3D QSAR and Molecular Docking Studies of Benzimidazole Derivatives as Hepatitis C Virus NS5B Polymerase Inhibitors

Pallav D. Patel,[†] Maulik R. Patel,[†] Neerja Kaushik-Basu,[‡] and Tanaji T. Talele*,[†]

Department of Pharmaceutical Sciences, College of Pharmacy and Allied Health Professions,
St. John's University, Jamaica, New York 11439, and Department of Biochemistry and Molecular Biology,
UMDNJ-New Jersey Medical School, 185 South Orange Avenue, Newark, New Jersey 07103

The urgent need for novel HCV antiviral agents has provided an impetus for understanding the structural requisites of NS5B polymerase inhibitors at the molecular level. Toward this objective, comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) of 67 HCV NS5B polymerase inhibitors were performed using two methods. First, ligand-based 3D QSAR studies were performed based on the lowest energy conformations employing the atom fit alignment method. Second, receptor-based 3D QSAR models were derived from the predicted binding conformations obtained by docking all NS5B inhibitors at the allosteric binding site of NS5B (PDB ID: 2dxs). Results generated from the ligand-based model were found superior (r^2_{cv} values of 0.630 for CoMFA and 0.668 for CoMSIA) to those obtained by the receptor-based model (r^2_{cv} values of 0.536 and 0.561 for CoMFA and CoMSIA, respectively). The predictive ability of the models was validated using a structurally diversified test set of 22 compounds that had not been included in a preliminary training set of 45 compounds. The predictive r^2 values for the ligand-based CoMFA and CoMSIA models were 0.734 and 0.800, respectively, while the corresponding predictive r^2 values for the receptor-based CoMFA and CoMSIA models were 0.538 and 0.639, respectively. The greater potency of the tryptophan derivatives over that of the tyrosine derivatives was interpreted based on CoMFA steric and electrostatic contour maps. The CoMSIA results revealed that for a NS5B inhibitor to have appreciable inhibitory activity it requires hydrogen bond donor and acceptor groups at the 5-position of the indole ring and an R substituent at the chiral carbon, respectively. Interpretation of the CoMFA and CoMSIA contour maps in context of the topology of the allosteric binding site of NS5B provided insight into NS5B-inhibitor interactions. Taken together, the present 3D QSAR models were found to accurately predict the HCV NS5B polymerase inhibitory activity of structurally diverse test set compounds and to yield reliable clues for further optimization of the benzimidazole derivatives in the data set.

INTRODUCTION

Hepatitis C virus (HCV), the etiological agent of parenteral non-A non-B hepatitis, often causes the development of malignant chronic disease, including liver cirrhosis and hepatocellular carcinoma, which frequently end in liver failure.[1–4] An estimated 200 million cases of HCV infection exist worldwide, of which over 4.1 million infections occur in the United States.[5] At present, neither a vaccine against HCV nor an effective therapy with an acceptable broad spectrum of action against all genotypes of HCV is available.[6,7] A currently approved HCV therapy, comprised of pegylated interferon α (PEG-IFN-α), either alone or in combination with ribavirin, a broad spectrum antiviral agent, has found limited patient compliance due to the severe side effects of these two drugs.[8–10] Therefore, the development of novel HCV antiviral agents with a high therapeutic index, reduced side effects, and an easier route of administration will be paramount to meeting an urgently needed therapeutic weapon against HCV.

HCV is an enveloped, positive-stranded RNA virus with a ~9.6 kb genome that encodes a single large polyprotein composed of ~3010 amino acids. This protein is processed by host and viral proteases into four structural (Core, E1, E2, and p7) and six nonstructural proteins (NS2, -3, -4A, -4B, -5A, and -5B).[11,12] Nonstructural protein 5B (NS5B), a 66 kDa RNA-dependent RNA polymerase (RdRp), has attracted the attention of medicinal chemists as a target for drug development since it plays a pivotal role in HCV replication and the host lacks a functional counterpart of NS5B.[10,13] Advances in high throughput screening assays have resulted in the recent identification of numerous anti-NS5B compounds broadly categorized as nucleoside (NIs) and non-nucleoside (NNIs) inhibitors.[14–17] NIs comprise ribonucleoside triphosphate substrate analogues which function as chain terminators. In contrast, NNIs represent a chemically diverse family of compounds which act by complex and diverse mechanisms by binding near the active site or discrete allosteric sites on NS5B.[18] These include compounds belonging to the benzimidazole,[19–23] indole,[24,25] thiophene,[26] phenylalanine,[27] dihydropyranone,[28] pyrano-indole,[29] benzothiadiazine,[30] proline sulfonamide,[31] benzylidene,[14,32] diketoacids,[33] acrylic acid,[34,35] tetracyclic indole,[36] quinoxaline,[37] and indole N-acetamide[38] scaffolds.

* Corresponding author phone: (718)990-5405; fax: (718)990-1877; e-mail: talelet@stjohns.edu.
† St. John's University.
‡ UMDNJ-New Jersey Medical School.
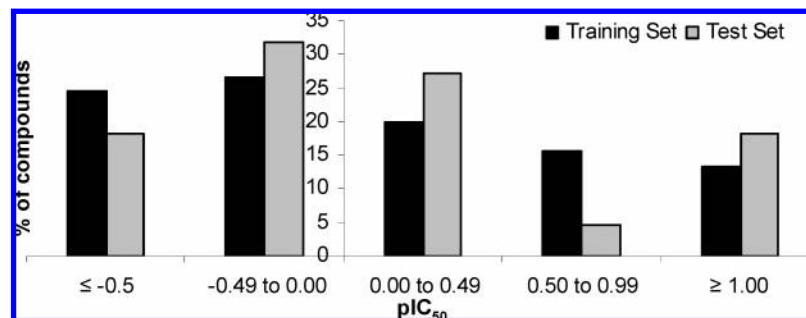
Figure 1. Distribution of biological activities for the training set and the test set versus % of compounds.

Among these, the aryl diketoacid (ADK) series has received much attention from a theoretical perspective and has been a subject of docking and 3D QSAR,[39] predictive topochemical modeling,[40] and 3D chemical-feature-based pharmacophore modeling studies,[41] all of which have provided new insights into their mechanism of action.

The benzimidazole series of NNIs represents a promising group of compounds which merit computational investigations. These compounds, originally screened from a combinatorial library as specific inhibitors of HCV NS5B activity in vitro,[20] exhibited marginal potency in cell-based HCV replicon assay,[23] presumably due to poor cellular permeability resulting from the presence of a 5-COOH group which is ionizable at physiological pH. Studies involving high-throughput parallel synthetic techniques have identified a 1,2-disubstituted benzimidazole-5-carboxylic acid scaffold as the minimum core for their biological activity, thereby paving the way for optimization and subsequent generation of analogues capable of inhibiting NS5B RdRp in vitro in nanomolar concentrations.[21] Efforts to improve the cellular permeability of these compounds has been achieved by modifying the benzimidazole structure through the attachment of functionalized amino acid residues to the ionizable −COOH group. These modifications resulted in decreased intrinsic potency but improved cell culture activity thereby yielding inhibitors of HCV subgenomic RNA replication ex vivo.[22] Modification of another −COOH group at the chiral center in these compounds further decreased enzymatic activity but enhanced cell based activity. Thus, ligand- and receptor-based molecular modeling studies are necessary to identify and optimize new benzimidazole-based inhibitors with potent in vitro and in vivo anti-NS5B activity.
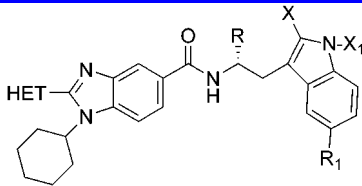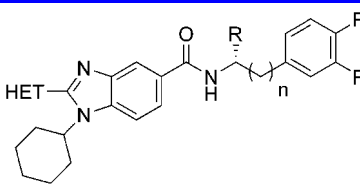
In the present study, we have applied comparative molecular field analysis (CoMFA)[42,43] and comparative molecular similarity indices analysis (CoMSIA)[44] 3D QSAR methodologies to the benzimidazole[21,22] scaffold of HCV NS5B polymerase inhibitors. In addition, we have included sets of structurally diverse test compounds representing tetracyclic indole,[36] quinoxaline,[37] and indole N-acetamide[38] scaffolds to broaden the scope of the current 3D QSAR investigation. To elucidate the probable binding modes of these inhibitors at the allosteric site of the enzyme, molecular docking simulations were performed using the X-ray crystallographic structure of a complex of NS5B with the tetracyclic indole inhibitor.[36] Here we report the topographical features of the allosteric binding site of NS5B in context of the best CoMFA and CoMSIA contour maps. This analysis provides a platform for the prediction of novel NS5B inhibitors and enables the interpretation of their binding modes to NS5B.

MATERIALS AND METHODS

Data Sets and Biological Activity. The training set and the test set used comprise a series of benzimidazole,[21,22] tetracyclic indole,[36] quinoxaline,[37] and indole N-acetamide[38] derivatives, which have been shown to possess HCV NS5B polymerase inhibitory activity. The IC_{50} values, in μM, were converted to pIC_{50} ($-\log IC_{50}$) values, which were used as dependent variables in the CoMFA and CoMSIA QSAR analyses. As a rule of thumb, the pIC_{50} values of the training data set should span approximately 3 log units. Accordingly, the pIC_{50} values of the training set described in this manuscript span 3.1 log units. Selection of the training set (45 compounds) and the test set (22 compounds) was done by considering the fact that the test set compounds represent structural diversity and a range of biological activities similar to that of the training set (Figure 1). Compounds in the test set allowed us to use one test compound per two training compounds, thus resulting in a more rigorous validation of the training model than would be possible with a smaller test set. In addition, a wide range of structural diversity of compounds in the test set permitted us to evaluate the extrapolative accuracy of the CoMFA and CoMSIA models. The mean (SD) of the NS5B inhibitory activity (pIC_{50}) in the training set and the test set were 0.13 (0.77) and 0.08 (0.72), respectively, which confirmed the test set as a true representative of the training set. The structures of the compounds in the training and test sets are shown in Tables 1 and 2.

