

# Exploring the P2 and P3 ligand binding features for Hepatitis C virus NS3 protease using some 3D QSAR techniques

Hsin-Yuan Wei, Chien-Sheng Lu, Thy-Hou Lin \*

Institute of Molecular Medicine & Department of Life Science, National Tsing Hua University, Hsinchu 30013, Taiwan

Received 10 August 2007; received in revised form 28 September 2007; accepted 3 October 2007

Available online 9 October 2007

---

## Abstract

Several three-dimensional quantitative structure–activity relationship (3D-QSAR) models have been constructed using the comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis (CoMSIA), and Catalyst pharmacophore feature building programs for a series of 26 truncated ketoacid inhibitors designed particularly for exploring the P2 and P3 binding pockets of HCV NS3 protease. The structures of these inhibitors were built from a structure template extracted from the crystal structure of HCV NS3 protease. The structures were aligned through docking each inhibitor into the NS3 active site using program GOLD. The best CoMSIA model was identified from the stepwise analysis results and the corresponding pharmacophore features derived were used for constructing a pharmacophore hypothesis by the Catalyst program. Pharmacophore features obtained by CoMFA and CoMSIA are found to be in accord with each other and are both mapped onto the molecular 5 K surface of NS3 active site. These pharmacophore features were also compared with those obtained by the Catalyst program and mapped onto the same NS3 molecular surface. The pharmacophore building process was also performed for 20 boronic acid based NS3 inhibitors characterized by a long hydrophobic side chain attached at position P2. This latter pharmacophore hypothesis built by the Catalyst program was also mapped onto the molecular surface of NS3 active site to define a second hydrophobic feature at position P2. The possibility of using the pharmacophore features mapped P2 and P3 binding pocket to design more potent depeptidized NS3 inhibitors was discussed.

© 2007 Elsevier Inc. All rights reserved.

**Keywords:** 3D QSAR; CoMFA; CoMSIA; Catalyst; Structure alignment; GOLD; Molecular docking; Pharmacophore model

---

## 1. Introduction

The hepatitis C virus (HCV) infects approximately 170,000,000 people worldwide [1]. The virus is transmitted primarily by blood and blood products. Most of the infected individuals have either received blood transfusions or have used intravenous drugs [1,2]. The HCV infection is more common in sexually promiscuous individuals but rare in monogamous couples [3]. Perinatal transmission from mother to fetus or infant is also relatively low but possible (less than 10%) [2,3]. Many individuals infected with HCV have no obvious risk factors [3]. Most of these persons have probably been inadvertently exposed to contaminated blood or blood products [3].

As a member of *Flaviviridae*, the complete genomes of various HCV isolates were cloned and sequenced by several

groups [4–7]. The HCV genome is a positive, single-stranded RNA of approximately 10,000 nucleotides and encodes a single polyprotein of about 3010 amino acids [8,9]. The polyprotein is processed by host cell and two viral proteases NS2/3 and NS3 into ten different products, with the structural proteins (core (C), E1 and E2) located in the N-terminal third and the nonstructural (NS2–5) replicative proteins in the remainder [10–12]. The role of NS2/3 protease appears to be limited to the autoproteolytic cleavage of the NS2-NS3 junction in *cis* [13]. The amino-terminal 180-amino-acid sequence of NS3 protein encodes a serine protease which cleaves at the NS3/4A junction in *cis*, which is followed by cleavage at the NS4/4B, NS4B/5A, and NS5A/B sites in *trans* [13]. The NS3 serine protease requires an accessory viral protein, NS4A, for optimal cleavage activity [14]. The contribution of NS4A to NS3 protease activity can be mimicked by a synthetic peptide encompassing amino acid residues 21–34 of NS4 [14,15].

The NS3 protease has been considered as one of the most attractive targets for anti-HCV therapy because it is essential

---

\* Corresponding author. Fax: +886 3 571 5934.

E-mail address: [thlin@life.nthu.edu.tw](mailto:thlin@life.nthu.edu.tw) (T.-H. Lin).

for viral replication and formation of infectious viral particles [5,7]. A 2.5 Å resolution structure of the NS3:NS4A complex reveals that the HCV NS3 serine protease domain adopts a chymotrypsin-like fold. The complex consists of two structural domains, each containing a twisted β sheet incorporating a “Greek key” motif while the C-terminal domain (residues 1120–1206) contains the conventional six-stranded β barrel, common to nearly all members of the chymotrypsin family, followed by a structurally conserved α helix [13]. The catalytic triad of His57, Asp81 and Ser139 is located in a crevice between the two domains [13]. The geometrical arrangement of the catalytic triad is similar to that of other serine proteases. Furthermore, a Zn ion is tetrahedrally coordinated with Cys97, Cys99, Cys145 and His149 residues of a site located opposite to the active site [13]. The Zn ion is believed to play a structural role because the removal of the ion causes unfolding and precipitation of the protein [13].

The NS3-dependent cleavage sites have been mapped by using amino-terminal sequence analysis of mature proteins expressed in cell culture [16]. The substrate specificity of the proteinase has been identified by comparing the cleavage sites on the HCV sequence. Sequences of these cleavage sites have been generalized as D/E-X-X-X-X-Cys/Thr↓Ser/Ala-X-X-L/W/Y, with cleavage occurring after cysteine or threonine [17]. According to the nomenclature of Schechter and Berger, [18] the newly generated carboxy terminus, after cleavage of the peptide bond, is designated P1, and it is preceded by the P2 residue, etc.; the newly generated amino terminus is designated P1', and it is followed by P2', etc. [19]. The preferred P1 residue is cysteine with an exception at the intramolecular NS3/4A junction where a threonine residue is preferred. Other conserved features are a negatively charged residue in the P6 position, an alanine or a serine in P1' and a hydrophobic residue in P4' position [19]. The corresponding binding subsites on the enzyme surface are denoted as S6 through S4' [20]. Many of the NS3 inhibitors designed are decapeptide or hexapeptide substrate analogues of the N-terminal cleavage (P6-P1) product [21]. Replacement of the P1 cysteine with other moieties, e.g. (S)-4,4-difluoro-2-aminobutyric acid and related ‘serine trap’ functional groups (cysteine mimetics) and substitution of the P3 unit with indoline-based structures (peptidomimetics) are the two most common approaches used for designing these inhibitors. Some rational designs such as the alternate P1 thiol chemistry in accordance with the lipophilicity of the S1 pocket formed by Leu135, Phe154 and Ala157, or the introduction of amphiphilic groups to indoline peptidomimetics in response to the presence of Arg161 and Lys136 proximal to the S3 site are conducted based on the NS3 structure determined [21]. The crystal structure of NS3/4A complex with either of two covalently bound α-ketoacid serine trap inhibitors determined by Di Marco et al. [22] reveals that the carbamate-protected tripeptide adducts bind in an extended conformation, occupying the S1–S4 sites and forming an antiparallel β-sheet with the E2 strand (Ala156-Val170). A similar overall binding configuration for a hexapeptide boronic acid inhibitor bound to NS3/4A was observed by NMR [23]. The boronic acid cysteine residue was found to interact with the

active site serine residue such that the P1 region of the bound inhibitor mimics the transition state of substrate hydrolysis [23].

A novel class of NS3 protease inhibitors has been made based on the C-terminal tetrapeptide cleavage product (P1'–P4') [17]. However, the most potent inhibitors reported to date contain either a 4-substituted proline or a 3,4-disubstituted proline as P2 residue [24]. The potency of these inhibitors are further enhanced through a depeptidize process using 2-azabicyclo [2.2.1]-heptane carboxylic acid as a surrogate [24]. A bicyclic pyrimidinone-based P2–P3 dipeptide replacement has been incorporated into a peptide boronic acid inhibitor to increase the potency [25]. Oligopeptide derivatives containing α-ketoamide electrophilic trap are also potent inhibitors of NS3 as have been reported [26]. The potency of these P3-capped inhibitors is also increased through a depeptidization process [26]. Another series of highly potent inhibitors are designed based on the trisubstituted cyclopentane moiety at the P2 position [27]. The P2 benzene ring has been macrocyclized with the P3 capping group through an aryl-alkyl ether linkage to depeptidize the P2–P3 moiety [28]. These are the conformationally preorganized inhibitors with better stability and strong potency [28–30]. Recently, a series of potent tripeptide truncated at the N-terminus from some hexapeptide ketoacid inhibitors are described to explore the P2–P3 binding features [31]. The IC<sub>50</sub> measured for the most potent inhibitor of this series is 0.38 μM. The N-terminal aminoacid of these α-ketotripeptide inhibitors are further replaced with α-hydroxy acid, leading to a series of capped dipeptide inhibitors [32]. The original P3 residue is replaced with a small hydroxylated one to alternate the original chirality at the P3 binding position from S-configuration to R-configuration [32]. Some even more potent β-sheet dipeptide mimetics aiming at the S2 binding pocket namely, 3-amino bicyclic pyrazinones, made from a hexapeptide boronic acid lead are also reported [25,33]. In general, the P2 binding feature is believed to be lipophilic while both hydrophobic and hydrophilic groups are accepted in P3. To characterize the P2–P3 binding features theoretically, we have employed several 3D QSAR techniques on the aligned structures of 26 truncated ketoacid inhibitors of NS3 designed by Nizi et al. [31] and Colarusso et al. [32] for constructing some 3D QSAR models for these compounds. The truncated ketoacid inhibitors are divided into the training and test set and the structures in each set are aligned and analyzed by the CoMFA [34] and CoMSIA [35] methods to derive the best 3D QSAR model for the truncated ketoacid inhibitors. Further, the pharmacophore features obtained from the best CoMSIA model are used to construct some pharmacophore hypotheses using the Catalyst 4.9 program [36]. The top hypothesis generated is mapped onto the structures of several highly active truncated ketoacid inhibitors selected from both the training and test sets. The predicted activities for both the training and test sets by the top hypothesis are found to be in good accord in statistics with those predicted by the best CoMSIA model. The pharmacophore hypotheses by Catalyst 4.9 program are also built for structures of 20 boronic acid based pyrimidinone and pyrazinone inhibitors made by Glunz et al. [25] and Zhang

et al. [33], respectively. These latter hypotheses are particularly aimed at a hydrophobic feature at position P2 and are compared with those built for the other P2–P3 positions. The feasibility of using the constructed 3D QSAR models to design more potent and less peptidic NS3 inhibitors is discussed.

