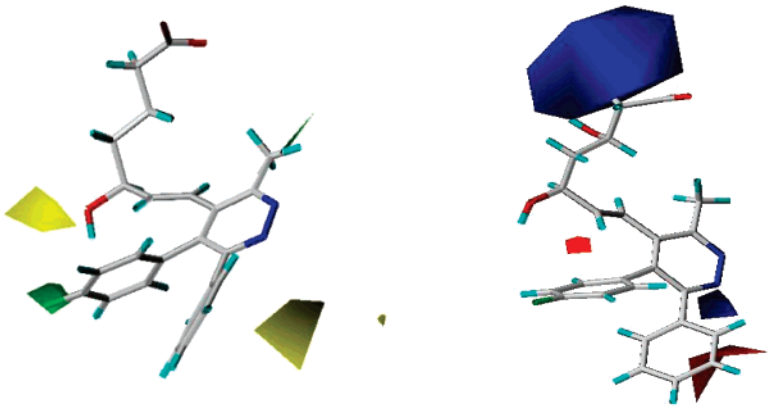


Figure 5. Steric and electrostatic maps from the CoMFA model using active conformation alignment. Compound 152938 is shown inside the field. Sterically favored areas (80% contribution) are represented by green polyhedra. Sterically disfavored areas (contribution level of 20%) are represented by yellow polyhedra. Blue contours (80% contribution) encompass regions where an increase of positive charge will enhance affinity, whereas in red contoured areas (contribution level of 20%) more negative charges are favorable for binding properties.

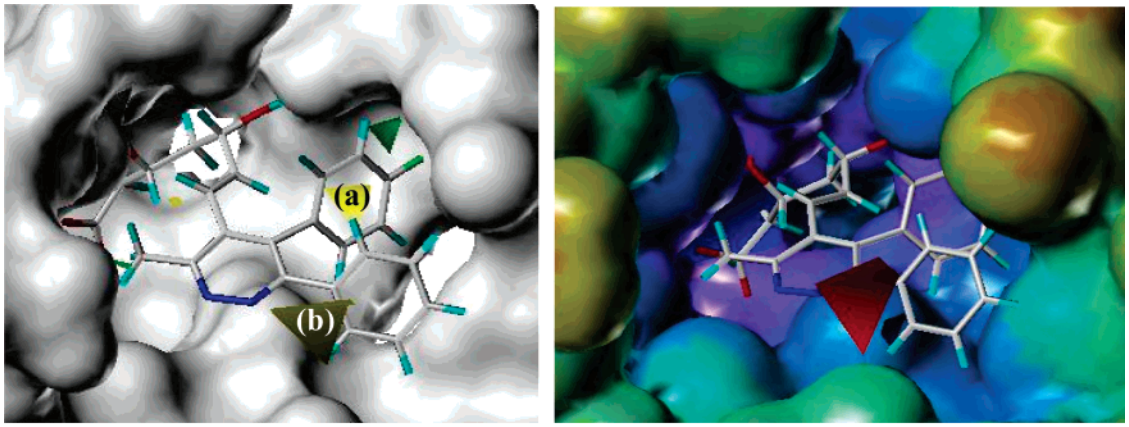


Figure 6. (A) Steric contours projected over the solvent accessible (connolly) topological surface (MOLCAD generated) of the HMGR active site. (B) Electrostatic contours projected over the electrostatic potential surface (blue, negative potential; red/brown, positive potential) of the HMGR active site.

drogen bonding (donor/acceptor) potential maps of the HMGR active site were generated on the solvent-accessible surface using the MOLCAD module[26] of SYBYL. These MOLCAD-generated property surface maps were compared and contrasted to those obtained by CoMFA modeling.