Molecular Modeling and Alignment. The 3D QSAR and molecular docking computations were carried out on a Dell Precision 470n workstation with the RHEL 4.0 operating system using SYBYL 7.2[45] and Glide 4.0,[46] respectively. The position of each atom is important for CoMFA and CoMSIA because the descriptors were calculated based on the 3D-space grid. Thus, the method to determine the conformation of each molecule and the way to align molecules together are two sensitive input parameters to build a reasonable model.[42] The 3D structures of the training and test set of compounds were constructed using the Sketch Molecule function in SYBYL. Since the (*R*)-enantiomer was found to be more potent than the (*S*)-enantiomer, we have modeled the training set and the test set compounds **1**−**57** as (*R*)-enantiomers;[21,22] however, test compound **63** was tested as the (*S*)-enantiomer hence it was modeled as the (*S*)-enantiomer.[37] Energy minimizations were performed using the Tripos force field[47] and Gasteiger−Marsili charges[48] with the steepest descent followed by the conjugate gradient method with a convergence criterion of 0.001 kcal/mol Å. In this study, two alignment methods were applied. In the

Table 1. Structures of the Training Set and the Test Set Compounds

HET = 3-furyl unless specified
X and X$_1$ = H unless specified

HET = 3-furyl unless specified
n = 1 unless specified

| compd	R	R$_1$	IC$_{50}$[a]	compd	R	R$_1$	R$_2$	IC$_{50}$[a]
				Training Set				
1	COOH	OCH$_2$COOH	0.027	26	COOH	OH	H	0.5
2	COOCH$_3$	OH	0.380	27	CH(CH$_3$)$_2$	OH	H	4.3
3	CONH$_2$	OH	0.220	28	CON(CH$_3$)$_2$	OH	H	4.9
4	thiazol-4-yl	OH	0.300	29	morpholine-4-carbonyl	OH	H	9.5
5	thiazol-2-amino-4-yl	OH	0.190	30	4-methylpiperazine carbonyl	OH	H	11.0
6	N-methylthiazol-2-amino-4-yl	OH	0.500	31	N-[2-(dimethylamino)ethyl]carboxamide	OH	H	2.1
7	N,N-dimethyl thiazol-2-amino-4-yl	OH	0.900	32	N-[(4-morpholinyl)ethyl]carboxamide	OH	H	3.2
8	thiazol-2-amino-4-yl (X$_1$ = CH$_3$)	OH	0.900	33	N-[2-(dimethylamino)propyl]carboxamide	OH	H	1.9
9[b]	COOH (HET = 2-pyridyl)	H	1.900	34	N-(3-pyridinylmethyl)carboxamide	OH	H	2.6
10	COOH (HET = 2-pyridyl)	OH	0.140	35	N-(4-pyridinylmethyl)carboxamide	OH	H	3.7
11	COOH	H	0.400	36	N-(3-pyridinyl)carboxamide	OH	H	3.0
12	COOH (X$_1$ = CH$_3$)	OH	0.160	37	2-methylthiazol-4-yl	OH	H	3.3
13	COOH (X = CH$_3$)	OH	0.460	38	thiazol-2-amino-4-yl	OH	H	1.3
14	H	OH	1.500	39	N,N-dimethylthiazol-2-amino-4-yl	OH	H	5.1
15	COOH	NO$_2$	2.000	40	N-acetyl thiazol-2-amino-4-yl	OH	H	4.8
16[b]	COOH	CH$_3$	2.200	41	H (HET = 2-pyridyl) (n = 0)	OCH$_3$	OCH$_3$	7.0
17	COOH	F	2.900	42	COOH (HET = 2-pyridyl) (n = 0)	OCH$_3$	OCH$_3$	2.0
18	COOH	OCH$_3$	0.400	43	COOH (HET = 2-pyridyl)	OH	H	1.5
19	COOH	NH$_2$	0.090	44	COOH (n = 0)	OCH$_3$	OCH$_3$	0.4
20	COOH	NHSO$_2$CH$_3$	0.120	45	COOH	H	H	4.0
21	COOH	NHSO$_2$CF$_3$	0.150					
22	COOH	NHCOCOOH	0.008					
23	COOH	COOH	0.020					
24	COOH	C$_5$−tetrazole	0.015					
25	COOH	OC(CH$_3$)$_2$COOH	0.100					
				Test Set				
46	COOH	OH	0.050	51	CONH$_2$	OH	H	0.8
47	2-methylthiazol-4-yl	OH	0.300	52	N-(2-pyridinylmethyl) carboxamide	OH	H	2.7
48	N-acetyl thiazol-2-amino-4-yl	OH	0.550	53	thiazol-4-yl	OH	H	6.6
49	COOH	NHAc	0.060	54	N-methylthiazol-2-amino-4-yl	OH	H	2.8
50	COOH	CONH$_2$	0.037	55	2-N-(acetamido)-1H-imidazole-4-yl)	OH	H	4.2
				56	CH$_3$ (HET = 2-pyridyl) (n = 0)	OCH$_3$	OCH$_3$	12.0
				57	COOH (HET = 2-pyridyl)	H	H	7.0

[a] IC$_{50}$ in μM. [b] Outliers.

first computational design, i.e., the ligand-based model, the most potent molecule (compound **22**) was chosen as a template to fit the remaining training and test set of compounds by using the fit atoms function available in SYBYL. The reference atoms in compound **22** used for alignment were as follows: (i) C$_1$ of the cyclohexyl ring; (ii) N$_2$, C$_3$ and C$_4$ of the benzimidazole ring; and (iii) N$_5$, C$_6$, and C$_7$ of the tryptophan moiety (Figure 2A). The alignment of compounds with distinct differences in chemical structure (as is the case of quinoxaline inhibitors included in the test set) usually poses difficulties. In such a situation, chemical structures will need to be aligned by making use of pharmacophorically identical atoms that approximately correspond to the reference atoms of compound **22** indicated by asterisks in Table 2. The resulting ligand-based alignment model is shown in Figure 2B. The ensuing alignment model was subjected to CoMFA and CoMSIA 3D QSAR studies.

In the second computational design, i.e., the receptor-based model, the molecules were aligned according to the bioactive conformations obtained from docking (Figure 2C). For docking experiments, all the compounds were constructed

using the fragment dictionary of Maestro 7.5 and geometry-optimized using the Optimized Potentials for Liquid Simulations-All Atom (OPLS-AA) force field[49] with the steepest descent followed by truncated Newton conjugate gradient protocol as implemented in Macromodel 9.5. Water molecules of crystallization were removed from the complex, and the protein was optimized for docking using the protein preparation and refinement utility provided by Schrödinger LLC and the Impact program (FirstDiscovery v4.0). Partial atomic charges for the protein were assigned according to the OPLS-AA force field. Next, the extra precision (XP) Glide docking method was used to dock all compounds into the allosteric site of HCV NS5B polymerase (PDB ID: 2dxs).[36] Although details on the methodology used by Glide are described elsewhere,[50−53] a short description is provided below. The binding site, for which the various energy grids were calculated and stored, is defined in terms of two concentric cubes: the bounding box, which must contain the center of any acceptable ligand pose, and the enclosing box, which must contain all ligand atoms of an acceptable pose. Cubes with an edge length of 12 Å and centered at the

Table 2. Structures of the Test Set Compounds with Significantly Different Structural Scaffolds[a]
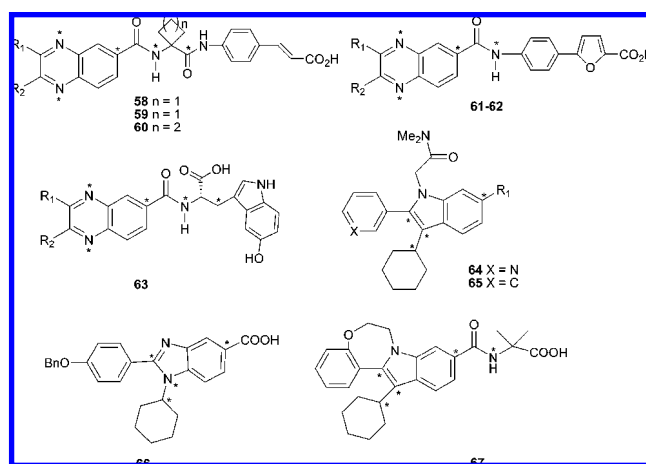
compd	R_1	R_2	IC_{50}[b]
58	4-F−Ph-	4-F−Ph-	0.690
59	cyclohexyl-	4-F−Ph-	1.300
60	4-F−Ph-	4-F−Ph-	1.200
61	Ph-	cyclohexyl-	1.800
62	cyclohexyl-	Ph-	1.600
63	cyclohexyl-	Ph-	0.600
64	1,2,4-oxadiazol-3-yl-5(4H)-one		0.780
65	N-(benzylsulfonyl)-N-methylcarboxamide		0.640
66			1.400
67			0.046

[a] Asterisks atoms were used for the alignment. [b] IC_{50} in μM.

midpoint of the longest atom−atom distance in the respective cocrystallized ligand defined the bounding box in the protein. Similarly, the larger enclosing box was defined in terms of the cocrystallized ligand with an edge length of 30 Å. Poses with a root-mean-square deviation (rmsd) of less than 0.5 Å and a maximum atomic displacement of less than 1.3 Å were eliminated as redundants to be able to increase diversity in the retained ligand poses. The scale factor for van der Waals radii was applied to those atoms with absolute partial charges less than or equal to 0.15 (scale factor of 0.8) and 0.25 (scale factor of 1.0) electrons for ligand and protein, respectively. The *maxkeep* variable which sets the maximum number of poses generated during the initial phase of the docking calculation was set to 5000, and the *keep best* variable which sets the number of poses per ligand that enters the energy minimization was set to 1000. Energy minimization protocol includes a dielectric constant of 4.0 and 1000 steps of conjugate gradient. Upon completion of each docking calculation, at most 100 poses per ligand were allowed to generate. Selection of top-scored conformations based on the Glide scoring functions[54] did not produce alignments resulting in statistically significant models. In the set of best-ranked conformations according to Glide, structural features of eight inhibitors (compounds **3**, **6**, **31**, **34**, **35**, **36**, **51**, and **52**) did not superimpose with the rest of the docked inhibitor set as an amide group next to the benzimidazole ring did not align, thereby the entire substituent on the amide nitrogen deviated significantly. These eight inhibitors were built in a sketch/ build option according to the conformation of the most structurally related docked inhibitor serving as a reference. Thus, compounds **3** and **6** were built based on compounds **2** and **5**, respectively, compounds **31**, **34**, **35**, **36**, and **52** were built using compound **32** as a reference, and compound **51**

was built based on compound **26**. The bioactive conformations resulting from docking of the training set and the test set inhibitors were imported into a SYBYL molecular database for CoMFA and CoMSIA studies without further energy minimization. Charge calculations were done using the Gasteiger−Marsili method as implemented in SYBYL.

CoMFA and CoMSIA 3D QSAR Models. To derive the CoMFA and CoMSIA descriptor fields, a 3D cubic lattice with grid spacing of 2 Å in *x*, *y*, and *z* directions, was created to encompass the aligned molecules. CoMFA descriptors were calculated using an sp³ carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 to generate steric (Lennard-Jones 6−12 potential) field energies and electrostatic (Coulombic potential) fields with a distance-dependent dielectric at each lattice point. The steric and electrostatic energy values were truncated at a default value of 30 kcal/mol. The CoMFA steric and electrostatic fields thus generated were scaled by the CoMFA-STD method in SYBYL.

