http://dx.doi.org/10.5012/bkcs.2013.34.3.837

Molecular Modeling of Small Molecules as BVDV RNA-Dependent RNA Polymerase Allosteric Inhibitors

Han-ha Chai,* Dajeong Lim, Hee-yeoul Chai,† and Eunkyoung Jung‡

*Division of Animal Genomics and Bioinformatics, National Institute of Animal Science, Rural Development Administration, Suwon 441-706, Korea. *E-mail: hanha@korea.kr*

†Division of Biosafety Evaluation and Control, Korea National Institute of Health, Chungcheongbuk-do 363-951, Korea

‡Insilicotech Co. Ltd., C-602 Korea Bio Park, Gyeonggi-Do 463-400, Korea

Received October 9, 2012, Accepted December 18, 2012

Bovine viral diarrhea virus (BVDV), a major pathogen of cattle, is a well-characterized pestivirus which has been used as a good model virus for HCV. The RNA-dependent RNA polymerase (RdRp) plays a key role in the RNA replication process, thus it has been targeted for antivirus drugs. We employed two-dimensional quantitative structure-activity relationship (2D-QSAR) and molecular field analysis (MFA) to identify the molecular substructure requirements, and the particular characteristics resulted in increased inhibitory activity for the known series of compounds to act as effective BVDV inhibitors. The 2D-QSAR study provided the rationale concept for changes in the structure to have more potent analogs focused on the class of arylazoenamines, benzimidazoles, and acridine derivatives with an optimal subset of descriptors, which have significantly contributed to overall *anti*-BVDV activity. MFA represented the molecular patterns responsible for the actions of antiviral compound at their receptors. We conclude that the polarity and the polarizability of a molecule play a main role in the inhibitory activity of BVDV inhibitors in the QSAR modeling.

Key Words : BVDV NS5B polymerase, Allosteric inhibitor, QSAR, MFA

Introduction

Bovine viral diarrhea virus (BVDV) is a major bovine pathogen that will become the most wide-spread virus in cattle. BVDV infection may also be one of the main reasons for economic losses in the livestock industry. For the United States alone, this translates roughly into an average loss of \$10 to \$40 per calving.[2] The infection can cause a decrease in milk production, reduced reproductive performance, growth retardation, and increased mortality among young stock causing immune system dysfunction and a predisposition to secondary viral and bacterial infections.[1-3] Currently, there is no specific antiviral agent directed against BVDV infection, although it may be contained by vaccines and control programs. Therefore, antiviral lead compounds that specifically inhibit the replication of the virus are very important to treat expensive animals in breeding programs, since they could be used to safeguard cattle that live on farms in close proximity to an infected farm with *anti*-BVDV drugs on hand; vaccines do not confer protection until 10 to 14 days after being administered. On the other hand, BVDV relies on the host cell for its replication and a single replication cycle of BVDV lasts 10 to 20 h.[24,25]

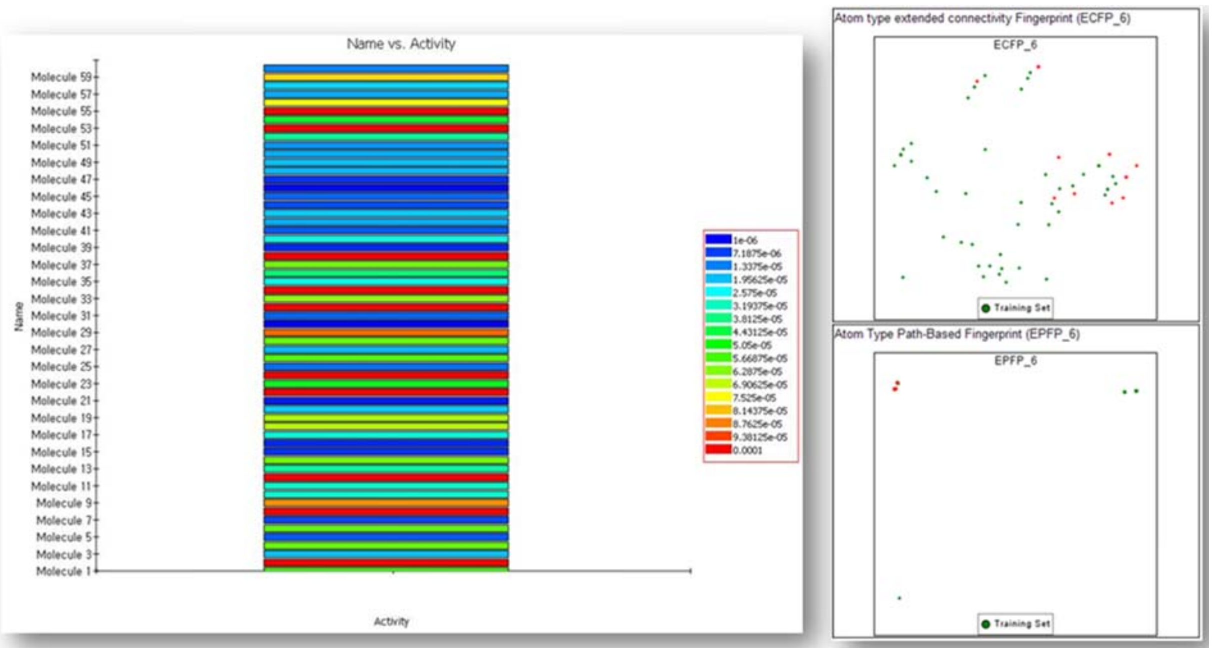
BVDV, a Flaviviridae Pestivirus, has been a good model virus for investigating HCV, which is a member of genus Hepacivirus, which belongs to the same family. BVDV, like HCV, is a small enveloped virus, with a diameter between 43 and 58 nm that has a single 12.6 kb plus-strand RNA genome encoded in a single polyprotein ($_{NH_2}$-N_{pro}-C-E_{rms}-E1-E2-P7-NS2-NS3-NS4A-NS4B-NS5A-NS5B-$_{COOH}$).[4,5] The prototypic representative of BVDV and HCV is different, but they largely share a similar replication cycle and molecular characteristics. The NS5B of both viruses has RNA-dependent RNA polymerase (RdRp) activity and the NS5B RdRp is responsible for genome replication, which alone is capable of RNA synthesis as a part of a larger membrane-associated replicase complex.[6] This is why the NS5B RdRp is the main target of antiviral compound research. Over the past eight years, there have been many screening efforts targeting BVDV RdRp, resulting in the identification of potent inhibitors. This has allowed researchers to understand the structure-activity relationship (SAR) in addition to developing novel, more effective lead compounds for inhibition of BVDV replication. Discovery of small molecule inhibitors of BVDV RdRp as a potential therapeutic target has been reported in the literature with various scaffolds, such as imidazopyridines,[7-11] benzimidazole derivatives,[12,13] arylazoenamines,[14,15] indole derivatives,[16] γ-carboline derivatives,[17] thiosemicabarzone,[18] diphenylmethane,[19,20] and aromatic cationic molecules.[21,22] The majority of *anti*-BVDV inhibitors could also be taken as an accurate measurement of antiviral activity against HCV or other Pestiviruses (*e.g.*, CSFV, BDV) and Flaviviruses (*e.g.*, YFV, WNV, DENV) in the same family for the purpose of exploiting approved treatments. However, the direct relationship between *anti*-BVDV activity and HCV blocking activity is not known, if BVDV was used as a surrogate system for HCV. Recently some *anti*-BVDV hits have been identified by both experimental approaches and molecular modeling of the pharmacophore and the binding interactions; for example arylazoenamines[14,15] and imidazopyridines.[10,11] Based on this scaffold, a series of compounds have been analyzed by ligand and structure-based design methods and will be further optimized for prevention and control of BVDV, HCV, and other single strand RNA viruses.

A three-dimensional quantitative structure-activity relationship (3D-QSAR) approach, such as molecular field analysis (MFA) or pharmacophore models have widely been used not only to find small molecules complementary in shape and charge to a biomolecular target but also to provide a molecular framework that carries the essential features responsible for biological activity.[23] In this research, we built several QSAR models for a series of arylazoenamine derivatives[14] using 2D-QSAR and MFA to understand the effects of structure on activity, which make it possible to predict the properties of basic aromatic analogues. Further results can be used to understand interactions between the molecule's functional groups of highest activity with those of BVDV RdRp. To obtain insight into the key structural features affecting the activity, we used the biological data of *anti*-BVDV to apply a few standards: (i) all activity values of a series molecule having various scaffolds in the data set must be obtained by definitive experimental conditions, methods, and procedures; (ii) biological data for the model should also include a wide range of activity; (iii) the data collected in the training set should reflect as much as possible the complete property space for the class of molecules since the QSAR results can be used to confidently predict the most likely compounds of the best activity; and (iv) the external validation set must be different from the training and test set of compounds, and satisfy conditions (i) and (ii).

Recently, M. Tonelli *et al.* reported that arylazoenamines have been synthesized and evaluated in cell-based assays for cytotoxicity and antiviral activity against most frequently affected viruses such as CVB-2 (Coxsackie virus type B2), RSV (Respiratory syncytial virus), and BVDV. They used the computational approach to identify predictive pharmacophore models for them,[14] and to estimate the docking procedure with previously known pharmacophore constants and binding affinity of active compounds for the RdRp.[15] Like M. Tonelli *et al*., we explore the potential of this class of molecules as starting inhibitors of BVDV RdRp, because arylazoenamines mimic the effect of aryl and basic moieties. Its basic derivatives can be considered as biological activity and correlating the variation in this activity to the changes in polarity, electronic distribution, and H-bonding or its electron withdrawing properties. The aromatic portion of this molecule shall prove to be essential for specificity and potency through hydrophobic interactions. Of particular importance is the role of aromatic ring stacking at the allosteric active site near the substrate-binding site of BVDV RdRp, which has the appropriate shape and size for specific binding of aromatic rings. Our predictive QSAR models were made in two stages: (i) a local model aimed at particular series of arylazoenamines for both training and test sets, and then (ii) the local model expanded to a global model, which should be validated with a more diverse set of basic aromatic derivatives as the prediction set.

