

# Possible allosteric interactions of monoindazole-substituted P2 cyclic urea analogues with wild-type and mutant HIV-1 protease

Rajni Garg · Barun Bhatarai

Received: 9 October 2007 / Accepted: 9 March 2008 / Published online: 27 March 2008  
© Springer Science+Business Media B.V. 2008

**Abstract** Our ongoing efforts to understand the difference in the binding pattern of HIV-1 protease inhibitor (HIVPI) with the wild-type and mutant HIV-1 protease (HIVPR) and to provide mechanistic insight are continued further. We report here the results of a recent quantitative structure–activity relationship (QSAR) study on monoindazole-substituted P2 analogues of cyclic urea HIVPIs. The QSAR models revealed an inverted parabolic relationship between biological activity and calculated molar refractivity (CMR). That is, biological activity first decreases with increase in CMR and at a certain minimum point (inversion point) it suddenly changes and increases with further increase in CMR. CMR is a measure of volume-dependent-polarizability and is an indication of the polar interactions between ligand and receptor. The results seem to be best rationalized by larger molecules inducing a change in a receptor unit that allows for a new mode of interaction. Similar QSAR models were also observed for the biological activity of these molecules tested against a panel of mutant viruses including mutant strains with single amino acid substitution (I84V), double amino acid substitutions (I84V/V82F), and multiple amino acid changes corresponding to mutations observed in clinical isolates of patients treated with Ritonavir<sup>®</sup>. Interestingly the inversion points for these mutant strains were found larger than for wild-type. The subtle but significant difference in the inversion point indicates change in the shape and size

of the binding pocket. Earlier QSAR studies have shown that the correlation of biological activity with an inverted parabola is an indicative of the ‘allosteric interaction’ of the ligands with the receptor. This report presents a detail analysis of these observations.

**Keywords** HIV-1 · Protease inhibitor · Wild-type HIV-1 protease · Mutant HIV-1 protease · Allosteric interaction · Monoindazole cyclic urea · QSAR

## Introduction

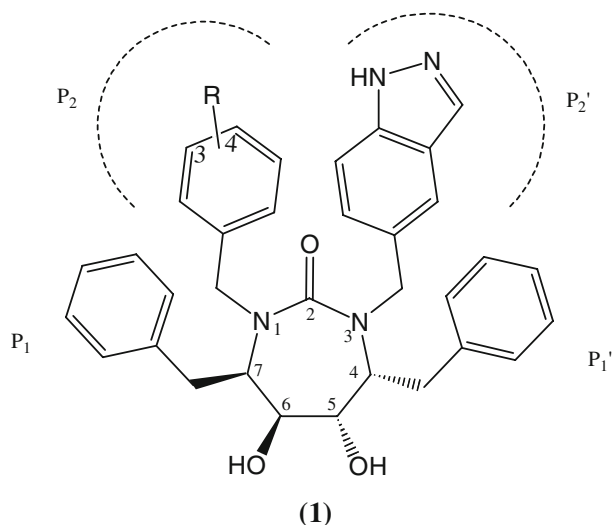
AIDS epidemic has led to recent advancement in the treatment of HIV infection [1]. In particular, the use of HIV-1 protease inhibitors (HIVPIs) in combination with reverse transcriptase (RT) inhibitors has revolutionized care of AIDS patients [1, 2]. However, the therapeutic utility of current drug regimen is often compromised by low oral bioavailability and rapid excretion [3, 4]. In addition, a number of patients have developed resistance to currently available drugs through viral mutations [5]. The successful outcome of future drug-combination-therapy will depend upon the development of new drugs active against wild-type (WT) as well as mutant protease. To expedite development of such HIVPIs, there is a clear need to understand the difference between the binding interactions of WT and the mutant variants with the inhibitors.

Several quantitative computer-aided drug design techniques and approaches are used in drug discovery, design, and development. One such approach is Quantitative Structure–Activity Relationship (QSAR), which has greatly benefited target drug discovery and development process [6]. It correlates biological activity of a molecule with its molecular descriptors and/or physico-chemical parameters.