CoMSIA descriptors were derived according to Klebe et al.[44] with the same lattice box as that used for the CoMFA calculations, with a grid spacing of 2 Å, and employing a C^{1+} probe atom with a radius of 1.0 Å as implemented in SYBYL. CoMSIA similarity indices (A_F) for a molecule *j* with atoms *i* at a grid point *q* were calculated using eq 1:

$$A_{F,k}^{q}(j) = -\sum \omega_{probe,k}\omega_{ik}e_{iq^2}^{-\alpha r} \qquad (1)$$

Being an extension to the CoMFA approach, which has two fields, the CoMSIA method incorporates five physico-chemical properties (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) denoted as *k* in eq 1, which were evaluated using the probe atom. A Gaussian type distance-dependence was used between the grid point *q* and each atom *i* in the molecule. A default value of 0.3 was used as the attenuation factor α. In CoMSIA, the steric indices were related to the third power of the atomic radii, the electrostatic descriptors were derived from partial atomic charges, the hydrophobic fields were derived from atom-based parameters,[55] and the hydrogen bond donor and acceptor atoms within a putative protein environment were derived from experimental values.[56,57]

In partial least-square (PLS) regression analyses, the CoMFA and CoMSIA descriptors were used as independent variables and the pIC_{50} values were used as dependent variables[58,59] to derive 3D QSAR models using the standard implementation in the SYBYL package. The predictive value of the models was evaluated first by leave-one-out (LOO) cross-validation.[43,60] The cross-validated coefficient, r^2_{cv}, was calculated using eq 2

$$r^2_{cv} = 1 - \frac{\sum(Y_{predicted} - Y_{observed})^2}{\sum(Y_{observed} - Y_{mean})^2} \qquad (2)$$

where $Y_{predicted}$, $Y_{observed}$, and Y_{mean} are the predicted, observed, and mean values of the target property (pIC_{50}), respectively. $\sum(Y_{predicted} - Y_{observed})^2$ is the predictive residual sum of squares (PRESS).

The optimal number of components (ONC) obtained from the cross-validated PLS analysis were used to derive the final QSAR model using the compounds in the training set
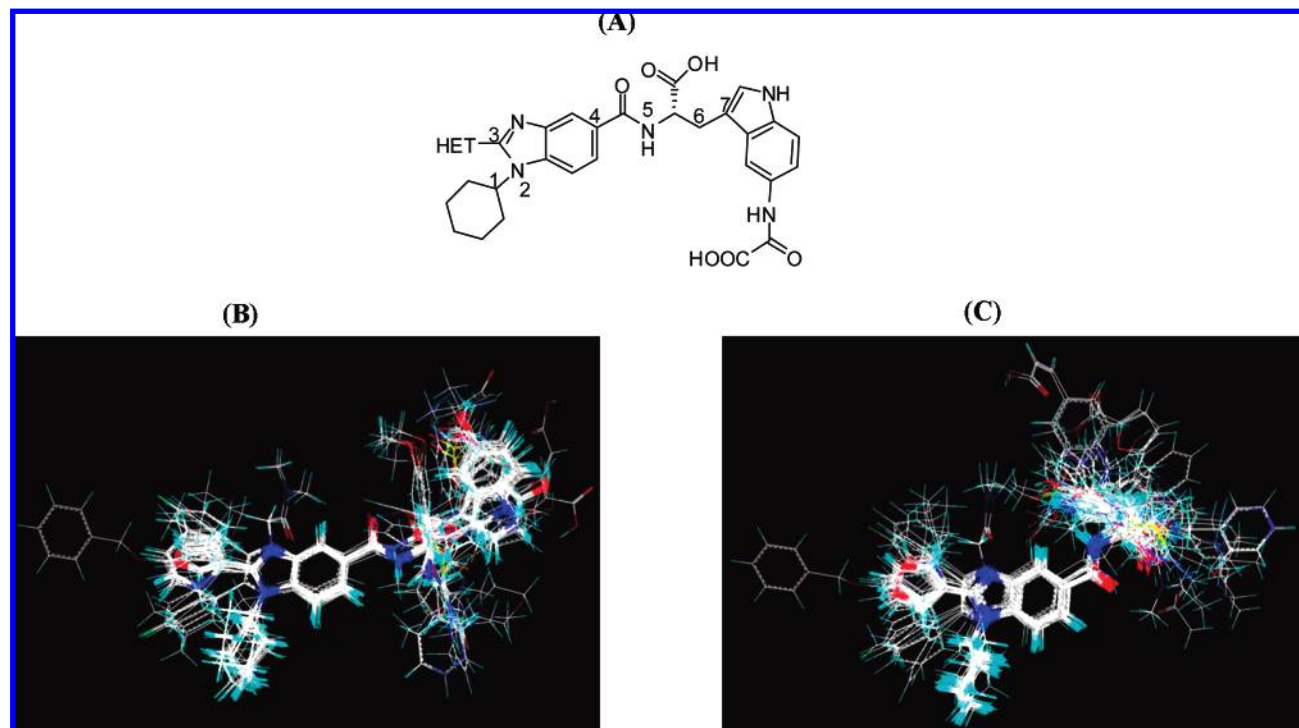
Figure 2. (A) Compound **22** used as a template for atom based alignment. The atoms for alignment are numbered 1−7 (Note: the atom numbering does not follow IUPAC rules). Ligand- and receptor-based alignments of all the compounds are shown in panels **(B)** and **(C)**, respectively.

(without cross-validation). The non-cross-validated correlation coefficient (r^2_{ncv}) served as a measure of the quality of the model.

The boot strapping analysis[61] for 100 runs and the number of cross-validations (e.g., two and five) were carried out and confirmed by the average value for 50 runs from each cross-validation. To test the utility of the model as a predictive tool, an external set of compounds with known activities but not used in model generation (the test set) were predicted. The predictive r^2, calculated by using eq 3, was based on molecules in the test set and was used to evaluate the predictive power of the CoMFA and CoMSIA models

$$\text{predictive } r^2 = 1 - (\text{"press"}/\text{SD}) \quad (3)$$

where SD is the sum of the squared deviations between the actual activities of the compounds in the test set and the mean activity of the compounds in the training set and "press" is the sum of the squared deviations between predicted and actual activities for every compound in the test set. The activity of the test set was predicted by the CoMFA and CoMSIA models using the predict property command. CoMFA and CoMSIA coefficient maps were generated by interpolation of the pairwise products between the PLS coefficients and the standard deviations of the corresponding CoMFA or CoMSIA descriptor values.

RESULTS AND DISCUSSION

CoMFA and CoMSIA Statistical Results. The CoMFA and CoMSIA analyses were performed for both the ligand- and receptor-based alignment models. For the 45 compounds, the ligand-based model yielded ($r^2_{cv} = 0.555$ and $r^2_{ncv} = 0.892$) and ($r^2_{cv} = 0.660$ and $r^2_{ncv} = 0.937$) values for the CoMFA and CoMSIA models, respectively, whereas the

receptor-based model gave ($r^2_{cv} = 0.474$ and $r^2_{ncv} = 0.784$) for CoMFA and ($r^2_{cv} = 0.548$ and $r^2_{ncv} = 0.916$) for CoMSIA. Thus, the ligand-based model performed better than the receptor-based model. Our next step was to generate statistically significant 3D QSAR models in terms of cross-validated correlation coefficients with the least standard deviation using a variety of 3D QSAR tools (such as region focusing, steric/electrostatic cutoffs, and different charge calculation methods) that are available in SYBYL. Use of region focusing on the 45 compound ligand-based alignment model yielded values of ($r^2_{cv} = 0.536$, SEP = 0.583, $r^2_{ncv} = 0.870$, and SEE = 0.298) and ($r^2_{cv} = 0.672$, SEP = 0.550, $r^2_{ncv} = 0.914$, and SEE = 0.243) for CoMFA and CoMSIA, respectively, whereas the receptor-based alignment model for the 45 compounds led to ($r^2_{cv} = 0.535$, SEP = 0.585, $r^2_{ncv} = 0.799$, and SEE = 0.360) for CoMFA and ($r^2_{cv} = 0.623$, SEP = 0.549, $r^2_{ncv} = 0.957$, and SEE = 0.170) for CoMSIA. Though the region focusing identified a slightly better model for receptor-based alignment, it was still not significant compared to the ligand-based atom fit alignment model. Even though the effect of steric and electrostatic cutoffs for both the ligand- and receptor-based models was studied, their statistical results obtained were inferior to the default values (data not shown). Furthermore we have computed charges such as Gasteiger−Huckel, Huckel, Delre, Pullman, and MMFF94 for molecules in the training set to be able to evaluate the effect of the charge calculation method on the statistical results. In this instance, the results were poorer than those obtained by the Gasteiger−Marsili method for the CoMFA and CoMSIA models. Our attention was mainly focused on the ligand-based alignment model since it produced satisfactory statistical data, and, in addition, the conformations used in this alignment were energetically more favorable than those of the docked conformations. Hence,

3D QSAR STUDIES ON HCV NS5B POLYMERASE INHIBITORS

J. Chem. Inf. Model., Vol. 48, No. 1, 2008 **47**

Table 3. Summary of CoMFA and CoMSIA Results

| PLS statistics	ligand-based model				receptor-based model			
	45-compound model		43-compound model		45-compound model		43-compound model	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
r^2_{ncv}[a]	0.892	0.937	0.932	0.943	0.784	0.916	0.883	0.888
SEE[b]	0.271	0.207	0.220	0.201	0.374	0.236	0.284	0.274
F_{test}[c]	64.30	115.59	100.68	121.93	49.70	109.43	71.67	103.28
r^2_{cv}[d]	0.555	0.660	0.630	0.668	0.474	0.548	0.536	0.561
SEP[e]	0.572	0.526	0.546	0.523	0.626	0.572	0.599	0.572
r^2_{pred}[f]			0.734	0.800			0.538	0.639
PLS components[g]	5	5	5	5	3	4	4	3
contribution:								
steric	0.560		0.560		0.490		0.460	
electrostatic	0.440		0.440		0.510		0.540	
hydrophobic		0.260		0.250		0.260		0.260
H-bond donor		0.490		0.490		0.450		0.460
H-bond acceptor		0.250		0.260		0.290		0.280
r^2_{boot}[h]			0.962	0.962			0.934	0.928
SEE_{boot}[i]			0.160	0.157			0.212	0.217
r^2_{LHO}[j]			0.543	0.555			0.412	0.426
SD_{LHO}[k]			0.088	0.090			0.100	0.119
r^2_{5cv}[l]			0.602	0.635			0.492	0.517
SD_{5cv}[m]			0.052	0.035			0.061	0.055

[a] Non-cross-validated correlation coefficient. [b] Standard error of estimate. [c] Ratio of r^2_{ncv} explained to unexplained $= r^2_{ncv}/(1-r^2_{ncv})$. [d] Cross-validated correlation coefficient after the leave-one-out procedure. [e] Standard error of prediction. [f] Predicted correlation coefficient for the test set of compounds. [g] Optimal number of principal components. [h] Average of correlation coefficient for 100 samplings using the bootstrapped procedure. [i] Average standard error of estimate for 100 samplings using the bootstrapped procedure. [j] Average cross-validated correlation coefficient for 50 runs using the leave-half-out (LHO) group. [k] Standard deviation of average cross-validated correlation coefficient for 50 runs. [l] Average cross-validated correlation coefficient for 50 runs using the five cross-validation group. [m] Standard deviation of average cross-validated correlation coefficient for 50 runs.

the training set of compounds for the ligand-based alignment model was examined for outliers. There are several reasons that may account for the outliers, including unique structural differences, different binding conformation, and a higher residual between the observed and the predicted biological activity of an inhibitor. For compounds **1**−**25**, the corresponding IC_{50} values somewhat reflected the differences in the substituent at the R_1-position of the indole ring (i.e., hydrogen bond donating/accepting functional groups) and which seems to play an important role in the HCV NS5B polymerase inhibitory activity. Compounds **9** (5-H) and **16** (5-CH$_3$), by virtue of lacking such functional groups, demonstrated a poor activity (IC_{50} values of 1.9 μM and 2.2 μM, respectively). On the premise that highly active compounds possess substituents at the R_1-position of the indole ring which are capable of forming a hydrogen-bonding network, predictions on compounds **9** and **16**, which lack such substituents, yielded high residual values. As far as substituents at the 5-position of the indole ring is concerned, unlike compound **9** (5-H), compound **11** (5-H) was better predicted. This could be due to the fact that compound **11** belongs to a 3-furyl (highly active) series, while compound **9** belongs to a 2-pyridyl (moderately active) series (for example, compounds **9** vs **11** and **10** vs **46**). In addition, the residual between the observed pIC_{50} and predicted pIC_{50} values for compounds **9** and **16** was approximately 1 logarithm unit, making these two compounds outliers. After dropping compounds **9** and **16** from the training set, 3D QSAR studies were reperformed on the remaining 43 compounds. In this manner, a significantly higher r^2_{cv} value of 0.630 with (ONC) 5 for CoMFA and 0.668 for CoMSIA with (ONC) 5 was attained (Table 3). Since statistically significant CoMFA and CoMSIA models were derived from the compounds remaining in the training set, these com-

pounds served as the basis for further assessment and discussion. The experimental and predicted pIC_{50} values derived from the training set by the best CoMFA model are reported with −0.04 of average value of the residuals in Supporting Information, Table S1 and Figure 3A.