Materials and Methods

Selection of Molecules and Biological Data. We took a data set of 60 arylazoenamines for QSAR models targeted to BVDV NS5B polymerase from the literature.[14,15] To validate the models, we tested 24 benzimidazoles[12] and 10 acridine



(a) (b)

Figure 1. (a) The distribution of biological data (EC_{50} in M) in a data set of 60 arylazoenamines and (b) the split data set in chemical spaces defined by atom type expended connectivity of fingerprint (ECFP6) and atom type path-based fingerprint (EPFP6) in 47 training (green) and 13 test (red) sets from the original data.

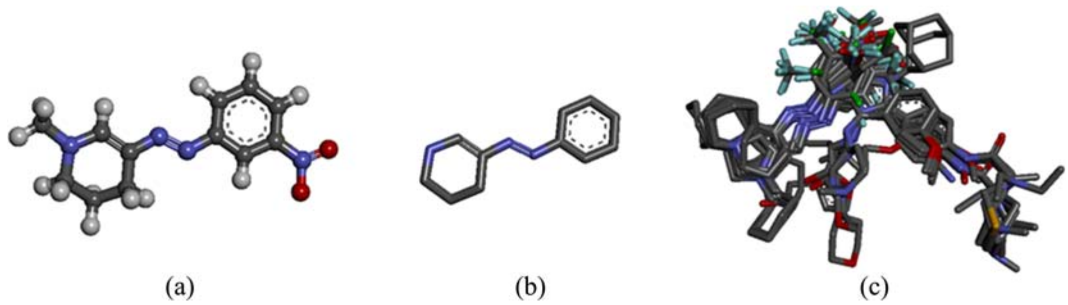
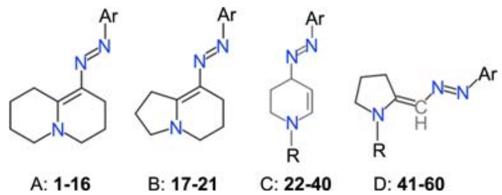


Figure 2. (a) The most active compound **30** structure in the arylazoenamines, (b) The region is used as a substructure for alignment, (c) all of the structures in the analogue series based on a defined substructure (b) in the active molecule are aligned.

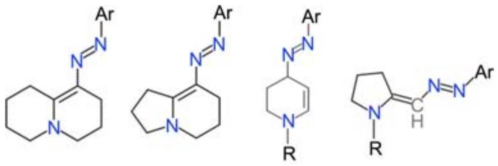
derivatives[26] against the same binding target on a prediction set. The biological data obtained is EC_{50}: 50% effective concentration in μM, which is the amount required to achieve 50% protection of MDBK cells from BVDV-induced cytopathogenicity, as determined by the MTT method. Of the 60 arylazoenamines, eleven compounds exhibited high activity with EC_{50} in the range of 0.8-10 μM, and the other 20 compounds had EC_{50} between 11 and 30 μM. Only 29 compounds had EC_{50} in the range 31-100 μM (Fig. 1(a)). They were converted into pEC_{50} (-log EC_{50}) values and used as dependent variables in QSAR models. To build QSAR models, the compounds from the original data were split into training and test sets. Since the predictive power of the QSAR model relies on the quality of the training and test sets, the distribution represented in Figure 1 can then be used to get an indication of biological activity and the structural similarity/diversity in the original data set of arylazoenamines before effectively dividing the data into a training set and a test set. In choosing a small subset to represent a large data set, both the similarity of the structural diversity (Fig. 1(b)) and the range of biological activity were clustered using the Pareto optimization method as shown Figure 1. In Fig. 1(b), the spilt analysis was given to evaluate the distribution of the two data sets in various chemical species using two principal components as ECFP6 and EPFP6 molecular fingerprints which encode both the electronic state and topological environments of an atom in a molecule [36], thus can measure the structural diversities for all data set. Therefore, this proved the information that molecules in training set may reflect the similar property spaces with those in a test set.

Table 1. Structure, actual and predicted activities for 63 arylazoenamines based on the best 2D/3D QSAR models



Comp.	Aryl=	Actual pEC_{50}	2D-QSAR Predicted	2D-QSAR Residual[a]	3D-QSAR (MFA) Predicted	3D-QSAR (MFA) Residual[a]	Test set[b]
1	4-F-Ph	4.252	3.917	0.335	4.089	0.163	
2	2-Cl-Ph	4.000	4.145	-0.145	3.934	0.066	
3	3-Cl-Ph	4.699	4.732	-0.033	4.655	0.044	
4	4-Cl-Ph	4.222	3.998	0.224	4.296	-0.074	
5	3-Br-Ph	5.000	4.829	0.171	4.933	0.067	
6	4-Br-Ph	4.229	4.029	0.200	4.231	-0.002	
7	3-CF_3-Ph	5.097	4.591	0.506	4.918	0.179	
8	3-NO_2-Ph	4.000	3.981	0.019	4.295	-0.295	Test
9	4-NO_2-Ph	4.066	3.785	0.281	4.507	-0.441	Test
10	3,4-DiCl-Ph	4.523	4.496	0.027	4.573	-0.050	
11	3,5-DiCF_3-Ph	4.509	4.341	0.168	4.318	0.191	Test
12	3-CF_3-4-F-Ph	4.000	4.469	-0.469	4.213	-0.213	
13	3-CF_3-4-Cl-Ph	4.456	4.323	0.133	4.261	0.195	
14	3-CF_3-4-Br-Ph	4.215	4.364	-0.149	4.101	0.114	Test
15	1-Naphthyl	5.222	5.377	-0.155	5.237	-0.015	
16	7-Chloro-1-quinolyl	5.301	5.410	-0.109	5.306	-0.005	
17	4-Cl-Ph	4.523	4.475	0.048	4.559	-0.036	

Table 1. Continued



A: **1-16** B: **17-21** C: **22-40** D: **41-60**

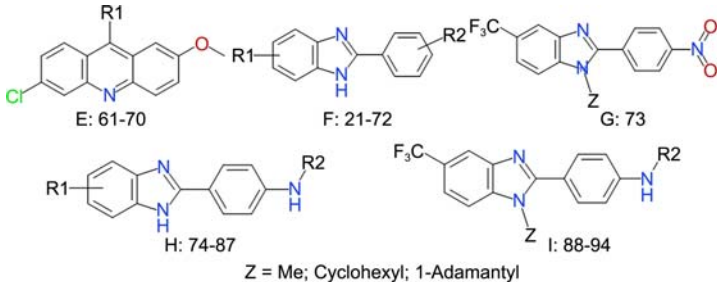
Comp.	Aryl=	Actual pEC_{50}	2D-QSAR		3D-QSAR (MFA)		Test set[b]
			Predicted	Residual[a]	Predicted	Residual[a]	
18	3-NO_2-Ph	4.167	4.333	-0.166	4.510	-0.343	Test
19	3,4-DiCl-Ph	4.174	4.651	-0.477	4.422	-0.248	
20	3-CF_3-4-Cl-Ph	4.678	4.656	0.022	4.617	0.060	
21	1-Naphthyl	5.456	5.344	0.112	5.418	0.038	
R = CH_3							
22	Phenyl	4.000	4.318	-0.318	4.067	-0.067	
23	4-F-Ph	4.319	4.145	0.174	4.443	-0.124	
24	2-Cl-Ph	4.000	4.455	-0.455	3.876	0.124	
25	3-Cl-Ph	4.921	4.513	0.408	4.768	0.153	
26	4-Cl-Ph	4.237	4.235	0.002	4.510	-0.274	
27	3-Br-Ph	4.721	4.560	0.161	4.507	0.214	Test
28	4-Br-Ph	4.222	4.269	-0.047	4.661	-0.439	Test
29	3-CF_3-Ph	4.046	4.545	-0.499	4.107	-0.061	
30	3-NO_2-Ph	6.097	6.074	0.023	5.946	0.151	
31	4-NO_2-Ph	4.959	4.996	-0.037	4.590	0.368	
32	4-CH_3-Ph	4.000	4.390	-0.390	4.040	-0.040	
33	3,5-DiCF_3-Ph	4.194	4.128	0.066	4.310	-0.116	
34	3-CF_3-4-F-Ph	4.000	4.272	-0.272	4.218	-0.218	
35	3-CF_3-4-Cl-Ph	4.553	4.343	0.210	4.323	0.230	
36	3-CF_3-4-Br-Ph	4.398	4.397	0.001	4.465	-0.067	
37	3-NO_2-4-Cl-Ph	4.215	4.285	-0.070	4.530	-0.315	
38	PentaF-Ph	4.000	3.938	0.062	4.223	-0.223	
39	1-Naphthyl	5.301	5.029	0.272	5.101	0.200	
R = CH_2-C_6H_5							
40	3-CF_3-4-Cl-Ph	4.553	4.299	0.254	4.546	0.007	Test
R = CH_3							
41	Phenyl	5.000	4.684	0.316	4.587	0.413	
42	4-F-Ph	4.745	4.526	0.219	4.717	0.028	
43	4-Cl-Ph	4.678	4.607	0.071	4.724	-0.046	
44	3-Br-Ph	5.046	4.984	0.062	4.839	0.207	Test
45	4-Br-Ph	4.959	4.635	0.324	4.634	0.325	Test
46	3-CF_3-Ph	4.678	5.084	-0.406	5.372	-0.695	
47	3-NO_2-Ph	5.260	5.185	0.075	5.186	0.074	
48	4-NO_2-Ph	4.721	4.395	0.326	4.808	-0.087	
49	4-CH_3-Ph	4.699	4.743	-0.044	4.533	0.166	
50	4-CH_3O-Ph	4.745	4.924	-0.179	4.663	0.082	
51	3,4-DiCl-Ph	4.796	4.573	0.223	4.634	0.162	Test
52	3,5-DiCF_3-Ph	4.444	4.420	0.024	4.514	-0.070	
53	3-CF_3-4-F-Ph	4.000	4.509	-0.509	4.060	-0.060	
54	3-CF_3-4-Br-Ph	4.337	4.596	-0.259	4.679	-0.342	Test
55	PentaF-Ph	4.000	4.122	-0.122	4.127	-0.127	
R = CH_2-C_6H_5							
56	Phenyl	4.131	4.575	-0.444	4.189	-0.059	
57	4-Cl-Ph	4.745	4.306	0.439	4.505	0.240	
58	4-CH_3-Ph	4.658	4.739	-0.081	4.643	0.015	
59	3,4-DiCl-Ph	4.102	4.346	-0.244	3.868	0.234	
60	3-CF_3-4-Cl-Ph	4.854	4.593	0.261	4.692	0.162	Test

[a]Residual = actual pEC_{50} – predicted pEC_{50}. [b]The test set molecules are represented as Test

Then 47 compounds (78% of the original data set) were selected for training set and 13 compounds (22% of the original data set) for the test set; the most active compound (compound **30**, Fig. 2(a)) was included in the training set. The structure and biological data and predicted activity of the compounds in the training and test set are listed in Table 1.

