Combining 3-D Quantitative Structure−Activity Relationship with Ligand Based and Structure Based Alignment Procedures for *in Silico* Screening of New Hepatitis C Virus NS5B Polymerase Inhibitors

Ira Musmuca,[†] Antonia Caroli,[†] Antonello Mai,[†] Neerja Kaushik-Basu,[‡] Payal Arora,[‡] and Rino Ragno*[,†]

Istituto Pasteur-Fondazione Cenci Bolognetti, Dipartimento di Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, P. le A. Moro 5, 00185, Rome, Italy and Department of Biochemistry and Molecular Biology, UMDNJ-New Jersey Medical School, 185 South Orange Avenue, Newark, NJ

The viral NS5B RNA-dependent RNA-polymerase (RdRp) is one of the best-studied and promising targets for the development of novel therapeutics against hepatitis C virus (HCV). Allosteric inhibition of this enzyme has emerged as a viable strategy toward blocking replication of viral RNA in cell based systems. Herein, we describe how the combination of a complete computational procedure together with biological studies led to the identification of novel molecular scaffolds, hitherto untested toward NS5B polymerase. Structure based 3-D quantitative structure−activity relationship (QSAR) models were generated employing NS5B non-nucleoside inhibitors (NNIs), whose bound conformations were readily available from the protein database (PDB). These were grouped into two training sets of structurally diverse NS5B NNIs, based on their binding to the enzyme thumb (15 NNIs) or palm (10 NNIs) domains. Ligand based (LB) and structure based (SB) alignments were rigorously investigated to assess the reliability on the correct molecular alignment for unknown binding mode modeled compounds. Both Surflex and Autodock programs were able to reproduce with minimal errors the experimental binding conformations of 24 experimental NS5B allosteric inhibitors. Eighty-one (thumb) and 223 (palm) modeled compounds taken from literature were LB and SB aligned and used as external validation sets for the development of 3-D QSAR models. Low error of prediction proved the 3-D QSARs to be useful scoring functions for the *in silico* screening procedure. Finally, the virtual screening of the NCI Diversity Set led to the selection for enzymatic assays of 20 top-scoring molecules for each final model. Among the 40 selected molecules, preliminary data yielded four derivatives exhibiting IC_{50} values ranging between 45 and 75 μM. Binding mode analysis of hit compounds within the NS5B polymerase thumb domain showed that one of them, **NSC 123526**, exhibited a docked conformation which was in good agreement with the thumb training set most active compound (**6**).

INTRODUCTION

Hepatitis C virus (HCV) infection represents an important global health-care burden, which is likely to increase over the coming years. There are approximately 3 to 4 million new cases of HCV infection each year, and current estimates suggest that a minimum of 3% of the worldwide population are chronically infected.[1,2] In most infected individuals, this remarkable RNA virus evades the immune system and establishes a chronic infection that can lead to cirrhosis, liver cancer, and death.[3] While advances have been made in treating HCV, the current therapy, a combination of pegylated alpha interferon and ribavirin, is unlikely to succeed in curing all patients.[4] In fact, besides the incomplete efficacy on genotypes 1, the combination therapy has significant adverse effects and is poorly tolerated in particular by individuals affected by other diseases.[5] The quest for new HCV therapies has driven research both in academia and industry, which has led to significant progress in understand-

ing basic virus replication mechanisms and in developing specific candidate antiviral compounds.

HCV belongs to the *Flaviviridae* family and possesses a positive, single-stranded RNA genome that encodes a single polyprotein composed of approximately 3000 amino acids comprising four structural proteins (core, E1, E2, and p7) and six nonstructural proteins (NS2, −3, −4A, −4B, −5A, and −5B).[6,7] These proteins, as well as the viral translation process using the internal ribosomal entry site and a range of host factors, are candidate targets for therapeutic intervention.[8,9] Following the remarkable paradigm established for HIV reverse transcriptase and protease inhibitors, initial efforts to develop HCV-directed antiviral agents have focused on the inhibition of essentially virally encoded enzymes.

Nonstructural protein 5B (NS5B) is the viral RNA-dependent RNA-polymerase (RdRp) at the core of the HCV replicative complex.[10] This viral enzyme has been extensively characterized both at biochemical and structural levels and has emerged as a major target for antiviral intervention. NS5B polymerase synthesizes RNA using an RNA template. This biochemical activity is not present in mammalian cells,

* Corresponding author phone: +396-4991-3937; fax: +396-4991-3627; e-mail: rino.ragno@uniroma1.it.
† Sapienza Università di Roma.
‡ UMDNJ-New Jersey Medical School.

offering the opportunity to identify very selective inhibitors of the viral enzyme. In common with other polymerases, the NS5B enzyme is imaginatively assimilated to that of right handed fingers, thumb, and palm domain structure.[11−13] The palm domain contains the active site of the enzyme, whereas the fingers and the thumb modulate the interaction with the RNA chain.

The rational search of substrate analogues of NS5B polymerase has led to the identification of several nucleoside inhibitors (NIs), whereas high-throughput screening efforts have uncovered a variety of non-nucleoside inhibitors (NNIs).[14] Nucleoside analogues need to be converted by the host-cell machinery to the corresponding nucleotides, which in turn inhibit synthesis of viral RNA as chain terminators.[10,15,16] Conversely, NNIs are almost invariably allosteric inhibitors[17] believed to block the enzyme, preventing a conformational transition needed for initiation of RNA synthesis.

The present paper summarizes efforts for the discovery of hit compounds with molecular scaffolds previously untested as NS5B inhibitors. To select potential new NS5B NNIs, three-dimensional quantitative structure−activity relationship (3-D QSAR), ligand based (LB), and structure based (SB) alignments methods and a LB-SB virtual screening (LB-SB-VS) protocol was set up. Further, the NCI Diversity Set was virtually screened employing the LB-SB-VS strategy, and 40 molecules were selected for enzyme based biological assays. Among the tested molecules, 10% resulted in inhibiting the NS5B RdRp activity at micromolar levels.

LIGAND BASED, STRUCTURE BASED, AND 3-D QSAR PROTOCOL

In any 3-D QSAR model preparation, after a careful selection of the training set, several steps need to be accomplished.[18−21] Among them, alignment rules (ligand based or structure based), statistical analysis (PLS and cross-validation),[22] and external validation (external test sets) are of fundamental importance. Once these steps have been assessed, the model can be applied to predict the activity of untested molecules. The following section describes how we have addressed these steps, in order to develop an adequate procedure to efficiently predict the activity of new NS5B NNIs.

Training Sets Selection. At the beginning of these studies, 27 NS5B/NNIs complexes were available in the Protein Data Bank.[23] Analysis of the downloaded crystal structures revealed that ligands could be classified into three groups on the basis of their different binding sites on the NS5B RdRp (Figure 1). The first class consists of 15 crystallized NNIs that are believed to occupy a common binding site in the thumb subdomain (∼35 Å from the polymerase active site), and thus, the ligands were collected and used as a thumb training set. This set of compounds is structurally related to *N,N*-disubstituted phenylalanine,[24,25] thiophene,[26,27] thiazolone,[28−30] dihydropyrones,[31,32] and acyl pyrrolidine[33,34] molecular scaffolds. PDB codes together with ligand names, structures and biological activities of thumb training set compounds are represented in Table 1. The second class consisting of 10 crystallized NNIs, appear to bind approximately 10 Å from the catalytic site in the palm domain. Members of this class composed of the following NNIs:
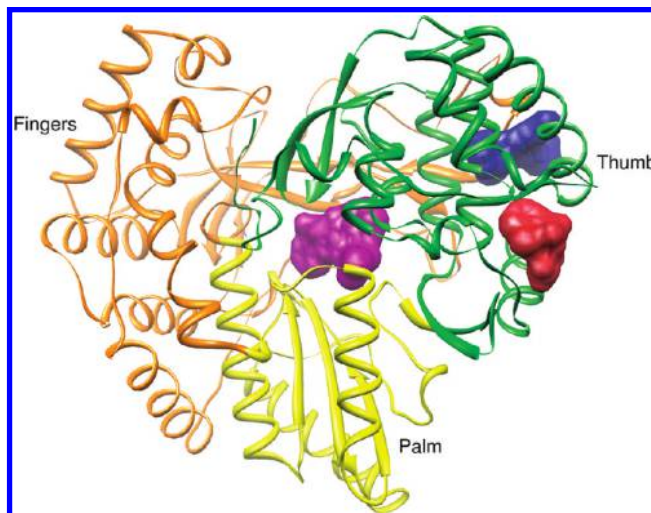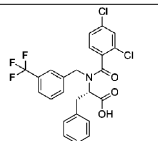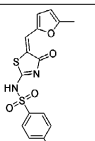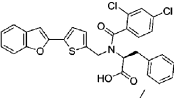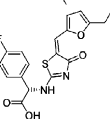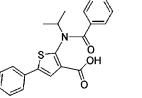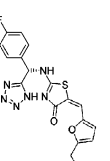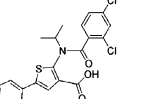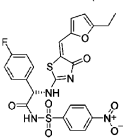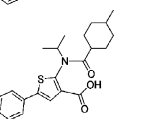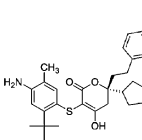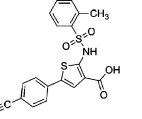

Figure 1. Ribbon representation of the overall structure of HCV NS5B polymerase with domains colored according to thumb (green), palm (yellow), and fingers (orange). Surface representations of three allosteric binding sites are also shown. The dark red colored surface corresponds to the thumb allosteric binding pocket; the dark violet colored surface corresponds to the palm allosteric binding site, and the dark blue colored surface corresponds to the allosteric binding site situated in the thumb domain, near but clearly distinct from the first one (dark red surface).

benzothiadiazine analogues,[35,36] acyl pyrrolidine analogues,[33,34] benzylidene analogues,[37,38] acrylic acid derivatives,[39,40] and proline sulfonamide analogues.[41] These ligands were collected and used as palm training set and their PDB codes, structures, and biological activities are represented in Table 2. It may be noted that compound **15**, an acyl pyrrolidine derivative (PDB entry code: 2JC0),[33,34] has been included in bo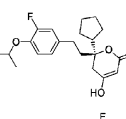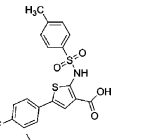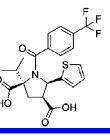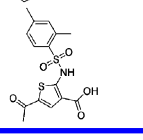th training sets, since the available cocrystallized structure of this compound showed it bound in both allosteric pockets.

A third class of compounds consists of indole-derived NNIs[42,43] that were reported to bind to an allosteric pocket in the thumb domain, close in space but clearly distinct (Figure 1) from the cleft occupied by the above-described thumb training set. In Table 3 are shown PDB codes, ligand names, structures and biological activities of the third class. However, due to the limited number (three) of crystallized ligands in this group, it was not possible to use them as a further training set. Thus, in this study, a total of 24 NNIs/ NS5B complexes were utilized, representing 15 thumb and 10 palm allosteric NNIs, with 2JC0 complex included in both categories. All subsequent 3-D QSAR studies, molecular docking simulations, and ligand based alignments were carried with these training sets.