It has been established that the five different descriptor fields are not totally independent of each other and that such dependency among individual fields usually decreases the statistical significance of the models.[62,63] An evaluation of which one of the five CoMSIA fields is actually needed for the generation of a predictive model was performed by computing all possible combinations of fields (Supporting Information, Table S2). The steric, hydrophobic, and hydrogen bond donor fields yielded the best individual field models with r^2_{cv} of 0.525, 0.550, and 0.585, respectively. In the combined models, the best r^2_{cv}, r^2_{ncv}, F value, and least standard error of estimation were obtained by the combination of hydrophobic, hydrogen bond donor, and acceptor fields ($r^2_{cv} = 0.668$, $r^2_{ncv} = 0.943$, $F = 121.93$, and SEE = 0.201). Thus, the model with hydrophobic, hydrogen bond donor, and acceptor fields appeared to be superior among all the models derived. The detailed experimental and predicted pIC_{50} values based on the selected CoMSIA model for the training set are shown with −0.02 of average value of the residuals in Supporting Information, Table S1 and Figure 3B.

To further assess the robustness and statistical confidence of the derived models, bootstrapping analysis for 100 runs was performed (Table 3). Bootstrapping involves the generation of many new data sets from the original data sets after randomly choosing samples from the original data set. The bootstrapped r^2 of 0.962 was obtained for both CoMFA and CoMSIA suggesting that a good internal consistency exists within the underlying data set.
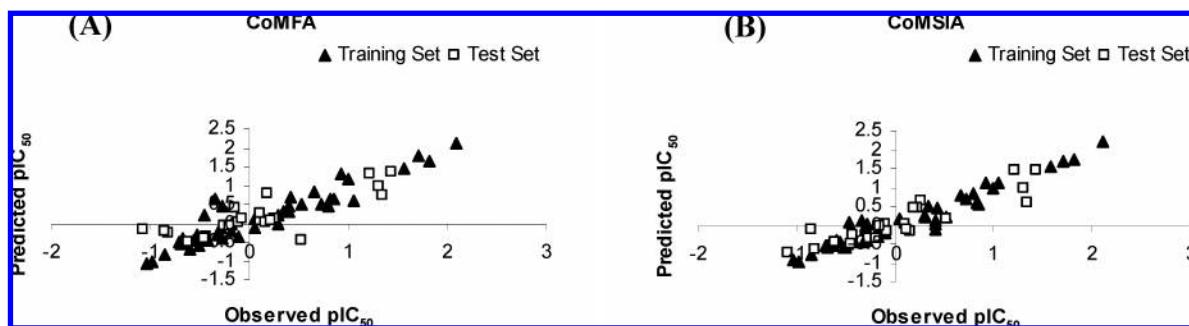
Figure 3. Plot of observed versus predicted pIC_{50} values for the training set and the test set compounds based on the CoMFA (**A**) and CoMSIA (**B**) models.
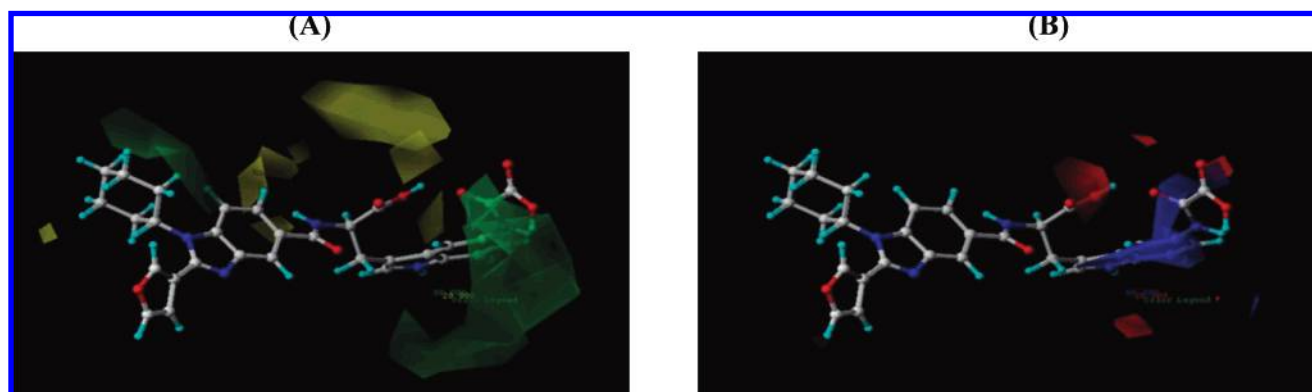
Figure 4. (**A**) CoMFA stdev*coeff steric contour map; green contours indicate regions where bulky groups increase activity, whereas yellow contours indicate regions where bulky groups decrease activity. (**B**) CoMFA stdev*coeff electrostatic contour map; blue contours indicate regions where electropositive groups increase activity, whereas red contours indicate regions where electronegative groups increase activity. The most potent compound **22** is displayed as a reference.

Next, a cross-validation analysis was applied to the set of compounds in the training set to investigate the stability of the CoMFA and CoMSIA models. The training set model was cross-validated using two (leave-half-out) and five (leave 20% out) cross-validation groups 50 times each. The average and standard deviation values of r^2_{cv} are shown in Table 3. When two cross-validation groups were used, the average r^2_{cv} values for CoMFA and CoMSIA were respectively 0.543 (standard deviation = 0.088) and 0.555 (standard deviation = 0.090). By using five cross-validation groups the average r^2_{cv} and SD values were 0.602 and 0.052 for CoMFA, and those for the CoMSIA were 0.635 and 0.035, respectively. Thus, two and five cross-validation analyses of the training set composition for each run were consistent for the CoMFA as well as CoMSIA models.

Validation of the 3D QSAR Models. The predictive abilities of the CoMFA and CoMSIA models were determined from a set of 22 test compounds not included in the model generation. The test inhibitors were accurately predicted by both models. The predicted r^2 values from the CoMFA and CoMSIA models were found to be 0.734 and 0.800, respectively (Table 3). The experimental and predicted pIC_{50} values based on the selected CoMFA and CoMSIA models for the test set compounds are listed with −0.08 and −0.02 of average value of the residuals, respectively, in Supporting Information, Table S1 and graphically shown in Figure 3A,B. The high predictive power of CoMFA and CoMSIA training models for sets of diverse structural scaffolds suggest that these models possess a high accommodating capacity and, hence, wide applicability in the

development of novel NS5B inhibitors. The CoMSIA model gave the best relationship between the predictive r^2 with less residual values of the test set compared to the CoMFA model (Supporting Information, Table S1).

CoMFA Contour Maps. In the CoMFA steric field, the green (sterically favorable) and yellow (sterically unfavorable) contours represent 80% and 20% level contributions, respectively. Similarly in the CoMFA electrostatic field the red (electronegative charge favorable) and blue (electropositive charge favorable) contours represent 80% and 20% level contributions, respectively.

The CoMFA contour map of steric contribution is depicted in Figure 4A. To aid the visualization, the most potent compound **22** is overlaid on the map. A small green contour around the cyclohexyl substituent at the N_1-position of the benzimidazole ring indicates that a sterically bulky group is favored in this region. In case of the highly active compounds **19**−**25**, **49**, and **50**, the cyclohexyl substituent touches this green contour, while many of the least active compounds including compounds **27**−**30**, **41**, and **56** tend to locate their cyclohexyl moiety away from this contour and approach a small sterically unfavorable yellow contour. The large green contour on the right side of the image encompasses the indole moiety along with its R_1 substituent suggesting the requirement of bulky substituents in this region for potent HCV NS5B polymerase inhibitory activity as exemplified by compounds exhibiting the greatest potency (i.e., **20**−**25**, **49**, and **50**) and having functional groups such as sulfonamide (**20**, **21**), acidic functional group (**22**−**25**), acetamide (**49**), and carboxamide (**50**). It is of interest to note that since these

3D QSAR Studies on HCV NS5B Polymerase Inhibitors

J. Chem. Inf. Model., Vol. 48, No. 1, 2008 **49**

functional groups are polar in nature, the requirement for potent inhibitory activity would be to have a bulky yet polar substituent. Two findings lend further support to this concept, namely the greater inhibitory potency of substituted tryptophan analogues than of unsubstituted tyrosine analogues and the reasonably good activity of the phenylglycine analogues **42** and **44**, by virtue of their *meta*- and *para*-OCH$_3$ groups occupying a large green contour. In contrast, compounds **41** and **56**, which are phenyl glycine analogues with *meta*- and *para*-OCH$_3$ groups, showed poor activity because of their lack of a $-$COOH group at the chiral carbon. A large yellow contour map seen in the vicinity of the R substituent at the chiral carbon indicates that occupancy of this sterically unfavorable contour would have a detrimental effect on the HCV NS5B polymerase inhibitory activity. Most potent compounds **19**$-$**25** have a sterically less bulky R ($-$COOH) group and are not seen near the yellow contour map. Less active compounds are found to orient their R groups into this yellow contour as exemplified by the tertiary amide group in compounds **28**, **29**, and **30**, the secondary amide group in compounds **33**, **36**, and **52**, the thiazole moiety in compounds **37**$-$**40**, **53**, and **54**, and the imidazole moiety in compound **55**. Increasing the bulkiness of the substituent at the 2-position of the thiazole and the imidazole rings decreased the inhibitory activity probably because of increased steric hindrance near this yellow contour. This correlation is verified with compound **38**, whose NH$_2$ group contributes to a higher activity than compounds **39**, **40**, **54**, and **55** which, instead, have a methyl or acetyl substituent on the amino nitrogen. Four small yellow contours were observed around the 6-position of the benzimidazole ring, $-$NH of the amide linkage, and bulky R substituents, which suggests that compounds with substituents entering these contours will be less active than those with substituents that do not. One of these yellow contours was found to be occupied by the R [N-(2-pyridinylmethyl) carboxamide] group of compound **52**, one of the least active compounds (IC$_{50}$ = 2.7 μM). A small yellow contour at the 4-position of the indole ring suggests that a substituent other than $-$H at this position would be detrimental to the inhibitory activity.