Molecular Modeling. Three-dimensional structure building and all computational studies were performed with the Discovery Studio (DS) 3.0 molecular modeling package[27] on a personal workstation. Molecular structures were built with Sketcher program, then the molecular geometric structures were generated into the local lowest energy conformation minimized by the CHARMm[28,29] force field with a distance-dependent dielectric function, steepest descent method, a convergence criterion of 0.001 kcal/mol. Partial atomic charges were also calculated using the CHARMm method.

Table 2. Structure and predicted activities of both 24 benzimidazoles and 10 acridine derivatives using 2D/3D QSAR models



Comp.	R1	R2	Z	Actual pEC_{50}	2D-QSAR Predicted	2D-QSAR Residual[c]	3D-QSAR(MFA) Predicted	3D-QSAR(MFA) Residual[c]
61	NH_2			7.000	5.172	1.828	-	-
62	$NHCH_2CH_2OH$			6.222	5.582	0.640	-	-
63	$NHCH(Et)CH_2OH$			5.097	5.556	-0.459	-	-
64	$NHCH(iPr)CH_2OH$			6.222	5.604	0.618	-	-
65	$NHCH_2(CH_2)_3\text{-}CH_2OH$			5.824	5.530	0.294	-	-
66	$NHCH_2\text{-}2\text{-}C_4H_4O$			5.886	5.883	0.003	-	-
67	$NHCH_2CH_2\text{-}N(CH_2CH_2)_2O$			5.699	5.685	0.014	-	-
68	$NHCH_2CH_2CH_2\text{-}N(CH_2CH_2)_2O$			5.301	5.585	-0.284	-	-
69	$NHN(CH_2CH_2)_2\text{-}NCH_3$			5.523	5.849	-0.326	-	-
70	$NH\text{-}3\text{-}N(CH)_5$			5.222	5.412	-0.190	-	-
71	$5\text{-}CF_3$	$4\text{-}OCH_3$		6.000	4.504	1.496	5.267	0.733
72	$5\text{-}COCH_3$	$2,4\text{-}DiOCH_3$		4.377	4.037	0.340	5.078	-0.701
73			Ch[a]	4.222	4.378	-0.156	5.265	-1.043
74	H	$COCH_3$		4.046	4.503	-0.457	5.225	-1.179
75	H	$COCH_2CH_3$		4.328	4.663	-0.335	5.123	-0.795
76	H	$COCH_2\text{-}N(CH_2)_4$		5.000	4.920	0.080	5.350	-0.350
77	H	$COCH_2\text{-}N(CH_2)_5$		5.155	4.873	0.282	5.376	-0.222
78	$5\text{-}CF_3$	H		4.824	3.989	0.835	5.248	-0.424
79	$5\text{-}CF_3$	$COCH_3$		5.886	4.865	1.021	5.120	0.766
80	$5\text{-}CF_3$	$COCH_2\text{-}N(C_2H_5)_2$		5.699	5.398	0.301	5.005	0.694
81	$5\text{-}CF_3$	$COCH_2\text{-}N(CH_2)_4$		5.523	5.399	0.124	5.036	0.487
82	$5\text{-}CF_3$	$COCH_2\text{-}N(CH_2)_5$		5.699	5.361	0.338	5.290	0.409
83	5,6-DiCl	$COCH_3$		6.000	5.002	0.998	5.689	0.311
84	5,6-DiCl	$COCH_2\text{-}N(CH_2)_4$		5.602	5.480	0.122	5.755	-0.153
85	5,6-DiCl	$COCH_2\text{-}N(CH_2)_5$		5.699	5.406	0.293	5.744	-0.045
86	5,6-DiCl	$COCH_2\text{-}N(CH_2CH_2)_2O$		5.620	5.181	0.439	5.574	0.046
87	5,6-DiCl	$COCH_2\text{- }N(CH_2CH_2)_2N\text{-}CH_3$		5.398	5.324	0.074	5.759	-0.361
88		H	CH_3	4.161	3.843	0.318	5.177	-1.016
89		$COCH_2\text{-}N(C_2H_5)_2$	CH_3	5.000	5.210	-0.210	5.123	-0.123
90		$COCH_2\text{-}N(CH_2)_4$	CH_3	5.155	5.109	0.046	5.065	0.090
91		$COCH_2\text{-}N(CH_2)_5$	CH_3	5.222	5.050	0.172	5.007	0.215
92		$COCH_2\text{- }N(CH_2CH_2)_2S$	CH_3	5.097	4.957	0.140	5.001	0.096
93		$COCH_2\text{- }N(CH_2CH_2)_2N\text{-}C_6H_5$	CH_3	4.824	4.969	-0.145	4.903	-0.079
94		H	Ada[b]	5.155	4.365	0.790	4.617	0.538

[a]Ch = Cyclohexyl. [b]Ada = 1-Adamantyl. [c]Residual = actual pEC_{50} – predicted pEC_{50}

To gain reliable information about the correlation between molecular structure and activity, the active structure for each molecule must be known; however, none of the compounds have the structure of a BVDV NS5B polymerase/inhibitor complex. Molecular flexibility was calculated for each compound using Poling method[30,31] and recorded as a collection of conformers over a 0-20 kcal/mol interval above the global energy minimum computed for each molecule, and was limited to a maximum of 250 in the conformational space. In this paper, we hypothesized that the active conformer was the minimum energy conformer of the most active compound (compound **30**, Fig. 2).

Molecular alignment in the studied compounds was achieved using a specified substructure (Fig. 2(b)) as the localized common subgroup with the active conformer of the most active compound, computed using the collective root mean square (RMS) of their atomic coordinates *via* producing the best superposition. In this method, the sum of squares of the distances is minimized between all atoms to be superimposed based on a defined substructure as a common feature.[32] For the substructure alignment, we quickly explore the core substructure which is common to all molecules in the total data set including 24 benzimidazoles and 10 acridine derivatives to use aromaticity as done by ISIS and then the molecules and the substructure are aligned using all the mapping to the one with the best RMSD is selected in the process. As the criteria for substructure searching was considered as these aromatic units are a key factor in the aromatic ring stacking at the allosteric active site of BVDV RdRp and are also found in many molecules from candidate inhibitors to BVDV and HCV target. So we think the all data sets collected in the study are giving us information that might allow us to improve how understanding the aromatic ring stacking and hydrophobic interactions are going to be important for that class of BVDV allosteric inhibitors. The resulting stereoview of the total set of aligned molecules is shown in Figure 2. The structure of each molecule was used to build the QSAR models. The 2D-QSAR study included 2D-descriptors (topological representation) and 3D-descriptors (geometrical representation). The descriptors were calculated after substructure alignment in all molecules.

2D-QSAR Models of Generation. The physicochemical properties of each molecule were quantified with the various types of descriptors using property calculations within the QSAR+ module. They were characterized by molecular formats, fragment counts, molecular structure, and properties that depended on their topological or geometrical features. The resulting physical and chemical properties for each molecule were thus reduced to numbers or sequences of numbers by utilizing atomic coordinates and connectivity. In generating a 2D-QSAR model, we excluded a descriptor as an independent variable if any of the variables were constant or highly correlated with another variable for all the compounds. A complete list of remaining descriptors in the study used for 2D-QSAR models is given and described in Table 3.

For the development of 2D-QSAR models, the statistical method used genetic function approximation (GFA)[33-35] to generate a population of equations for the correlation between biological activity and physicochemical properties. GFA uses multiple models to provide different insights into the inquiry system using an evolutionary algorithm that combines Holland's genetic algorithm with Friedman's multivariate adaptive regression splines (MARS).[37,38] In our present study, the GFA method works with a randomly generated population of 45,000 equations using a measure of the fitness of each model. Then, pairs of parent equations are chosen from this set of 45,000 equations and mutation operations are performed to create 45,000 new children equations by repeatedly replacing the worst rated models with better rated models with at most 600,000 evolutionary steps. A key feature of GFA is that it uses linear splines, which produce an accurate interpolation model for the data set. Other default settings were maintained, including the smoothing parameter (d) value of 0.5. We used 4 and 8 for the initial and maximum equation length value, respectively, instead of using constant equation length. The goodness of each generated equation was evaluated on how well it fit the data and the model's predictive power with Friedman's lack of fit (LOF) score.[38]

Molecular Field Analysis. MFA provides valuable information about the correlation between the fields and the activity using energy grids, since descriptors are computed by interaction energy both as a set of aligned molecules and as a probe designed to measure steric (van der Waals carbon CH_3) and electrostatic (positive point charge H^+) effects at a

Table 3. A list of remaining descriptors and their types used in building 2D-QSAR models by GFA method