Polymerase-Inhibitor Complex Structures Preparation. The NS5B RNA polymerase structures cocrystallized with allosteric inhibitors were downloaded from the Brookhaven Protein Data Bank. Of the 24 complexes, 15 were cocrystallized with ligands that bind the thumb subdomain of NS5B. The remaining 10 complexes were cocrystallized with palm binding site NNIs. In order to reduce the geometric inaccuracy of the downloaded structures, the complexes were submitted to a molecular modeling protocol. All complexes were arbitrary superimposed using as template 2D3U, because it was one of the better resolved ($R = 2.0$ Å) and with the smallest number of missing residues. The superimposition of the NS5B complexes was carried out by

Table 1. PDB Code, Ligand Name, and Ligand Structures of the 15 NS5B Polymerase NNIs That Represent the Thumb Training Set

PDB Code	Ligand Name	Entry	Ligand Structure	IC_{50} (μM)	PDB Code	Ligand Name	Entry	Ligand Structure	IC_{50} (μM)
1NHU	(2S)-2-[(2,4-dichloro-benzoyl)-(3-trifluoromethyl-benzyl)-amino]-3-phenyl-propionic acid	1		1.7	2HWH	4-methyl-N-{(5E)-5-[(5-methyl-2-furyl)methylene]-4-oxo-4,5-dihydro-1,3-thiazol-2-yl} benzenesulfonamide	9		2.0
1NHV	(2S)-2-[(5-benzofuran-2-yl-thiophen-2-yl methyl)-(2,4-dichloro-benzoyl)-amino]-3-phenyl-propionic acid	2		8.6	2HWI	(2S)-({(5Z)-5-[(5-ethyl-2-furyl)methylene]-4-oxo-4,5-dihydro-1,3-thiazol-2-yl}amino)(4-fluorophenyl)acetic acid	10		3.0
1YVX	3-[isopropyl(4-methylbenzoyl)amino]-5-phenylthiophene-2-carboxylic acid	3		8.0	2I1R	(5Z)-5-[(5-ethyl-2-furyl)methylene]-2-{[(S)-(4-fluorophenyl)(1H-tetrazol-5-yl)methyl]amino}-1,3-thiazol-4(5H)-one	11		9.7
1YVZ	3-[(2,4-dichlorobenzoyl)(isopropyl)amino]-5-phenylthiophene-2-carboxylic acid	4		4.4	2O5D	(2S)-2-({(5Z)-5-[(5-ethyl-2-furyl)methylene]-4-oxo-4,5-dihydro-1,3-thiazol-2-yl}amino)-2-(4-fluorophenyl)-N-[(4-nitrophenyl)sulfonyl] acetamide	12		7.0
2GIR	3-{isopropyl[(trans-4-methylcyclohexyl)carbonyl]amino}-5-phenylthiophene-2-carboxylic acid	5		1.5	1OS5	3-(4-amino-2-tert-butyl-5-methyl-phenylsulfanyl)-6-cyclopentyl-4-hydroxy-6-[2-(4-hydroxy-phenyl)-ethyl]-5,6-dihydro-pyran-2-one	13		0.93
2D3U	5-(4-cyanophenyl)-3-{[(2-methylphenyl)sulfonyl]amino}thiophene-2-carboxylic acid	6		0.2663	2HAI	(6S)-6-cyclopentyl-6-[2-(3-fluoro-4-isopropoxyphenyl)ethyl]-4-hydroxy-5,6-dihydro-2H-pyran-2-one	14		0.53
2D3Z	5-(4-fluorophenyl)-3-{[(4-methylphenyl)sulfonyl]amino}thiophene-2-carboxylic acid	7		0.2918	2JC0	(2S,4S,5R)-2-isobutyl-5-(2-thienyl)-1-[4-(trifluoromethyl)benzoyl]pyrrolidine-2,4-dicarboxylic acid	15		20.0
2D41	5'-acetyl-4-{[(2,4-dimethylphenyl)sulfonyl]amino}-2,2'-bithiophene-5-carboxylic acid	8		0.3073					

means of the program Chimera[44] using the command-line implementation of MatchMaker.[45] All crystal waters were discarded, and hydrogen atoms were added using the leap module of AMBER suite.[46,47] All ligands were individually inspected, and the correct protonation states at pH 7.4 were considered; i.e., lysines, arginines, aspartates, and glutamates were assumed to be in the ionized form, and the parameters were calculated by means of the Antechamber module of AMBER suite. The complexes were solvated (SOLVATEOCT command) in a box extending 10 Å with water molecules (TIP3 model) and neutralized with Na^+ and Cl^- ions. The solvated complexes were then refined by a single point minimization using the Sander module of AMBER suite. The minimized complexes were realigned with MatchMaker using the same reference complex. The minimized ligands were extracted from the minimized complexes, and maintenance of the coordinates (experimental alignments) was used as obtained for the generation of two structure based statistical models (3-D QSARs).

Molecular Interaction Fields Calculation. The GRID force field[48,49] was used to describe the previously superimposed molecular structures. Interaction energies between the selected probes and each molecule were implemented using a grid spacing of 1 Å. The *xyz* coordinates (in angstroms) of the grid rectangular box used for the computation are $X_{min}/X_{max} = -4.0/52,0$, $Y_{min}/Y_{max} = 6.0/78.0$, $Z_{min}/Z_{max} = 13.0/80.0$. Initially, many probes were tested (C1=,

sp^2 carbon atom bonded to one hydrogen; C3, sp^3 aliphatic carbon atom bonded to three hydrogen atoms; OH, oxygen atom bearing one acidic hydrogen in phenols or carboxy −COOH; O, explicit resonating sp^2 oxygen with 2 lone pairs in carboxy acid anion or pyridine oxide C_5H_5NO; N1, planar nitrogen bonded to one hydrogen atom and two other atoms, no lone pair; N1=, cationic sp^2 nitrogen with one hydrogen, one other single, and one double bond; N3+, cationic sp^3 nitrogen bonded to 3 hydrogens and with one other rotatable bond; DRY, hydrophobic). Preliminary r^2, q^2, and external standard deviation errors of prediction (SDEP) values by PLS, cross-validation, and external validations (see Tables S1, S2, S3, and S4 in the Supporting Information) suggested the use of C1= probe for final 3-D QSAR interpretations of both training sets. Attempts to use a combination of different GRID probes did not lead to any improvement of the statistical models.

Statistical Analyses. The MIFs of the training sets were imported into the GOLPE program.[21] The associated biological activities, expressed as IC_{50} (concentrations reflecting 50% inhibition of HCV NS5B RdRp activity) values (in μM), were converted to molar pIC_{50} (-log IC_{50}) and used as dependent variables in 3-D QSAR analysis. The correlation between biological activities and MIFs was carried out by a statistical PLS analysis[22] as implemented in GOLPE program.[20,21] Thus, a series of 3-D QSAR models for the thumb and the palm training sets were obtained (Tables S1

IN SILICO SCREENING OF NEW HEPATITIS C VIRUS

J. Chem. Inf. Model., Vol. 50, No. 4, 2010 **665**

Table 2. PDB Code, Ligand Name, and Ligand Structures of the 10 NS5B Polymerase NNIs That Represent the Palm Training Set
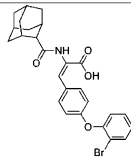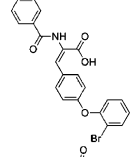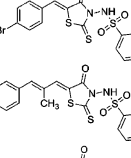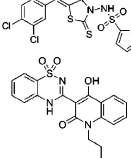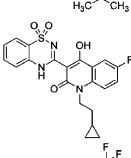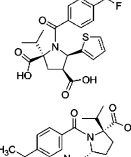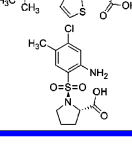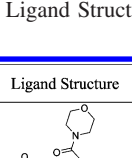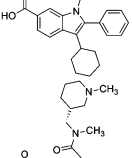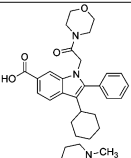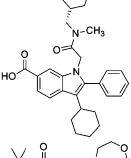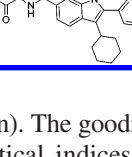
PDB Code	Ligand Name	Entry	Ligand Structure	IC_{50} (μM)
1Z4U	(2Z)-2-[(1-adamantylcarbonyl)amino]-3-[4-(2-bromophenoxy)phenyl]prop-2-enoic acid	16		0.07
1YVF	(2Z)-2-(benzoylamino)-3-[4-(2-bromophenoxy)phenyl]- 2-propenoic acid	17		0.10
2AWZ	(5R)-(4-bromophenylmethyl)-3-(benzenesulfonylamino)- 4-oxo-2-thionothiazolidine	18		0.70
2AX0	(5R)-(2E-methyl-3-phenyl-allyl)-3-(benzene sulfonylamino)-4-oxo-2-thionothiazolidine	19		0.20
2AX1	(5R)-(3,4-dichlorophenylmethyl)-3-(2-thiophenesulfonylamino)-4-oxo-2-thionothiazolidine	20		0.20
2FVC	3-(1,1-dioxido-4H-1,2,4-benzothiadiazin-3-yl)-4-hydroxy-1-(3-methyl--butyl)quinolin-2(1H)-one	21		0.032
2GIQ	1-(2-cyclopropylethyl)-3-(1,1-dioxido-2H -1,2,4-benzothiadiazin-3-yl)-6-fluoro-4-hydroxyquinolin-2(1H)-one	22		0.010
2JC0	(2S,4S,5R)-2-isobutyl-5-(2-thienyl)-1-[4 -(trifluoromethyl)benzoyl]pyrrolidine-2,4-dicarboxylic acid	15		20.0
2JC1	(2S,4S,5R)-1-(4-tert-butylbenzoyl)-2-isobutyl- 5-(1,3-thiazol-2-yl)pyrrolidine-2,4-dicarboxylic acid	23		3.80
2GC8	1-[(2-amino-4-chloro-5-methylphenyl)sulfonyl]- l-proline	24		3.10

Table 3. PDB Code, Ligand Name, and Ligand Structures of the Indole Based NNIs

PDB Code	Ligand Name	Ligand Structure	IC_{50} (μM)
2BRK	3-cyclohexyl-1-(2-morpholin-4-yl-2-oxoethyl)-2-phenyl-1H-indole-6-carboxylic acid		0.026
2BRL	3-cyclohexyl-1-(2-{methyl[(1-methylpiperidin-3-yl)methyl]amino}-2-oxoethyl)-2-phenyl-1H-indole-6-carboxylic acid		0.018
2DXS	N-[(13-cyclohexyl-6,7-dihydroindolo[1,2-[1,4]benzoxazepin-10-yl)carbonyl]-2-methyl-l-alanine		0.046

Table 4. PLS Analysis Results for the Thumb and the Palm Structure Based 3-D QSAR Models[a]

model	N	GRID probe	V	PC	r^2	q^2	$SDEP_{CV\text{-}L5O}$
thumb training set	15	C1=	5133	3	0.99	0.69	0.31
palm training set	10	C1=	3848	3	0.99	0.55	0.66

[a] N, number of compounds in the training set; V, number of GOLPE variables; PC, optimal number of principal components; r^2, conventional square correlation coefficient; q^2, 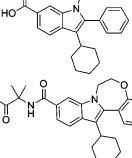cross-validation correlation coefficient; SDEP, cross-validated standard error of prediction using the leave-five-out cross-validation method.