## 2. Methodology

The structures and activities of 26 truncated ketoacid inhibitors of HCV NS3 protease studied were listed in Table 1. The X-ray and molecular modeling determined conformation [22] of compound **002** (BOC-GLU-LEU-FKI) was used as the structural template to build structures for other compounds of the same series. No conformational search was conducted for the template structure since it was known that the experimentally determined structure was similar to that determined theoretically from molecular modeling techniques [37,38]. The structure of each other compounds of the series was constructed within the active site of 1DY8 [22] by replacing the side chains of compound **002** with other groups as was described previously [39] (Table 1). The hydrogen atoms were added for each structure. The IC<sub>50</sub> measured for these truncated ketoacid inhibitors was 0.3–100 μM [31,32]. These inhibitors were divided into the training and test set with each consisting of 16 inhibitors and covering roughly the same activity range. Each structure constructed was subjected to a brief energy minimization with the receptor together using the SYBYL 6.9.1 program [40]. Then, the crystal ligand was docked into the active site of HCV NS3 protease using the GOLD V3.1 program [41] and the corresponding RMSD between the crystal and docked conformation was computed. The covalent bond between the catalytic serine (Ser139) and ketone group of inhibitors of HCV NS3 protease was fixed during the GOLD docking process. The major docking parameters namely, number of operations and population size chosen were 100,000 and 100, respectively. The energy-minimized structures were subsequently docked into the active site of HCV NS3 protease. The MMFF94 [42] charges were deployed for each inhibitor.

The steric and electrostatic potential fields of CoMFA were calculated by the SYBYL 6.9 program using a regularly spaced lattice of 2.0 Å. The lattice was extended to 4 Å units beyond the van der Waals volume of each molecule in the X, Y, and Z directions. A sp<sup>3</sup> carbon atom of radius 1.52 Å and charge +1.0 was used as a probe to calculate both steric and electrostatic fields. Truncation for both steric and electrostatic contributions was set at ±30 kcal/mol. The electrostatic contribution at the lattice intersections where maximum steric interactions were computed was ignored. Both CoMFA steric and electrostatic fields computed were scaled by the standard option given in the program. The same lattice where each molecule was submerged for CoMFA was used for CoMSIA. A sp<sup>3</sup> atom of radius 1.0 Å and charge +1 was used as the probe to compute the CoMSIA similarity indices defined by Klebe et al. [35]. The similarity indices were calculated using the Gaussian-type distance dependence between the probe and atoms of the molecules of a data set. This functional form requires no arbitrary definition of cutoff limits and the similarity indices can be calculated at all

Table 1  
Structure and activities of the 26 truncated ketoacid inhibitors against the HCV NS3 protease studied

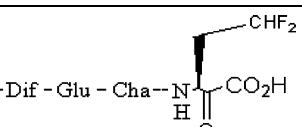
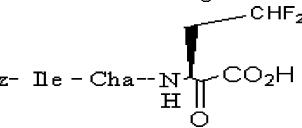
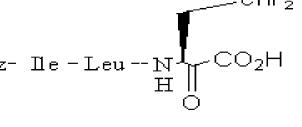
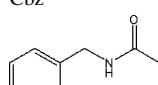
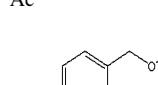
Inh.	Structure	IC <sub>50</sub> (μM)		
<b>015</b>		5.3		
<b>007</b>		1.4		
<b>009</b>		1.7		
Inh.	X	P3	P2	IC <sub>50</sub> (μM)
<b>009</b>	Cbz	Ile	Leu	1.7
<b>014</b>		Ile	Leu	3.3
<b>006</b>		Ile	Leu	1.0
<b>023</b>	Ac	Ile	Leu	16
<b>010</b>			Leu	1.7
<b>004</b>		Ile	Leu	0.44
<b>031</b>				>100
<b>022</b>	Cbz	Ile		15.6
<b>013</b>	Gly			3.28
<b>019</b>	Cbz	Ile		9.4
<b>024</b>	Ala			16.0
<b>001</b>	Cbz	Ile		0.30
<b>002</b>	Abu			0.33
<b>005</b>	Cbz	Ile		0.46
<b>011</b>	Phe			2.10
<b>008</b>	Cbz	Ile		1.63
	Val			
	Boc	Glu	Difluro	
	Abu			
	Boc		Glu	
	Leu			
	Boc		Val	
	Leu			
	Boc		Gln	
	Leu			
	Boc		Asp	
	Leu			

Table 1 (Continued)

Inh.	X	P3	P2	IC <sub>50</sub> (μM)
Inh.	R			IC <sub>50</sub> (μM)
025				18
016				7
Inh. conf.	R	P3	P2	IC <sub>50</sub> (μM)
029		(R)		41
027		(R)		28
020		(R)		13
021		(R)		13
026		(R)		18
017		(R)		7

lattice points inside and outside of a molecule [28]. The attenuation factor  $\alpha$  was set as 0.3. In the SYBYL CoMSIA module, [35] the third power of the atomic radii was computed as the steric indices, the atomic partial charges were treated as the electrostatic indices, the atom-based parameters developed by Viswanadhan et al. [43] were used as the hydrophobic indices, and a rule-based method derived experimentally [44] was used as the hydrogen bond donor and acceptor indices.

In proceeding with CoMFA and CoMSIA, the structures were aligned based on the docked conformations generated by the GOLD V3.1 program. The CoMFA and CoMSIA results were cross-validated using the SYBYL PLS module [40]. The CoMFA and CoMSIA descriptors were treated as the independent variables while the IC<sub>50</sub> values (Table 1) were treated as the dependent ones in all the PLS regression analyses for deriving the 3D QSAR models. The optimum number of components used to derive a nonvalidated model was defined as the number of components leading to the highest crossvalidated  $r^2$  ( $q^2$ ) and lowest standard error of prediction. The goodness of nonvalidated models was judged by the conventional correlation coefficient  $r^2$ , standard error of estimate, and  $F$ -values. The

results of nonvalidated analyses were used to make prediction for the binding affinities of the test set inhibitors and also to display the coefficient contour maps.

The same training set used in CoMFA and CoMSIA was also used for constructing some pharmacophore models by the Catalyst 4.9 program. All the parameters used were default settings except that Unc (uncertainty) was set at 1.4. There were five pharmacophore features namely, HD (hydrogen-bond donor), NEG (negative charge), HY (hydrophobic), and HR (hydrophobic aromatic) selected for the hypothesis generation process. The pharmacophores were automatically generated by the HypoGen module [29] of Catalyst 4.9 program for the training set. The top 10 scored hypotheses generated for each inhibitor of the set were exported. The Catalyst 4.9 program was also applied on structures of 20 boronic acid based pyrimidinone [25] and pyrazinone [33] inhibitors. There were 250 conformations generated for each boronic acid based inhibitor using the Catalyst/ConFirm module. The best conformation was selected from each conformational pool and then subjected to the hypothesis generation process using the Catalyst/HypoGen module.

### 3. Results and discussion

The binding positions P1, P1', P2, and P3 of a truncated  $\alpha$ -ketoacid inhibitor with the corresponding binding pocket of HCV NS3 protease are highlighted respectively in Fig. 1a and b where the structure of truncated  $\alpha$ -ketoacid inhibitor is engulfed in the surface of binding pocket. This X-ray structure [22] clearly shows that the capping group of inhibitor is pointed toward the hydrophobic residue Val158 while the acid one is toward Arg123. Fig. 1b also shows that residues around position P3 and P2 are His57, Arg123, Val132, Lys136, Ala156, Val158 and Cys159, respectively. This reflects the fact that position P3 and P2 are both hydrophobic and hydrophilic. Functional groups placed at position P2 and P3 of the 26 truncated ketoacid inhibitors studied accompanying with their IC<sub>50</sub> values measured are listed in Table 1. Structure features of these compounds already indicate that placing a hydrophilic group at P3 and a hydrophobic one at P2 gives rise to compounds of higher potency (Table 1).

To generate rational structures for 3D QSAR analyses, the parameters of GOLD V3.1 program are fine tuned to give a RMSD of 0.65 between the docked and crystal structures of 1DY8 [22]. The docked structures are aligned using some backbone atoms selected as the correspondence points. The aligned structural sets are then analyzed by the SYBYL CoMFA, CoMSIA and PLS programs and results with significant statistics obtained are kept for further analyses. The statistics of the best CoMFA and CoMSIA results obtained for the training set inhibitors is presented in Table 2. The best CoMFA result yields a leave-one-out (loo) validated  $r^2$  ( $q^2_{\text{loo}}$ ) of 0.749, crossvalidated  $r^2$  ( $q^2$ ) of 0.752, and conventional  $r^2$  ( $r^2$ ) of 0.998 (Table 2). The CoMSIA is conducted in a stepwise manner namely, a single field index, a combination of any two field indices, a combination of any three field indices, a combination of any four field indices, and then a combination of

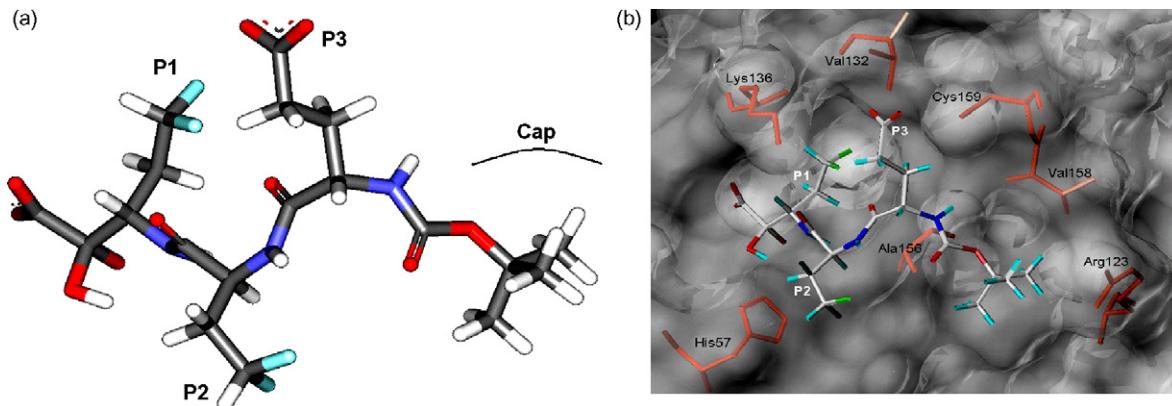


Fig. 1. (a) The binding positions of HCV NS3 active site reported by Di Marco et al. [22] are labeled around the structure of inhibitor 001 (Table 1). Position P1 is near the catalytic serine (Ser139), position P2 and P3 are difluoroAbu and Glu groups (Table 1), and position P6 is capped by a Boc group (Table 1), respectively. (b) The active site of HCV NS3 protease where the structure of inhibitor 001 (Table 1) is docked. The corresponding binding pockets/positions described in (a) are marked.

all field indices selected is employed step-by-step in the analyses. There are five different field indices namely, steric (S), electrostatic (E), hydrophobic (H), H-bond acceptor (A), and H-bond donor (D), being chosen for the stepwise CoMSIA and the results are presented in Table 2. We compare the CoMSIA results using values of  $r^2$ ,  $q_{\text{loo}}^2$ ,  $q^2$ , Standard Error of Prediction (SEP), and  $F$ -statistics, (Table 2). Apparently, with values of SEP of 0.020,  $F$  of 3932.34,  $q_{\text{loo}}^2$  of 0.76, the best  $q^2$  of 0.79, and relative contribution of steric/electrostatic field of 0.745:0.255 obtained, we identify the best CoMSIA result as the one where S (steric), E (electrostatic), H (hydrophobic), and D (H-bond donor) indices are selected in the analysis (Table 2). This shows that the corresponding P2–P3 binding features are rather versatile.