R. Garg (✉)  
Chemistry and Biochemistry Department, California  
State University, San Marcos, 333S. Twin Oaks valley Road,  
San Marcos, CA 92006, USA  
e-mail: rgarg@csusm.edu; rg004747@yahoo.com

R. Garg · B. Bhatarai  
Chemistry Department, Clarkson University, Potsdam, NY, USA

QSAR models lead to assessment of the specific effects of various types of substituents reducing trial experiments and investments. It helps in identifying possible expensive failures in early stage of drug design and discovery process and eliminates them as outliers. Extensive work has been published using QSAR to predict pharmacokinetic properties such as absorption, distribution, binding mechanism, metabolism, etc. [7–9]. QSAR studies have provided valuable insight in the design and development of several drug targets including HIV-1 protease [10, 11]. Recently, we have been interested in understanding the mechanism of interaction of HIVPIs with its receptor using this technique [12, 13]. Our ongoing efforts to understand the difference in the binding pattern of HIVPI with the WT and mutant HIVPR using QSAR approach and to provide mechanistic insight are continued further. We report here the results of a recent study on *hexahydro-5,6-dihydroxy-1-(R-benzyl)-3-(indazole-5-yl-methyl)-4,7-bis-(phenyl-methyl)-2H-1,3-diazepin-2-one* substituted non-symmetric analogues of cyclic urea (**1**) [14].



The QSAR models revealed an inverted parabolic relationship between biological activity and calculated molar refractivity (CMR) of the molecules. That is, biological activity first decreases with increase in CMR and at a certain minimum point (inversion point) it suddenly changes and increases with further increase in CMR. CMR is a measure of ‘volume-dependent-polarizability’ and an indication of the polar interactions between ligand and receptor. Thus, the results seem to be best rationalized by larger molecules inducing a change in a receptor unit that allows for a new mode of interaction. Similar QSAR models were also observed for the biological activity of these molecules tested against a panel of mutant viruses including mutant strains with single amino acid substitutions (I84V), double amino acid substitutions (I84V/V82F),

and multiple amino acid changes corresponding to mutations observed in clinical isolates of patients treated with Ritonavir<sup>®</sup>. Interestingly the inversion points for these mutant strains were found larger than for wild-type.

Hansch et al. have reported several instances where QSAR based on inverted parabola/bilinear correlation were shown to be related to the possibility of ‘allosteric interaction’ [15]. QSAR was successfully applied and validated for the well-understood allosteric binding interaction of hemoglobin and a change in the mechanism of the reaction was proposed to occur at the inversion point [15]. QSAR was also used to successfully address the mechanism of receptor-ligand interaction despite different changes occurring at more than one position of the ligand [16]. All these studies show that the correlation of biological activity with an inverted parabola may indicate ‘allosteric interaction’ of the ligands with the receptor. Our present study of non-symmetric analogues of cyclic urea (**1**) also unravels dependence of biological activity on inverted parabolic term in QSAR models. The mechanistic insight provided by the correlation in the binding pattern of these ligands is discussed in this report.

## Materials and methods

All the biological data were taken from De Lucca et al. [14]. They reported [14] inhibition of HIVPR ( $K_i$ ) measured by assaying the cleavage of a fluorescent peptide substrate using HPLC. The antiviral potency ( $IC_{90}$ ) of molecules as reported [14] was assessed by measuring their effect on the accumulation of viral RNA transcripts 3 days after infection of MT-2 cells with HIV-1 RF [17].

QSAR studies were performed using the structure–activity data reported [14] for congeneric series of cyclic urea (**1**). In all the QSAR reported here,  $n$  is the number of data points,  $r$  is the correlation coefficient,  $s$  is the standard deviation,  $q$  is the quality of fit and calculated as described by Cramer et al. [18], and the data within parentheses are for the 95% confidence intervals. The QSAR multiple linear regression analyses were executed with the CQSAR program [19].  $ClogP$  is the calculated log of octanol/water partition coefficient and is a measure of hydrophobicity of the molecule [20]. CMR is the calculated molar refractivity for the whole molecule, and is a measure of volume dependent polarizability [21].  $CMR_{R,3}$  is the calculated molar refractivity for the substituents at the third position (meta) of P2 benzyl of cyclic urea (**1**).  $\sigma_1$  refers to the field/inductive effect for the particular substituent [21]. All these parameters were calculated and auto loaded using CQSAR.  $ClogP$  and CMR are normally for the neutral form of partially ionized molecules. An

indicator variable  $I$  was assigned the value of 1 or 0 for special effects that cannot be parameterized. A positive coefficient of  $I$  indicates that moieties with a value of 'one' would be favorable for effective binding and vice-versa. Outliers shown in the tables are congeners that were not included in the derivation of QSAR. A discussion about the specific indicators used and the occurrence of outliers in QSAR is given later.