The resulting probe−target interaction energies for each compound were first unfolded to produce one-dimensional vector variables and then assembled in the so-called X matrix. The matrix was pretreated using a cutoff of 5 kcal/mol, in order to produce a more symmetrical distribution of energy values. Variable preselection was operated by zeroing values with absolute values smaller than 0.01 kcal/mol and removing variables with standard deviation below 0.05. In addition, variables taking only two and three distributions were also removed. Attempts to use other settings in the variable preselection did not lead to any improvement of the models.

The smart region definition (SRD) algorithm[50] was applied. This is a GOLPE implementation aimed at selecting cluster of variables, rather than the single variable mainly responsible for activity. The SRD technique seems less prone to change correlation than any single variable selection and improves the interpretability of the models.[51] A number of seeds (1000) were selected using a D-optimal design criterion in the weight space. Structural differences between different molecules in the series will be reflected in groups of variables and, therefore, groups were generated around each seed in the 3-D space. Variables with a critical distance cutoff of 2 Å to the seeds were included in the groups. If two neighboring groups (with a distance smaller than 10 Å) contained the same information, the groups were collapsed. The groups were used in the variable selection procedure replacing the original variables. The effect of the groups on the predictability was evaluated, and groups instead of the individual variables were removed from the data file. Attempts to use other settings in the SRD did not lead to any improvement of the models.

All 3-D QSARs were internally tested by cross-validation using the LSO-5 method with 100 randomizations. The optimum number of principal components used in any 3-D QSAR model was chosen on the basis of the different cross-validated PLS analyses, setting to 5 the maximum number of principal components to be extracted. In the present study, the optimal number of principal components refers to the ones where $SDEP_{CV}$ assumes the minimum value and the cross-validation correlation coefficient q^2 assumes the maximum value. Attempts to increase the maximum principal components during CVs did not lead to any substantially different model.

Models Interpretations. The GRID/GOLPE method used within this study led to the definition of two final SB 3-D QSAR models. The statistical results obtained using the C1= GRID probe are reported in Table 4. The internal predictive power of each model was checked by the leave-some-out
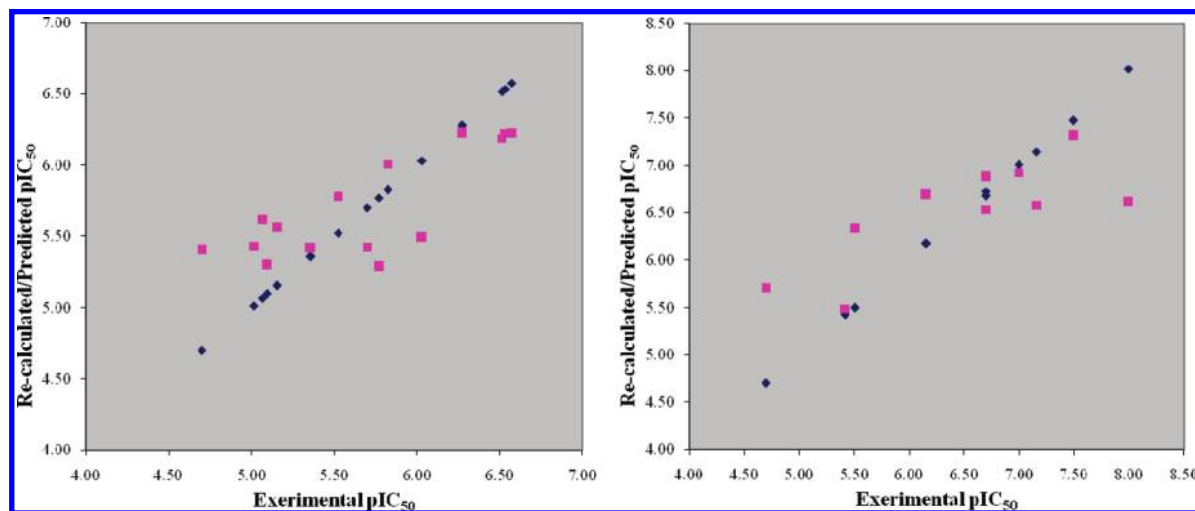
and S2 in the Supporting Information). The goodness of each model was measured by the statistical indices r^2, q^2, and SDEP using leave-some-out with five random groups (LSO-5) and cross-validation method.

Figure 2. Fitting and Cross-Validation plots for the thumb (left) and palm (right) training sets.

with five random groups (LSO-5) cross-validation method (Figure 2). The latter indicated that the optimal principal component was the third one. In fact, the second, fourth, and fifth principal components showed lower q^2 and higher SDEP values (data not shown).

As represented in Table 4, the statistical analysis yielded highly significant statistical coefficients with r^2 values of 0.99 for both models and q^2 of 0.69 for the thumb model and of 0.55 for the palm model, thus providing the goodness of the system.

Moreover, an accurate graphical analysis of GOLPE contour maps led to the identification of the most representative areas selected by the models. In fact, one of the most interesting features of a GRID/GOLPE 3-D QSAR analysis is the possibility of translating back the PLS coefficients assigned to each variable to the 3-D positions they occupy in real space. These values can be contoured at a particular significant level and can be displayed as a grid plot of PLS coefficients. The contour coefficient maps indicate those areas in which the model has found a high correlation between the ligand−probe interaction energy and the biological activity. However, the signs of PLS coefficients can induce errors in this plot. Coefficients could have opposite meaning depending on the positive or negative field values produced by the compound in the same area. In Figure 3 are shown the PLS coefficient plots for the thumb training set analyzed with the C1= GRID probe. One of the most important areas selected by the model is represented by a big red polyhedron located between the *p*-cyanophenyl thiophene and the 2-methylphenyl moieties of the most active compound (**6**), corresponding to the thiazole and the 4-F phenyl group of one of the lowest active compounds (**11**). For chemical interpretation, one must remember that red contours in coefficient maps suggest that repulsive interactions (positive GRID energies) increase affinity, while attractive interactions (negative GRID energies) decrease activity. The sp^2 aromatic carbon atom used for the calculation of the molecular field is a prevalently hydrophobic group. Thus, it might be suggested that these areas correspond to hydrophobic surface of the thumb domain where van der Waals interactions as well as $\pi-\pi$ interactions with binding site residues positively contribute to the biological activity. On the other hand, a blue colored region around the tetrazole portion of compound
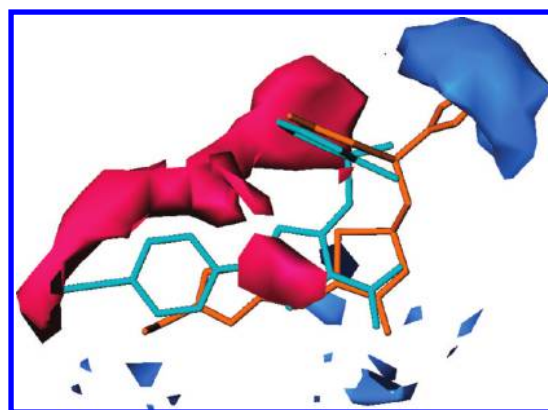
Figure 3. Contour maps of the PLS coefficients derived from C1= GRID probe analysis using the 15 compounds of the thumb training set (contour levels: 0.0008 red, −0.0008 blue). To aid in interpretation, only the highest active (**6** in cyan) and one of the lowest active (**11** in orange) compounds are shown. Hydrogen atoms are omitted for the sake of clarity.

Figure 4. Contour maps of the PLS coefficients derived from OH GRID probe using the 15 compounds of the thumb training set (contour levels: 0.0008 red, −0.0008 blue). To aid in interpretation, only the highest active (**6** in cyan) and one of the lowest active (**11** in orange) compounds are shown. Hydrogen atoms are omitted for the sake of clarity.

11 clearly indicates that the steric hindrance in this area is not suitable for potency.

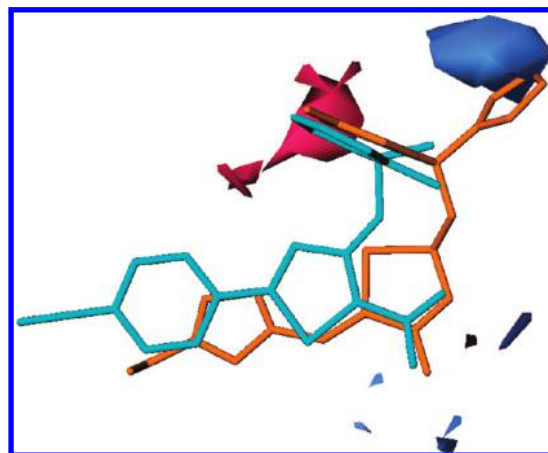In Figure 4, some small blue polyhedra near the carboxylate group of compound **6** are also depicted. To shed light
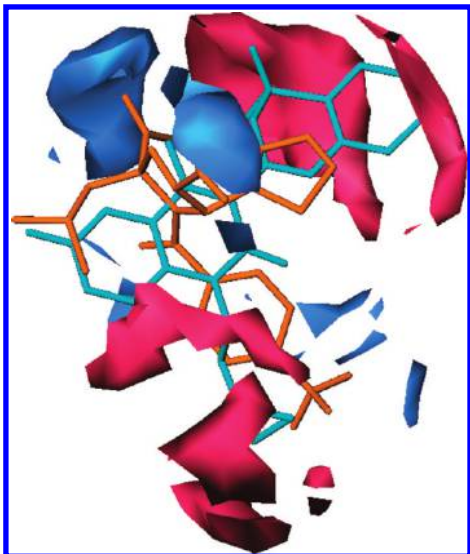
Figure 5. Contour maps of the PLS coefficients derived from C1=
Grid probe using the 10 compounds of the palm training set (contour
levels: 0.0010 red, −0.0010 blue). To aid in interpretation, only
the highest active (**22** in cyan) and the lowest active (**15** in orange)
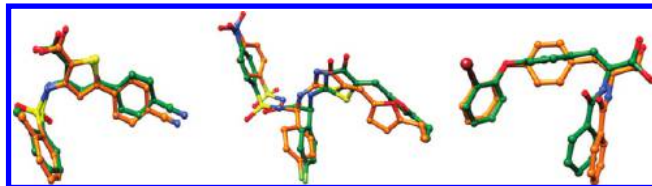are shown. Hydrogen atoms are omitted for the sake of clarity.