The electrostatic contour map of the CoMFA model is shown in Figure 4B. To facilitate the visualization, the most potent compound **22** is overlaid on the map. A large red contour map around the R substituent at the chiral carbon indicates that an electronegative group needs to be present at this position for inhibitory activity. Most active compounds have a $-$COOH group at this position, and the least active compounds lack the electronegative group at this position as exemplified by compounds **14** and **41**, whereas compounds **27** (R = isopropyl) and **56** (R = $-$CH$_3$) have an electropositive group. The thiazole substituent at the R position in compounds **39** and **53** was mapped in the vicinity of this red contour. More specifically, the less electronegative S$_1$ atom compared to the corresponding carbonyl oxygen atom and electropositive C$_5$ of the thiazole ring were found to be in close contact with this red contour. Hence, thiazole analogues tend to be less active than $-$COOH analogues. A small red contour seen around the terminal carbonyl oxygen of the oxalic amide substituent suggests that electronegative groups at this position are favorable for inhibitory activity. For example, the highly active compounds **20** ($-$NHSO$_2$CH$_3$), **21** ($-$NHSO$_2$CF$_3$), and **22** ($-$NHCOCOOH)

have electronegative (R$_1$) groups located in this red contour. In contrast, the tryptophan-containing compounds **15** and **17** by having the electronegative groups $-$NO$_2$ and $-$F as R$_1$ substituents, respectively, are relatively less active. These findings point to the need for electronegative groups with hydrogen bond donating capacity. One medium size red contour observed behind the plane of the indole ring are contributed by the N$_1$ and C$_2$ atoms of the indole ring in compounds **4**$-$**8** and is taken as an indication of the role of electronegative groups at these positions in increasing the inhibitory activity. These two atoms of the indole ring have an overall negative charge and hence show reasonable activity. Similarly, this red contour was found to be occupied by the *ortho*-position of the tyrosine ring, and thus substitution of electronegative group at this position in tyrosine analogues may lead to an increase in HCV NS5B polymerase inhibitory activity. Another small red contour was located on the NH$_2$ function of the thiazol-2-amino-4-yl (R substituent) at the chiral carbon and indicates that the electronegative group at this position is essential for activity. For example, reasonably good potency of compounds **5**$-$**8** and **48** could be due to the presence of electron-rich functional groups such as primary amine, secondary amine/amide, and tertiary amine. A small red contour was noticed in front of the plane of the indole ring as a result of the *meta*-OCH$_3$ substituent in the tyrosine analogues **41**, **42**, **44**, and **56**. Compounds **42** and **44** exhibited reasonably good activity as a result of localization of their *meta*-OCH$_3$ group into this red contour. Despite the presence of the *meta*-OCH$_3$ group in compounds **41** and **56**, they were poorly active; this could be mainly due to the fact that they lack the essential carboxyl function at the R position. Change in the heteroaryl moiety at the 2-position of the benzimidazole ring resulted in a tiny red contour suggesting that the electronegative group is required at this position. The oxygen atom of the furyl moiety was localized in close proximity of this red contour thus explaining the higher activity of furyl analogues compared to pyridyl analogues which orient the pyridine ring nitrogen away from this electronegative favorable contour. A large blue contour map wrapped around the R$_1$-position of the indole ring indicated that the electropositive group would increase the HCV NS5B polymerase inhibitory activity. Compounds **22**$-$**25**, **49**, and **50** possess the electropositive carbon atom such as carbonyl carbon, carbon of the tetrazole, and C$_5$ of the indole ring may be responsible for the high activity of these compounds. The electropositive nature of these carbon atoms could be due to their direct attachment with the electron withdrawing groups.

CoMSIA Contour Maps. In the CoMSIA hydrophobic field, the white (hydrophobic favorable) and orange (hydrophobic unfavorable or hydrophilic favorable) contours represent 90% and 10% level contributions, respectively. Similarly, the cyan (hydrogen bond donor favorable) and purple (hydrogen bond donor unfavorable) contours represent 90% and 10% level contributions, respectively, in the hydrogen bond donor fields. In the CoMSIA hydrogen bond acceptor field, the magenta (hydrogen bond acceptor favorable) and red (hydrogen bond acceptor unfavorable) contours represent 90% and 10% level contributions, respectively.

The hydrophobic contour map of the CoMSIA model in the presence of the most active compound **22** is displayed in Figure 5A. The white and orange contour maps highlight
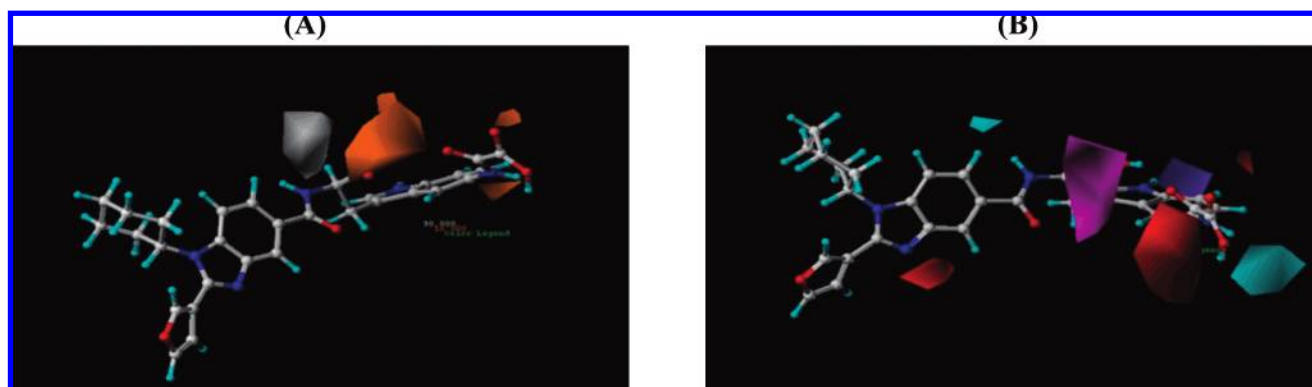
Figure 5. **(A)** CoMSIA stdev*coeff hydrophobic contour maps; white and orange contours indicate favorable and unfavorable hydrophobic groups, respectively, for HCV NS5B polymerase inhibitory activity. **(B)** CoMSIA stdev*coeff hydrogen bond donor and hydrogen bond acceptor contour maps; cyan and purple contours indicate favorable and unfavorable hydrogen bond donor groups, respectively, whereas magenta and red contours indicate favorable and unfavorable hydrogen bond acceptor groups, respectively, for HCV NS5B polymerase inhibitory activity. The most potent compound **22** is displayed as a reference.

areas where hydrophobic and hydrophilic properties are preferred. The presence of a large orange contour map around the R substituent at the chiral carbon suggests that its occupancy by hydrophilic groups would favor an inhibitory activity. In the most active compounds (i.e., **19**−**24**) R was a polar (−COOH) group, whereas in the least active ones R was a nonpolar isopropyl (i.e., **27**) or methyl (i.e., **56**) group. Another orange contour map was located around the hydrophilic group at the R_1-position of the indole ring. Most of the highly active compounds **19**−**25**, **49**, and **50** possess bulky polar functional groups at this position which creates a hydrophilic environment around this contour. The moderate activity of compounds **15**−**17** can be explained based on their hydrophobic R_1 substituent entering this contour. A tiny orange contour falls at the terminal carbonyl of the R_1 substituent of the indole ring suggesting the occupancy of this contour by hydrophilic group would be favorable for the activity as exemplified by the most potent compound **22**. The white contour suggests that the hydrophobic group at this position would increase HCV NS5B polymerase inhibitory effect. Even though the −COOH group of the most potent compound **22** localizes in the vicinity of this contour, it orients in the direction of the orange contour. On the other hand, the less active compounds **29**−**40** and **51**−**55** orient the hydrophilic N−H function of the R substituent toward the white contour. The reasonably good inhibitory activity of compounds **4**−**8**, **47**, and **48** is probably due to occupancy of the white contour by the S_1 and C_5 atoms of the thiazole ring.

The hydrogen bond donor contour map of the CoMSIA model in the presence of the most potent compound **22** is shown in Figure 5B. A cyan contour seen near the R_1-position of the indole ring indicates that the hydrogen bond donor group is favorable for activity. The most potent compounds exhibited a sulfonamide (**20**, **21**), acetamide (**49**), carboxamide (**50**), and acidic functional group (**22**−**25**) as a substituent at the R_1-position of the indole ring. All these groups possess hydrogen bond donor atoms which seem to play a key role in their predominantly positive effect on HCV NS5B polymerase inhibitory activity. Compounds **15** and **17** have hydrogen bond acceptor groups at the 5-position of the indole ring such as −NO_2 and F, respectively, whereas compound **16** possesses a non-hydrogen bond forming −CH_3 group, thus accounting for their relatively poor activity. A

small cyan contour at the −NH of the amide linking bridge indicates that the hydrogen bond donor is necessary for HCV NS5B polymerase inhibitory activity. A purple contour distantly located from the −NH of the indole ring suggests a possible positive effect on activity by having a hydrogen bond acceptor in this region.

The hydrogen bond acceptor contour map of the CoMSIA model in the presence of the most potent compound **22** is depicted in Figure 5B. A magenta contour around the R substituent at the chiral carbon represents the higher activity of compounds having a hydrogen bond acceptor group at this position. Since the most active compounds contained a hydrogen bond acceptor group such as the carbonyl of the −COOH and the least active compounds **14**, **27**, **41**, and **56** did not, this group must be required for a strong inhibitory activity. However, in the poorly active compounds **37**−**40** and **53**−**55**, the R substituent is a thiazole ring whose weak hydrogen bond acceptor sulfur atom is closer to the magenta contour. A large red contour map seen near the R_1-position of the indole ring indicates that hydrogen bond acceptor groups are unfavorable (i.e., hydrogen bond donor group favorable). The hydrogen bond donating groups at this position of the most active compounds **20**−**25** are responsible for the potent activity. A medium size red contour located not so close to any of the atoms of the compounds suggests that occupancy of this spatial region by hydrogen bond acceptor group would cause a decrease in activity. A small hydrogen bond acceptor unfavorable red contour map on the right side of the image resulted from the *para*-position of the tyrosine analogues. The oxygen atom of the −OH function in the tyrosine series tends to be closer than its −H to the red contour and thus explains the poor activity of tyrosine analogues.