Virtual Screening. Before the crystallographic studies of Istavan and Deisenhofer,[4] the intuitional pharmacophore, which can be employed in attempts for virtual screening for better HMGR inhibitors, is the HMG-like moiety (Figure 1). The work Deisenhofer and co-workers further solidified the structure-based virtual screening.[27] In this study, a generalized definition of active pharmacophore model was

Table 3. Total Scores of FlexE and the Predicted pIC_{50} of CoMFA (Model-dock-RF) for 10 Selected Compounds[a]

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
total score of FlexE	−16.29	−9.04	−17.54	−33.78	−30.84	−27.83	−10.05	−13.55	−15.98	−15.55
pIC_{50} of CoMFA	7.86	8.28	7.93	7.17	6.34	5.64	6.04	5.79	6.11	6.17

[a] **1**−**3** are compounds with low total score of FlexE, but high pIC_{50} of CoMFA; whereas **4**−**6** are compounds with high total score of FlexE, but low pIC_{50} of CoMFA. **7**−**10** are compounds with low scores of both Methods. MDL database, test set: **1**, 142913; **2**, 172873; **3**, 140110. LeadQuest: **4**, 1525-04329; **5**, 1465-00080; **6**, 1525-03160; **7**, 1502-12305; **8**, 1534-04609; **9**, 1502-01284; **10**, 1534-04900.

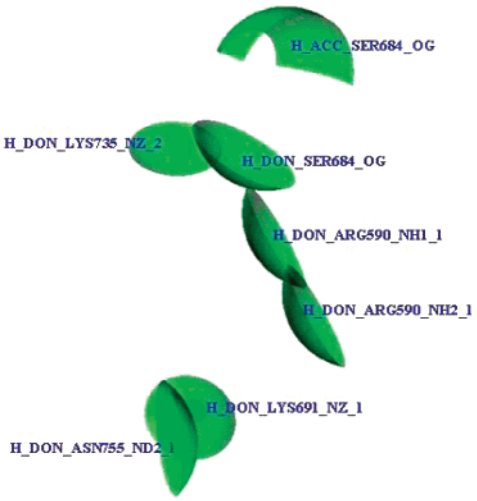


Figure 7. Pharmacophore constraints constructed by FlexX-Pharm. H_ACC denotes H-bond acceptor, H_DON denotes H-bond donor.

generated by the fingerprints hypothesis technique of Tuplets.[6] Tuplets (a ligand structure-based) virtual screening was carried out followed by FlexE and FlexX-Pharm. We performed the CoMFA modeling (Models-dock-RF) before FlexX-Pharm,[28] which complemented and validated the TScore results of FlexE. By jointly using the crystal structure of human HMGR active site, the electronic structure of statins, and the HMGR-statin interaction information, H-bond donor and H-bond acceptor constraints of five residues were generated by FlexX-Pharm.[28] FlexX-Pharm enables pharmacophore-type constraints to be used in FlexX to guide ligand docking.

Results and Discussion

CoMFA 3D QSAR Modeling. Table 1 summarizes the PLS analysis results for two types of CoMFA 3D QSAR modelings: Models-fit and Models-dock. Models-fit has cross-validated coefficients (q^2) up to 0.652 and regression coefficients (r^2) up to 0.977; Models-dock has cross-validated coefficients (q^2) up to 0.731 and regression coefficients (r^2) up to 0.947. These lead to two best models based on the RF (Region focusing) method.

Table 2 presents the predicted activity values and their residues from the experimentally measured pIC_{50} for the 35 inhibitors in the training set and the 12 inhibitors in the testing set (different choices of training set and testing set lead to virtually the same results). The numerical results in Table 2 show that our present CoMFA models have good predictive power, the predicted pIC_{50} values are generally in good agreement with the experiment data and the residuals are all small. Compared with the results of Models-fit, Models-dock has higher cross-validated coefficients, lending credit to the reliability of active conformations obtained by FlexE.