Figure 6. Examples of Surflex alignment. Superimposition of the
modeled ligand conformations (carbon atoms in green) to the
experimental ones (carbon atoms in orange) of three compounds
of training sets (from left to right, compounds **6**, **12**, and **16**). Atom
bonds are in ball and stick fashion. Hydrogen atoms are omitted
for the sake of clarity.

on the nature of the interactions made by this substituent,
we decided to analyze it by means of OH GRID probe
derived 3-D QSAR (for statistical data see Tables S1 and
S2 in the Supporting Information). The hydroxyl probe used
for the calculation of the molecular field is a prevalently polar
group with the ability to participate in hydrogen bonds and
electrostatic interactions. Thus, areas containing negative
coefficients (blue polyhedrons) correspond to areas in the
binding pocket where energetically favorable interactions
produce an increase in the activity. These regions correspond
to position of polar groups in the binding site.[50] Conse-
quently, blue polyhedra in this area mainly reflect hydrophilic
features, which may be due to the presence of hydrogen bond
donor residues in the binding pocket.

In Figure 5, the PLS coefficients plots for the palm training
set are shown. The 3-D QSAR model graphical analyses was
carried out by means of C1= probe. The red polyhedron
situated around the benzo portion of the benzothiadiazine
system of compound **22** appears to be relatively important
for the biological activity. In fact, as previously reported,[36]
a key feature observed in all of the benzothiadiazine
structures determined is an apparent edge-to-face π-interac-
tion between Phe193 and the benzene ring. Compound **15**,
an acyl pyrrolidine derivative, did not show any particular
interaction in this area. This may be due to the relevant
differences in the chemical structures of these molecules.
Nevertheless, the red polyhedra located in proximity of the
N-1 (2-cyclopropyl)ethyl moiety (**22**) and the 4-trifluoro-
methyl benzoyl group of compound **15** imply that space
filling in both positions might play a key role in further
potency enhancements.

Alignment Rules. In the present study, ligand based as
well as structure based alignment methodologies were
explored with the aim to determine the alignment of
molecules with unknown binding mode (test set). Several
types of software that perform diverse alignment procedures
are available, and among them, Surflex-Sim[52] and Au-
todock[53] were chosen since they are free for those in
academics and are of good quality as assessed by a

considerable number of references.[54−60] In particular, for the
LB approach, the application of the concept of morphological
similarity, implemented in Surflex-Sim, was used, while for
the SB approach the program Autodock was selected.

In the following paragraphs, different levels of alignment
assessment are reported for either LB or SB methods.
Regarding the LB approach, randomly built conformation
for each ligand was generated and then aligned to the
experimental binding conformation (resurflex). This way, the
LB alignment procedure was first tested for self-alignment
and was due to different topological atom descriptions
between the experimental and modeled structures; the surflex
aligned conformation was retained as a reference conforma-
tion in the subsequent calculations involving the use of
modeled structures. Although arbitrary, this assumption was
supported by the visual inspection of superimposed experi-
mental versus the resurflexed conformations (Figure 6). In
the next step, each randomly built conformation is aligned
to a target list which contains all ligands (in their binding
conformation), except the corresponding experimental one
(cross-surflex).

Regarding the SB approach, a four step procedure was
set up. In the first step, the Autodock program was tested
for the ability to reproduce the bound state of a certain ligand
using the respective receptor and the known binding con-
formation as the starting point (redocking). In the second
step, the same redocking was repeated using as the starting
point a randomly generated conformation (redocking mod-
eled). The third level of docking assessment was achieved
by a cross-docking procedure using the experimental con-
formation docked in all the available receptors while exclud-
ing the native one (cross-docking). In the last step, the cross-
docking procedure was repeated using the modeled ligands
(cross-docking modeled). In either cross-docking procedures,
the output conformations were merged in a single file and
then clustered using the internal clustering procedure of the
Autodock program. The rationale of the docking assessment
resides in the fact that first the docking program was tested
for the ability to correctly reposition the native ligand
conformation within the respective experimental receptor
pocket. Besides, a modeled ligand conformation was used
in order to measure how the starting docking conformation
can influence the docking program in the reproduction of
the bound ligand conformation. Analogously the cross-
docking experiments, using the experimental and modeled
ligand conformations, were carried out to evaluate the
docking program aptitude to predict the correct binding
conformation using an ensemble of receptor conformations
while keeping out the native cocrystallized one. In the cross-
docking assessments, the clustering of poses coming from

docking runs using different receptor conformations can be assumed a sort of full flexible docking simulation in which both the ligand and the receptor are allowed to move. In the case of the ligand, the flexibility is achieved by the Solis−Wets algorithm,[61] while for the receptor the experimental NS5B snapshots give the induced ligand conformational changes. An attempt to use also the Autodock side-chain flexibility feature was not successful (data not shown).

LB Alignment: Surflex Assessment. The ligand based alignment of the molecules was achieved using Surflex-Sim.[52] This method optimizes the pose of a query molecule to an object molecule in order to maximize 3-D similarity. Morphological similarity is defined as a Gaussian function of the differences in molecular surface distances of two molecules at weighted observation points on a uniform grid, thus yielding a value from 0 to 1. The function is dependent on the relative alignment of two molecules.

The first step of the Surflex-Sim assessment was checking the goodness of the alignment when a randomly built ligand conformation was generated and then overlapped to the experimental binding conformation (self-alignment). For this purpose, each of the 24 modeled compounds of the training sets were superimposed to the corresponding minimized ligands as extracted from their respective complex (resurflex). Due to the differences in topological description of the molecules, the goodness of the self-alignment was visually inspected. As examples of alignment goodness, in Figure 6 are depicted the aligned conformations of three different modeled compounds (**6**, **12**, and **16**) superimposed on their respective experimental bound conformations. In the next alignment assessment step, each of the 24 modeled ligand conformations were overlapped to a target list containing all the ligands (in their SB experimental aligned binding conformations), except the corresponding experimental one (cross-surflex). Here, the self-aligned surflex conformation was employed as a reference for measuring the root-mean-square deviation (rmsd) values between the cross-surflex aligned and the resurflex one. As seen in Table S7 (Supporting Information), the rmsd values for all 24 NS5B NNIs were below the cutoff value of 1, thus suggesting good alignment. This validation of Surflex-Sim's ability to accurately align randomly built conformations suggests that it would exhibit a similar accuracy even with molecules with unknown binding conformation and, thus, can be considered as a useful tool for LB alignment of NS5B inhibitors (see external test set below) and to align molecules in a virtual screening protocol (see below).

SB Alignment: General Docking Settings. Autodock 4 was used for all docking calculations. The AutoDockTools (ADT) package was employed to generate the docking input files and analyze the docking results. The 24 ligands extracted from the minimized complexes (see Polymerase-Inhibitor Complex Structures Preparation section) were separated in two groups on the basis of their different binding sites on the NS5B polymerase receptor. The first group contains 15 NNIs that are believed to occupy a common binding site located in the thumb domain. The second group contains a total of 10 NNIs reported as binding an allosteric pocket located near the active site in the palm domain.

Two different grid boxes, one for each binding site, were centered on the average mass center of the ligands. Thus, a grid box of $91 \times 91 \times 86$ points and a grid spacing of 0.375

Å was set in order to accommodate the NNIs that bind the thumb domain of NS5B polymerase. The second grid box (spacing, 0.375 Å) of $62 \times 75 \times 75$ points was implemented in such a way to accommodate the NNIs that bind the palm allosteric site of NS5B. Several atom probes (characterized by the same stereoelectronics properties as the atoms constituting the inhibitors) were moved on the grid nodes, while interaction energy between the probe and the inhibitors were calculated at each node. The grid maps were generated for each atom probe, describing its interactions with the inhibitors. Autogrid 4, as implemented in the Autodock software package, was used to generate grid maps.

The Lamarckian genetic algorithm (LGA)[62] was employed to generate orientations or conformations of the ligands within the binding site. The global optimization started with a population of 150 randomly positioned individuals, a maximum of 2.5×10^6 energy evaluations, and a maximum of 27 000 generations. A total of 100 runs was performed, while all the remaining run parameters were maintained at their default settings. A cluster analysis was carried out using 2 Å as the root-mean-square deviation tolerance. Docking experiments were also tried using a single grid box comprising the two binding sites. Although Autodock performed quite well in selectively positioning the ligands in the right pockets, the results in term of root-mean-square-deviation (rmsd) values were not fully satisfying (data not shown). This is likely due to the fact that with only 100 runs such a large region is not sufficiently explored. Nevertheless, a higher number of conformation sampling, although not assuring a better result, would have been too computationally demanding to be pursued.

Assessment of Docking: Redocking. In order to check for the reliability of the docking protocol, we performed docking studies on 24 NS5B polymerase complexes. The docking results were evaluated through a comparison of the predicted docked positions of the ligand and the experimental ones. As a measure of docking reliability, the rmsd between the positions of heavy atoms of the ligand in the calculated and experimental structures was considered. The choice of the best conformation was based on the assumption that, although for high-throughput screening protocols, only the first ranked docked conformation should be considered (that is, the conformer characterized by the lowest estimated free energy of binding)[63] in other cases, the lowest energy conformer of the most populated cluster should also be taken into account.[64] In Table 5, the rmsd for the best docked, best cluster, and lowest values obtained are reported. Analyses of these results show that, in cases where the best cluster conformer does not coincide with the best docked conformer (1NHU, 1NHV, 1YVZ, 2GIR, 2HWH, 2I1R, 2O5D, 1YVF, 2AWZ, 2AX1, and 2GIQ), a better rmsd value is observed, except for 1NHV, increasing the percentage of conformations found below the threshold value of 2 Å from 41.6% (best docked) to 70.8% (best cluster) to 100% (best fitted cluster). While the best docked is the lowest energy docked conformation of the first Autodock generated cluster, the best cluster is the lowest energy docked conformation of the most populated cluster and best fitted cluster is the lowest energy docked conformation of the cluster showing the lowest rmsd value. In the ideal case, i.e., the error free docking program, the three conformations would have coincided. The fact that Autodock reproduces the correct conformation (best fitted

IN SILICO SCREENING OF NEW HEPATITIS C VIRUS

J. Chem. Inf. Model., Vol. 50, No. 4, 2010 **669**

Table 5. Assessment of the Autodock Program in the Redocking Stage[a]

binding site	PDB	ligand entry	best docked (rmsd)	best cluster		best fitted cluster	
				cluster N°	rmsd	cluster N°	rmsd
thumb	1NHU	**1**	3.17	3	2.17	3	0.89
	1NHV	**2**	4.13	2	4.75	26	1.86
	1OS5	**13**	3.46	1	3.46	9	1.50
	1YVX	**3**	3.81	1	3.81	6	1.58
	1YVZ	**4**	3.74	4	1.92	4	0.78
	2D3U	**6**	0.71	1	0.71	1	0.44
	2D3Z	**7**	0.75	1	0.75	1	0.60
	2D41	**8**	1.43	1	1.43	1	0.58
	2GIR	**5**	5.70	2	1.12	2	0.70
	2HAI	**14**	2.05	1	2.05	1	0.92
	2HWH	**9**	9.86	2	2.13	2	0.79
	2HWI	**10**	0.34	1	0.34	1	0.24
	2I1R	**11**	5.84	2	1.67	2	0.73
	2JC0	**15**	0.85	1	0.85	1	0.68
	2O5D	**12**	5.74	3	2.78	7	1.28
	average rmsd		3.44		2.00		0.90

binding site	PDB	ligand entry	best docked (rmsd)	best cluster		best fitted cluster	
				cluster N°	rmsd	cluster N°	rmsd
palm	1YVF	**17**	3.38	4	1.12	4	0.93
	1Z4U	**16**	0.89	1	0.89	1	0.57
	2AWZ	**18**	3.53	3	1.72	3	1.11
	2AX0	**19**	0.84	1	0.84	1	0.52
	2AX1	**20**	3.24	2	0.99	2	0.61
	2FVC	**21**	1.04	1	1.04	1	0.74
	2GC8	**24**	2.03	1	2.03	2	1.82
	2GIQ	**22**	3.49	5	1.91	5	1.81
	2JC0	**15**	0.65	1	0.65	1	0.42
	2JC1	**23**	0.74	1	0.74	1	0.47
	average rmsd		1.98		1.19		0.9

[a] rmsd values for the first ranked pose (best docked), the lowest energy docked conformation of the most populated cluster (best cluster), and the one closest to the experimentally bound conformation (best fitted cluster).

cluster) by 100% reflects the intrinsic lack of accuracy (at least for NS5B) of the implemented scoring function in selecting the right pose, and this is at the basis for the use of the above 3-D QSAR models (as external scoring function) during the virtual screening reported below. As proof-of-concept, we compared the Autodock predicted pK_i values with the experimental pIC_{50} and only a poor correlation was found (data not shown).