The CoMFA and CoMSIA contour maps agree with each other on the identification of favor regions for steric interaction

(displayed with green contours by both CoMFA and CoMSIA) which are around position P2 (Fig. 2a and b). As displayed in yellow contours, both CoMFA and CoMSIA maps also agree with each other that there are disfavor regions for steric interaction around position P3 (Fig. 2a and b). The favor region for negative charge interaction identified by CoMFA (displayed in red contours) around position P3 (Fig. 2a) agrees with that displayed in the same color of contours by CoMSIA around the same position (Fig. 2b). However, the CoMSIA contours identify some favor regions for hydrophobic interaction around position P2, P3 and the capping group (displayed in orange color) (Fig. 2b) and also a favor region for H-bond donor interaction around position P3 (displayed in cyan color) (Fig. 2c). These latter two CoMSIA identifications agree with those reported by others around the same positions. A projection of the CoMFA contours over the electrostatic

Table 2  
A summary of CoMFA and stepwise CoMSIA statistics for the training set inhibitors

CoMFA		CoMSIA	Leave one out		Cross validation		No validation		
			NC <sup>a</sup>	$q_{\text{loo}}^2$	NC	$q^2$	SEP <sup>a</sup>	$r^2$	$F$
Leave one out $q_{\text{loo}}^2$	0.749	S	4	0.558	2	0.603	0.187	0.965	41.318
Cross-validation ( $q^2$ )	0.752	E	4	0.497	4	0.497	0.058	0.997	445.052
Conventional $r^2$	0.998	H	6	0.659	6	0.657	0.036	0.999	1161.017
Standard Error	0.039	D	1	0.102	1	0.163	0.285	0.919	16.996
Principal components	5	A	5	0.515	5	0.534	0.274	0.925	18.415
$F$ -value	978.518	H + S	3	0.648	3	0.710	0.046	0.998	695.699
Relative contribution of steric/electrostatic field	0.745:0.255	H + E	6	0.743	4	0.742	0.030	0.999	1686.120
		H + D	6	0.772	6	0.766	0.020	1.000	3684.186
		H + A	5	0.765	5	0.755	0.029	0.999	1736.800
		H + D + S	6	0.759	6	0.773	0.018	1.000	4642.943
		H + D + E	6	0.747	6	0.711	0.018	1.000	4623.887
		H + D + A	6	0.753	6	0.768	0.021	1.000	3540.102
		H + A + S	6	0.771	6	0.755	0.033	0.999	1345.065
		H + A + E	4	0.760	6	0.769	0.033	0.999	1401.305
		H + D + S + E	6	0.760	6	0.788	0.020	1.000	3932.338
		H + D + S + A	6	0.760	6	0.701	0.023	0.999	2848.304
		All fields	5	0.754	6	0.753	0.015	1.000	6474.435

Abbreviations: SEP, Standard Error of Prediction; NC, number of components; S, steric; E, electrostatic; H, hydrophobic; A, H-bond acceptor; D, H-bond donor indices.

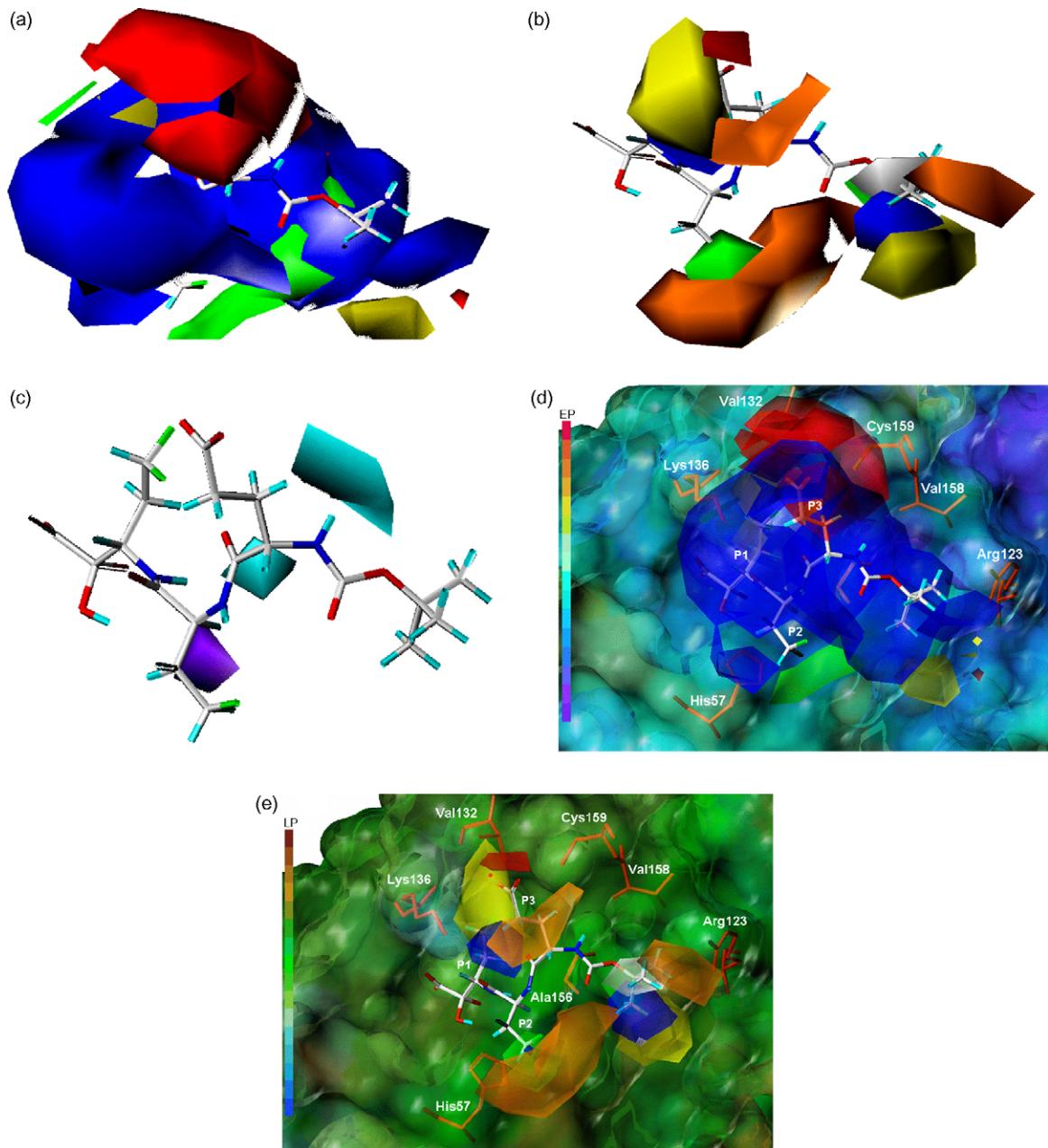


Fig. 2. (a) The CoMFA contours (blue contours favor positive charge, red contours favor negative charge, green contours favor steric, and yellow contours disfavor steric) of the best CoMFA model (Table 2) of the training set inhibitors is mapped with the structure of inhibitor 001 (Table 1). (b) The CoMSIA steric, electrostatic, and hydrophobic contours (blue contours favor positive charge, red contours favor negative charge, green contours favor steric; yellow contours disfavor steric, orange contours favor hydrophobic, and white contours disfavor hydrophobic) of the best CoMSIA result (Table 2) of the training set inhibitors is mapped with the structure of inhibitor 001 (Table 1). (c) The CoMSIA H-bond contours (cyan contours favor H-bond donor and purple contours disfavor H-bond donor) of the best CoMSIA result (Table 2) of the training set inhibitors is mapped with the structure of inhibitor 001 (Table 1). (d) Projection of the CoMFA contours depicted in (a) over the electrostatic potential surface of the HCV NS3 active site. (e) Projection of the CoMSIA contours depicted in (b) over the lipophilic potential surface of the HCV NS3 active site.

potential surface map of the NS3 active site is presented in Fig. 2d. Note that the CoMFA red contours are directly pointed toward Lys136 (Fig. 2d). A favor region for hydrophobic interaction at position P3 and capping group are identified by the CoMSIA orange contours mapped onto the lipophilic surface of NS3 active site (Fig. 2e). These contours are pointed toward some nearby hydrophobic residues such as Val132, Cys159, and Val158 highlighted (Fig. 2e). Some other CoMSIA orange contours are also being mapped at position P2 (Fig. 2e)

but no apparent hydrophobic residues are identified around the position except Ala156. This shows that the hydrophobic pocket at position P2 is somewhat buried.

Based on the stepwise CoMSIA results obtained (Table 2), the structural features selected for constructing a pharmacophore for the truncated ketoacid inhibitors using the Catalyst 4.9 program are HD, NEG, HY, and HR. A comparison for the statistical significance of the top 10 hypotheses generated by the CatSramble module of Catalyst program is given in Table 3.