Type	Descriptors
Structural	Number of aromatic bonds, number of heavy atoms, number of hydrogen-bond acceptors, number of atoms with positive charge atoms, number of rings, number of rotatable bonds, number of 5 rings and 6 rings
Spatial	Jurs descriptors, principal moments of inertia, Shadow indices, radius of gyration
Electronic	Dipole moment, sum of atomic polarizabilities, the pK_a of all ionizable sites, Electrotopological-state indices
Topological	Connectivity indices, wiener index, Zagreb index, kappa shape indices, Graph-theoretical Info content descriptors, Balaban indices, Subgraph counts
Thermodynamic	Log of the partition coefficient, molecular refractivity and AlogP for each molecule and partition the atomic surface areas, atomic surface area for each atom in the molecule, molecular solubility
Shape and volume	The total solvent accessible surface area and volume using 3D coordinates, Polar surface area, the ratio of the polar surface area divided by the total surface area, fractional polar solvent accessible surface area

series of points.[39,40] Regression models built from whole molecular steric and electrostatic fields can be useful for predicting activity and for visualizing favorable and unfavorable interactions. MFA attempts to represent the essential features of a receptor site from the aligned common features of molecules in a 3D isosurface. MFA would be very useful in the current study with available activity data assuming that each structure exhibits the same binding site on BVDV NS5B RdRp. Since the accuracy of MFA depends on the activity and the diversity of the compounds, poor results are possible if used on compounds with a narrow range of activity or that were highly diverse. If all compounds are aligned in a pharmacologically active orientation, diverse molecules may have very different orientations, and thus generated features may not be reliable. This was carefully considered during the current study. For a set of pre-aligned structures, the energy fields were generated through the H^+ and CH_3 probes using a grid spacing of 1.5 Å around the structure. The MFA first calculates the smallest lattice box containing energy grids around the structure, then extends each side of the box by 8.0 (half the grid extension in each direction). A correlation threshold is applied to avoid extremely large or small grid energy values which are less than 1.5 Å or greater than 4.0 Å from any compound atom. Then, an energy cutoff of 5.0 kcal was applied for the remaining grids when assigning the CHARMm atom partial charges to be used in a set of bond-charge-increment rules. A field with 1314 CHARMm energy grids was generated. We used the genetic partial least squares (G/PLS) method with a maximum of 600,000 iterations and a population size of 45,000. The smoothing parameter (d) was kept at 0.5. The optimum number of components was set to 4 and equation length was fixed at 9, including a constant. We validated both the 2D-QSAR models and MFA analysis with a pre-divided test set of 13 arylazoenamines and performed external validation by applying the model to a prediction set of basic aromatic class molecules. Both sets of molecules have known activities but were not used in model generation.

Results and Discussion

In the QSAR study, the models were selected in an attempt to discover which main substituents of molecules affect biological activity and which do not. In replacing functional groups, we observed the quantitative effect of them on the biological activity to establish their role in the molecules and BVDV NS5B RdRp interactions. The important functional groups that are required for the *anti*-BVDV NS5B RdRp activity and their relative positions in space with respect to each other should be identified. We also investigated the influence of portions of parts of the basic aromatic molecules on the BVDV inhibitor activity.

2D-QSAR Models Analysis. We performed QSAR models with the assumption of one-to-one correspondence between *anti*-BVDV activity of the 47 arylazoenamines and their molecular structures. The 2D-QSAR models generated five population equations with various lengths of analyzed molecular properties, molecular descriptors or their combinations on the conformation of dependent molecules using GFA linear regression and linear splines for effective search of the best multiparameter correlations in large spaces. The 2D-QSAR models are shown in Table 4 with varying number of terms along with the statistical parameters. The models were selected at 95% confidence level. The 2D-QSAR models built on the training set were judged with their suitability based on (i) Friedman LOF score adapted to measure the fitness of a GFA model during the evolution process and (ii) the predicted capacity of the models using the predictive r^2 equivalent to r^2_{CV} from a leave-1-out (LOO) internal cross validation and r^2_{pred} external validation. The score of a 2D-QSAR model is likely to be optimistic if used to assess the predictive performance of a model since the r^2_{CV} value in the

Table 4. Statistical evaluations of 2D-QSAR models for *anti*-BVDV RdRp with varying number of descriptors in the training set (n=47)

Eq. No	Descriptor equation	LOF	r^2	r^2_{adj}	r^2_{CV}	RMS	S.O.R. p-value
1	pEC_{50} = 2.2024 + 3.8867(BIC) – 0.089619(VSA_AtomicAreas[1]) + 261.56 (Jurs_FNAS_3 + 0.012306) + 1.4765(396.12-Jurs_TASA)	0.196	0.643	0.609	0.565	0.320	$5.869e^{-9}$
2	pEC_{50} = 1.8106 + 0.19323(Dipole_Y) + 1.0934(IC) – 0.1165(VSA_AtomicAreas[15]) + 199.69(Jurs_FNSA_3+0.01209) + 1.0625(396.63-Jurs_TASA)	0.194	0.707	0.671	0.640	0.293	$5.506e^{-10}$
3	pEC_{50} = 4.9507 + 0.18411(Dipole_Y) – 0.15005(VSA_Atomic Areas[15]) + 0.015678(VSA_PartialCharge[7]) + 2.5157(IC-3.6072) + 1.005(396.73-Jurs_TASA)	0.196	0.703	0.667	0.626	0.295	$7.097e^{-10}$
4	pEC_{50} = 5.5774 + 0.25189(Dipole_Y) – 0.18114(VSA_AtomicAreas[15]) – 0.024799 (VSA_AtomicArea[18]) + 0.013908 (VSA_PartialCharge[7]) + 6.2608(Jurs_RPCG-0.22359) + 2.198(395.88-Jurs_TASA)	0.199	0.755	0.718	0.687	0.271	$8.233e^{-11}$
5	pEC_{50} = 6.9303 + 0.25999(Dipole_Y) – 4.5083(Jurs_RNCG) – 0.10485 (VSA_AtomicAreas[13]) – 0.17367(VSA_AtomicAreas[15] + 0.020278 (VSA_PartialCharge[7]) + 2.6747(0.6457-CIC) + 1.1944(396.63-Jurs_TASA)	0.194	0.809	0.775	0.723	0.243	$3.553e^{-12}$

Friedman LOF is measured as LOF=SSE/$\{1-(c+dp)/m\}^2$ where SSE is the sum of squares of errors, c is the number of basis functions other than the constant term, d is a smoothing parameter, p is the total number of features and M is the number of samples in the training set. r^2 is the square of correlation coefficient as SSR/SST where SST is the total sum of squares and SSR is deviation between SST and SSE. r^2_{adj} is square of adjusted correlation coefficient as 1–[{SSE/(p–1)}/{SST/(n–p)}] where n is the number of data points and this penalize equations containing too many terms to justify their quality of fit. r^2_{CV} is leave-1-out (LOO) internal variance. RMS is squared root mean residual error. S.O.R. p-value is the p-value for significance or regression.

training set and the predictive performance (r^2_{pred}) are best assessed with a test set separated from the training data. Their predictive power was calculated by r^2_{CV} = 1-PRESS$_{(training)}$/SST$_{(training)}$, where PRESS$_{(training)}$ is the predicted sum of squares of a model, and SST$_{(training)}$ is the mean-corrected sum of squares of responses over the training set on LOO internal cross-validation. The r^2_{pred} value is (SD-PRESS$_{test}$)/SD in the test set, in which SD[41] is the sum of squares of deviations between the biological activity of each molecule and the mean activity of the training set, and PRESS$_{test}$ is the sum of squares of deviations between the predicted and actual activity values for every molecule in the test set. Occasionally, the value of both r^2_{CV} and r^2_{pred} should be more than 0.5, which is considered an acceptable model.

As shown Table 4, the LOF measure cannot be reduced by adding more terms to the regression model; unlike the least squares measure, in which new terms may reduce the sum of squares of errors (SSE), increasing both the number of terms and the smoothing parameter tends to increase the LOF score. In other words, the LOF measure resists over-fitting for the addition of basic functions to the equation in such a way to penalizing through control over the smoothing parameter. As a result, the LOF score can detect an over-fitting problem better than the SSE measure, so it was selected as a score function during GFA to assess the goodness of fit of each progeny equation. This indicates that Lower values of Friedman's LOF for each generated equation are less likely to the GFA model will fit the data. Moreover, if r^2_{CV} is much less than r^2, the model equation will be probably over-fit the data. The measures are used to determine whether the regression is statistically significant or not. The statistically significant 2D-QSAR model in the training set for *anti*-BVDV NS5B RdRp is shown as follows:

$$\begin{aligned} pEC_{50} = {} & 1.8106 + 0.19323(\text{Dipole Y}) + 1.0934(\text{IC}) \\ & - 0.1165(\text{VSA AtomicAreas[15]}) \\ & + 0.9969(\text{Jurs FNSA3} + 0.01209) \\ & + 1.0625(396.63\text{-Jurs TASA}) \end{aligned} \tag{1}$$

$n_{training}$ = 47, LOF = 0.194, r^2 = 0.707, r^2_{adj} = 0.671, r^2_{CV} = 0.640, RMS = 0.293; n_{test} = 13, r^2_{pred} = 0.655, PRESS$_{test}$ = 0.542, SDEP$_{test}$ = 0.204, $S_{PRESS,test}$ = 0.301; n_{pred} = 34, r^2_{valid} = 0.666, PRESS$_{pred}$ = 11.538, SDEP$_{pred}$ = 0.583, $S_{PRESS,pred}$ = 0.654.

Other statistical measures were also used in the test and prediction set. The prediction error of the measure is the standard deviation of the error of prediction (SDEP) as $(PRESS/n)^{1/2}$, and S_{PRESS} is standard deviation based on PRESS(S_{PRESS}) as $[PRESS/(n-p-1)]^{1/2}$, in which n is the number of compounds and p is the number of predictor variables.[42,43] The above model could explain 67.1% of the adjusted coefficient of variation. The LOO in internal cross validation found predicted variance to be 64.0%. The predicted power on external validation for this model was observed to be 65.5% and 66.6% in the test and prediction set, respectively. This is based on the model, which can be inferred to be viable because the differences between the values of r^2_{CV}, r^2_{pred} and r^2_{valid} were much less than 0.3. In the criteria mentioned above, Eq. (1) was selected for the final 2D-QSAR model and it found satisfactory results for predicting the activities of the test set (Table 1) further validating the prediction set (Table 2).