Ligand-based methods of analysis such as CoMFA and CoMSIA are widely used not only because they are not very computationally intensive but also because they can lead to the rapid generation of QSARs from which the biological activity of newly designed compounds can be predicted. In contrast, an accurate prediction of activity of untested compounds based on the computation of binding free energies is both complicated and lengthy. Thus, based on the findings derived from the 3D QSAR studies, several analogues of the most potent molecule (compound **22**) were designed (Table 4). Due to the high predictive power of the

3D QSAR Studies on HCV NS5B Polymerase Inhibitors

J. Chem. Inf. Model., Vol. 48, No. 1, 2008 51

Table 4. Structures and Predicted pIC$_{50}$ Values of the Designed Molecules (**22a**−**22h**)
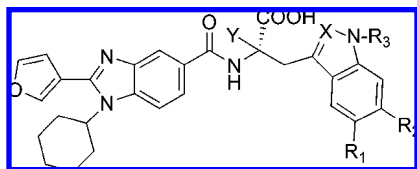
compd	R$_1$	R$_2$	R$_3$	X	Y	pIC$_{50}$[a]
22	NHCOCOOH	H	H	C	H	2.10
22a	NHCOCONH$_2$	H	H	C	H	2.40
22b	NHCOCOOH	OH	H	C	H	2.73
22c	NHCOCOOH	H	COCH$_3$	C	H	2.52
22d	NHCOCOOH	H	H	N	H	2.60
22e	NHCOCOOH	H	H	C	CH$_3$	2.30
22f	NHCOCONH$_2$	OH	COCH$_3$	C	CH$_3$	2.83
22g	NHCOCONH$_2$	OH	COCH$_3$	C	H	2.90
22h	NHCOCONH$_2$	OH	COCH$_3$	N	H	2.98

[a] pIC$_{50}$ values predicted by CoMSIA model.

CoMSIA model, CoMSIA contour maps were used for the design of new analogues with superior pIC$_{50}$ values. In this manner, several key structural features important to the inhibitory activity of NS5B were identified. First, replacement of the −NHCOCOOH group by −NHCOCONH$_2$ at the R$_1$-position of the indole/indazole ring significantly improved the pIC$_{50}$ values due to better placement of the terminal −NH$_2$ group into a hydrogen bond donor-favorable cyan contour (e.g., compounds **22a**, **22f**, **22g**, and **22h**). Second, the unfavorable hydrogen bond acceptor red contour suggested the substitution of the hydrogen bond donor group at the R$_2$-position of the indole ring (e.g., compounds **22b**, **22f**, **22g**, and **22h**). Third, the unfavorable hydrogen bond donor purple contour suggested the possibility of substituting the hydrogen bond acceptor group (−COCH$_3$) at the R$_3$-position of the indole ring for obtaining potent analogues (e.g., compounds **22c**, **22f**, **22g**, and **22h**). Fourth, replacement of the C$_2$ atom of the indole ring by a N atom led to the indazole analogues with improved activity (e.g., compounds **22d** and **22h**). Fifth, at the chiral carbon favorable hydrophilic orange contour on one side and a favorable small hydrophobic white contour on the other side can be exploited to yield potent analogues (e.g., compounds **22e** and **22f**). In summary, a five-step optimization of the most potent compound **22** was performed through substitution at the chiral carbon and N$_1$, C$_5$, and C$_6$ of the indole ring and replacement of the indole ring by an indazole moiety. All of the designed analogues led to improved predicted pIC$_{50}$ values compared to the reference compound **22** (Table 4).

Docking Study. *Selection of X-ray Crystal Structure and Validation of the Docking Protocol.* Comparison of the NS5B-GTP complex structure[64] with the NS5B-indole-based inhibitor complex structure[24] shows that the binding sites for GTP and for the indole-based inhibitors are close in space but clearly distinct. The GTP molecule binds on a shallow surface in proximity to the fingertip α-helix A, whereas the inhibitor fills the pocket occupied in the apoenzyme by the same α-helix. In addition, the binding site for the indole-based inhibitors is also distinct from the binding site for another group of non-nucleoside inhibitors based on thiophene, phenylalanine, or dihydropyranone scaffolds. These compounds bind to a common site on an elongated hydrophobic

cleft in the thumb domain near the C-terminus of the polymerase.[26,28,64] Although the binding sites for this group of compounds and for the indole-based inhibitors are 14 Å apart, they reside on opposite sides of the thumb domain. The recent high-resolution crystal structure of the NS5B polymerase complexed with an indole-based tetracyclic inhibitor[36] further supports the existence of an allosteric binding site which is different from that of the GTP and phenylalanine binding sites. In addition, enzyme-bound conformation of a benzimidazole inhibitor, as determined by NMR experiments, suggests that indole- and benzimidazole-based inhibitors bind to the same allosteric binding site of HCV NS5B polymerase.[65] Chemical and topological similarities between indole-based tetracyclic inhibitor and benzimidazole, quinoxaline, and indole N-acetamide inhibitors prompted us to use this recently reported crystal structure for carrying out structure-based molecular modeling of these derivatives.

The most straightforward method of evaluating the accuracy of a docking procedure is to determine how closely the lowest energy pose (binding conformation) predicted by the object scoring function, Glidescore (Gscore) in our case, resembles an experimental binding mode as determined by X-ray crystallography. In the present study, the extra precision (XP) Glide docking procedure was validated by removing a crystallographic bound tetracyclic indole inhibitor (compound **67**)[36] from the binding site and redocking it to the binding site of HCV NS5B polymerase. Our analysis suggests an excellent agreement between the localization of the inhibitor upon docking and from the crystal structure as evident from the 0.827 Å root-mean-square deviations. Thus, the present study suggests the high docking reliability of Glide in reproducing the experimentally observed binding mode for HCV NS5B polymerase inhibitor and the parameter set for the Glide docking reasonably reproduces the X-ray structure.

Binding Mode of the Most Potent HCV NS5B Polymerase Inhibitor (22). The binding mode of the most potent compound **22** is displayed in Figure 6A. The benzimidazole ring forms hydrophobic contacts with residues identified as His428, Trp500, Pro495, and Val494. The cyclohexyl group attached to the N$_1$ position of the benzimidazole ring is located in a deep hydrophobic pocket formed by Leu392, Ala395, Ala396, Ile424, Leu425, and Phe429. It is noteworthy to note that on one side of the cyclohexyl ring there is sufficient room for accommodating a large hydrophobic substituent, which is inferred from the steric green contour seen by CoMFA, whereas on the other side extension of the cyclohexyl moiety into a larger group would potentially lead to steric clash with Leu424 and His428, which is evident from the presence of a small yellow contour. These observations were corroborated by finding that increasing the ring size to a seven-membered ring (cycloheptyl) lowered the inhibitory activity and replacing the cyclohexyl ring with a cyclopentyl ring resulted in only a 3-fold decrease in potency.[20] The methyl substitution at the 2-position of the cyclohexane ring (potentially occupying a green contour map) was found to be 4-fold more potent than the corresponding methyl group at the 4-position (potentially occupy the sterically forbidden site).[20] The furyl moiety at the 2-position of the benzimidazole ring fits nicely into a narrower hydrophobic pocket formed by Val37, Leu392,
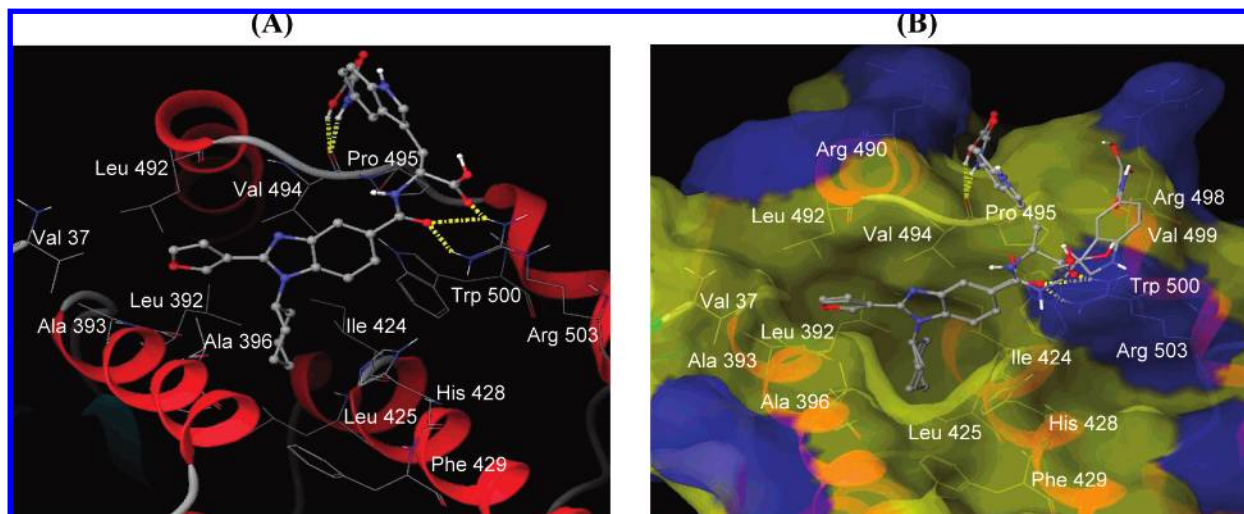
Figure 6. **(A)** Docked conformation derived for compound **22** with the allosteric binding site of HCV NS5B polymerase. Hydrogen bonds are shown as dotted yellow lines. Active site amino acid residues are represented as sticks, while the inhibitor is shown as ball and stick model. **(B)** Comparison of the docked (ball and stick model) and the free-state lowest energy (stick model) conformations of compound **22** overlaid within the allosteric binding site of NS5B. The protein is represented as Macromodel surface.

Ala393, Leu492, and Val494, whereas a pyridyl substituent at the same position extends its hydrogen atoms into close proximity (2.0−2.5 Å) of the side chains of Val37 and Leu392 (potential steric clash). This interaction profile of furyl (highly active) and pyridyl (less active) analogues explains the variations in biological activities that exist between them. Various types of heterocyclic rings have been placed at the 2-position of the benzimidazole ring among which only 5-membered heterocycles such as furan, pyrrole, imidazole, thiophene, and thiazole were found to be more active than a 6-membered heterocycle such as pyridine.[20] The presence of a hydrophilic favorable, orange contour for the indole ring and the oxalic amide function indicates that they are fairly exposed to the solvent. Furthermore, no sterically unfavorable yellow contour was seen around the R_1-substituent of the indole ring.