## Results and discussion

De Lucca et al. [14] reported structure–activity data of *hexahydro-5,6-dihydroxy-1-(R-benzyl)-3-(indazole-5-ylmethyl)-4,7-bis-(phenyl-methyl)-2H-1,3-diazepin-2-one* substituted non-symmetric cyclic urea (**1**) analogues. Different congeners were synthesized by varying 3rd and 4th position substituents on P2 benzyl ring aiming at the S2/S3 pocket of the receptor. We derived several QSAR models not only for different biological endpoints measured for the inhibition of wild-type (WT) HIVPR but also for different types of mutant HIVPRs. After examinations of all the possible parameters, models obtained with the best correlations are reported and discussed.

Activity data for wild-type (WT) virus

Enzyme inhibition activity ( $K_i$ ) [14] (Table 1)

The analysis of enzyme inhibitory activity of analogues of **1** gave QSAR 1a.

$$\begin{aligned} \text{Log}1/K_i = & -2.557(\pm 1.399) \text{CMR} \\ & + 0.072(\pm 0.038) \text{CMR}^2 \\ & - 0.360(\pm 0.096) C \log P \\ & + 0.746(\pm 0.321) \sigma_{1,4} + 35.142(\pm 12.958) \\ n = 31, r = 0.887, r^2 = 0.787, q^2 = 0.690, s = 0.142 \end{aligned} \quad (1a)$$

For CMR, inversion point: 17.812 (17.11–18.19)

Three parameters CMR,  $C \log P$  and  $\sigma_{1,4}$  were found most significant.  $C \log P$  indicates the importance of hydrophobicity of the molecule. Its negative sign indicates that highly hydrophobic groups studied in this series are not good for improving the activity of the molecules. They are all in higher range compared to the optimum value that we observed for other 3-aminoindazole cyclic urea analogues [22].  $\sigma_{1,4}$ , the inductive effect of ligands at the 4th position of benzene ring at P2 position was also found significant. The inductive effect of substituents at this position may be assisting in the polarization of the ligand by withdrawing electron (positive  $\sigma_{1,4}$ ) and help to increase the biological

activity. These physico-chemical parameters were also found significant in our previous studies on other classes of non-peptidic HIVPR inhibitors [12, 13].

An inverted parabola with CMR indicates that at first the enzyme inhibitory activity decreases with the increase in CMR parameter value, but after attaining a minimum called 'inversion point' the activity increases again. The inversion point predicts the optimum value of a parameter for the molecules under study.

Since, the only variation in the series was due to different substituents on 3rd and 4th position of the P2 benzyl ring; we calculated the CMR value of the substituents and obtained QSAR 1b.

$$\begin{aligned} \text{Log}1/K_i = & -0.194(\pm 0.147) \text{CMR}_{R,3} \\ & + 0.064(\pm 0.036) \text{CMR}_{R,3}^2 \\ & + 0.776(\pm 0.324) \sigma_{1,4} \\ & - 0.353(\pm 0.098) C \log P + 12.470(\pm 0.601) \\ n = 31, r = 0.882, r^2 = 0.778, q^2 = 0.683, s = 0.145 \end{aligned} \quad (1b)$$

For  $\text{CMR}_{R,3}$ , inversion point: 1.515 (0.726–1.937)

This QSAR shows that the biological activity of these molecules is significantly dependent on the CMR value of the 3rd position substituents of P2 benzyl. It is important to note that whereas QSAR 1a correlates to the CMR value of the whole molecule, QSAR 1b correlates to the substituents on the P2 benzyl ring that is designed for S2/S3 pocket of the enzyme. Since, the overall parent structure of the molecules under study is same and the only variation was in the P2 benzyl ring with different substitution on 3rd and 4th position it seems highly possible that the change in biological activity corresponds to the change in these substituents (CMR versus  $\text{CMR}_{R,3}$ ,  $r^2 = 0.983$ ). The sufficient variation in the substituents at the 3rd position of the ring resulted in good spread in the parameter value for  $\text{CMR}_{R,3}$ . The substituents at the 4th position of P2 benzyl contain only amino-, methyl- and fluoro-derivatives. Thus, with less variation in the parameters range the parameter  $\text{CMR}_{R,4}$  was not found very significant.