The contribution of the steric field to the activity is 75.0% and that of the electrostatic field is 25.0% based on Model-fit-RF. For the Model-dock-RF, these numbers are 58.0% and 42.0%, respectively. Figures 4 and 5 show the 3D steric and electrostatic maps with the compound 152938, the most potent inhibitor in MDDR,[8,9] shown inside the fields. As two models lead to very similar results, only results from Model-dock-RF will be discussed here. The 3D contour maps show that the changes of molecular fields are associated with the differences of biological activity. The steric fields are in green and yellow. The region of green contour suggests that more bulky substituents in these positions will improve the biological activity, while the yellow region indicates that an increased steric bulk is unfavorable for the inhibitory activity. The *p*-phenyl-monofluoride in R_1 of **4** in Figure 3 is in the green region, and the *p*-phenyl-monofluoride of R_2 is far from the yellow region, resulting in a higher activity of 9.0. The *p*-phenyl-monofluoride in R_2 of **15** is in the yellow region, leading to a lower activity of 5.7.

Figure 6A shows the steric contours projected over the solvent accessible topological surface (MOLCAD generated) of the HMGR active site. The yellow region (a) flanks the imidazole of the His752 residue, while the yellow region (b) is located along the surface of residues toward the outside of active site cavity. The green region is located on the top right corner with larger room of the active site cavity. These steric field distributions of CoMFA model (Model-dock-RF) guarantee a larger percentage of the ligand volume buried inside the binding pocket.[29,30]

Figure 6B shows the electrostatic contours projected over the electrostatic potential surface (MOLCAD generated) of the HMGR active site. The CoMFA electrostatic fields in blue suggest that the positively charged substations may increase the inhibitory activity, while the red region indicates that a high electron density may play a favorable role in inhibitory potency. The blue regions of the MOLCAD generated contour maps represent a negative electrostatic potential; the red regions represent a positive electrostatic potential. The large blue region of CoMFA contour matches well with the blue (highly electronegative) surface of the binding site provided by the carboxyl group of the Asp690 residue. The red region of the CoMFA contour map is located on the surface of residues toward outside of the active site pocket. The comparison of the two types of contour maps shows that the present CoMFA model generally matches well with the active site pocket of HMGR.

Virtual Screening. (1) Selection of the Testing Set and Initial Filtering To Satisfy the Lipinski Rules. To control the performance of our strategy, six well-known HMGR inhibitors[3,4] were added to the search samples whose appearance in the hit list served to calibrate and validate the