Assessment of Docking: Redocking Modeled. In addition, the reliability of the docking protocol was evaluated by a redocking modeled simulation. Our aim was to check more realistically Autodock's ability to reproduce binding mode conformations of molecules with no experimental data. In this stage, therefore, the same modeled conformations used in the Surflex assessment were used to test the Autodock ability. Similarly, as reported in the cross-surflex assessment section, the docking assessment here was evaluated using as reference conformation the resurflexed one.

The modeled ligands were submitted to a redocking simulation using the same docking settings as for the above-mentioned redocking protocol. In Table S5 of the Supporting Information, the best docked, best cluster, and best fitted cluster rmsd values found are reported. When one compares the values of Table S5 (redocking modeled; Supporting Information) with Table 5 (redocking), at first glance they do not appear to confirm the ability of Autodock to reproduce the correct binding conformation. However, a visual inspection of the redocked modeled conformation revealed that using the resurflexed conformation could lead to a misin-
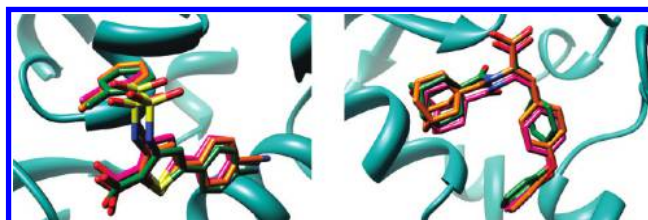
Figure 7. Examples of Redocking Modeled. Superimposition of the Surflex-aligned conformation (carbon atoms in green) and the redocked conformation (carbon atoms in magenta) to the experimental conformation (in orange) of one thumb NNI (**6**, in the left) and one palm NNI (**16**, in the right) within the NS5B (cyan colored ribbons). Atom bonds are in stick fashion. Hydrogen atoms are omitted for the sake of clarity.

terpretation of the Autodock potentials. In fact, upon comparing the three conformations (experimental, resurflexed, and redocked modeled), a better agreement between the experimental and the redocked modeled conformations versus the resurflexed and redocked modeled (the tabled rmsd) one was noted. As an example of these results, Figure 7 depicts the best docked conformations, the reference conformations (Surflex-Sim aligned), and those extracted from the minimized complexes (experimental), for two arbitrarily chosen representatives of thumb (**6**) and palm (**16**) binding NS5B NNIs.

Assessment of Docking: Cross Docking. Along with redocking experiments (self-docking), cross-docking simulations (docking each ligand into the binding pocket of a series of different ligand−receptor complexes) were also per-

Table 6. RMSD Values Obtained by Cross-Docking Simulations

binding site	PDB	ligand entry	best docked (rmsd)	best cluster cluster N°	best cluster rmsd	best fitted cluster cluster N°	best fitted cluster rmsd
thumb	1NHU	**1**	2.25	1	2.25	3	1.87
	1NHV	**2**	4.37	14	3.68	19	2.54
	1OS5	**13**	7.59	5	3.61	32	1.81
	1YVX	**3**	3.51	2	1.77	10	1.58
	1YVZ	**4**	1.73	1	1.73	1	1.67
	2D3U	**6**	5.79	2	5.06	14	1.79
	2D3Z	**7**	5.36	2	3.75	8	1.22
	2D41	**8**	6.49	2	1.51	2	1.51
	2GIR	**5**	1.57	1	1.57	1	1.57
	2HAI	**14**	1.45	1	1.45	1	1.45
	2HWH	**9**	3.60	7	2.9	7	1.71
	2HWI	**10**	3.61	3	2.20	13	1.40
	2I1R	**11**	2.27	2	1.77	3	1.61
	2JC0	**15**	4.26	11	9.7	3	3.38
	2O5D	**12**	3.48	1	3.48	18	1.73
	average rmsd		3.82		3.1		1.79

binding site	PDB	ligand entry	best docked (rmsd)	best cluster cluster N°	best cluster rmsd	best fitted cluster cluster N°	best fitted cluster rmsd
palm	1YVF	**17**	2.56	2	3.43	9	2.52
	1Z4U	**16**	2.26	2	3.87	12	1.78
	2AWZ	**18**	1.64	1	1,.64	1	1.64
	2AX0	**19**	1.85	1	1.85	3	1.58
	2AX1	**20**	1.68	1	1.68	1	1.68
	2FVC	**21**	1.88	1	1.88	3	1.06
	2GC8	**24**	11.13	2	2.08	6	1.84
	2GIQ	**22**	1.84	1	1.84	11	1.64
	2JC0	**15**	1.21	1	1.21	1	1.21
	2JC1	**23**	0.97	1	0.97	1	0.97
	average rmsd		2.7		2.09		1.59

formed. The cross-docking method was based on docking each experimental conformation in all the available receptors, while excluding the native one (cross-docking). Then, all output conformations were merged in a single file and clustered. The aim of such a cross-docking procedure was to evaluate the docking program's ability in proposing correct binding conformations, even if no structural data were available for a given modeled and yet untested ligand. In this docking assessment step, similar to the redocking section, the inclusion of known experimental binding conformation is taken into account.

Additional parameters included evaluation of conformational change during receptor−ligand docking simulations, as well as the consideration of the concept of ligand-induced fit.[64] Autodock 4 was used for all cross-docking calculations employing the same grid boxes parameters (thumb and palm binding sites) and Autogrid maps previously used for redocking calculations. The numbers of runs of each docking experiment was set so that the number of generated conformations was comparable to that in the redocking experiments.

In Table 6, the rmsd values for the best docked, best cluster, and best fitted cluster conformations are reported. As already outlined in the previous paragraph, the best cluster selection versus the best docked criteria gave the best results. Indeed, from analyzing Table 6, one can clearly see that the number of correctly docked conformations for the best cluster selection is higher than those obtained by the best docked criteria.

Assessment of Docking: Cross-Docking Modeled. Cross-docking experiments of modeled training set ligands were

also carried out. Grid box parameters and all docking settings were the same as for the cross-docking simulations. In Table S6 of the Supporting Information, the rmsd for the best docked, best cluster, and lowest rmsd values found are reported. As outlined above, we have evaluated the docking program's ability in proposing correct binding conformations even if no structural data are available for a given modeled and yet untested ligand. Here, the docking program is tested as in a real case situation where a randomly chosen ligand conformation is being docked into a receptor binding pocket whose conformation does not bear any information on the actual ligand inducing fit.

Table S6 of the Supporting Information again confirms that the best cluster conformation has to be considered instead of the best docked. Here again, similarly, though it is less evident compared to the redocking modeled assessment, the rmsd values seem to mislead in the abilities of Autodock to correctly reproduce acceptable docked conformations. Visual inspection (data not shown) revealed that there is a better agreement between the experimental and the cross-docked modeled conformations than those observed between the resurflexed and cross-docked modeled conformation (rmsd values in Tables S5 and S6 in the Supporting Information).

Similarly, as stated above for the cross-surflex experiments, in this last SB alignment assessment, the ability to accurately align randomly built conformations, suggested that Autodock can be used as an useful tool for SB alignment of NS5B inhibitors (see external test set below) and, to some extent, to align molecules in virtual screening protocols (see below).

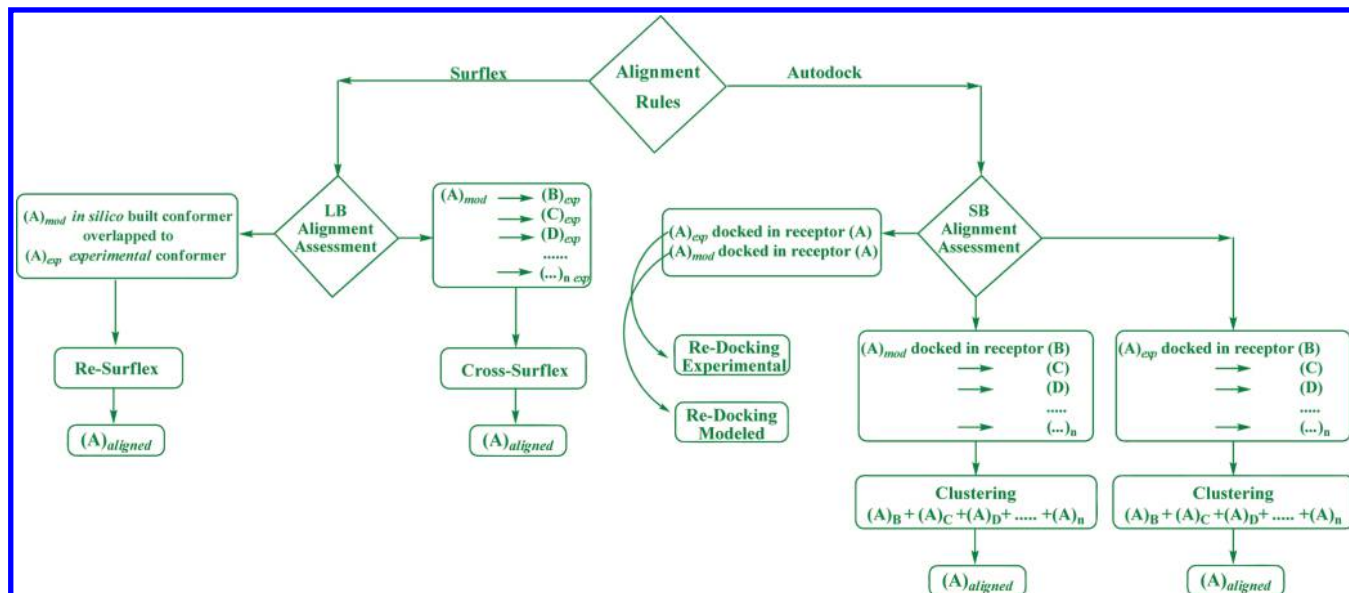Description of the workflow of the alignment process is shown in Figure 8.