Table 3

Validation of the Hypo1 and Hypo2 hypothesis using the CatScramble program implemented in the Catalyst program

Hypothesis no.	Hypo1 (HD + NEG + HY + HY + HY)					Hypo2 (HD + HR + HY + HY)				
	Total cost	$\Delta$ cost1	$\Delta$ cost1	rms deviation	Correlation (r)	Total cost	$\Delta$ cost1	$\Delta$ cost2	rms deviation	Correlation (r)
1	66.31	46.60	52.91	0.88	0.95	83.67	42.81	52.33	0.886	0.943
2	73.17	39.74	52.91	1.28	0.90	87.97	38.51	52.33	1.134	0.902
3	73.58	39.33	52.91	1.24	0.91	90.24	36.24	52.33	1.261	0.876
4	79.59	33.32	52.91	1.53	0.85	97.59	28.89	52.33	1.526	0.812
5	81.44	31.47	52.91	1.58	0.84	98.80	27.68	52.33	1.453	0.841
6	84.57	28.34	52.91	1.59	0.84	98.89	27.59	52.33	1.523	0.816
7	106.45	6.46	52.91	2.27	0.64	99.05	27.43	52.33	1.577	0.798
8	107.44	5.47	52.91	2.30	0.63	99.44	27.04	52.33	1.558	0.805
9	109.05	3.86	52.91	2.34	0.61	100.87	25.61	52.33	1.553	0.812
10	112.91	0	52.91	2.94	0	102.12	24.36	52.33	1.671	0.769

 $\Delta$ cost1 = Null-total cost,  $\Delta$ cost2 = Null-fixed cost, HD: Hydrogen bond donor, NEG: negative charge, HY: hydrophobic, and HR: hydrophobic aromatic.

The Config cost obtained is 15.07 and Unc obtained is 1.8. Apparently, with a cost difference between null and total one being 46.6 and that between null and fixed one being 52.9, the top hypothesis (designated as Hypo 1 thereafter) generated meets the criteria of being a good hypothesis (Table 3). The actual and predicted IC<sub>50</sub> by the best CoMFA (Table 2 and Fig. 2a) and CoMSIA (Table 2 and Fig. 2b) models and the Hypo1 hypothesis for each training set inhibitor are listed and compared in Table 4 while those for each test set inhibitor are given in Table 5, respectively. In these tables, all the actual and predicted IC<sub>50</sub> are listed in an ascending order from top to bottom. Apparently, most of the predicted IC<sub>50</sub> of the training set by the best CoMFA and CoMSIA models are in good accord with the actual ones (Table 5). Linear regression of the predicted versus actual IC<sub>50</sub> yields a regression coefficient of 0.998 and 1.000 for both the CoMFA and CoMSIA results, respectively. However, the predicted and actual IC<sub>50</sub> given by the best CoMFA and CoMSIA model on the test set inhibitors are somewhat discrepant (Table 5). A linear regression

coefficient of 0.6 and 0.7 is obtained for the test set inhibitors given by the best CoMFA and CoMSIA model, respectively (Table 5). The discrepancy may be ascribed to some outstanding outliers such as inhibitor **023** and **025** present in the analysis (Table 5). The predicted IC<sub>50</sub> for both the training and test set inhibitors given by the Hypo1 hypothesis are also listed in Tables 4 and 5, respectively. The goodness of these predictions can be judged through the activity scale namely, the '+' sign labeled for each actual and predicted IC<sub>50</sub> listed in both Tables 4 and 5. While highly active inhibitors predicted are labeled with the '+++' sign, those less and least active ones predicted are labeled with the '++' and '+' signs, respectively. Apparently, there are two differences in the activity scale labeled between the actual and predicted IC<sub>50</sub> for the training set inhibitors by the Hypo1 hypothesis (Table 4). The differences are due to inhibitor **004** and **010** where the actual IC<sub>50</sub> labeled are either higher or lower than the predicted one (Table 4). Therefore, the prediction accuracy by the Hypo1 hypothesis for the training set inhibitors estimated is 88%.

Table 4

The actual (Act) and predicted (Pred) IC<sub>50</sub> by the best CoMFA and CoMSIA models and by the Hypo1 hypothesis for the training set inhibitors are compared

Inh.	Act IC <sub>50</sub> ( $\mu$ M)	CoMFA	CoMSIA H + D + S + E	Catalyst Pharmacophore Hypothesis Hypo1					
				Pred IC <sub>50</sub> ( $\mu$ M)	Pred IC <sub>50</sub> ( $\mu$ M)	Act IC <sub>50</sub> ( $\mu$ M)	Pred IC <sub>50</sub> ( $\mu$ M)	Error	Act activity scale <sup>a</sup>
<b>001</b>	0.30	0.29	0.30		0.30	0.89	3.0	+++	+++
<b>002</b>	0.33	0.33	0.33		0.33	0.28	-1.2	+++	+++
<b>004</b>	0.44	0.45	0.45		0.44	1.10	2.5	+++	++
<b>005</b>	0.46	0.52	0.49		0.46	0.29	-1.6	+++	+++
<b>006</b>	1.00	1.01	0.92		1.00	1.40	1.4	++	++
<b>008</b>	1.63	1.65	1.67		1.63	2.10	1.3	++	++
<b>009</b>	1.70	1.60	1.65		1.70	2.80	1.6	++	++
<b>010</b>	1.70	1.64	1.65		1.70	0.74	-2.3	++	+++
<b>014</b>	3.30	3.38	3.44		3.30	2.20	-1.5	++	++
<b>019</b>	9.40	8.85	9.25		9.40	11.00	1.2	++	++
<b>020</b>	13.00	13.65	12.82		13.00	10.00	-1.3	++	++
<b>021</b>	13.00	13.03	13.34		13.00	13.00	-1.0	++	++
<b>022</b>	15.60	15.85	15.38		15.60	13.00	-1.2	++	++
<b>024</b>	16.00	13.90	16.00		16.00	9.70	-1.7	++	++
<b>027</b>	28.00	25.76	27.73		28.00	16.00	-1.8	++	++
<b>031</b>	100.00	112.72	99.77		100.00	110.00	1.1	+	+

<sup>a</sup> Definition of the activity scale is given by the Catalyst program and described in the text.

Table 5

The actual (Act) and predicted (Pred) IC<sub>50</sub> by the best CoMFA and CoMSIA models and by the Hypo1 hypothesis for the test set inhibitors are compared

Inh.	Act IC <sub>50</sub> (μM)	CoMFA	CoMSIA H+D+S+E	Catalyst Pharmacophore Hypothesis Hypo1				
				Pred IC <sub>50</sub> (μM)	Pred IC <sub>50</sub> (μM)	Act IC <sub>50</sub> (μM)	Pred IC <sub>50</sub> (μM)	Error
<b>007</b>	1.40		0.77	1.36		1.40	1.80	1.3
<b>011</b>	2.10		1.87	0.54		2.10	1.40	-1.5
<b>013</b>	3.28		4.00	3.58		3.28	3.30	1.0
<b>015</b>	5.30		2.17	5.22		5.30	3.60	-1.5
<b>016</b>	7.00		6.50	6.28		7.00	16.00	2.2
<b>017</b>	7.00		10.16	9.51		7.00	13.00	1.9
<b>023</b>	16.00		6.34	5.73		16.00	3.80	-4.2
<b>025</b>	18.00		2.32	4.63		18.00	5.90	-3.1
<b>026</b>	18.00		11.00	7.57		18.00	13.00	-1.4
<b>029</b>	41.00		12.53	15.38		41.00	15.00	-2.8

<sup>a</sup> Definition of the activity scale is given by the Catalyst program and described in the text.

However, no conflict is found in the activity scale labeled between the actual and predicted IC<sub>50</sub> for the test set inhibitors by the Hypo1 hypothesis as shown in Table 5. A linear regression of the predicted against actual IC<sub>50</sub> for the training

set inhibitors yields a regression coefficient of 0.95, revealing the feasibility of using the structural features selected by the stepwise CoMSIA to construct a pharmacophore hypothesis. The same linear regression procedure on the test set inhibitors

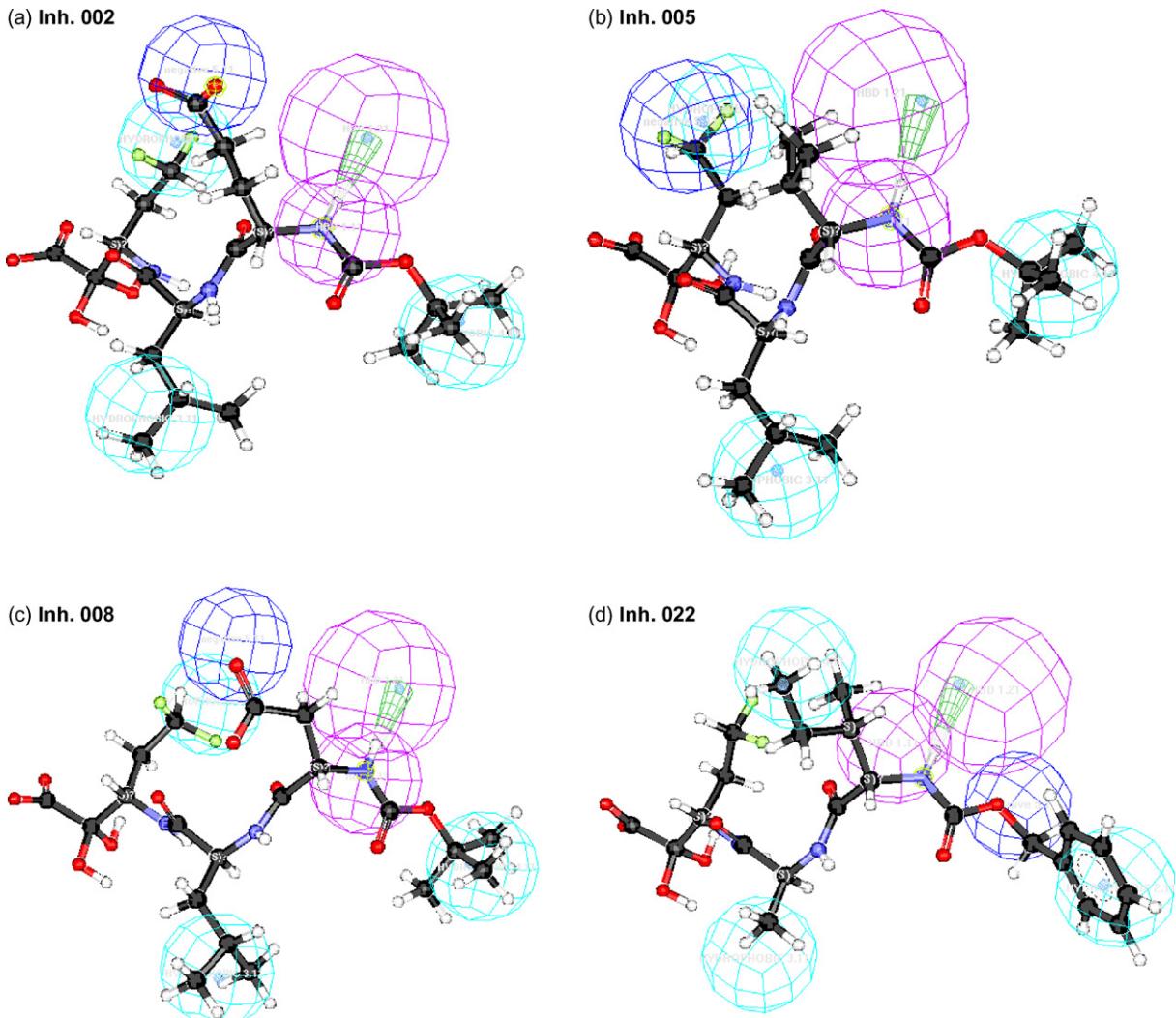


Fig. 3. Mapping of the Hypo1 hypothesis onto the structures of four truncated ketoacid inhibitors: (a) 002, (b) 005, (c) 008, and (d) 022 (Table 1) selected from the training set. The pharmacophore features are color coded as follows: cyan spheres for hydrophobic (H), violet spheres for hydrogen-bond donor (D), and blue spheres for negative charge (E).