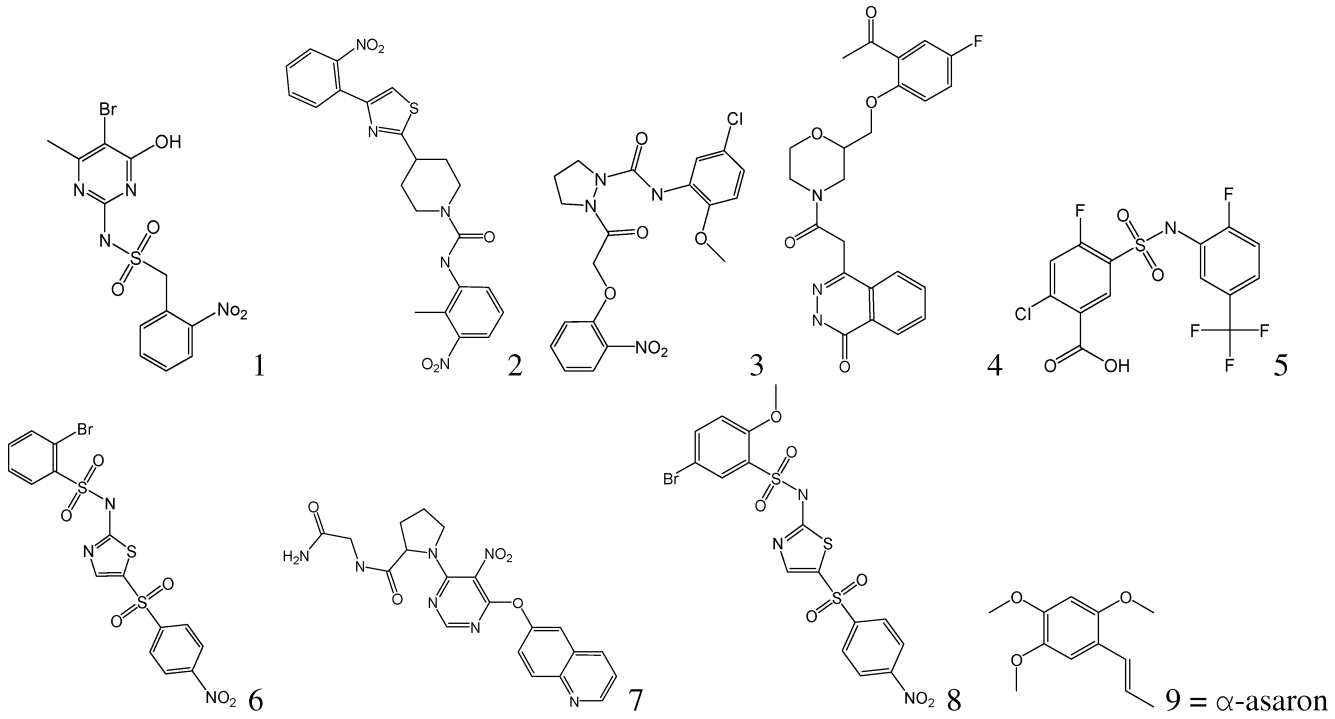


Figure 8. Eight potential hit compounds with high total scores of FlexE and FlexX-Pharm as well as high predicted activities of pIC_{50} from CoMFA (Model-dock-RF). The ninth one is α-asarone, which is the main biological active component of the bark extract of *Guatteria gaumeri* that Greenman utilized in Mexico to treat hypercholesterolemia and cholelithiasis.[35] Note that structures of all these compounds are not statin-like.

Table 4. Total Scores of FlexE and FlexX-Pharm As Well as the Predicted Activities (pIC_{50}) from CoMFA (Model-dock-RF) for the Eight Hit Compounds Whose Structures Are Not Statin-like (cf. Figure 8)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
total score of FlexE	−30.86	−26.52	−28.87	−25.88	−24.74	−28.76	−26.07	−25.16
pIC_{50} of CoMFA	7.94	7.68	8.08	7.94	7.81	7.85	7.83	7.77
total score of Flex-Pharm	−29.57	−38.58	−30.06	−28.51	−26.42	−28.49	−22.19	−24.13

approach at each stage. Search entries for our analysis were taken from the LQSample/LeadQuest databases (41 393 entries). We used the criteria defined by the Lipinski rules[31] (≤5 H-bond donors (no. of OH and NH groups), ≤10 H-bond acceptors (no. of O or N atoms), MW ≤ 500Da, *M* log *P* ≤ 5) to pre-select the database.

(2) Ligand-Derived Pharmacophore. According to the information of a set of compounds which are known to bind to HMGR, we obtained a flexible and effective pharmacophore using the Tuplets module of SYBYL. With this new pharmacophore definition, all six statins as test samples were retrieved successfully from the database. The new pharmacophore served as query to screen the database preselected out of LeadQuest at the first step. Of the pre-selected database, 4138 entries satisfied the pharmacophore query in the Tuplets screening.

(3) Docking with FlexE. In this step, the 4138 entries were docked into the binding pocket using FlexE docking program; all the parameters used were the same as those of FlexE mentioned above. The results were retrieved by selecting the best docking solution for each ligand according to TScore.

(3′) CoMFA Modeling. At this stage, we also performed the CoMFA modeling. On one hand, FlexE score is sensitive to the hydrogen bond formation, putting higher weights on this term than to the steric effect on the activity; on the other hand, CoMFA is more sensitive to the steric field, being directly related to the activity. Thus the predicted pIC_{50} value by CoMFA complimented and validated the TScore results of FlexE. Table 3 listed a few examples. We put more confidence on those entries with high scores from both CoMFA modeling and FlexE docking and discarded those with low scores from both methods. We emphasized not to overlook those entries evaluated high only by one model.