Figure 8. Description of the workflow of the alignment protocol. Key steps for the assessment of LB and SB alignments processes.

External Validation of the 3-D QSAR Models. To evaluate the predictive ability of the 3-D QSAR models, two series of test sets were used. Compounds of the first test set include 81 NS5B NNIs selected from literature, that bind in the thumb subdomain of the receptor. Molecular structures of thumb test set are represented in Tables S8−S19 of the Supporting Information. Compounds of the second test set include 223 NS5B ligands reported in literature as palm subdomain allosteric inhibitors. Molecular structures of palm test set are represented in Tables S20−S36 of the Supporting Information. Test set compounds were selected so that they span the whole descriptor (chemical) space occupied by the training set and cover as much as possible the same range of bioactivities.

All 304 test set molecules were generated by means of molecular mechanics of Chemaxon Marvin software (http://

www.chemaxon.com/), and all molecular graphics images were produced using a UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, on a 3 Ghz AMD CPU equipped IBM compatible workstation with the Debian 5.0 version of the Linux operating system. Then, test set compounds were submitted to the molecular alignment using the LB (cross-surflex) and SB (cross-docking) protocols as reported above. In the alignment of the external test sets, all the experimental ligands were used as target list (for each receptor site) in Surflex and all the receptor structures were used in the cross-docking experiments.

Each test set of compounds was then aligned twice (cross-surflex conformation and best cluster conformation from cross-docking), and the respective MIFs were calculated. Application of the two SB 3-D QSAR models to the external
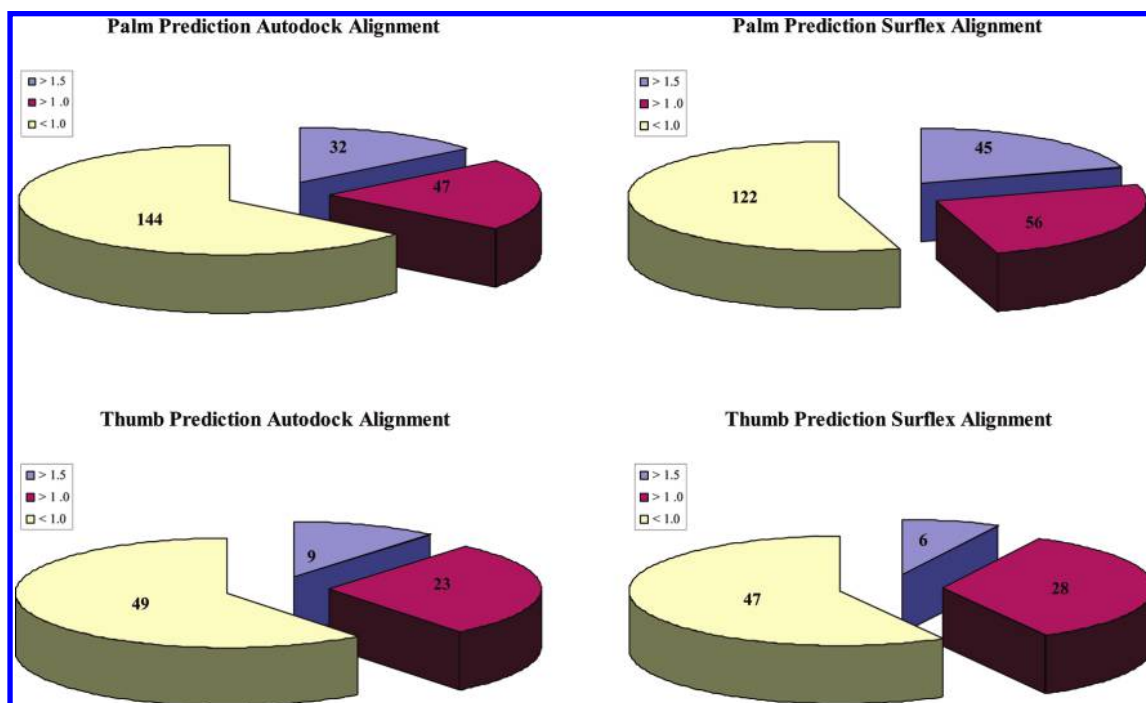

Figure 9. Diagrams of number of compounds with an error of prediction greater than 1.5 (violet) and 1.0 (dark red) and lower than 1.0 (pale yellow).
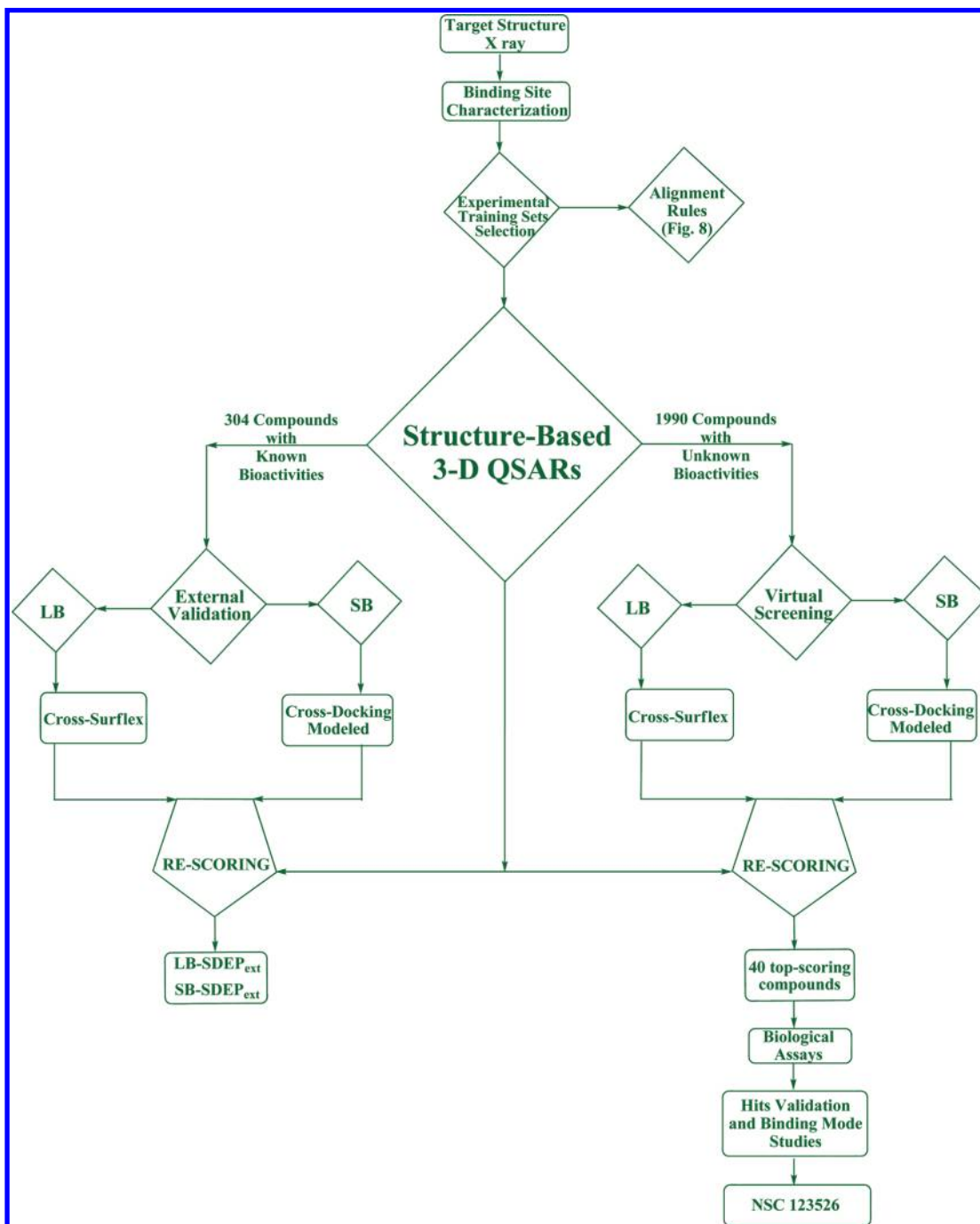
Figure 10. Workflow of 3-D QSAR studies and alignment rules integration in the virtual screening process used in the present paper.
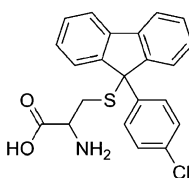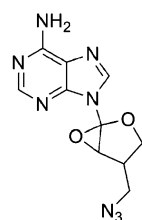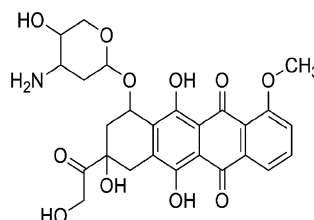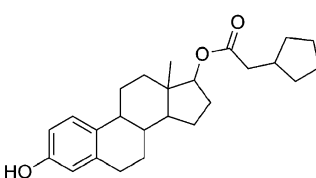
test sets (C1= GRID probe) yielded low standard deviation error of prediction values, shown in Tables S1 and S2 of the Supporting Information for all GRID probes.

Analysis of individual error of predictions are graphically reported in Figure 9. Here, it is clearly visible how the 3-D QSAR models, coupled to LB and SB alignment methods, can be used as an external scoring function. In particular, the use of quite large external test sets is a sort of a virtual screening simulation (although all molecules are known to bind NS5B). In particular from Figure 9, it is possible to observe a good agreement between the two alignment methods, and setting to 1 an arbitrarily threshold value of acceptable prediction error, the percentage of compounds well predicted ranged from 54% to 63% and rose to a maximum of 92% if the cutoff value would be set to 1.5.

These interesting results of the above mixed protocol pushed us toward the application of a virtual screening approach, which is reported in the next section. It has to be stressed that upon a detailed check, either training set or test set compounds were all biologically tested with similar condition assays.

Virtual Screening (VS). Structure based virtual screening has become an integral part of the drug design process and is now the most widely used approach to leverage structure for ligand discovery. In receptor based virtual screening, the aim is to find a ligand that has a high binding affinity to the target receptor whose 3-D structure is known. Ligand similarity searches also provide a very powerful method of quickly screening large databases of ligands to identify possible hits with activity against the target of interest. There

Table 7. Molecular Structure and Antiviral Activity[a] of the Compounds Selected by VS Protocol

Thumb Domain		Palm Domain	
NSC 123526	NSC 125626	NSC 169534	NSC 3354
IC_{50} = 46.0 μM	IC_{50} = 73.3 μM	IC_{50} = 64.5 μM	IC_{50} = 54.3 μM

[a] The data represents an average of at least two independent experiments.

are many tools available for performing these computational analyses, and they can be categorized as being either ligand based or receptor based. For ligand based methods, the strategy is to perform a similarity search if one or more active molecules are known. This can be done by a variety of methods, QSAR and 3-D QSAR models as well as pharmacophore matching.