The most accurate mapping descriptors formatted to Eq. (1) that were deemed important for explaining BVDV antiviral activity were selected in the mode share, a common property of the molecules. The inhibition activity of BVDV RdRp can be described as the molecular descriptors like Dipole Y (the Y component of the dipole moment), IC (the overall information content as the graph-theoretical molecular descriptor defined on basis of the Shannon information theory[50]), VSA AtomicAreas[15] (the van der waals surface area on atomic number 15), Jurs FNSA3 (the atomic charge weighted fractional partial negative surface areas), and Jurs TASA (total hydrophobic surface area). This analysis allows for identifying molecular properties positively and negatively related to the activity in question. Among them, the dipole Y component and IC descriptor are directly related in that the positive coefficients in Eq. (1) are conducive for *anti*-BVDV activity. The related importance of the descriptors appearing in the GFA equation based on their regression coefficients is in order to highly correlated Jurs TASA, VSA AtomicAreas [15], Dipole Y, IC, and Jurs FNSA3 with changes in the antiviral activity. For Eq. (1), the polarity and polarizability of a molecule are most closely correlated with the overall patterns of *anti*-BVDV activity. As a measure of the polarity of a compound, the Y component of the molecular dipole moment reflects only a partial of polarities of a molecule along the Y axis, while local polarity is represented by Jurs FNSA3 calculated for a fragment of a molecule. The polarizability has also been related to hydrophobicity as Jurs TASA and thus to the *anti*-BVDV activity. The dipole moment as a descriptor related to 3D charge distribution captures a special electronic polarization related to the strength and spatial orientation of a molecule.[44,45] This dipole property may be correlated to long range electrostatic interaction as the driving force of recognition and subsequent binding to BVDV RdRp polymerase. Charged partial surface area (CPSA) descriptors, which map atomic partial charges on the solvent-accessible surface area (SASA) of individual atoms of the whole molecule or a fragment thereof were invented by Jurs *et al.*[47,48] and can be encoded as the features responsible for polar interactions between molecules. Jurs FNSA3 has been defined as a description of the fractional atomic charged weighted partial negative surface area and is a measure of the polarity of a molecule, and the atomic total hydrophobic surface area (Jurs TASA) descriptor is the sum of atoms with an absolute value of partial charges less than 0.2.

The prediction of this series of arylazoenamine revealed some interesting trends. The presence of a naphthyl group at the Aryl substituent generally confers higher *anti*-BVDV activity than the phenyl group for the arylazohexahydroquinolizines **A** and arylazohexahydroindolizines **B** and arylazotetrahydropyridines **C**, as seen by comparing compounds

22 and **39** with their corresponding compounds **15** and **21**. The presence of aromatic moieties of some size in a flat plane at the position seems to be important for antiviral activity, which has been an expressive descriptor of Jurs TASA values and which increase in good correlation with the actual activity. In another location (R-group), increasing

Table 5. The values of selected descriptors used in 2D-QSAR models

Cpd. no.	Dipole Y	IC	VSA Atomic area[15]	Jurs FNSA3	Jurs TASA	Cpd. no.	Dipole Y	IC	VSA Atomic area[15]	Jurs FNSA3	Jurs TASA
Training set						Test set					
1	-1.188	3.156	9.564	-0.026	454.866	**8**	-5.091	3.690	7.552	-0.078	444.005
2	-1.198	3.366	9.564	-0.014	511.096	**9**	-3.277	3.404	9.564	-0.078	439.666
3	-1.105	3.471	5.672	-0.019	519.513	**11**	-1.571	3.277	6.427	-0.068	603.828
4	-0.768	3.156	9.564	-0.020	519.953	**14**	-2.910	3.534	6.427	-0.039	575.589
5	-1.390	3.471	4.365	-0.016	535.429	**18**	-2.242	3.722	9.564	-0.081	424.828
6	-0.607	3.156	9.564	-0.016	526.546	**27**	-1.226	3.750	9.564	-0.019	488.208
7	-1.950	3.572	6.427	-0.042	549.924	**28**	-0.607	3.375	9.564	-0.019	477.500
10	-1.830	3.384	5.672	-0.025	545.748	**40**	-2.029	3.654	9.564	-0.044	606.247
12	-2.363	3.534	6.427	-0.038	565.504	**44**	-0.887	3.875	7.657	-0.020	493.108
13	-3.123	3.534	6.427	-0.040	570.942	**45**	-0.569	3.500	7.657	-0.020	482.190
15	0.125	3.114	6.427	-0.008	579.185	**51**	0.608	3.735	7.657	-0.024	507.195
16	-1.920	3.175	4.066	-0.007	594.350	**54**	-2.221	3.735	7.657	-0.021	514.262
17	-0.729	3.170	5.672	-0.020	499.279	**60**	-2.111	3.734	7.657	-0.040	608.262
19	-1.151	3.406	5.672	-0.027	522.332	Prediction set					
20	-1.943	3.550	5.672	-0.047	552.206	**61**	2.570	3.503	8.291	-0.045	409.709
21	0.017	3.154	6.427	-0.008	549.797	**62**	1.336	3.821	5.700	-0.056	442.856
22	-0.061	3.323	9.564	-0.014	449.644	**63**	1.345	3.795	5.700	-0.053	488.243
23	-1.251	3.375	9.564	-0.030	412.793	**64**	1.315	3.844	5.700	-0.045	526.558
24	-1.058	3.625	9.564	-0.018	476.184	**65**	1.457	3.752	5.700	-0.047	528.396
25	-1.468	3.750	9.564	-0.018	479.172	**66**	3.757	3.668	5.700	-0.038	529.355
26	-0.784	3.375	9.564	-0.023	472.424	**67**	1.845	3.825	5.700	-0.042	548.088
29	-1.510	3.787	9.564	-0.048	500.934	**68**	1.590	3.778	5.700	-0.040	598.351
30	-3.106	3.948	9.564	-0.087	395.061	**69**	2.745	3.924	6.716	-0.032	542.680
31	-3.414	3.614	9.564	-0.087	395.677	**70**	1.498	3.637	5.700	-0.044	489.026
32	0.020	3.375	9.564	-0.012	477.720	**71**	1.341	3.065	7.867	-0.066	461.579
33	-1.376	3.382	9.564	-0.075	556.494	**72**	-1.364	3.296	9.564	-0.033	484.500
34	-2.554	3.722	9.564	-0.058	472.183	**73**	-0.383	3.220	7.552	-0.126	402.379
35	-2.190	3.722	9.564	-0.051	522.481	**74**	1.026	3.301	9.564	-0.043	415.36
36	-1.907	3.722	9.564	-0.047	522.278	**75**	1.167	3.422	9.564	-0.037	450.993
37	-3.680	3.932	9.564	-0.086	420.736	**76**	1.104	3.668	9.564	-0.035	519.632
38	-1.388	3.061	8.162	-0.021	484.681	**77**	0.938	3.654	9.564	-0.033	548.103
39	0.048	3.432	9.564	-0.009	524.574	**78**	0.177	2.766	7.552	-0.068	434.463
41	-0.069	3.457	7.657	-0.015	452.425	**79**	2.419	3.171	7.552	-0.072	464.747
42	-1.137	3.500	7.657	-0.032	417.370	**80**	2.357	3.669	7.552	-0.059	587.698
43	-0.716	3.500	7.657	-0.024	481.144	**81**	2.363	3.669	7.552	-0.059	578.994
46	-0.470	3.892	7.657	-0.048	505.959	**82**	2.196	3.664	7.552	-0.056	605.551
47	-0.888	4.059	7.657	-0.088	400.207	**83**	2.915	3.423	9.564	-0.054	454.041
48	-3.087	3.725	7.657	-0.087	400.057	**84**	2.863	3.869	9.564	-0.044	575.890
49	-0.013	3.500	7.657	-0.013	477.906	**85**	2.569	3.853	9.564	-0.042	600.497
50	0.279	3.614	7.657	-0.018	525.126	**86**	1.410	3.853	9.564	-0.056	557.543
52	-1.504	3.469	7.657	-0.075	560.117	**87**	1.921	3.893	9.564	-0.044	598.246
53	-2.553	3.735	7.656	-0.030	458.195	**88**	-2.452	3.096	7.552	-0.062	460.469
55	-1.306	3.161	7.657	-0.021	491.462	**89**	0.042	3.096	7.552	-0.053	625.966
56	-0.012	3.346	7.657	-0.014	536.534	**90**	-0.478	3.096	7.552	-0.055	609.545
57	-1.732	3.304	7.657	-0.016	568.783	**91**	-0.719	3.895	7.552	-0.053	625.479
58	0.191	3.404	7.657	-0.012	567.639	**92**	-1.204	3.895	7.552	-0.063	628.069
59	-2.671	3.607	7.657	-0.020	596.417	**93**	-0.952	3.935	7.552	-0.055	636.375
						94	-2.596	3.600	7.552	-0.047	596.644

the Jurs TASA values reduces the inhibitor effect of molecules, and the R group's size in arylazomethylenepyridines **D** (compounds **41** and **56**) are smaller and have more inhibitory activity than the corresponding aromatic ring congeners. Also, an NO_2 functional group bound on an Aryl substituent conducts electron redistribution with conjugated π-bond and leads to varied growth in its π-electron density and electron donating properties. The stronger H-bonding acceptor ability for arylazotetrahydropyridines C compounds **30**, **31**, **47**, and **48** results in analogues with increased inhibitory activity. Introduction of an NO_2 group bound phenyl ring for the arylazohexahydroquinolizines A and arylazohexahydroindolizines B resulted in reduced activity (8, 9, and 18), which was predicted by the calculated dipole moment Y component (Dipole Y), Jurs FNSA3, and Info Content (IC)[46] descriptors. The reduced activity is particularly expressed by the IC descriptor because of the constitution and topology of the molecules. The induction effect at the position was considered with the electrostatic interaction between the bond dipoles and the partial charges of the atom, so the magnitude of this polarization depends on the molecular structure according to the size, degree of branching and overall shape. The IC descriptor uses graph-theory concepts to represent different molecular structures and is a very good description of the induction effect, the van der Waals effect, and hydrogen bonding using a topological approach for special molecular connectivity terms.[49] The induction effect may depend on the solvent in which the process or property of compounds would be indirectly founded, which complicates these descriptors. The van der Waals surface area (VSA) for each atom (VSA AtomicAreas) in a molecule is related to the intra-molecular dispersion effect. In Eq. (1), the VSA on atomic number 15 (VSA AtomicAreas [15]) indirectly reflects the effects of a substituent bound to the common subgroup at a particular location; see, for example, the dispersion effect of the *meta*-substituted Aryl group in arylazohexahydroquinolizines A compounds **3**, **5**, **7**, and **8**. Comparing these compounds, as the size of the substituent's bound atom number 15, VSA decreases (VSA AtomicAreas [15]). This causes substantial changes in the polarizability and dose dispersion effect of the molecule and thus affects inhibitory activity. Descriptor values appearing in the 2D-QSAR models of the training and test set molecules are shown Table 5. In the three isomeric compounds **24**, **25**, and **26**, because of the complexity of SAR, the inhibitory activity was not explained by the effect that depended on the relative position of the substituents in the best 2D-QSAR model. However, the varied activities for these kinds of compounds were very well-explained in the MFA analysis in which the molecular field is characterized by respective property values in the pre-determined grid points of three dimensional space, each of which is responsible for individual intra- or intermolecular interactions.