Our docked model revealed that hydrogen bonding is an important interaction between the inhibitor and the HCV NS5B polymerase. A total of five hydrogen bonds are formed between compound **22** and the HCV NS5B polymerase. The carbonyl group of the amide linkage forms a hydrogen bond with −NH_2 of the guanidine group of Arg503 (C=O−H_2N, 2.25 Å, 154.1°) and (C=O−H_2N, 2.38 Å, 148.4°). Similarly, the carbonyl oxygen of the −COOH group at the R position forms a hydrogen bond with the guanidine group of Arg503 (C=O−H_2N, 1.96 Å, 149.1°). Thus, Arg503 creates a positive electrostatic potential favoring its binding to the −COOH group at the chiral carbon of the most potent compounds **22** and **24**. The presence of a hydrophilic favorable orange contour for the −COOH group at the chiral center indicates that this group is solvent-exposed. Further evidence of the essential role of the −COOH group at the R-position is shed by experiments in which compounds without this group tended to be less active than those with it.[21,22] The CoMSIA hydrogen bond acceptor favorable magenta contour around the carbonyl oxygen atom of the −COOH group further supports the structure-based analysis. The oxalic amide function at the R_1-position of the indole ring was seen involved in hydrogen bonding interaction with the backbone of Val494 (NH−O=C, 1.90 Å, 163.9° and

OH−O=C, 1.92 Å, 161.5°). These results, suggesting the importance of hydrogen bond donor groups at the R_1-position of the indole ring, are supported by the presence of a CoMSIA hydrogen bond donor favorable cyan contour. However, the formation of such hydrogen bonds should be interpreted with caution since Val494 is entirely solvent-exposed, thus creating a potential competition between water molecules and the inhibitor for forming the hydrogen bonds. Despite the presence of hydrogen bond donor groups at the *para*-position of the tyrosine moiety, these analogues were found to be less active compared to their tryptophan counterparts since the tyrosine moiety was not properly positioned to form a hydrogen bond with the backbone of the Val494 residue. Thus, the 3D QSAR contour maps and allosteric binding site topography complements each other.

The majority of the compounds in the training set and the test set have an extension at the 5-COOH function of the benzimidazole and at the 6-COOH function of the quinoxaline and indole N-acetamide inhibitors (analogous to the 6-carboxamide function of the bound tetracyclic indole inhibitor);[36] therefore, this extended moiety requires an additional binding site on NS5B. This site, being relatively shallow and significantly solvent-exposed, will stabilize the extended moiety of the inhibitors in either of the two possible orientations (Figure 6B). One, seen in the docked conformation, extends near the guanidine group of Arg490 and forms a hydrogen bond with the backbone of Val494. The other, present in a ligand-based alignment conformation, would engage in electrostatic interactions with the guanidine group of Arg498 and the backbone of Val499. Structural information on the benzimidazole and quinoxaline inhibitors in the training set and the test set binding to the HCV NS5B polymerase is not yet available. The present docking studies suggest binding modes for these inhibitors at the allosteric site of the enzyme. Half of the structure of the inhibitors (which includes the N_1 cyclohexyl ring, C_2 furyl or pyridinyl ring, 5-substituent of the benzimidazole, and benzimidazole itself) was predicted according to the experimentally determined tetracyclic indole analogue.[36] A docking study indicates that all the members of the training set and test set

3D QSAR Studies on HCV NS5B Polymerase Inhibitors

J. Chem. Inf. Model., Vol. 48, No. 1, 2008 **53**

inhibitors adopt an unfavorable conformation with high internal ligand energy compared to the free-state (ligand-based conformation).

CONCLUSION

The ligand- and receptor-based 3D QSAR study of 67 benzimidazole, tetracyclic indole, quinoxaline, and indole N-acetamide inhibitors of HCV NS5B polymerase was carried out using CoMFA and CoMSIA tools. A high bootstrapped r^2 value for CoMFA and CoMSIA models with a small standard deviation indicated the existence of a similar relationship among all of the compounds used to build the model. In addition to steric and electrostatic fields, hydrophobic and hydrogen-bond donor/acceptor fields were also found to be important for HCV NS5B polymerase inhibitory activity as exemplified by the higher predictive power of the CoMSIA model relative to the CoMFA model. The CoMFA results suggested that the tryptophan derivatives are more active than the tyrosine derivatives. The results obtained from the CoMSIA model revealed the importance of hydrogen bond donor and acceptor favorable contours at the R_1-position of the indole ring and at the R-position at the chiral carbon, respectively, for potent HCV NS5B polymerase inhibitory activity. Docking studies revealed that the N_1-cyclohexyl and the C_2-furyl substituents of the benzimidazole ring were binding to a deep and to a narrower hydrophobic pocket, respectively. The carbonyl oxygen atoms of the amide linkage and the $-COOH$ group at the chiral carbon were found to be involved in a hydrogen bond network with the guanidine of Arg503. The acidic H-bond donor group at the R_1-position of the indole ring was found to be solvent-exposed and forms hydrogen bonding with the backbone of Val494. Interpretation of the CoMFA and CoMSIA contour maps in context of the topology of the allosteric binding site of HCV NS5B polymerase provided a better understanding of the inhibitor-HCV NS5B polymerase interactions. The results obtained from 3D QSAR models were found to accurately predict the HCV NS5B polymerase inhibitory activity of structurally diverse test set compounds and to yield reliable clues for further optimization of the benzimidazole derivatives in the data set.

ACKNOWLEDGMENT

Supporting Information Available: Observed versus predicted pIC_{50} values for the training and test set compounds (Table S1) and results of combinations of different CoMSIA fields (Table S2). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Choo, Q. L.; Kuo, G.; Weiner, A. J.; Overby, L. R.; Bradley, D. W.; Houghton, M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **1989**, *244*, 359−362.

(2) Alter, M. J.; Margolis, H. S.; Krawczynski, K.; Judson, F. N.; Mares, A.; Alexander, W. J.; Hu, P. Y.; Miller, J. K.; Gerber, M. A.; Sampliner, R. E. The natural history of community-acquired hepatitis C in the United States. The Sentinel Counties Chronic non-A, non-B Hepatitis Study Team. *N. Engl. J. Med.* **1992**, *327*, 1899−1905.

(3) Alter, M. J. The detection, transmission, and outcome of hepatitis C virus infection. *Infect. Agents Dis.* **1993**, *2*, 155−166.

(4) Di Bisceglie, A. M.; Order, S. E.; Klein, J. L.; Waggoner, J. G.; Sjogren, M. H.; Kuo, G.; Houghton, M.; Choo, Q. L.; Hoofnagle, J. H. The role of chronic viral hepatitis in hepatocellular carcinoma in the United States. *Am. J. Gastroenterol.* **1991**, *86*, 335−338.

(5) Wasley, A.; Alter, M. J. Epidemiology of hepatitis C: geographic differences and temporal trends. *Semin. Liver Dis.* **2000**, *20*, 1−16.

(6) Wang, Q. M.; Heinz, B. A. Recent advances in prevention and treatment of hepatitis C virus infections. *Prog. Drug Res.* **2000**, *55*, 1−32.

(7) Abrignani, S.; Houghton, M.; Hsu, H. H. Perspectives for a vaccine against hepatitis C virus. *J. Hepatol.* **1999**, *31 Suppl 1*, 259−263.

(8) Ferenci, P.; Fried, M. W.; Shiffman, M. L.; Smith, C. I.; Marinos, G.; Goncales, F. L.; Haeussinger, D.; Diago, M.; Carosi, G.; Dhumeaux, D.; Craxi, A.; Chaneac, M.; Reddy, K. R. Predicting sustained virological responses in chronic hepatitis C patients treated with peginterferon alfa-2a (40 KD)/ribavirin. *J. Hepatol.* **2005**, *43*, 425−433.

(9) Lauer, G. M.; Walker, B. D. Hepatitis C virus infection. *N. Engl. J. Med.* **2001**, *345*, 41−52.

(10) Myles, D. C. Recent advances in the discovery of small molecule therapies for HCV. *Curr. Opin. Drug Discovery Dev.* **2001**, *4*, 411−416.

(11) Hijikata, M.; Shimotohno, K. Processing of HCV precursor polyprotein. *Nippon Rinsho* **2004**, *62 Suppl 7*, 54−58.

(12) Tanji, Y.; Hijikata, M.; Hirowatari, Y.; Shimotohno, K. Hepatitis C virus polyprotein processing: kinetics and mutagenic analysis of serine proteinase-dependent cleavage. *J. Virol.* **1994**, *68*, 8418−8422.

(13) Wang, Q. M.; Heinz, B. A. Recent advances in prevention and treatment of hepatitis C virus infections. *Prog. Drug Res.* **2001**, Spec No 79−110.

(14) Powers, J. P.; Piper, D. E.; Li, Y.; Mayorga, V.; Anzola, J.; Chen, J. M.; Jaen, J. C.; Lee, G.; Liu, J.; Peterson, M. G.; Tonn, G. R.; Ye, Q.; Walker, N. P.; Wang, Z. SAR and mode of action of novel non-nucleoside inhibitors of hepatitis C NS5b RNA polymerase. *J. Med. Chem.* **2006**, *49*, 1034−1046.

(15) Bartenschlager, R. Candidate targets for hepatitis C virus-specific antiviral therapy. *Intervirology* **1997**, *40*, 378−393.

(16) De Francesco, R.; Tomei, L.; Altamura, S.; Summa, V.; Migliaccio, G. Approaching a new era for hepatitis C virus therapy: inhibitors of the NS3−4A serine protease and the NS5B RNA-dependent RNA polymerase. *Antiviral Res.* **2003**, *58*, 1−16.

(17) Ni, Z. J.; Wagman, A. S. Progress and development of small molecule HCV antivirals. *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 446−459.

(18) Beaulieu, P. L.; Tsantrizos, Y. S. Inhibitors of the HCV NS5B polymerase: new hope for the treatment of hepatitis C infections. *Curr. Opin. Invest. Drugs* **2004**, *5*, 838−850.

(19) Tomei, L.; Altamura, S.; Bartholomew, L.; Biroccio, A.; Ceccacci, A.; Pacini, L.; Narjes, F.; Gennari, N.; Bisbocci, M.; Incitti, I.; Orsatti, L.; Harper, S.; Stansfield, I.; Rowley, M.; De Francesco, R.; Migliaccio, G. Mechanism of action and antiviral activity of benzimidazole-based allosteric inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. *J. Virol.* **2003**, *77*, 13225−13231.

(20) Beaulieu, P. L.; Bos, M.; Bousquet, Y.; Fazal, G.; Gauthier, J.; Gillard, J.; Goulet, S.; LaPlante, S.; Poupart, M. A.; Lefebvre, S.; McKercher, G.; Pellerin, C.; Austel, V.; Kukolj, G. Non-nucleoside inhibitors of the hepatitis C virus NS5B polymerase: discovery and preliminary SAR of benzimidazole derivatives. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 119−124.

(21) Beaulieu, P. L.; Bos, M.; Bousquet, Y.; DeRoy, P.; Fazal, G.; Gauthier, J.; Gillard, J.; Goulet, S.; McKercher, G.; Poupart, M. A.; Valois, S.; Kukolj, G. Non-nucleoside inhibitors of the hepatitis C virus NS5B polymerase: Discovery of benzimidazole 5-carboxylic amide derivatives with low-nanomolar potency. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 967−971.

(22) Beaulieu, P. L.; Bousquet, Y.; Gauthier, J.; Gillard, J.; Marquis, M.; McKercher, G.; Pellerin, C.; Valois, S.; Kukolj, G. Non-nucleoside benzimidazole-based allosteric inhibitors of the hepatitis C virus NS5B polymerase: inhibition of subgenomic hepatitis C virus RNA replicons in Huh-7 cells. *J. Med. Chem.* **2004**, *47*, 6884−6892.

(23) McKercher, G.; Beaulieu, P. L.; Lamarre, D.; LaPlante, S.; Lefebvre, S.; Pellerin, C.; Thauvette, L.; Kukolj, G. Specific inhibitors of HCV polymerase identified using an NS5B with lower affinity for template/primer substrate. *Nucleic Acids Res.* **2004**, *32*, 422−431.