(4) Docking with FlexX-Pharm. In the final step, the 2291 best-ranked hits from the FlexE and/or CoMFA filtering were docked into the bindig sites by using the FlexX-Pharm modeling. FlexX-Pharm constricted the number of possible poses and increased the inactive drop-out rate. In the present study, FlexX-Pharm constraints were constituted by five residues with H-bond donor and H-bond acceptor based on known crystal structures of HMGR complexed with statins. These residues are LYS735, SER684, ARG590, ASN755, and LYS691, respectively. The H-bond donors of LYS735 and ASN755 were set as essential, while the H-bond donors of ARG590, SER684, and LYS691 and the H-bond acceptor of SER684 were set as optional. Only when the compounds fitted three or more of these constraints (as illustrated in Figure 7), they were considered as being docked successfully. This is an important step. As we are looking for potent HMGR inhibitors which are competitive to the substrate, it

is vital that the ligands of the inhibitors interact well with these key residues.

Figure 8 depicted eight representative compounds with high total scores of FlexX-Pharm (see Table 4). We emphasized here that structures of all these compounds are not statin-like. Table 4 also summarized the total scores of FlexE as well as the predicted bioactivities of pIC_{50} from CoMFA (Model-dock-RF). The ninth one is α-asarone, which is the main biological active component of the bark extract of *Guatteria gaumeri* that Greenman utilized in Mexico to treat hypercholesterolemia and cholelithiasis.[32] The total scores of FlexE are within the range from −24.74 to −30.86; while the predicted bioactivities of pIC_{50} from CoMFA are between 7.68 and 8.08. For the ninth compound (α-asarone), our present CoMFA binding model predicted a higher pIC_{50} value of 7.41. Thus the ranks of the predicted bioactivities by CoMFA were generally in good agreement with those of FlexE and FlexX-Pharm virtual screening, which implies that our virtual screening strategy with a new pharmacophore definition is most likely practicable and that the union of using CoMFA models and the structure-based virtual screening procedure is promising for rational quest and optimization of potential novel HMGR inhibitors.

Conclusion

We have established predictive CoMFA 3D QSAR models for the human HMGR inhibitors with two different molecular alignment strategies. Models-fit was built by conventional fit-atom molecular alignment rule of SYBYL. Models-dock was obtained using active conformation alignment obtained by FlexE molecular docking. The latter was found to perform better than the former in the present study. Molecular surface property (electrostatic and steric) mapping and structure-based (both ligand and receptor) virtual screening have been integrated with CoMFA 3D QSAR predictive models, leading to eight potential novel hits for the HMGR inhibitors, which possess no common HMG-like moiety but with high pIC_{50} value.

This work was supported by the Natural Science Foundation of China (Nos. 20672041, 20525311, 20423002, and 20021002), the Natural Science Foundation of Hubei Province (No. 2005ABB012 for Distinguished Young Scholar, No. 2004ABC002 for Innovation Group), the Ministry of Science and Technology of China (Nos. 2004CCA00100, 2004CB719902, and 2001CB61506), the Science and Technology Research Project of Ministry of Education (No. 106116), CNGI-04-15-7A, and China National Technology Platform.