Molecular Field Analysis. The MFA was developed from 84 compounds (60 arylazoenamines in the training and test set and 24 benzimidazoles in the prediction set) using pre-aligned structures at 1.5 Å grid spacing. In the process of developing quantitative relationships between *anti*-BVDV activity and continuous distribution of energy fields such as electrostatic and hydrophobic effects, the results depended strongly on the molecular alignment. For molecular alignment, the spatial group of active fragments was applied to molecules with diverse structures, which unfortunately resulted in much more ambiguity for 10 acridines. This type of alignment used in the study is adequate for a data set that involves structurally closely related molecules. However, the 10 acridines, which are related to the presence and nature of the tricyclic system and the amino group in the prediction set, do not have a structural fragment or shape similar to the other molecules in the current study. Thus, the higher number of independent variables may create ambiguities in the extraction of the chemical information relevant to the MFA. All the significant structural diversity is necessary in such regions to compare possible active regions in the molecules to determine the inhibitory activity of the molecules. In an attempt to further prove the impact of basic aromatic analogues on *anti*-BVDV activity, four structurally different 2-phenylbenzylimidazole derivatives (F-I subgroups in Table 2) were validated in the MFA together with the original set of arylazoenamines. Therefore, the MFA model would be much more significant.

The QSAR equation of the MFA model involved eight field descriptors, in which the steric (CH_3) and electrostatic (H^+) field descriptors essentially determined the changes in the inhibitory activity, and the constant consisted of nine total parameters. The equation generated by the MFA model using G/PLS regression method is given by:

$$pEC_{50} = 4.697 + 1.071(H^+/661, \mathrm{Ele\_9\_11\_7}) + 0.3324(H^+/165, \mathrm{Ele\_12\_7\_3}) + 1.006(H^+/244, \mathrm{Ele\_14\_9\_4}) - 0.3813(H^+/245, \mathrm{Ele\_14\_9\_5}) + 0.9012(CH_3/262, \mathrm{VdW\_3\_7\_2}) - 0.6232(CH_3/317, \mathrm{VdW\_4\_8\_7}) - 0.5888(CH_3/355, \mathrm{VdW\_5\_7\_9}) - 0.5342(CH_3/401, \mathrm{VdW\_6\_7\_12}) \quad (2)$$

$n_{training} = 47$, $r = 0.919$, $r^2 = 0.844$, $r^2_{adj} = 0.830$, $r^2_{CV} = 0.722$, RMS = 0.189; $n_{test} = 13$, $r^2_{pred} = 0.348$, $PRESS_{test} = 1.005$, $SDEP_{test} = 0.278$, $S_{PRESS,test} = 0.579$; $n_{pred} = 24$, $r^2_{test} = 0.551$, $PRESS_{pred} = 7.625$, $SDEP_{pred} = 0.564$, $S_{PRESS,pred} = 0.738$

The following MFA model could explain 83.0% of variance and predict 72.2% of LOO internal variance. The model showed good internal validation ($r^2_{CV} > 0.5$), but the predictive power of the MFA model for external validation on the test compound and prediction set was not better than the 2D-QSAR model. Tables 1 and 2 show the predicted inhibitory activity values and validated values obtained from the MFA model for the training and test set and prediction set molecules, respectively. The graph of actual versus predicted pEC_{50} values of *anti*-BVDV activity of the total data set for both the MFA model and the 2D-QSAR model is also shown in Figure 3. In comparing and analyzing the predictive power the 2D-QSAR model is better than the MFA model considering external validation (prediction set) whereas the

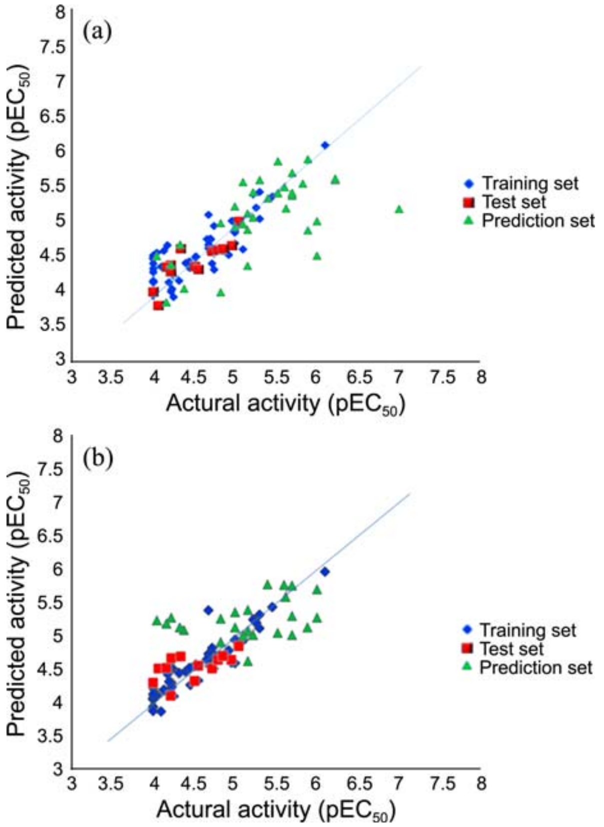


Figure 3. Graphs of actual versus predicted *anti*-BVDV activity for all three data sets for training, internal test, and external prediction set molecules based on both (a) 2D-QSAR model and (b) MFA model.

letter is better in internal validation (test set) than the former. Some potential limitations of the MFA model would be arising from the structural diversity in the validation set and then pre-aligned molecules are less likely to a good alignment for the analysis of molecular fields. Furthermore, the MFA model is the number of field descriptors dependent tends to have higher changes in the inhibition activity than 2D-QSAR model for an identical molecule validated. In the MFA model, the descriptors H^+/x and Ele_a_b_c represent the electrostatic field created by a molecule at the rectangular points a, b, c in its number x respectively. The number associated with the energy grids represents the location in the 3D-grid around the molecule to specify their presence at special positions, but not every grid point can be so expressed. The 3D-isosurface as a shape field for one set of energy grids could give insight into what modifications to the molecules would enhance inhibitory activity. A positive coefficient on the electrostatic descriptor indicates a region favorable for an electropositive group (electron donating group), while a negative coefficient indicates an electronegative group (electron withdrawing group) is required at the position. For instance, the term of 1.071(H^+/661, Ele_9_11_7) has a positive coefficient in Eq. (2), which means at this position an electron donating group (NH_2 or CH_3) will increase the activity. The term of 0.9012(CH_3/262, VdW_3_7_2) has a positive coefficient, which means that at this position a large group

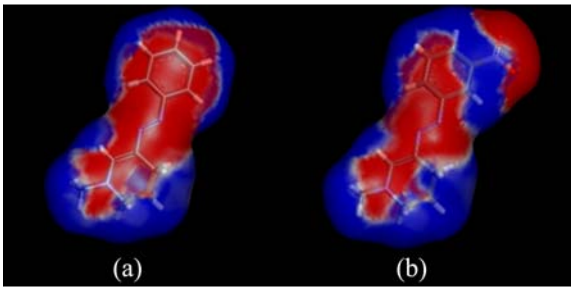


Figure 4. The 3D view of MFA model coefficients of electrostatic interactions in the molecules of the least active compound **22** and most active compound **30** is shown. The red area shows the negative coefficient and the blue area shows the positive coefficient for (a) compound **22** (4.000 in pEC_{50}) and (b) compound **30** (6.097 in pEC_{50}).

will increase activity; on the other hand, terms with negative coefficients mean at this position, a small group will increase the activity. The blue isosurface represents a contour for those points that correspond to a given positive contribution of the electrostatic descriptors (H^+ probe); the red surface, on the other hand, represents those contour points corresponding to a given negative contribution. Consequently, the terms H^+/661, H^+/165, and H^+/244 are distinguished in the variable of the blue isosurface from H^+/245 of red isosurface with respect to their activity.