Investigation of Table 1 shows that the range in  $K_i$  data reported [14] is not very wide. Normally, a good spread in parameter value is required for deriving meaningful QSAR models. Standing alone QSAR 1a and 1b may not be meaningful. However, similarity in these models (with others reported later), a good spread in the parameter value for  $\text{CMR}_{R,3}$  and their statistical robustness makes them meaningful. It is to be noted that although QSAR 1a with CMR (and other models discussed later) has large confidence intervals and intercepts, this error does not affect the quality of QSAR model as indicated by high correlation coefficient ( $r^2 = 0.787$ ,  $q^2 = 0.690$ ) and low standard deviation ( $s = 0.142$ ) in QSAR 1a. Furthermore, QSAR 1b

**Table 1** The enzyme inhibition activity ( $K_i$ ), antiviral activity ( $IC_{90}$ ) and physicochemical parameters used for deriving QSAR 1a, 1b, 2a, and 2b

S. no.	R	ClogP	CMR	$I_{BENZ}$	$I_{HARA}$	CMR <sub>R,3</sub>	$\sigma_{1,4}$	Log $1/K_i$			Log $1/IC_{90}$		
								Obsd. <sup>14</sup>		Calcd.	Obsd. <sup>14</sup>		Calcd.
								QSAR 1a	QSAR 1b		QSAR 2a	QSAR 2b	
1 <sup>b,c</sup>	3-Carboxy	6.46	16.86	0	0	0.65	0	10.31	10.10	10.09	6.12	7.90	7.76
2 <sup>b,c</sup>	3-Hydroxy	6.05	16.36	0	0	0.15	0	10.35	10.34	10.31	7.44	8.25	8.01
3	3-(2-Hydroxy)ethyl	5.99	17.29	0	0	1.08	0	10.30	10.23	10.22	7.77	7.71	7.64
4	3-(Carbomethoxy)methyl	6.40	17.78	0	0	1.58	0	9.96	10.06	10.07	7.49	7.60	7.58
5 <sup>a</sup>	3-Acetyl	6.16	17.17	0	0	0.96	0	10.44	10.18	10.17	7.82	7.75	7.66
6	4-Fluoro-3-carboxamido	5.13	17.09	0	0	0.87	0.52	10.74	10.94	10.94	7.85	7.78	7.69
7	3-Amino	5.49	16.57	0	0	0.37	0	10.52	10.50	10.47	7.96	8.08	7.89
8	4-Amino	5.49	16.57	0	0	0.00	0.12	10.52	10.59	10.62	7.96	8.08	8.10
9	3-Amino-4-fluoro	5.84	16.59	0	0	0.37	0.52	10.74	10.76	10.75	7.85	8.07	7.89
10	4-Amino-3-fluoro	5.84	16.59	0	0	0.02	0.12	10.44	10.46	10.50	8.15	8.07	8.09
11	3-Amino-4-methyl	5.94	17.04	0	0	0.37	-0.04	10.30	10.24	10.28	7.82	7.81	7.89
12	3-( <i>N,N</i> -dimethyl)amino	6.89	17.50	0	0	1.30	0	9.64	9.89	9.90	7.41	7.64	7.60
13 <sup>a,c</sup>	3-( <i>N</i> -methyl)amino	6.22	17.04	0	0	0.83	0	10.41	10.17	10.16	8.10	7.81	7.70
14 <sup>c</sup>	3-( <i>N</i> -methyl)amino-4-fluoro	6.66	17.05	0	0	0.83	0.52	10.57	10.40	10.40	8.15	7.80	7.70
15	3-(Pyrazol-1-yl)methyl	6.79	18.40	0	0	2.19	0	9.77	9.94	9.96	7.34	7.64	7.66
16	3-(1,2,4-Triazol-1-yl)methyl	5.79	18.19	0	0	1.98	0	10.27	10.29	10.29	7.46	7.60	7.62
17 <sup>a</sup>	3-(1,2,3-Triazol-1-yl)methyl	5.82	18.19	0	0	1.98	0	9.96	10.28	10.28	7.52	7.60	7.62
18	3-(1,2,3-Triazol-2-yl)methyl	6.31	18.19	0	0	1.98	0	10.09	10.11	10.11	7.49	7.60	7.62