References and Notes

(1) Goldstein, J. L.; Brown, M. S. *Nature* **1990**, *343*, 425−430.
(2) Nakaishi, M.; Goldstein, J. L.; Brown, M. S. *J. Biol. Chem.* **1988**, *263*, 8929−8937.
(3) Istvan, E. S.; Deisenhofer, J. *Science* **2001**, *292*, 1160−1170.
(4) Istvan, E. S.; Palnitkar, M.; Buchanan, S. K.; Deisenhofer, J. *EMBO J.* **2000**, *19*, 819−830.
(5) Cramer, M.; Cramer, R. D., III; Jones, D. M. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(6) Abrahamian, E.; Fox, P. C.; Naerum, L.; Christensen, I. T.; Thogersen, H.; Clark, R. D. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 458−468.
(7) *LeadQuest Chemical Compound Libraries*; Tipos, Inc.: St.Louis, MO, 2000; Vols. 1−3.
(8) MDL Information Systems, Inc.; http://www.mdli.com.
(9) Prous Science; http://www.prous.com.
(10) Chan, C.; Bailey, E. J.; Hartley, C. D.; Hayman, D. F.; Hutson, J. L.; Ingalis, G. G. A.; Jones, P. S.; Keling, S. E.; Kirk, B. E.; Lamont, R. B.; Lester, M. G.; Prithard, J. M.; Barry, C. P.; Scicinski, J. J.; Spooner, S. J.; Smith, G.; Steeples, I. P.; Watson, N. S. *J. Med. Chem.* **1993**, *36*, 3646−3657.
(11) Procopiou, P. A; Draper, C. D.; Hutson, J. L.; Ingalis, G. G. A.; Ross, B. C.; Watson, N. S. *J. Med. Chem.* **1993**, *36*, 3658−3662.
(12) *Sybyl7.0*; Tripos Inc.: St. Louis, MO, 2003; http://www.tripos.com.
(13) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219−3228.
(14) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982−1012.
(15) Powell, M. J. D. *Programming* **1977**, *12*, 241−254.
(16) ClauBen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol.* **2001**, *308*, 377−395.
(17) Kramer, B.; Rarey, M.; Lengauer, T. *Proteins: Struct. Funct. Genet.* **1999**, *37*, 228−241.
(18) Tripos Associates. *SYBYL Molecular Modeling Software*, version 6.x; Tripos Assiociates, Inc.: St. Louis, MO, 1994.
(19) Knegtel, R.; Kuntz, I.; Oshiro, C. *J. Mol. Biol.* **1997**, *266*, 424−440.
(20) Nilsson, J. Multiway Calibration in 3D QSAR. Applications to Dopamine Receptor Ligands; http://www.ub.rug.nl/eldoc/dis/science/j.nilsson.
(21) Wold, S.; Rhue, A.; Wold, H.; Dunn, W. J. I. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735−743.
(22) Wold, S.; Albano, C.; Dunn, W. J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johanson, E.; Lindberg, W.; Sjostrom, M. *NATO ASI Ser., Ser. C* **1984**, *138*, 17−95.
(23) Clark, M.; Cramer, R. D., III. *Quant. Struct.-Act. Relat.* **1993**, *12*, 137−145.
(24) Bush, B. L.; Nachbar, R. B. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587−619.
(25) Lindgren, F.; Geladi, P.; Rännar, S.; Wold, S. *J. Chemom.* **1994**, *8*, 349−363.
(26) Waldherr-Teschner, M.; Goetze, T.; Heiden, W.; Knoblauch, M.; Vollhardt, H.; Brickmann, J. *J. Adv. Sci. Visualization* **1992**, 58−67.
(27) Gruneberg, S.; Stubbs, M. T.; Klebe, G. *J. Med. Chem.* **2002**, *45*, 3588−3602.
(28) Hindle, S. A.; Rarey, M.; Buning, C.; Lengaue, T. *J. Comput.-Aided. Mol.Des.* **2002**, *16*, 129−149.
(29) Lyne, P. D. *Drug Delivery* **2002**, *7* (20).
(30) Stahl, M.; Böhm, H.-J. *J. Mol. Graphics Modell.* **1998**, *16*, 121−132.
(31) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(32) Medina-Franco, J .L.; López-Vallejo, F.; Rodríguez-Morales, S.; Castillo, R.; Chamorro, G.; Tamariz, J. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 989−994.

CC060101E