(24) Di Marco, S.; Volpari, C.; Tomei, L.; Altamura, S.; Harper, S.; Narjes, F.; Koch, U.; Rowley, M.; De Francesco, R.; Migliaccio, G.; Carfi, A. Interdomain communication in hepatitis C virus polymerase

abolished by small molecule inhibitors bound to a novel allosteric site. *J. Biol. Chem.* **2005**, *280*, 29765−29770.

(25) Harper, S.; Pacini, B.; Avolio, S.; Di Filippo, M.; Migliaccio, G.; Laufer, R.; De Francesco, R.; Rowley, M.; Narjes, F. Development and preliminary optimization of indole-N-acetamide inhibitors of hepatitis C virus NS5B polymerase. *J. Med. Chem.* **2005**, *48*, 1314−1317.

(26) Biswal, B. K.; Cherney, M. M.; Wang, M.; Chan, L.; Yannopoulos, C. G.; Bilimoria, D.; Nicolas, O.; Bedard, J.; James, M. N. Crystal structures of the RNA-dependent RNA polymerase genotype 2a of hepatitis C virus reveal two conformations and suggest mechanisms of inhibition by non-nucleoside inhibitors. *J. Biol. Chem.* **2005**, *280*, 18202−18210.

(27) Chan, L.; Reddy, T. J.; Proulx, M.; Das, S. K.; Pereira, O.; Wang, W.; Siddiqui, A.; Yannopoulos, C. G.; Poisson, C.; Turcotte, N.; Drouin, A.; Alaoui-Ismaili, M. H.; Bethell, R.; Hamel, M.; L'Heureux, L.; Bilimoria, D.; Nguyen-Ba, N. Identification of N,N-disubstituted phenylalanines as a novel class of inhibitors of hepatitis C NS5B polymerase. *J. Med. Chem.* **2003**, *46*, 1283−1285.

(28) Love, R. A.; Parge, H. E.; Yu, X.; Hickey, M. J.; Diehl, W.; Gao, J.; Wriggers, H.; Ekker, A.; Wang, L.; Thomson, J. A.; Dragovich, P. S.; Fuhrman, S. A. Crystallographic identification of a noncompetitive inhibitor binding site on the hepatitis C virus NS5B RNA polymerase enzyme. *J. Virol.* **2003**, *77*, 7575−7581.

(29) Gopalsamy, A.; Lim, K.; Ciszewski, G.; Park, K.; Ellingboe, J. W.; Bloom, J.; Insaf, S.; Upeslacis, J.; Mansour, T. S.; Krishnamurthy, G.; Damarla, M.; Pyatski, Y.; Ho, D.; Howe, A. Y.; Orlowski, M.; Feld, B.; O'Connell, J. Discovery of pyrano[3,4-*b*]indoles as potent and selective HCV NS5B polymerase inhibitors. *J. Med. Chem.* **2004**, *47*, 6603−6608.

(30) Dhanak, D.; Duffy, K. J.; Johnston, V. K.; Lin-Goerke, J.; Darcy, M.; Shaw, A. N.; Gu, B.; Silverman, C.; Gates, A. T.; Nonnemacher, M. R.; Earnshaw, D. L.; Casper, D. J.; Kaura, A.; Baker, A.; Greenwood, C.; Gutshall, L. L.; Maley, D.; DelVecchio, A.; Macarron, R.; Hofmann, G. A.; Alnoah, Z.; Cheng, H. Y.; Chan, G.; Khandekar, S.; Keenan, R. M.; Sarisky, R. T. Identification and biological characterization of heterocyclic inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. *J. Biol. Chem.* **2002**, *277*, 38322−38327.

(31) Gopalsamy, A.; Chopra, R.; Lim, K.; Ciszewski, G.; Shi, M.; Curran, K. J.; Sukits, S. F.; Svenson, K.; Bard, J.; Ellingboe, J. W.; Agarwal, A.; Krishnamurthy, G.; Howe, A. Y.; Orlowski, M.; Feld, B.; O'Connell, J.; Mansour, T. S. Discovery of proline sulfonamides as potent and selective hepatitis C virus NS5b polymerase inhibitors. Evidence for a new NS5b polymerase binding site. *J. Med. Chem.* **2006**, *49*, 3052−3055.

(32) Lee, G.; Piper, D. E.; Wang, Z.; Anzola, J.; Powers, J.; Walker, N.; Li, Y. Novel inhibitors of hepatitis C virus RNA-dependent RNA polymerases. *J. Mol. Biol.* **2006**, *357*, 1051−1057.

(33) Summa, V.; Petrocchi, A.; Pace, P.; Matassa, V. G.; De Francesco, R.; Altamura, S.; Tomei, L.; Koch, U.; Neuner, P. Discovery of alpha, gamma-diketo acids as potent selective and reversible inhibitors of hepatitis C virus NS5b RNA-dependent RNA polymerase. *J. Med. Chem.* **2004**, *47*, 14−17.

(34) Pfefferkorn, J. A.; Nugent, R.; Gross, R. J.; Greene, M.; Mitchell, M. A.; Reding, M. T.; Funk, L. A.; Anderson, R.; Wells, P. A.; Shelly, J. A.; Anstadt, R.; Finzel, B. C.; Harris, M. S.; Kilkuskie, R. E.; Kopta, L. A.; Schwende, F. J. Inhibitors of HCV NS5B polymerase. Part 2: Evaluation of the northern region of (2Z)-2-benzoylamino-3-(4-phenoxy-phenyl)-acrylic acid. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2812−2818.

(35) Pfefferkorn, J. A.; Greene, M. L.; Nugent, R. A.; Gross, R. J.; Mitchell, M. A.; Finzel, B. C.; Harris, M. S.; Wells, P. A.; Shelly, J. A.; Anstadt, R. A.; Kilkuskie, R. E.; Kopta, L. A.; Schwende, F. J. Inhibitors of HCV NS5B polymerase. Part 1: Evaluation of the southern region of (2Z)-2-(benzoylamino)-3-(5-phenyl-2-furyl)acrylic acid. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2481−2486.

(36) Ikegashira, K.; Oka, T.; Hirashima, S.; Noji, S.; Yamanaka, H.; Hara, Y.; Adachi, T.; Tsuruha, J. I.; Doi, S.; Hase, Y.; Noguchi, T.; Ando, I.; Ogura, N.; Ikeda, S.; Hashimoto, H. Discovery of conformationally constrained tetracyclic compounds as potent hepatitis C virus NS5B RNA polymerase inhibitors. *J. Med. Chem.* **2006**, *49*, 6950−6953.

(37) Rong, F.; Chow, S.; Yan, S.; Larson, G.; Hong, Z.; Wu, J. Structure−activity relationship (SAR) studies of quinoxalines as novel HCV NS5B RNA-dependent RNA polymerase inhibitors. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1663−1666.

(38) Stansfield, I.; Pompei, M.; Conte, I.; Ercolani, C.; Migliaccio, G.; Jairaj, M.; Giuliano, C.; Rowley, M.; Narjes, F. Development of carboxylic acid replacements in indole-N-acetamide inhibitors of hepatitis C virus NS5B polymerase. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5143−5149.

(39) Jinyoung, K.; Jin Hee, H.; Youhoon, C. A structure-based 3D QSAR (CoMSIA) study on a series of aryl diketoacids (ADK) as inhibitors of HCV RNA-dependent RNA polymerase. *Bull. Kor. Chem. Soc.* **2006**, *27*, 1919−1922.

(40) Bajaj, S.; Sambi, S. S.; Madan, A. K. Topochemical models for predicting the activity of α,γ-diketo acids as inhibitors of the HCV NS5B RNA-dependent RNA polymerase. *Pharm. Chem. J.* **2006**, *40*, 650−654.

(41) Di Santo, R.; Fermeglia, M.; Ferrone, M.; Paneni, M. S.; Costi, R.; Artico, M.; Roux, A.; Gabriele, M.; Tardif, K. D.; Siddiqui, A.; Pricl, S. Simple but highly effective three-dimensional chemical-feature-based pharmacophore model for diketo acid derivatives as hepatitis C virus RNA-dependent RNA polymerase inhibitors. *J. Med. Chem.* **2005**, *48*, 6304−6314.

(42) Cramer, R. D., 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.

(43) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18−25.

(44) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.

(45) *SYBYL 7.2*; Tripos Inc.: 1699 South Hanley Rd., St. Louis, MO 63144.

(46) Schrodinger, LLC: New York.

(47) Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. Validation of the general-purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982−1012.

(48) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity: A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.

(49) Jorgensen, W. L.; Maxwell, D.; Tirado-Rives, J. Development and testing of the OPLS-All Atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(50) Bytheway, I.; Cochran, S. Validation of molecular docking calculations involving FGF-1 and FGF-2. *J. Med. Chem.* **2004**, *47*, 1683−1693.

(51) Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. On evaluating molecular-docking methods for pose prediction and enrichment factors. *J. Chem. Inf. Model.* **2006**, *46*, 401−415.

(52) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(53) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750−1759.

(54) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.

(55) Viswanadhan, V. N.; Ghose, A. K.; Revenkar, G. R.; Robins, R. K. Atomic physicochemical parameters for three-dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superimposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163−172.

(56) Klebe, G. The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. *J. Mol. Biol.* **1994**, *237*, 212−235.

(57) Klebe, G.; Mietzner, T.; Weber, F. Methodological developments and strategies for a fast superposition of drug-size molecules. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 35−49.

(58) Wold, H.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbenson, K.; Geladi, P.; Hellberg, S.; Lindburg, W.; Sjostrom, M. In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel: Dordrecht, The Netherlands, 1984; pp 17−95.

(59) Staahle, L.; Wold, S. Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study. *J. Chemom.* **1987**, *1*, 185−196.

(60) Wold, S. Cross validatory estimation of the number of components in factor and principal components models. *Technometrics* **1978**, *4*, 397−405.

(61) Vong, R.; Geladi, P.; Wold, S.; Esbensen, K. Source contributions to ambient aerosol calculated by discriminant partial least squares regression (PLS). *J. Chemom.* **1988**, *2*, 281−296.

(62) Bringmann, G.; Rummey, C. 3D QSAR Investigations on antimalarial naphthylisoquinoline alkaloids by comparative molecular similarity

indices analysis (CoMSIA), based on different alignment approaches. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 304−316.

(63) Bohm, M.; Stürzebecher, J.; Klebe, G. 3D QSAR analyses using CoMFA and CoMSIA to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J. Med. Chem.* **1999**, *42*, 458−477.

(64) Bressanelli, S.; Tomei, L.; Rey, F. A.; De Francesco, R. Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *J. Virol.* **2002**, *76*, 3482−3492.

(65) Lapante, S.; Jakalian, A.; Aubry, N.; Bousquet, Y.; Ferland, J. -M.; Gillard, J.; Lefebvre, S.; Poirier, M.; Tsantrizos, Y. S.; Kukolj, G.; Beaulieu, P. L. Drug design: Binding mode determination of benz-imidazole inhibitors of the hepatitis C virus RNA polymerase by a structure and dynamics strategy. *Angew. Chem., Int. Ed.* **2004**, *43*, 4306−4311.