yields a regression coefficient of 0.77. To examine the effectiveness of Hypo1 hypothesis constructed, the structures of some highly potent inhibitors of the training set namely, **002**, **005**, **008**, and **022** are mapped onto the Hypo1 features and presented in Fig. 3a–d, respectively. The backbone NH group of all the four structures at position P3 is mapped correctly with the hydrogen-bond donor feature (displayed by violet spheres) of Hypo1 hypothesis (Fig. 3a–d). There is also a negative charge feature (displayed by blue spheres) identified by the Hypo1 hypothesis around position P3 which is best mapped

with structures of **002** and **008** (Fig. 3a and c). This not only correctly reflects the fact that both inhibitors carry a negative charge group at position P3 (Table 1) but also agrees with the red CoMFA contours obtained at the same position (Fig. 2a). The favor regions for hydrophobic interaction around position P2 and P3 are identified by the Hypo1 hypothesis as cyan spheres which agree with the orange contours given by CoMSIA at the same positions (Fig. 2b). Note that mutation of Leu to Ala at position P2 for **022** (Table 1) causes a bad mapping of the corresponding structure onto cyan spheres, the

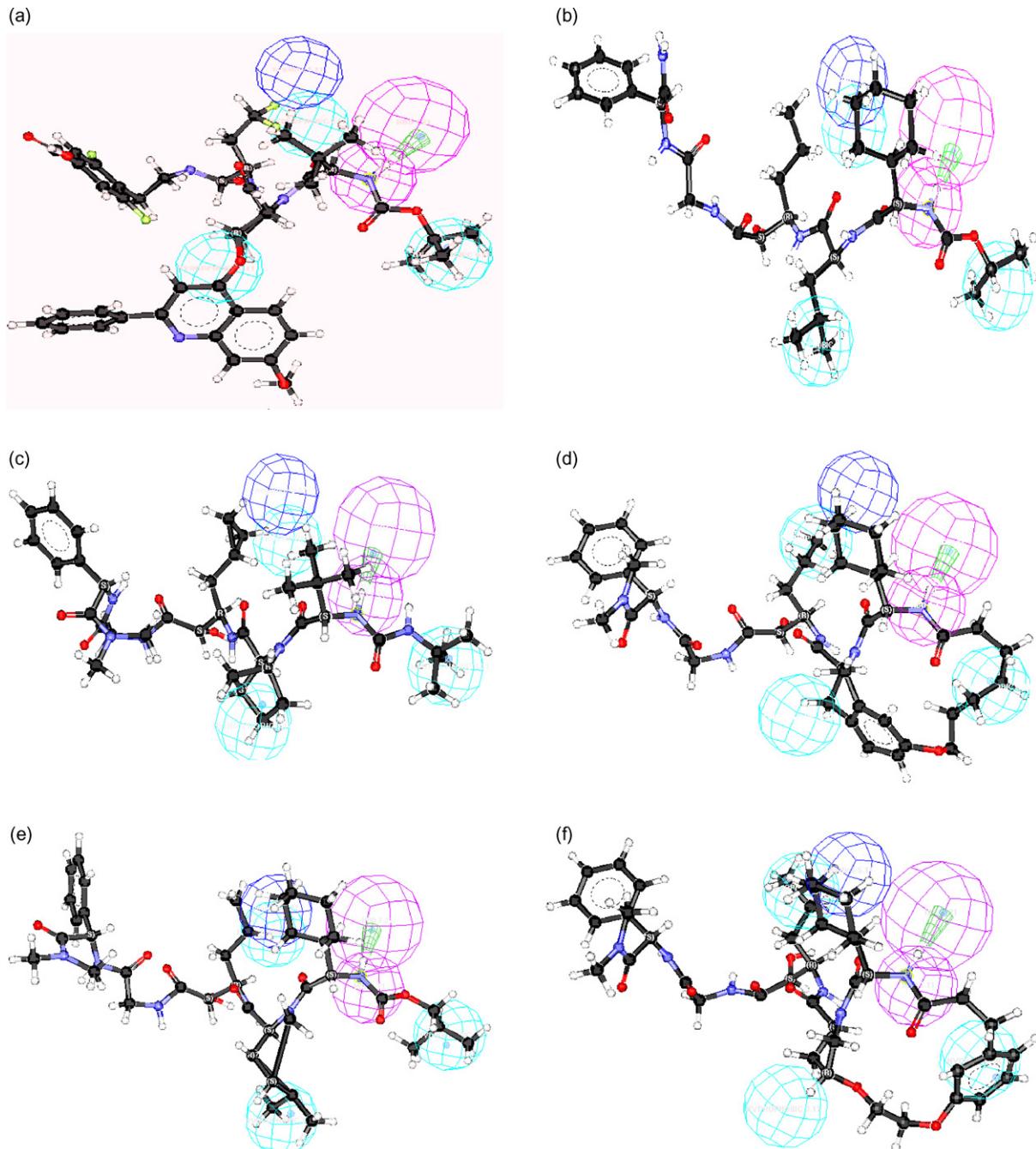


Fig. 4. Mapping of the Hypo1 hypothesis onto the structures of some highly potent inhibitors, namely: (a) inhibitor 7 of IC<sub>50</sub> 0.054 μM [45], (b) inhibitor **43** of K<sub>i</sub>\* 0.066 μM [46], (c) inhibitor **37** of K<sub>i</sub>\* 0.010 μM [24], (d) inhibitor **23** of K<sub>i</sub>\* 0.021 μM [47], (e) inhibitor **19** of K<sub>i</sub>\* 0.010 μM [30] and (f) inhibitor **38** of K<sub>i</sub>\* 25 nM [48] described in the literature. The pharmacophore features are color coded as those described in Fig. 3.

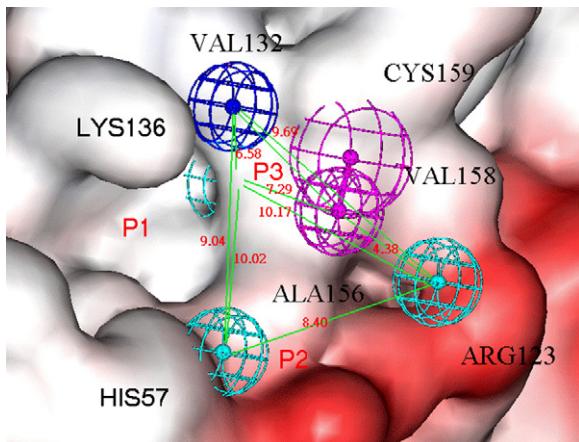


Fig. 5. Mapping of the Hypo1 hypothesis (the distances between pharmacophore features are marked) generated onto the molecular surface of HCV NS3 active site.

hydrophobic feature, displayed by the Hypo1 hypothesis (Fig. 3d) which also results in decrease of activity for the inhibitor (Table 1).

Mapping of GOLD docked structures of some highly potent inhibitors aimed at position P2 or P3 synthesized by several groups onto the Hypo1 hypothesis is also conducted. As shown in Fig. 4a, both NEG and HD features of inhibitor **007** with IC<sub>50</sub> of 54 nM measured and having the 4,4-difluoroaminobutyric acid (difluoroAbu) phenethylamides as P1-P1' and quinolyloxyprolines as P2 fragments, [45] are correctly mapped onto the Hypo1 hypothesis. The difluoroAbu group of this compound is mapped onto blue while the backbone amino group is mapped onto violet spheres (Fig. 4a) of position P3 of the Hypo1 hypothesis. The backbone HD and NEG features of inhibitor **43** of K<sub>i</sub><sup>\*</sup> 0.066 μM measured and characterized by a cyclohexylglycine at P3, a leucine at