From Eq. (2) above, larger positive values of pEC_{50} value indicate more active compounds. To increase the activity (a larger positive value of predicted activity), a molecule should have a functional group on the isosurface that shows electrostatic potential with a positive coefficient (the blue area), and on the isosurface that shows negative electrostatic potential with a negative coefficient (the red area). As shown in Figure 4, when we compare the compound **30** with the highest activity and **22** with the compound with lowest activity, we notice that this can be achieved by strategically adding polar groups bound on an Aryl substituent in the electrostatic interactions in order to increase the inhibitory activity. For example, the 3D-isosurface of the coefficients of electrostatic interaction in the most active compound (Fig. 4(b)) show that the red area is visible the middle and the top right, while the blue area is visible in between the red zone and the bottom of the aligned molecule. Electronegative atoms (*e.g.* $-NO_2$) adjacent to the π system in compound **30** strongly deactivate the aromatic ring by decreasing electron density on the ring though a resonance-withdrawing effect, making it less nucleophilic in the blue area (negative charge unfavorable). However, compound **22** has a conjugated π-bond and the highest π-electron density in the blue area compared to compound **30**; this produces lower inhibitory activity. Also, the relative position of halogen substituents bound to the aromatic ring with π-bonds may affect the activity differently because they both induce electronegativity and resonance donation (lone pair donating). The inductive effect lowers the reactivity but the resonance effect controls the stability of the intermediates. With respect to the position of the halogen substituents on the benzyl group, *meta*-, *para*-, and

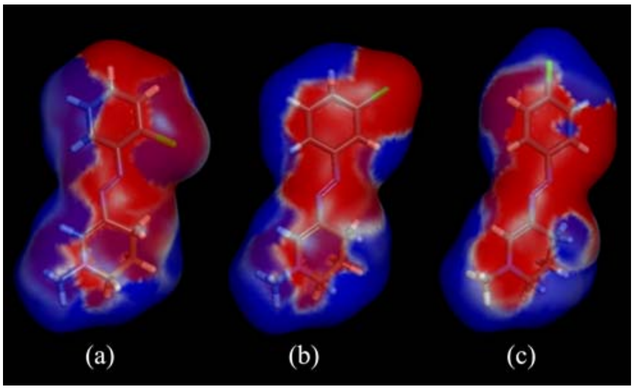


Figure 5. Molecules **24**(a), **25**(b), and **26**(c) within the 3D-isosurface of the electrostatic interaction, where the introduction of chlorine substituent bound on benzyl group has increased negative electrostatic interaction. Also, the inhibitory activity and their relative positions in space with respect to each other plays an important role in the inhibition of BVDV RdRp.

ortho-substitutions are preferred (compound **24**, **25**, and **26**, respectively) for improvements of activity, in particular chlorine and bromine in position **3**, which is as also shown in Figure 5. It is always difficult to determine the effect of electrophilic aromatic substituents on electrostatic interaction and to predict the impact on *anti*-BVDV activity. The van der Waals interactions in the two positions have different effects on the activity: the 1-naphethyl analogue 39 had excellent activity compared the arylazotetrahydropyridines C compounds **22** and **39**, but the arylazomethylenepyridines D compounds **41** and **56**, where the methyl group on the R-group was replaced by a methyl phenyl ring, had reduced activity that agreed very well with the 2D-QSAR model. Figure 6 represents the 3D-isosurface of model coefficients of van der Waals interactions with two colors: green indicates a positive coefficient and yellow indicate a negative coefficient. To increase activity as predicted by Eq. (2), the molecule should have strong van der Waals attraction in the green area and weak van der Waals attraction in the yellow area. In a comparison of compounds **41** and **56**, replacing the methyl group with a benzyl group had a stronger van der Waals attraction in the yellow area, which resulted in reduced activity (Fig. 6(c)). In contrast, exchanging the phenyl with 1-naphethyl in an analogue (39) gave stronger hydrophobic attraction in the green area and enhanced activity (Fig. 6(b)). The MFA model obtained from this study is statically reliable and would be useful for designing potent inhibitors of BVDV NS5B RdRp polymerase, and together with the 2D-QSAR model would provide information to characterize and differentiate their binding sites and describe substitutional requirements for favorable quantitative interaction.

Extended Prediction Set Analysis. To further evaluate the availability of the QSAR models, we used a difficult external prediction set with an initial series of training and test sets to validate the model (Table 2). Modified substituents of the acridine scaffold (compounds **61** to **70**) that may result in improved activity can be explained with particular molecular descriptors that indicate *anti*-BVDV activity. When the chain is branched on hydroxyalkyl derivatives **62** to **64**, a strikingly different effect, depending on their degree of branching, might reveal Jurs TASA values. However, dramatic changes in antiviral activity were not observed. The other effect of introducing a heterocyclic ring on the substituents for compounds **66** to **70** was to reduce the activity as compared with compound **61**. Our 2D-QSAR model predicted experimental data well, except for molecule **61**. Moreover, where the length of the methylene linker increased (compounds **67** and **68**), inhibition activity decreased, total hydrophobic surface increased as Jurs TASA, and dipole moment Y decreased. The activities of molecules **62**, **64**, **65**, **66**, and **67** had a side chain terminated with an oxygenated group and were generally observed with higher values (EC_{50} in the range 0.6-2 μM); the 2D-QSAR model was in good correlation with biological data as a consequence of minor changes in their range of activities.

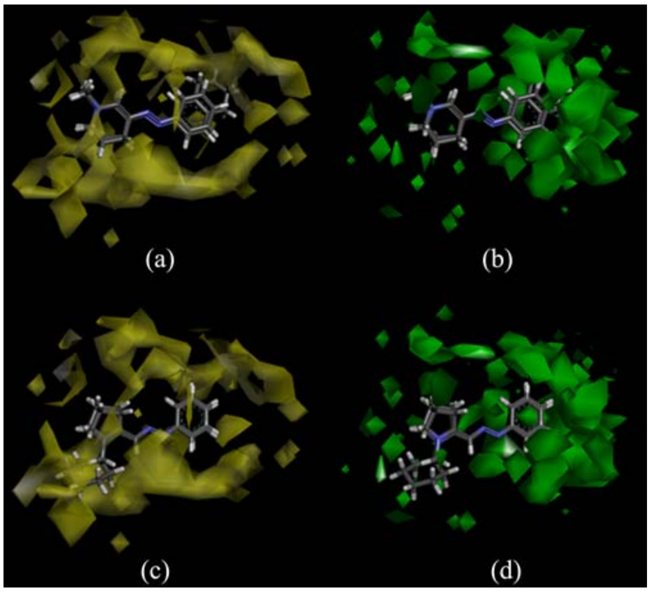


Figure 6. The 3D view of model coefficients of van der Waals interactions for pre-aligned compounds **22** *vs*. **39** (a, b) and **41** *vs*. **56** (c, d) is shown with two colors: green indicates positive coefficients and yellow indicates negative coefficients. Compound **56** whose phenyl ring is closer to large yellow contour led to decreased activity (c), but compound **39** of bulky substituent is green region near that is beneficial to the inhibitory activity (b).

In the case of 2-phenylbenzimidazole derivatives, the main scaffold is a hydrophobic aromatic ring of benzimidazole moiety, and nitrogen atoms of the azole ring are involved in H-bond formation. Modification of the substitution pattern on 2-phenylbenzimidazole derivatives of structure H resulted in introducing substituents into 5-CF_3 with higher activity; additional analogues I in which methyl was replaced in position Z had similar or slightly reduced antiviral activities (compounds **32**, **42**, **90** corresponding to **33**, **43**, **91**, Table 2). Both compounds **42** and **43** were more active than their non-substituted analogues in position 5 (compounds **32** and **33** respectively) because an asymmetric polar group (-CF_3) with a high dipole Y value has more inhibitory activity. Thus compounds **90** and **91** which have hydrophobic substituents

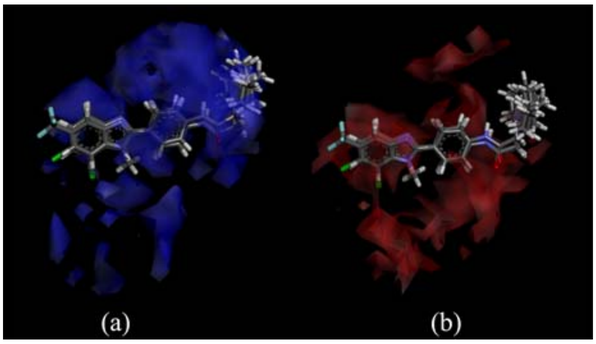


Figure 7. The electrostatic coefficient contour map of the MFA model in the presence of these compounds **76**, **77**, **81**, **82** and **84**, **85** with two colors; (a) positive electrostatic coefficient in blue region and (b) negative electrostatic coefficient in red region.

in position Z, have negative coefficients on their JursTASA terms and hence, according to Eq. (1), less inhibitory activity than the corresponding hydrogen congeners. These molecular characters such as the dipole moment represent global polarities in a molecule that may be of critical value in the *anti*-BVDV activity of this class of compounds, and their proven activity against BVDV can be arranged in the following order of decreasing dipole Y values: **84** > **81** > **76**. Indeed, compounds **76**, **77**, **81**, **82**, **84**, and **85** bear a basic head on the acyl moiety and exhibit potent activity against BVDV. In such cases, the effect of the groups on the compounds was assessed with the MFA model. The MFA electrostatic 3D-isosurface for those compounds is shown in Figure 7. A red contour region was found around the benzimidazole moiety where electronegative groups could have a positive influence on the inhibitory activity. Also, large blue regions near the 2-phenyl analogues indicate that electropositive substituents would enhance the activity. This may perhaps explain the fact that these compounds having both a carbonyl group and a basic amine group on the 2-pheyl substituent were a better match in the blue regions than only acyl moiety, and are predicted to be more active in compounds **74** and **75** compared to compounds **76** and **77**. That is also consistent with the important fact that the existence of a polar group at the R1 position and basic groups at the R2 position on 2-phenylbenzimidazole derivatives of structure **H** increase activity.

Conclusions

In an attempt to understand structural requirements for antiviral activity, we studied the particular characteristics depending on the structure that may affect the inhibitory activity of these *anti*-BVDV compounds. This was corroborated using complementary 2D-QSAR and MFA models to select and characterize chemical features that may be responsible for the activity of the inhibitors. Interestingly, 2D-QSAR models have been shown to be an important concept which acts as a common model of physical and chemical properties that could possibly explain the increased activity in various classes of structurally unrelated molecules. After identification of a major factor of action against BVDV, we also explored whether changes, different from the positions may explain the antiviral activity and would stabilize or disrupt the potential interaction of the inhibitors and BVDV NS5B RdRp polymerase. The variation of activity caused by those changes may be quantified with the MFA model. Overall, there appears to be a relatively good correlation between the actual and predicted activities of the selected compounds with different core structures and basic aromatic class molecules when combined with the methods. That also showed trends similar to SAR analysis and the pharmacophore prediction models as reported by M. Tonelli *et al.* Our MFA models have a possible reliability problem related to the diversity of the data set which might cause prediction error; we believe, however, that we can avoid that limitation by predicting more biological data for model training by combining these models. The QSAR models should be used for screening a large library of small molecules for potential inhibitors of BVDV replication; those models will be useful in attempting to identify features that enhance the inhibitory activity while reducing one or more undesirable effects (*e.g.* cytotoxicity), resulting in increased selectivity.