19 <sup>b,c</sup>	3-(Pyrazol-1-yl)-4-amino	5.96	18.30	0	0	1.73	0.12	10.35	10.33	10.32	8.22	7.62	7.59
20	3-(1,2,4-Triazol-1-yl)-4-amino	4.92	18.09	0	0	1.52	0.12	10.54	10.69	10.68	7.46	7.59	7.59
21	3-(1,2,3-Triazol-2-yl)-4-amino	6.73	18.09	0	0	1.52	0.12	10.23	10.04	10.04	7.92	7.59	7.59
22	3-(1,2,3-Triazol-1-yl)-4-amino	5.45	18.09	0	0	1.52	0.12	10.47	10.50	10.49	7.42	7.59	7.59
23	3-(1,2,4-Triazol-3-yl)	5.23	17.63	0	0	1.52	0	10.39	10.49	10.48	7.62	7.62	7.59
24	3-(Pyrazol-3-yl)-4-amino	5.86	18.30	0	0	1.73	0.12	10.60	10.36	10.35	7.66	7.62	7.59
25	3- <i>N</i> -(5-methylpyrid-2-yl)carboxamido	7.06	19.84	0	0	3.63	0	10.32	10.12	10.12	8.40	8.50	8.46
26	3-( <i>N</i> -pyrid-2-yl)carboxamido	6.56	19.37	0	0	3.17	0	10.29	10.18	10.18	8.30	8.11	8.11
27 <sup>a</sup>	3- <i>N</i> -(6-methylpyrid-2-yl)carboxamido	7.06	19.84	0	0	3.63	0	10.40	10.12	10.12	8.52	8.50	8.46
28 <sup>a</sup>	3-( <i>N</i> -thiazol-2-yl)carboxamido	6.40	18.80	0	0	2.59	0	10.80	10.13	10.14	8.10	7.77	7.80
29	3-(Pyrid-2-yl)acetyl	6.43	19.47	0	1	3.26	0	10.14	10.25	10.25	7.82	7.82	7.93
30	3-(Thien-2-yl)acetyl	7.57	19.49	0	1	3.28	0	9.64	9.84	9.85	7.74	7.83	7.95
31 <sup>b</sup>	3-(Pyrazol-1-yl)acetyl	6.08	18.90	0	1	2.69	0	10.20	10.26	10.27	7.82	7.46	7.60
32	3-(1,2,3-Triazol-1-yl)acetyl	5.43	18.69	0	1	2.48	0	10.68	10.46	10.47	7.52	7.37	7.52
33 <sup>b</sup>	3-(1,2,3-Triazol-2-yl)acetyl	5.77	18.69	0	1	2.48	0	10.34	10.34	10.35	7.82	7.37	7.52
34	3-(1,2,4-Triazol-1-yl)acetyl	5.08	18.69	0	1	2.48	0	10.66	10.59	10.59	7.29	7.37	7.52
35 <sup>a</sup>	3- <i>N</i> -(benzimidazol-2-yl)methylamino	7.09	20.45	1	0	4.25	0	10.80	10.31	10.30	7.20	7.35	7.36
36	3- <i>N</i> -(benzimidazol-2-yl)methylamino-4-fluoro	7.54	20.47	1	0	4.25	0.52	10.55	10.55	10.55	7.52	7.37	7.36
37 <sup>b,c</sup>	3- <i>N</i> -(benzthiazol-2-yl)methylamino	7.61	20.83	0	0	4.63	0	10.23	10.28	10.26	8.30	9.73	9.51

<sup>a</sup> Not included in deriving QSAR 1a and 1b<sup>b</sup> Not included in deriving QSAR 2a<sup>c</sup> Not included in deriving QSAR 2b

(and other models reported later) derived for the same data provides lateral support for QSAR 1a.

QSAR 1a and 1b with  $-CMR + CMR^2$  terms hint that the bulk of the substituent may assist in holding the ligand in place or causes a conformational change in receptor structure. If this is attributed to a change in the structure of the receptor that occurs with ligand binding then possibly it may be due to ‘allosteric interactions’. The inverted parabolic relationship with CMR has been earlier proposed to correlate with the change in the mechanism of the reaction due to allosteric interactions [15, 16].