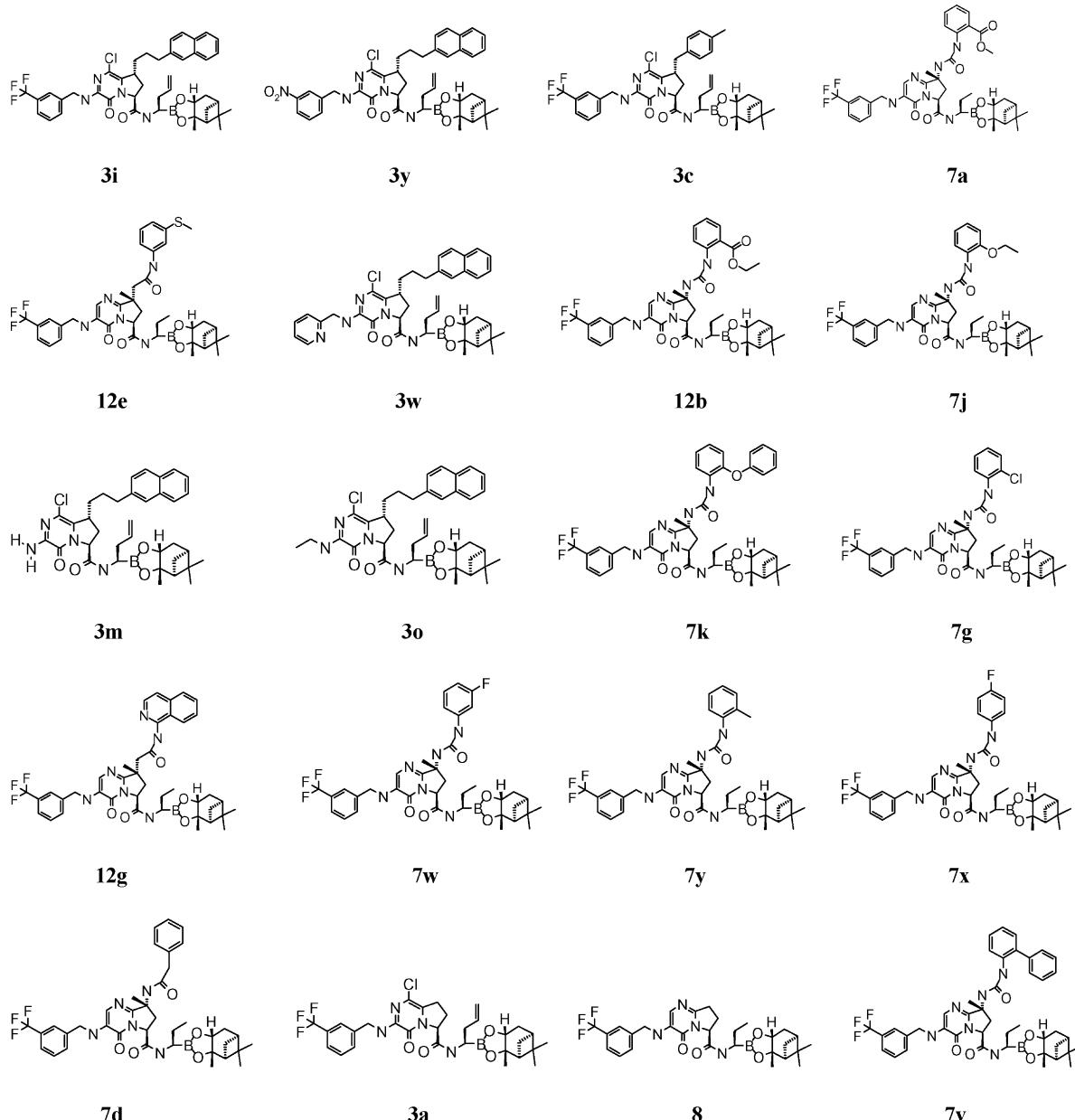


Fig. 6. Structures of the 20 boronic acid based inhibitors studied.

P2, and a phenyl glycine at P2' position [46] is also mapped correctly onto the corresponding P3 features of Hypo1 hypothesis (Fig. 4b). The P3 features of other highly potent inhibitors namely, inhibitor **37** of  $K_i^*$  0.010  $\mu\text{M}$  measured and characterized by a *tert*-butyl urea at P3, 2-aza-bicyclo [2.2.1]-heptane carboxylic acid at P2, and phenylglycine dimethylamide group at P2'; [24] inhibitor **23** of  $K_i^*$  0.021  $\mu\text{M}$  measured and where the *tert*-butyl urea capping group at P3 is cyclized with the 1,2,3,4-tetrahydroisoquinoline-3-carboxyamid moiety at P2 to make the inhibitor as a macrocyclic compound; [47] inhibitor **19** of  $K_i^*$  0.010  $\mu\text{M}$  measured and characterized by a P2' dimethylcarboxamide cap, P2 2,2-dimethylcyclopropyl proline, and P3 *tert*-leucine; [30] and inhibitor **38**, a yet another macrocyclic compound of  $K_i^*$  0.010  $\mu\text{M}$  measured and cyclized with a four-carbon linker between P2 moiety and P3 capping group, [48] are all correctly mapped onto the corresponding features of Hypo1

hypothesis as shown in Fig. 4c–f, respectively. Except the two macrocyclic inhibitors, the P2 groups of all these highly potent inhibitors are also mapped correctly onto the hydrophobic features (cyan spheres) of the Hypo1 hypothesis (Fig. 4a–d). Note that all these six inhibitors mapped are much more potent than the four inhibitors mapped with the same hypothesis and displayed in Fig. 3a–d, respectively. The P2 and P3 structural features of these six inhibitors (Fig. 4a–d) are substantially modified from those of the same positions of the four inhibitors (Fig. 3a–d). By mapping the pharmacophore features of Hypo1 hypothesis onto the molecular surface of NS3 active site as shown in Fig. 5, we find that the binding pocket of position P3 is large and versatile because it is surrounded by Lys136, Val132, and Cys159 while that of P2 appears to be buried somewhat. Based on this mapping (Fig. 5), we believe that there are still rooms around P3 for designing even more potent inhibitors for HCV NS3 protease.

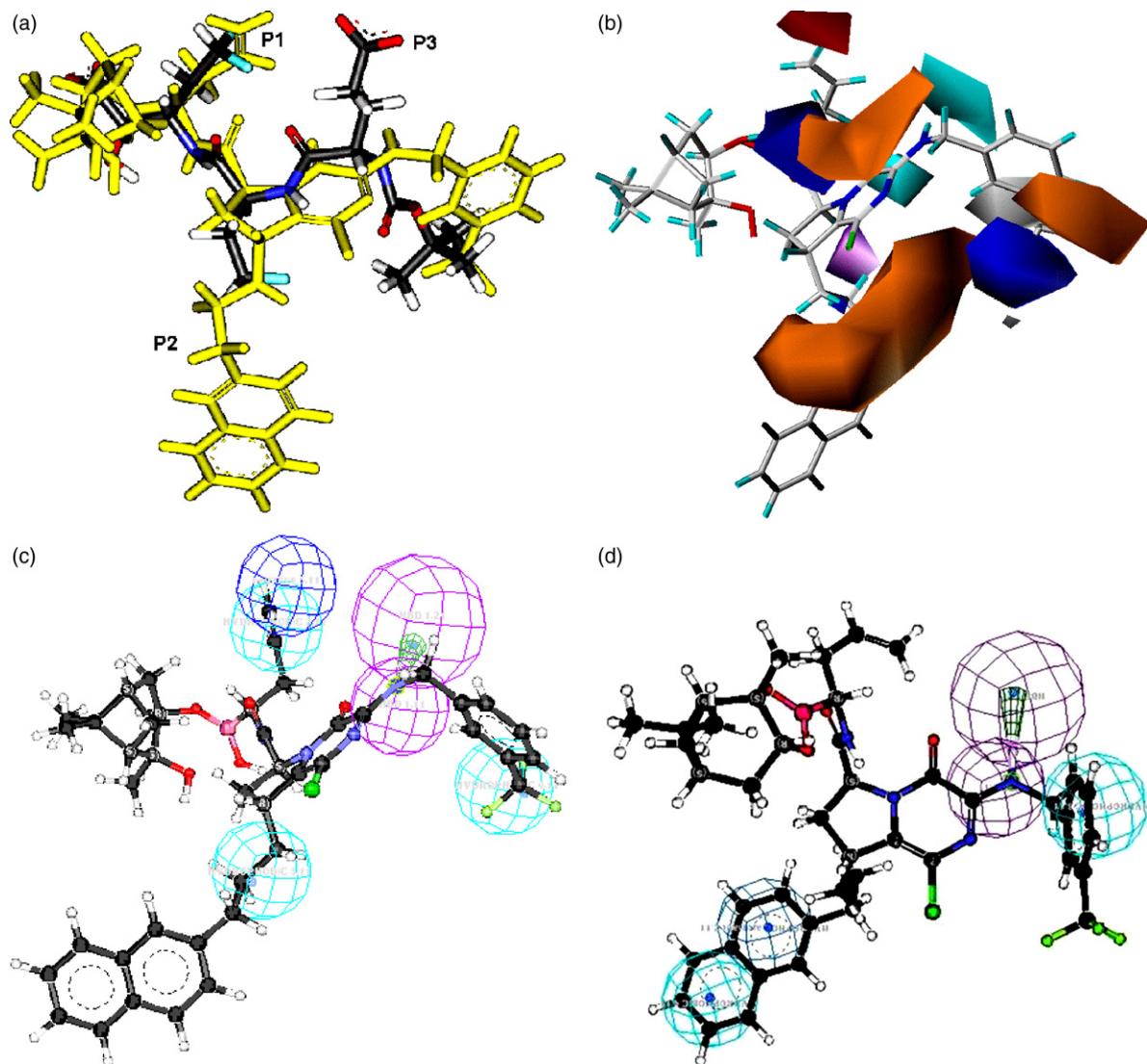


Fig. 7. (a) Superposition of the structure of boronic acid inhibitor **3i** (Table 6 and Fig. 6) with that of truncated ketoacid inhibitor **001** (Table 1). (b) Mapping of the CoMSIA contours described in Fig. 2 (b) onto the structure of boronic acid inhibitor **3i** (Table 6 and Fig. 6). (c) Mapping of the Hypo1 hypothesis described in Fig. 3 onto the structure of boronic acid inhibitor **3i** (Table 6 and Fig. 6). (d) Mapping of the Hypo2 hypothesis generated by the structures of 20 boronic acid inhibitors (Table 6 and Fig. 6) onto the structure of boronic acid inhibitor **3i** (Table 6 and Fig. 6).