Receptor based computational methods can be employed when the 3-D structure of the target protein is known. These involve docking algorithms to put ligands into the binding site of the target, and thus producing a predicted binding mode for each database compound. The application of scoring functions which in turn produce a ranking list of compounds is necessary to identify potential hits.

The VS protocol reported in this study is based on the application of two receptor based 3-D QSAR models, on external test sets containing 81 molecules for the thumb model and 223 molecules for the palm model (see Figure 10 for a detailed description of the VS workflow).

This procedure proved them to be useful scoring functions for the virtual screening approach, using either a ligand based (Surflex-Sim) or a structure based (Autodock) molecular alignment. Consequently, the NCI Diversity Set (DS) (http://dtp.nci.nih.gov), a database consisting of over 1990 compounds, was virtually screened. For each statistical model, the first 20 most active predicted molecules were selected for biological assays against recombinant NS5BCΔ21 (genotype 1b). Among the selected compounds, preliminary data yielded four derivatives (Table 7) with IC_{50} values between 45 and 75 μM. The selected compounds have been tested for their activity using similar biological conditions as for training set and external test set compounds (see biological data).

It could be argued that the NCI DS may have limitations to screen for compounds active against NS5B. However, the NCI database is one of the most widely used as it exhibits molecular diversity. Moreover, the implicit limitation of such a choice further strengths the applied approach considering that 10% of the tested molecules were indeed found active in the enzyme based assay.

BIOLOGICAL DATA

NS5B RdRp Assay. Recombinant NS5BCΔ21 was purified from the plasmid pThNS5BCΔ21 expressed in *Escherichia coli* DH5α by Ni-NTA affinity chromatography and was employed for evaluating the inhibitory activity of the compounds on poly rA/U$_{12}$ template-primer (TP) as previously described.[65,66] NS5B RdRp activity was determined in the absence or presence of the compounds at 30 °C for 60 min in a reaction buffer (25 μL) containing 20 mM Tris−HCl [pH 7.0], 100 mM NaCl, 100 mM sodium glutamate, 0.5 mM DTT, 0.01% BSA, 0.01% Tween-20, 5% glycerol, 20 U/mL of RNase Out, 0.5 μM poly rA/U$_{12}$ (preannealed), 25 μM UTP, 2−5 μCi [α-^{32}P]UTP, 0.5 mM MnCl$_2$, and 300 ng of NS5BCΔ21. Reactions were terminated by the addition of ice-cold 10% (v/v) trichloroacetic acid (TCA) containing 0.5 mM pyrophosphate. The quenched reaction mixtures were spotted on GF-B filters, washed with 5% (v/v) TCA-0.5 mM pyrophosphate to remove unincorporated rNTPs, rinsed, and quantified on a liquid scintillation counter (Packard). Activity of NS5B in the absence of the inhibitor was set at 100% and that in the presence of the inhibitor was calculated relative to this control. The 50%
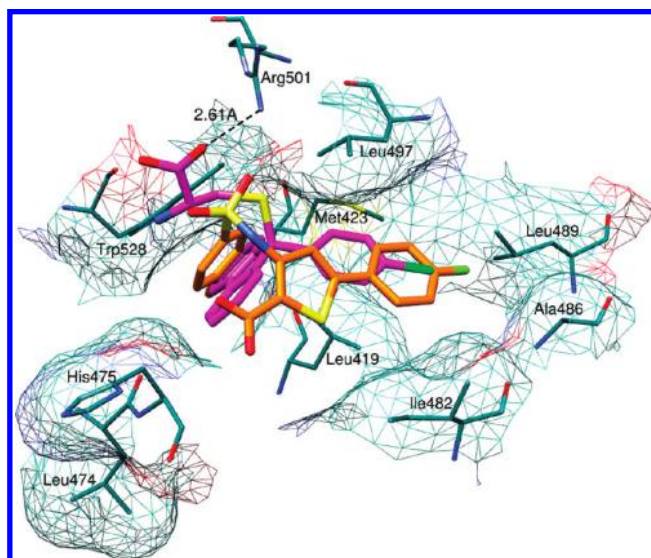
Figure 11. Superimposition of inhibitor **6** (orange) and hit compound **NSC 123526** (magenta) structures, showing a comparable mode of binding within the NS5B thumb domain. The residues in the binding site are shown by stick representation with carbon, nitrogen, oxygen, and sulfur atoms in cyan, blue, red, and yellow, respectively. For the sake of clarity, hydrogen bond interactions are represented by black dotted lines, and only a few residues of the binding site were displayed. Hydrogen atoms are omitted.

inhibition value (IC$_{50}$) for the inhibitors was evaluated from the dose response curve employing typically 9−12 concentrations of the compound and GraphPad Prism 5.0 soft.

N,N-disubstituted phenylalanine derivative 14 (as numbered in the original paper),[25] a documented NS5B inhibitor, was included as positive control in each set of experiments.

Binding Mode Analysis. Analysis of the binding mode of selected compounds (Table 7) within the NS5B RdRp polymerase thumb domain was performed with the aim of elaborating a putative way by which they inhibit the enzyme and, thus, get insights on the peculiar chemical features that influenced their activity. Among the four active compounds, only one draws (**NSC 123526**) particular attention from a medicinal chemistry point of view. In fact, compound **NSC 125626** is an oxirane (thus likely an irreversible inhibitor) and its structure is similar to a nucleoside, thus indicating it as a nucleoside inhibitor; compounds **NSC 169534** and **NSC 3354** are structurally too rigid and structurally resemble a tetracycline and a colestane related compound, respectively. As a matter of fact, the resulting four complexes were visually inspected, and only **NSC 123526**, endowed with the lowest inhibitory activity (IC$_{50}$ = 46.0 μM), was found in good agreement with the orientation of the most active thumb binding compound **6**. Moreover, a search for compound **NSC 123526** in the literature highlighted its effective activity against other polymerases diverse from NS5B RdRp.[67]

Figure 11 shows the predicted binding mode of the minimum energy conformation of the hit compound **NSC 123526** and the experimental conformation of the thumb inhibitor **6**. In detail, the *para*-chlorine-phenyl ring of **NSC 123526** is located in the region occupied by the fluoro-phenyl-thiophene moiety of inhibitor **6** and is characterized by van der Waals interactions with Leu419, Met423, Ile482, Ala486, Leu489, and Leu497. As reported,[68] the fluoro-phenyl-thiophene moiety (**6**) is bound to a hydrophobic surface of the thumb domain. In a similar fashion to the 4-methyl-phenyl group of inhibitor **6**, the tricyclic ring system of compound **NSC 123526** points toward the interior of the thumb domain, filling the primary binding cavity by making extensive van der Waals interactions with Leu419, Met423, Leu474, His475, and Trp528. In addition, one of the carboxylic oxygen atoms of **NSC 123526** may be involved in a hydrogen bond with the guanidinium group of Arg501. This is in full agreement with experimental data that show one of the sulfonyl oxygen atoms of compound **6** bonded to the same group of Arg501.

CONCLUSIONS

Two structure based 3-D QSAR models coupled with both ligand based and structure based alignment methods were developed for NS5B non-nucleoside inhibitors. The overall protocol was extensively assessed for both alignment and predictive ability. As a final test, a virtual screening application using the NCI Diversity Set 1990 compounds led to the discovery of a new active compound (**NSC 123526**) with an unprecedented scaffold among NS5B known inhibitors. Further virtual screening is in due progress on a wider and more appropriate containing structure database (i.e., Maybridge); furthermore, a focused virtual screening will be set up on a library of compounds structurally related to **NSC 123526**.

Supporting Information Available: Statistical results for the thumb and the palm training sets analyzed by different GRID probes, rmsd values obtained by redocking simulations of modeled compounds, rmsd values obtained by cross-docking simulations of modeled compounds, results from the assessment of Surflex-Sim. Molecular structures of 223 NNIs used as test set for palm TS Sim LB alignment protocol and molecular structures of 223 NNIs used as test set for palm TS. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Manns, M. P.; Foster, G. R.; Rockstroh, J. K.; Zeuzem, S.; Zoulim, F.; Houghton, M. The way forward in HCV treatment--finding the right path. *Nat. Rev. Drug Discov.* **2007**, *6*, 991–1000.

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

(3) Hoofnagle, J. H. Hepatitis C: the clinical spectrum of disease. *Hepatology* **1997**, *26*, 15S–20S.

(4) Feld, J. J.; Hoofnagle, J. H. Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature* **2005**, *436*, 967–972.

(5) Fried, M. W. Side effects of therapy of hepatitis C and their management. *Hepatology* **2002**, *36*, S237–244.

(6) 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.

(7) Reed, K. E.; Rice, C. M. Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties. *Curr. Top. Microbiol. Immunol.* **2000**, *242*, 55–84.

(8) Brass, V.; Blum, H. E.; Moradpour, D. Recent developments in target identification against hepatitis C virus. *Expert Opin. Ther. Targets* **2004**, *8*, 295–307.

(9) Tan, S. L.; Pause, A.; Shi, Y.; Sonenberg, N. Hepatitis C therapeutics: current status and emerging strategies. *Nat. Rev. Drug Discov.* **2002**, *1*, 867–881.

(10) 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.

(11) Bressanelli, S.; Tomei, L.; Roussel, A.; Incitti, I.; Vitale, R. L.; Mathieu, M.; De Francesco, R.; Rey, F. A. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13034–13039.

(12) Lesburg, C. A.; Cable, M. B.; Ferrari, E.; Hong, Z.; Mannarino, A. F.; Weber, P. C. Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nat. Struct. Biol.* **1999**, *6*, 937–943.

(13) Ago, H.; Adachi, T.; Yoshida, A.; Yamamoto, M.; Habuka, N.; Yatsunami, K.; Miyano, M. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Structure* **1999**, *7*, 1417–1426.

(14) Koch, U.; Narjes, F. Recent progress in the development of inhibitors of the hepatitis C virus RNA-dependent RNA polymerase. *Curr. Top. Med. Chem.* **2007**, *7*, 1302–1329.

(15) Carroll, S. S.; Tomassini, J. E.; Bosserman, M.; Getty, K.; Stahlhut, M. W.; Eldrup, A. B.; Bhat, B.; Hall, D.; Simcoe, A. L.; LaFemina, R.; Rutkowski, C. A.; Wolanski, B.; Yang, Z.; Migliaccio, G.; De Francesco, R.; Kuo, L. C.; MacCoss, M.; Olsen, D. B. Inhibition of hepatitis C virus RNA replication by 2′-modified nucleoside analogs. *J. Biol. Chem.* **2003**, *278*, 11979–11984.

(16) De Francesco, R.; Carfi, A. Advances in the development of new therapeutic agents targeting the NS3−4A serine protease or the NS5B RNA-dependent RNA polymerase of the hepatitis C virus. *Adv. Drug Deliv. Rev.* **2007**, *59*, 1242–1262.

(17) Koch, U.; Narjes, F. Allosteric inhibition of the hepatitis C virus NS5B RNA dependent RNA polymerase. *Infect. Disord.: Drug Targets* **2006**, *6*, 31–41.

IN SILICO SCREENING OF NEW HEPATITIS C VIRUS

J. Chem. Inf. Model., Vol. 50, No. 4, 2010 **675**

(18) 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.