Acknowledgments. We thank Accelrys Korea for the support of Discovery Studio software.

References

1. Peterhans, E.; Bachofen, C.; Stalder, H.; Schweizer, M. *Veterinary Research* **2010**, *41*, 44.
2. Houé, H. *Biologicals* **2003**, *31*, 137.
3. Baier, A. *Current Enzyme Inhibition* **2009**, *5*, 209.
4. Lindenbach, B. D.; Thiel, H. J.; Rice, C. M. *Fields Virology* **2007**, *1*, 1101.
5. Wegelt, A.; Reimann, I.; Granzow, H.; Beer, M. *Journal of General Virol*ogy **2011**, *92*, 1352.
6. Lindenbach, B. D.; Rice, C. M. In *Fields Virology*, 4th ed.; Knipe, D. M., Howley, P. M., Eds.; Philadelphia: Lippincott Williams and Wilkins: 2001; p 991.
7. Puerstinger, G.; Paeshuyse, J.; De Clercq, E.; Neyts, J. *Bioorganic & Medicinal Chemistry Letters* **2007**, *17*, 390.
8. Puerstinger, G.; Paeshuyse, J.; Heinrich, S.; Mohr, J.; Schraffl, N.; Clercq, E.; Neyts, J. *Bioorganic & Medicinal Chemistry Letters* **2007**, *17*, 5111.
9. Paeshuyse, J.; Leyssen, P.; Mabery, E.; Boddeker, N.; Vrancken, R.; Froeyen, M.; Ansari, I. H.; Dutartre, H.; Rozenski, J.; Gil, L. H. V. G.; Letellier, C.; Lanford, R.; Canard, B.; Koenen, F.; Kerkhofs, P.; Donis, R. O.; Herdewijn, P.; Watson, J.; De Clercq, E.; Puerstinger, G.; Neyts, J. *Journal of Virology* **2006**, *80*, 149.
10. Paeshuyse, J.; Chezal, J. M.; Froeyen, M.; Leyssen, P.; Dutartre, H.; Vrancken, R.; Canard, B.; Letellier, C.; Li, T.; Mittendorfer, H.; Koenen, F.; Kerkhofs, P.; De Clercq, E.; Herdewijn, P.; Puerstinger, G.; Gueiffier, A.; Chavignon, O.; Teulade, J. C.; Neyts, J. *Journal of Virology* **2007**, *81*, 11046.
11. Paeshuyse, J.; Letellier, C.; Froeyen, M.; Dutartre, H.; Vrancken, R.; Canard, B.; De Clercq, E.; Gueiffier, A.; Teulade, J. C.; Herdewijn, P.; Puerstinger, G.; Koenen, F.; Kerkhofs, P.; Baraldi, P. G.; Neyts, J. *Antiviral Research* **2009**, *82*, 141.
12. Tonelli, M.; Boido, V.; La Colla, P.; Loddo, R.; Posocco, P.; Paneni, M. S.; Fermeglia, M.; Pricl, S. *Bioorganic & Medicinal Chemistry* **2010**, *18*, 2304.
13. Tonelli, M.; Simone, M.; Tasso, B.; Novelli, F.; Boido, V.; Sparatore, F.; Paletti, G.; Pricl, S.; Giliberti, G.; Blois, S.; Ibba, C.; Sanna, G.; Loddo, R.; La Colla, P. *Bioorganic & Medicinal Chemistry* **2010**,

18, 2937.
14. Tonelli, M.; Boido, V.; Canu, C.; Sparatore, A.; Sparatore, F.; Paneni, M. S.; Fermeglia, M.; Pricl, S.; La Colla, P.; Casula, L.; Ibba, C.; Collu, D.; Loddo, R. *Bioorganic & Medicinal Chemistry* **2010**, *18*, 8447.
15. Giliberti, G.; Ibba, C.; Marongiu, E.; Loddo, R.; Tonelli, M.; Boido, V.; Laurini, E.; Posocco, P.; Fermeglia, M.; Pricl, S. *Bioorganic & Medicinal Chemistry* **2010**, *18*, 6055.
16. Giampieri, M.; Balbi, A.; Mazzei, M.; La Colla, P.; Ibba, C.; Loddo, R. *Antiviral Research* **2009**, *83*, 179.
17. Sako, K.; Aoyama, H.; Sato, S.; Hashimoto, Y.; Baba, M. *Bioorganic & Medicinal Chemistry* **2008**, *16*, 3780.
18. Castro, E. F.; Fabian, L. E.; Caputto, M. E.; Gagey, D.; Finkielsztein, L. M.; Moltrasio, G. Y.; Moglioni, A. G.; Campos, R. H.; Cavallaro, L. V. *Journal of Virology* **2011**, *85*, 5436.
19. Salim, M. T. A.; Okamoto, M.; Hosoda, S.; Aoyama, H.; Hashimoto, Y.; Bada, M. *Antiviral Chemistry & Chemotherapy* **2010**, *20*, 193.
20. Hosoda, S.; Aoyama, H.; Goto, Y.; Salim, M. T. A.; Okamoto, M.; Bada, M.; Hashimoto, Y. *Bioorganic & Medicinal Chemistry Letters* **2009**, *19*, 3157.
21. Givens, M. D.; Dykstra, C. C.; Brock, K. V.; Stringfellow, D. A.; Kumar, A.; Stephens, C. E.; Goker, H.; Boykin, D. W. *Antimicrobial Agents and Chemotherapy* **2003**, *47*, 2223.
22. Givens, M. D.; Galik, P. K.; Riddell, K. P.; Dykstra, C. C.; Brock, K. V.; Stringfellow, D. A. *Theriogenology* **2005**, *63*, 1984.
23. Ehrlich, P. *Dtsch. Chem. Ges.* **1909**, *42*, 17.
24. Gong, Y.; Trowbridge, R.; Macnaughton, T. B.; Westaway, E. G.; Shannon, A. D.; Gowans, E. J. *Journal of General Virology* **1996**, *77*, 2729.
25. Li, Y.; Mcnally, J. *Vrius Genes* **2001**, *23*, 149.
26. Tonelli, M.; Vettoretti, G.; Tasso, B.; Novelli, F.; Boido, V.; Sparatore, F.; Busonera, B.; Ouhtit, A.; Farci, P.; Blois, S.; Giliberti, G.; La Colla, P. *Antiviral Research* **2011**, *91*, 133.
27. Discovery studio version 3.0 [http://www.acceryls.com/products/discovery-studio]
28. Momany, F. A.; Rone, R. *Journal of Computational Chemistry* **1992**, *13*, 888.
29. Neria, E.; Fischer, S.; Karplus, M. *Journal of Chemical Phys*ics **1996**, *105*, 1902.
30. Smellie, A.; Kahn, S. D.; Teig, S. L. *Journal of Chemical Information Computer Sciences* **1995**, *35*, 285.
31. Smellie, A.; Kahn, S. D.; Teig, S. L. *Journal of Chemical Information Computer Sciences* **1995**, *35*, 295.
32. Chen, K. X.; Xie, H. Y.; Li, Z. G.; Gao, J. R *Bioorganic & Medicinal Chemistry Letters* **2008**, *18*, 5381.
33. Rogers, D.; Hopfinger, A. J. *Journal of Chemical Information Computer Sciences* **1994**, *34*, 854.
34. Fan, Y.; Shi, L. M.; Kohn, K. W.; Pommier, Y.; Weinstein, J. N. *Journal of Medicinal Chemistry* **2001**, *44*, 3254.
35. Deb, K.; Pratab, A.; Agarwal, S.; Meyarivan, T. *IEEE Transactions on Evolutionary Computation* **2002**, *6*, 182.
36. Bender, A.; Mussa, H. Y.; Glen, R. C.; Reilin, S. *Journal of Chemical Information Computer Sciences* **2004**, *44*, 1708.
37. Shi, L. M.; Fan, Y.; Myers, T. G.; O'conner, P. M.; Paull, K. D.; Friend, S. H.; Weinstein, J. N. *Journal of Chemical Information Computer Sciences* **1998**, *38*, 189.
38. Friedman, J. H. *Anal. Stat.* **1991**, *19*, 1.
39. Hirashima, A.; Morimoto, M.; Kuwano, E.; Eto, M. *Bioorganic & Medicinal Chemistry* **2003**, *11*, 3753.
40. Hirashima, A.; Eiraku, T.; Kuwano, E.; Eto, M. *Internet Electronic Journal of Molecular Design* **2003**, *2*, 511.
41. Equbal, T.; Silakari, O.; Ravikumar, M. *European Journal of Medicinal Chemistry* **2008**, *43*, 204.
42. Wold, S., Eriksson, L. H., Eds.; VCH: Weinheim, Germany, 1995; p 312.
43. Debnath, A. K. In *Combinatorial Library Design and Evaluation*; Ghosh, A. K., Viswanadhan, V. N., Eds.; Marcel Dekker Inc.: New York, 2001; p 73.
44. Re, D. G.; Pullman, B.; Yonezawa, T. *Biochemica et Biophysica Acta* **1963**, *75*, 153.
45. Marsili, M.; Gasteiger, J. *Croatica Chemica Acta* **1980**, *53*, 601.
46. Bonchev, D. *Chemometrics Series*; Bawden, D. D., Ed.; 1983; 5, New York: Research Studies Press Ltd.
47. Stanton, D. T.; Jurs, P. C. *Analytical Chemistry* **1990**, *62*, 2323.
48. Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Martin, G. H. *Journal of Chemical Information Computer Sciences* **1992**, *32*, 306.
49. Subhash, C. B.; Gregory, D. G.; Gerald, J. N. *From Chemical Topology to Three-Dimensional Geometry Topics in Applied Chemistry* **2002**, 73.
50. Mekenyan, O.; Bonchev, D.; Trinajstiæ, N. *Int. J. Quantum Chem.* **1989**, *18*, 369.