A pharmacophore model for 20 boronic acid based pyrimidinone [25] and pyrazinone [33] inhibitors (Fig. 6) are also constructed using the Catalyst program [36]. The IC<sub>50</sub> measured for these boronic acid based pyrimidinone [25] and pyrazinone [33] inhibitors was 20–30,000 nM. The structures of these 20 boronic acid based inhibitors are characterized by a long hydrophobic side chain attached at position P2 (Fig. 6). The pharmacophore model is built using the Catalyst/ConFirm and Catalyst/Hypogen modules. The pharmacophore features selected for constructing a hypothesis are HD, HY, and HR. The Config cost obtained is 14.96 while Unc obtained is 2.0. With a cost difference of 52.33 obtained between the null and fixed one, and a cost difference of 42.81 obtained between the null and total one, the top hypothesis generated for this series of inhibitors is designated as Hypo2 (Table 3). The predicted IC<sub>50</sub> given by the Hypo2 hypothesis for all the 20 inhibitors are listed and compared in Table 6. Apparently, the prediction accuracy given by the Hypo2 hypothesis is 80% because there are four conflicts between the actual and predicted activities obtained (Table 6). A linear regression of predicted versus actual IC<sub>50</sub> gives a regression coefficient of 0.65 (Table 6). The potency of some of these inhibitors are high even the side chain at position P3 is deprived as shown in Fig. 7a where structure of inhibitor **3i** of IC<sub>50</sub> 20 nM measured (Table 6) is superimposed on that of truncated ketoacid inhibitor **001** (Table 1). The boronic acid, P1 allyl, and backbone amino group near the bicyclic pyrazinone ring of **3i** are all mapped correctly onto the CoMSIA blue, orange, and cyan contours (Fig. 7b). However, the big and long hydrophobic naphthylpropyl side chain of **3i** at position P2 is only partially mapped by the P2 CoMSIA orange contours

Table 6

The actual (Act) and predicted (Pred) IC<sub>50</sub> of 20 boronic acid based inhibitors given by the Hypo2 hypothesis are compared

Inh.	Catalyst	Pharmacophore	Hypothesis	Hypo2	
	Act IC <sub>50</sub> (nM)	Pred IC <sub>50</sub> (nM)	Error	Act activity scale <sup>a</sup>	Pred activity scale <sup>a</sup>
<b>3i</b>	20	19	-1.1	++	++
<b>3y</b>	40	50	1.3	++	++
<b>3c</b>	60	49	-1.2	++	++
<b>7a</b>	80	57	-1.4	++	++
<b>12e</b>	90	160	1.8	++	+
<b>3w</b>	90	110	1.2	++	+
<b>12b</b>	100	360	3.6	+	+
<b>7j</b>	160	93	-1.7	+	++
<b>3m</b>	190	170	-1.1	+	+
<b>3o</b>	230	280	1.2	+	+
<b>7k</b>	300	490	1.6	+	+
<b>7g</b>	370	390	1.1	+	+
<b>12g</b>	440	570	1.3	+	-
<b>7w</b>	500	420	-1.2	+	+
<b>7y</b>	780	950	1.2	-	-
<b>7x</b>	1030	960	-1.1	-	-
<b>7d</b>	2400	4200	1.8	-	-
<b>3a</b>	4600	4500	-1.0	-	-
<b>8</b>	7000	3900	-1.8	-	-
<b>7v</b>	30000	4200	-7.2	-	-

<sup>a</sup> Definition of the activity scale is given by the Catalyst program and described in the text.

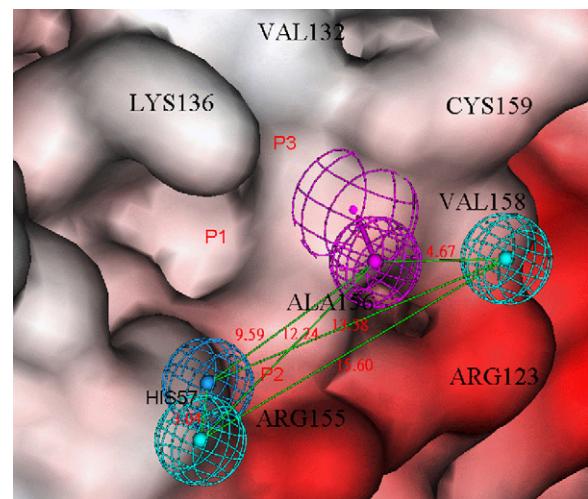


Fig. 8. Mapping of the Hypo2 hypothesis (the distances between pharmacophore features are marked) generated onto the molecular surface of HCV NS3 active site.

(Fig. 7b). Similarly, the P2 cyan spheres of Hypo1 hypothesis are badly mapped by the long P2 hydrophobic side chain of **3i** as shown in Fig. 7c. On the other hand, the P2 cyan spheres of Hypo2 hypothesis is correctly mapped by the P2 long hydrophobic side chain of **3i** as shown in Fig. 7d. The backbone amino group of this structure is also correctly mapped by the violet spheres of Hypo2 hypothesis (Fig. 7d). A mapping of Hypo2 hypothesis onto the molecular surface of NS3 active site shows that while Arg155 is nearby there is no hydrophobic residue around the hypothesis feature identified (Fig. 8). It has been addressed in literature [49] that conformational changes may be induced by binding with an inhibitor such as **007** so that a salt bridge is formed between side chains of Arg155 and Asp168 which would facilitate an interaction between the quanidine group of Arg155 and the P2 methoxy quinoline group of the inhibitor [49]. However, this does not reflect the existence of the hydrophobic feature identified by the Hypo2 hypothesis shown in Figs. 7d and 8 around Arg155.

#### 4. Conclusion

We identified in this work the P2 and P3 pharmacophore features of HCV NS3 active site by using several 3D QSAR techniques and the structures of two series of inhibitors. We found that structures of both truncated ketoacid [31,32] and boronic acid inhibitors [25,33] can be mapped onto the same P3 pharmacophore features built separately from each structure set though there is significant difference in structural features of the two series at this position. The P3 binding pocket is big and versatile as revealed by a mapping of P3 pharmacophore features onto the molecular surface of the pocket. We expect that more potent depeptidized inhibitors aiming at this position may still be designed. We have also shown that structural features selected by a stepwise CoMSIA can be plugged into the Catalyst program for an automatic generation of conformations for constructing the pharmacophore models

for the inhibitors studied. The stepwise CoMSIA sweeps all single or possible combinations of structural features involved in the binding process and identifies only some of them to be statistically important in the pharmacophore construction process. As characterized by nearby Ala156, the P2 binding pocket is presumably hydrophobic though it is somewhat under explored in the literature. The P2 binding pocket was envisaged by us as somewhat buried as revealed by mapping of the pharmacophore features built specifically from some boronic acid inhibitors [25,33] carrying the extended P2 hydrophobic side chains on their structures. The depth of this buried P2 binding pocket was probably the same length as the P2 naphthylpropyl side chain of the most potent boronic acid inhibitors mapped. The activity of these boronic acid inhibitors was reduced substantially if the long hydrophobic side chain was deprived away from position P2 as that on inhibitor **3a** and **8** (Fig. 6). We expect that more potent depeptidized HCV NS3 inhibitors may be designed if both pharmacophore features mapped P2 and P3 binding pockets are taken into accounted.

## Acknowledgment

This work is supported in part by a grant from the National Science Council (NSC95-2313-B007-002), Taiwan. The GOLD, SYBYL CoMFA, CoMSIA and PLS, plus the Catalyst 4.9 studies were conducted at the National Center for High Performance Computing, Taiwan.

## References

- [1] L.B. Seeff, J.H. Hoofnagle, Appendix: the National Institute of Health Consensus Development Conference management of hepatitis C, *Clin. Liver Dis.* 7 (2002) 261–287.
- [2] J. McHutchison, K. Patel, Future therapy of hepatitis C, *Hepatology* 36 (2002) S245–S252.
- [3] D.B. Strader, T. Wright, D.L. Thomas, L.B. Seeff, Diagnosis, management, and treatment of hepatitis C, *Hepatology* 39 (2004) 1147–1171.
- [4] N. Kato, M. Hijikata, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi, K. Shimotohno, Sequence diversity of hepatitis C viral genomes, *Mol. Biol. Med.* 7 (1990) 495–501.
- [5] R.H. Miller, R.H. Purcell, Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups, *Proc. Natl. Acad. Sci. U.S.A.* 87 (1990) 2057–2061.
- [6] J. Bukh, R.H. Miller, R.H. Purcell, Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. [Review], *Semin. Liver Dis.* 15 (1995) 41–63.
- [7] P. Simmonds, Variability of hepatitis C virus. [Review], *Hepatology* 21 (1995) 570–583.
- [8] B.D. Lindenbach, C.M. Rice, Flaviviridae; the viruses and their replication, in: D.M. Knipe, P.M. Howley, D.E. Griffin (Eds.), *Fields Virology*, fourth ed., Lippincott Williams & Wilkins, Philadelphia, PA, 2001, pp. 991–1041.
- [9] A.A. Kolykhalov, K. Mihalik, S.M. Feinstone, C.M. Rice, Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo, *J. Virol.* 74 (2000) 2046–2051.
- [10] R.L. Bartenschlager, J.M. Ablorn-Laake, H. Jacobsen, Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 junctions, *J. Virol.* 67 (1993) 3835–3844.
- [11] A. Grakoui, D.W. McCourt, C. Wychowski, S.M. Feinstone, C.M. Rice, Characterization of the hepatitis C virus-encoded serine protease: determination of proteinase-dependent polyprotein cleavage sites, *J. Virol.* 67 (1993) 2832–2843.
- [12] M. Hijikata, H. Mizushima, Y. Tanji, Y. Komoda, Y. Hirowatari, T. Akagi, N. Kato, K. Kimura, K. Shimotohno, Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus, *Proc. Natl. Acad. Sci. U.S.A.* 90 (1993) 10773–10777.
- [13] J.L. Kim, K.A. Morgenstern, C. Lin, T. Fox, M.D. Dwyer, J.A. Landro, S.P. Chambers, W. Markland, C.A. Lepre, E.T. O'Malley, S.L. Harbeson, C.M. Rice, M.A. Murcko, P.R. Caron, J.A. Thomson, Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide, *Cell* 87 (1996) 343–355.
- [14] C. Lin, J.A. Thomson, C.M. Rice, A central region in the hepatitis C virus NS4A protein allows formation of an active NS3-NS4A serine proteinase complex in vivo and in vitro, *J. Virol.* 69 (1995) 4373–4380.
- [15] L. Tomei, C. Failla, E. Santolini, R. De Francesco, N. La Monica, NS3 is a serine protease required for processing of hepatitis C virus polyprotein, *J. Virol.* 67 (1993) 4017–4026.
- [16] K. Lin, R.B. Perni, A.D. Kwong, C. Lin, VX-950, a novel hepatitis C virus (HCV) NS3-4A protease inhibitor, exhibits potent antiviral activities in HCV replicon cells, *Antimicrob. Agent Chemother.* 50 (2006) 1813–1882.
- [17] C. Steinkuhler, U. Koch, F. Narjes, V.G. Matassa, Hepatitis C virus serine protease inhibitors: current progress and future challenges, *Curr. Med. Chem.* 8 (2001) 919–932.
- [18] I. Schechter, A. Berger, On the size of the active site in proteases I. papain, *Biochem. Biophys. Res. Commun.* 27 (1967) 157–162.
- [19] E. Pizzi, A. Tramontano, L. Tomei, N. La Monica, C. Failla, M. Sardana, T. Wood, R. De Francesco, Molecular model of the specificity pocket of the hepatitis C virus protease: implications for substrate recognition, *Proc. Natl. Acad. Sci. U.S.A.* 91 (1994) 888–892.
- [20] S. Venkatraman, F.G. Njoroge, V.M. Girijavallabhan, V.S. Madison, N.H. Yao, A.J. Prongay, N. Butkiewicz, J. Pichardo, Design and synthesis of depeptidized macrocyclic inhibitors of hepatitis C NS3-4A protease using structure-based drug design, *J. Med. Chem.* 48 (2005) 5088–5091.
- [21] F. Narjes, U. Koch, C. Steinkuhler, Recent developments in the discovery of hepatitis C virus serine protease inhibitors-towards a new class of antiviral agents, *Expert Opin. Investig. Drugs* 12 (2003) 153–163.
- [22] S. Di Marco, M. Rizzi, C. Volpari, M.A. Walsh, F. Narjes, S. Colarusso, R. De Francesco, V.G. Matassa, F. Sollazzo, S. Colarusso, R. De Francesco, V.G. Matassa, M. Sollazzo, Inhibition of the hepatitis C virus NS3/4A protease, *J. Biol. Chem.* 275 (2000) 7152–7157.
- [23] S.J. Archer, D.M. Camac, Z.J. Wu, N.A. Farrow, P.J. Domaille, Z.R. Wasserman, M. Bukhtiyarova, C. Rizzo, S. Jagannathan, L.J. Mersinger, C.A. Kettner, Hepatitis C virus NS3 protease requires its NS4A cofactor peptide for optimal binding of a boronic acid inhibitor as shown by NMR, *Chem. Biol.* 9 (2002) 79–92.
- [24] S. Venkatraman, F.G. Njoroge, W. Wu, V. Girijavallabhan, A.J. Prongay, N. Butkiewicz, J. Pichardo, Novel inhibitors of hepatitis C NS3-NS4A serine protease derived from 2-aza-bicyclo [2.2.1] heptane-3-carboxylic acid, *Bioorg. Med. Chem. Lett.* 16 (2006) 1628–1632.
- [25] P.W. Glunz, B.D. Douty, C.P. Decicco, Design and synthesis of bicyclic pyrimidinone-based HCV NS3 protease inhibitors, *Bioorg. Med. Chem. Lett.* 13 (2003) 785–788.
- [26] C. Lin, A.D. Kwong, R.B. Perni, Discovery and development of VX-950, a novel, covalent, and versatile inhibitor of hepatitis C virus NS3, 4A serine protease, *Infect. Disord. Drug Targets* 6 (2006) 3–16.
- [27] P.O. Johansson, M. Back, I. Kvarnstrom, K. Jansson, L. Vrang, E. Hamelink, A. Hallberg, A. Rosenquist, B. Samuelsson, Potent inhibitors of the hepatitis C virus NS3 protease: use of a novel P2 cyclopentane-derived template, *Bioorg. Med. Chem. Lett.* 14 (2006) 5136–5151.
- [28] D. Lamarre, P.C. Anderson, M. Bailey, P. Beaulieu, G. Bolger, P. Bonneau, M. Bos, D.R. Cameron, M. Cartier, M.G. Cordingley, A.M. Faucher, N. Goudreau, S.H. Kawai, G. Kukolj, L. Laqace, S.R. LaPlante, H. Narjes, M.A. Poupart, J. Rancourt, R.E. Sentjens, R. St. George, B. Simoneau, G. Steinmann, D. Thibeault, Y.S. Tsantrizos, S.M. Weldon, C.L. Yong, M. Llinas-Brunet, A NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus, *Nature* 426 (2003) 186–189.
- [29] A. Arasappan, F.G. Njoroge, K.X. Chen, S. Venkatraman, T.N. Parekh, H. Gu, J. Pichardo, N. Butkiewicz, A. Prongay, V. Madison, V. Girijavallabhan, Inhibition of hepatitis C virus serine protease by a macrocyclic inhibitor, *Antimicrob. Agent Chemother.* 50 (2006) 1883–1889.