(19) Clark, M.; Cramer, R. D., III; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative molecular field analysis (CoMFA). 2. Toward its use with 3D-structural databases. *Tetrahedron Comput. Methodol.* **1990**, *3*, 47–59.

(20) Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struct-Act. Relat.* **1993**, *12*, 9–20.

(21) Cruciani, G.; Watson, K. A. Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J. Med. Chem.* **1994**, *37*, 2589–2601.

(22) Wold, S. J. E.; Cocchi, M. PLS - Partial Least-Squares Projections to Latent Structures. In *3D-QSAR in Drug Design*; Kubiniyi, H., Ed. ESCOM Science Publishers: Leiden, 1993; pp 523−550.

(23) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(24) Wang, M.; Ng, K. K.; Cherney, M. M.; Chan, L.; Yannopoulos, C. G.; Bedard, J.; Morin, N.; Nguyen-Ba, N.; Alaoui-Ismaili, M. H.; Bethell, R. C.; James, M. N. Non-nucleoside analogue inhibitors bind to an allosteric site on HCV NS5B polymerase. Crystal structures and mechanism of inhibition. *J. Biol. Chem.* **2003**, *278*, 9489–9495.

(25) 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.

(26) Chan, L.; Das, S. K.; Reddy, T. J.; Poisson, C.; Proulx, M.; Pereira, O.; Courchesne, M.; Roy, C.; Wang, W.; Siddiqui, A.; Yannopoulos, C. G.; Nguyen-Ba, N.; Labrecque, D.; Bethell, R.; Hamel, M.; Courtemanche-Asselin, P.; L'Heureux, L.; David, M.; Nicolas, O.; Brunette, S.; Bilimoria, D.; Bedard, J. Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCV NS5B polymerase and HCV subgenomic RNA replication. Part 1: Sulfonamides. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 793–796.

(27) Chan, L.; Pereira, O.; Reddy, T. J.; Das, S. K.; Poisson, C.; Courchesne, M.; Proulx, M.; Siddiqui, A.; Yannopoulos, C. G.; Nguyen-Ba, N.; Roy, C.; Nasturica, D.; Moinet, C.; Bethell, R.; Hamel, M.; L'Heureux, L.; David, M.; Nicolas, O.; Courtemanche-Asselin, P.; Brunette, S.; Bilimoria, D.; Bedard, J. Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCV NS5B polymerase and HCV subgenomic RNA replication. Part 2: tertiary amides. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 797–800.

(28) Yan, S.; Appleby, T.; Larson, G.; Wu, J. Z.; Hamatake, R.; Hong, Z.; Yao, N. Structure-based design of a novel thiazolone scaffold as HCV NS5B polymerase allosteric inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5888–5891.

(29) Yan, S.; Larson, G.; Wu, J. Z.; Appleby, T.; Ding, Y.; Hamatake, R.; Hong, Z.; Yao, N. Novel thiazolones as HCV NS5B polymerase allosteric inhibitors: Further designs, SAR, and X-ray complex structure. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 63–67.

(30) Yan, S.; Appleby, T.; Larson, G.; Wu, J. Z.; Hamatake, R. K.; Hong, Z.; Yao, N. Thiazolone-acylsulfonamides as novel HCV NS5B polymerase allosteric inhibitors: convergence of structure-based drug design and X-ray crystallographic study. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1991–1995.

(31) 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.

(32) Li, H.; Tatlock, J.; Linton, A.; Gonzalez, J.; Borchardt, A.; Dragovich, P.; Jewell, T.; Prins, T.; Zhou, R.; Blazel, J.; Parge, H.; Love, R.; Hickey, M.; Doan, C.; Shi, S.; Duggal, R.; Lewis, C.; Fuhrman, S. Identification and structure-based optimization of novel dihydropyrones as potent HCV RNA polymerase inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4834–4838.

(33) Burton, G.; Ku, T. W.; Carr, T. J.; Kiesow, T.; Sarisky, R. T.; Lin-Goerke, J.; Baker, A.; Earnshaw, D. L.; Hofmann, G. A.; Keenan, R. M.; Dhanak, D. Identification of small molecule inhibitors of the hepatitis C virus RNA-dependent RNA polymerase from a pyrrolidine combinatorial mixture. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1553–1556.

(34) Slater, M. J.; Amphlett, E. M.; Andrews, D. M.; Bravi, G.; Burton, G.; Cheasty, A. G.; Corfield, J. A.; Ellis, M. R.; Fenwick, R. H.; Fernandes, S.; Guidetti, R.; Haigh, D.; Hartley, C. D.; Howes, P. D.; Jackson, D. L.; Jarvest, R. L.; Lovegrove, V. L.; Medhurst, K. J.; Parry, N. R.; Price, H.; Shah, P.; Singh, O. M.; Stocker, R.; Thommes,

P.; Wilkinson, C.; Wonacott, A. Optimization of novel acyl pyrrolidine inhibitors of hepatitis C virus RNA-dependent RNA polymerase leading to a development candidate. *J. Med. Chem.* **2007**, *50*, 897–900.

(35) 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.

(36) Tedesco, R.; Shaw, A. N.; Bambal, R.; Chai, D.; Concha, N. O.; Darcy, M. G.; Dhanak, D.; Fitch, D. M.; Gates, A.; Gerhardt, W. G.; Halegoua, D. L.; Han, C.; Hofmann, G. A.; Johnston, V. K.; Kaura, A. C.; Liu, N.; Keenan, R. M.; Lin-Goerke, J.; Sarisky, R. T.; Wiggall, K. J.; Zimmerman, M. N.; Duffy, K. J. 3-(1,1-dioxo-2H-(1,2,4)-benzothiadiazin-3-yl)-4-hydroxy-2(1H)-quinolinones potent inhibitors of hepatitis C virus RNA-dependent RNA polymerase. *J. Med. Chem.* **2006**, *49*, 971–983.

(37) 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.

(38) 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.

(39) 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.

(40) 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.

(41) 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.

(42) 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.

(43) Ikegashira, K.; Oka, T.; Hirashima, S.; Noji, S.; Yamanaka, H.; Hara, Y.; Adachi, T.; Tsuruha, J.; 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.

(44) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(45) Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Huang, C. C.; Ferrin, T. E. Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* **2006**, *7*, 339.

(46) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A. A package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics, and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.

(47) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(48) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.

(49) Wade, R. C.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the

ability to form more than two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 148–156.

(50) Pastor, M.; Cruciani, G.; Clementi, S. Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *J. Med. Chem.* **1997**, *40*, 1455–1464.

(51) Pastor, M.; Cruciani, G.; Watson, K. A. A strategy for the incorporation of water molecules present in a ligand binding site into a three-dimensional quantitative structure--activity relationship analysis. *J. Med. Chem.* **1997**, *40*, 4089–4102.

(52) Jain, A. N. Ligand-based structural hypotheses for virtual screening. *J. Med. Chem.* **2004**, *47*, 947–961.

(53) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1–5.

(54) Cleves, A. E.; Jain, A. N. Robust ligand-based modeling of the biological targets of known drugs. *J. Med. Chem.* **2006**, *49*, 2921–2938.

(55) Holt, P. A.; Ragazzon, P.; Strekowski, L.; Chaires, J. B.; Trent, J. O. Discovery of novel triple helical DNA intercalators by an integrated virtual and actual screening platform. *Nucleic Acids Res.* **2009**, *37*, 1280–1287.

(56) Langham, J. J.; Cleves, A. E.; Spitzer, R.; Kirshner, D.; Jain, A. N. Physical binding pocket induction for affinity prediction. *J. Med. Chem.* **2009**, *52*, 6107–6125.

(57) Ragno, R.; Mai, A.; Massa, S.; Cerbara, I.; Valente, S.; Bottoni, P.; Scatena, R.; Jesacher, F.; Loidl, P.; Brosch, G. 3-(4-Aroyl-1-methyl-1H-pyrrol-2-yl)-N-hydroxy-2-propenamides as a new class of synthetic histone deacetylase inhibitors. 3. Discovery of novel lead compounds through structure-based drug design and docking studies. *J. Med. Chem.* **2004**, *47*, 1351–1359.

(58) Wang, M.; Zhang, J.; Andrei, D.; Kuczera, K.; Borchardt, R. T.; Wnuk, S. F. Are L-adenosine and its derivatives substrates for S-adenosyl-L-homocysteine hydrolase. *J. Med. Chem.* **2005**, *48*, 3649–3653.

(59) Taylor, C. M.; Barda, Y.; Kisselev, O. G.; Marshall, G. R. Modulating G-protein coupled receptor/G-protein signal transduction by small

molecules suggested by virtual screening. *J. Med. Chem.* **2008**, *51*, 5297–5303.

(60) Radi, M.; Falciani, C.; Contemori, L.; Petricci, E.; Maga, G.; Samuele, A.; Zanoli, S.; Terrazas, M.; Castria, M.; Togninelli, A.; Este, J. A.; Clotet-Codina, I.; Armand-Ugon, M.; Botta, M. A multidisciplinary approach for the identification of novel HIV-1 non-nucleoside reverse transcriptase inhibitors: S-DABOCs and DAVPs. *ChemMedChem* **2008**, *3*, 573–593.

(61) Solis, F. J.; Wets, R. J. B. Minimization by Random Search Techniques. *Math. Oper. Res.* **1981**, *6*, 19–30.

(62) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

(63) Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; et al. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.* **1995**, *2*, 293–302.

(64) Ragno, R.; Frasca, S.; Manetti, F.; Brizzi, A.; Massa, S. HIV-reverse transcriptase inhibition: inclusion of ligand-induced fit by cross-docking studies. *J. Med. Chem.* **2005**, *48*, 200–212.

(65) Kaushik-Basu, N.; Bopda-Waffo, A.; Talele, T. T.; Basu, A.; Costa, P. R.; da Silva, A. J.; Sarafianos, S. G.; Noel, F. Identification and characterization of coumestans as novel HCV NS5B polymerase inhibitors. *Nucleic Acids Res.* **2008**, *36*, 1482–1496.

(66) Kaushik-Basu, N.; Bopda-Waffo, A.; Talele, T. T.; Basu, A.; Chen, Y.; Kucukguzel, S. G. 4-Thiazolidinones: a novel class of hepatitis C virus NS5B polymerase inhibitors. *Front. Biosci.* **2008**, *13*, 3857–3868.

(67) Silverman, J. E.; Ciustea, M.; Shudofsky, A. M.; Bender, F.; Shoemaker, R. H.; Ricciardi, R. P. Identification of polymerase and processivity inhibitors of vaccinia DNA synthesis using a stepwise screening approach. *Antiviral Res.* **2008**, *80*, 114–123.

(68) Biswal, B. K.; Wang, M.; Cherney, M. M.; Chan, L.; Yannopoulos, C. G.; Bilimoria, D.; Bedard, J.; James, M. N. Non-nucleoside inhibitors binding to hepatitis C virus NS5B polymerase reveal a novel mechanism of inhibition. *J. Mol. Biol.* **2006**, *361*, 33–45.