- labhan, P2-P4 Macrocyclic inhibitor of hepatitis C virus NS3-4A serine protease, *Bioorg. Med. Chem. Lett.* 16 (2006) 3960–3965.
- [30] S.L. Bogen, A. Arasappan, F. Bennett, K. Chen, E. Jao, Y.T. Liu, R.G. Lovey, S. Venkatraman, W. Pan, T. Parekh, R.E. Pike, S. Ruan, R. Liu, B. Baroudy, S. Agrawal, R. Chase, P. Ingravallo, J. Pichardo, A. Prongay, J.M. Brisson, T.Y. Hsieh, K.C. Cheng, S.J. Kemp, O.E. Levy, M. Lim-WILBY, S.Y. Tamura, A.K. Saksena, V. Girijavallabhan, F.G. Njoroge, Discovery of SCH446211 (SCH6): a new ketoamide inhibitor of the HCV NS3 serine protease and HCV subgenomic RNA replication, *J. Med. Chem.* 49 (2006) 2750–2757.
- [31] E. Nizi, U. Koch, S. Ponzi, V.G. Matassa, C. Gardelli, Capped dipeptide  $\alpha$ -ketoacid inhibitors of the HCV NS3 protease, *Bioorg. Med. Chem.* 12 (2002) 3325–3328.
- [32] S. Colarusso, B. Gerlach, U. Koch, E. Muraglia, I. Conte, I. Stansfield, V.G. Matassa, F. Narjes, Evolution, synthesis and SAR of tripeptide  $\alpha$ -ketoacid inhibitors of the hepatitis C virus NS3/NS4A serine protease, *Bioorg. Med. Chem. Lett.* 12 (2002) 705–708.
- [33] X. Zhang, A.C. Schmitt, W. Jiang, Z. Wasserman, C.P. Decicco, Design and synthesis of potent, non-peptide inhibitors of HCV NS3 protease, *Bioorg. Med. Chem. Lett.* 13 (2003) 1157–1160.
- [34] R.D. Cramer 3rd, D.E. Patterson, J.D. Bunce, Recent advances in comparative molecular field analysis (CoMFA), *Prog. Clin. Biol. Res.* 291 (1989) 161–165.
- [35] G. Klebe, U. Abraham, T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity, *J. Med. Chem.* 37 (1994) 4130–4146.
- [36] Catalyst, Version 4.9 ed., Accelrys Inc. (previously known as Molecular Simalations Inc.), San Diego, CA.
- [37] S. Durdagi, A. Kapou, T. Kourouli, T. Andreou, S.P. Nikas, V.R. Nahmias, D.P. Papahatjis, M.G. Papadopoulos, T. Mavromoustakos, The application of 3D-QSAR studies for novel cannabinoid ligands substituted at the C1' position of the alkyl side chain on the structural requirements for binding to cannabinoid receptors CB1 and CB2, *J. Med. Chem.* 50 (2007) 2875–2885.
- [38] A. Kapou, N.P. Benetis, N. Avlonitis, T. Calogeropoulou, M. Koufaki, E. Scoulica, S.S. Nikolopoulos, T. Mavromoustakos, 3D-Quantitative structure–activity relationships of synthetic antileishmanial ring-substituted ether phospholipids, *Bioorg. Med. Chem.* 15 (2007) 1252–1265.
- [39] Y.Y. Ke, T.H. Lin, Modeling the ligand-receptor interaction for a series of inhibitors of the capsid protein of Eroterovirus 71 using several three-dimensional quantitative structure–activity relationship techniques, *J. Med. Chem.* 49 (2006) 4517–4525.
- [40] SYBYL, Version 6.9.1, Tripos Associates, Inc., St. Louis.
- [41] GOLD is distributed by the Cambridge Crystallographic Data Center, Cambridge, UK, [www.ccdc.ac.uk/prods/gold.html](http://www.ccdc.ac.uk/prods/gold.html).
- [42] M. Mobil, R.J. Abraham, Quantum vs. Classical models of the nitro group for proton chemical shift calculation and conformational analysis, *J. Comput. Chem.* 26 (2005) 389–398.
- [43] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, R.K. Robins, Atomic physicochemical parameters for three dimensional structure directed quantitative structure–activity relationships 4 Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring antibiotics, *J. Chem. Inf. Comput. Sci.* 29 (1989) 163–172.
- [44] G. Klebe, The use of composite crystal-field environments in molecular recognition and the de novel design of protein ligands, *J. Mol. Biol.* 237 (1994) 212–235.
- [45] F. Orvieto, U. Koch, V.G. Matassa, E. Muraglia, Novel potent phenethylamide inhibitors of the hepatitis C virus (HCV) NS3 protease: probing the role of P2 aryloxyprolines with hybrid structure, *Bioorg. Med. Chem. Lett.* 13 (2003) 2745–2748.
- [46] A. Arasappan, F.G. Njoroge, T.Y. Chan, F. Bennett, S.L. Bogen, K. Chen, H. Gu, L. Hong, E. Jao, Y.T. Liu, R.G. Lovey, T. Parekh, R.E. Pike, P. Pinto, B. Santhanam, S. Venkatraman, H. Vaccaro, H. Wang, X. Yang, Z. Zhu, B. McKittrick, A.K. Saksena, V. Girijavallabhan, J. Pichardo, N. Butkiewicz, R. Ingram, B. Malcolm, A. Prongay, N. Yao, B. Marten, V. Madison, S. Kemp, O. Levy, M. Lim-Wilby, S. Tamura, A.K. Ganguly, Hepatitis C virus NS3-4A serine protease inhibitors: SAR of P2' moiety with improved potency, *Bioorg. Med. Chem. Lett.* 15 (2005) 4180–4184.
- [47] K.X. Chen, F.G. Njoroge, J. Pichardo, A. Prongay, N. Butkiewicz, N. Yao, V. Madison, V. Girijavallabhan, Potent 7-hydroxy-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid-based macrocyclic inhibitors of hepatitis C virus NS3 protease, *J. Med. Chem.* 49 (2006) 567–574.
- [48] A. Arasappan, F.G. Njoroge, K.X. Chen, S. Venkatraman, T.N. Parekh, H. Gu, J. Pichardo, N. Butkiewicz, A. Prongay, V. Madison, V. Girijavallabhan, P2-P4 Macrocyclic inhibitors of hepatitis C virus NS3-4A serine protease, *Bioorg. Med. Chem. Lett.* 16 (2006) 3960–3965.
- [49] Y.S. Tsantrizos, G. Bolger, P. Bonneau, D.R. Cameron, N. Goudreau, G. Kukolj, S.R. LaPlante, M. Llinàs-Brunet, H. Nar, D. Lamarre, Macro cyclic inhibitors of the NS3 protease as potential therapeutic agents of hepatitis C virus infection, *Angew. Chem. Int. Ed.* 42 (2003) 1356–1360.