Allosteric mechanism involves coupling of conformational changes between two widely separated binding sites [23]. They are also said to occur due to a change in the structure of the receptor [24]. The common view holds that allosteric proteins are symmetric oligomers [25]. HIV protease is a symmetric oligomer [26] and there is a possibility of allosteric mechanism involved in its binding pattern. In HIV-1 protease each subunit exists in two conformational states with different affinity for ligands. An unsymmetrical ligand may induce different binding sites in the symmetrical HIV-1 protease. Although it is generally accepted that the binding of ligand to the receptor pocket is functionally dependent on the binding affinity and proximity of the ligand, it is also true that individual proteins can exhibit a wide variety of allosteric binding phenomena in the presence of appropriate binding partners. Several other recent computational studies have also suggested that the binding of inhibitors in allosteric sites might affect HIV-1 protease flexibility and disrupt its function [27–29].

The HIVPR is a dimer with an elliptical open-ended cylinder active site formed as well-defined sub-sites  $S_1/S_1'$ ,  $S_2/S_2'$  and  $S_3/S_3'$  containing mostly hydrophobic residues such as Glycine, Alanine and Valine. HIV protease binding site has a hydrophobic binding domain [30]. However, the site seems to have an optimum size as indicated by the presence of parabolic and bilinear correlations of  $ClogP$  in earlier reports [10, 22] and our recent publications [12, 13]. The active site of the enzyme pocket also accommodates 6–8 amino acids with its 18 potential H-bond donors/acceptors [31]. The presence of these active acidic/basic amino acids at the specific enzyme pockets has H-bonding abilities with or without the help of water molecules. The lone pair of electron of the polar groups such as oxygen and nitrogen atom of the ligand will have an appreciable effect on H-bond donating and accepting abilities with surrounding amino acids. Polar interaction of the inhibitor with enzyme pocket is critical for potent binding. The CMR is a measure of ‘volume-dependent-polarizability’ term, and is an indication of the polar interactions between ligand and receptor. The allostery that we observe is ligand induced and may not be true for all protease inhibitors but for the type of substituents analyzed here it seems possible.

#### Antiviral activity data ( $IC_{90}$ ) [14] (Table 1)

We also analyzed the antiviral data reported by De Lucca et al. [14] and obtained QSAR 2a and 2b.

$$\begin{aligned} \text{Log1/IC}_{90} = & -9.270(\pm 2.638) \text{ CMR} \\ & + 0.258(\pm 0.073) \text{ CMR}^2 \\ & - 0.362(\pm 0.230) I_{\text{HARA}} \\ & - 1.851(\pm 0.483) I_{\text{BENZ}} + 90.806(\pm 23.855) \\ n = 31, r = 0.858, r^2 = 0.737, q^2 = 0.632, s = 0.186 \end{aligned} \quad (2a)$$

For CMR, inversion point = 17.955 (17.771–18.111)

$$\begin{aligned} \text{Log1/IC}_{90} = & -0.654(\pm 0.238) \text{ CMR}_{\text{R},3} \\ & + 0.207(\pm 0.062) \text{ CMR}_{\text{R},3}^2 \\ & - 0.238(\pm 0.200) I_{\text{HARA}} \\ & - 1.701(\pm 0.463) I_{\text{BENZ}} + 8.101(\pm 0.200) \\ n = 31, r = 0.845, r^2 = 0.714, q^2 = 0.581, s = 0.186 \end{aligned} \quad (2b)$$

For  $\text{CMR}_{\text{R},3}$ , Inversion point: 1.579 (1.344–1.768)

The antiviral activity was also found equally dependent on inverted parabolic CMR term for the whole molecule (QSAR 2a) and for the 3rd position P2 benzyl substituent (QSAR 2b). In order to have good HIV protease inhibitory activity, the molecule needs to have good antiviral activity. QSAR 1 and 2 are quiet parallel and suggests that the enzyme inhibitory activity ( $K_i$ ) translates well to antiviral activity ( $IC_{90}$ ).

In addition, two other indicator parameters,  $I_{\text{HARA}}$  and  $I_{\text{BENZ}}$  used for the presence of *heteroaryl-acetyl* and *benzimidazole-2-yl groups*, respectively, were found important in QSAR 2a and 2b. These indicators were assigned a value of unity if the specific group is present and zero if absent. Their negative coefficients indicate that these groups of molecule are not very favorable for increasing antiviral activity. Most of these groups are highly hydrophobic and it is possible that they are not very good for achieving good antiviral activity.

Resistance data ( $IC_{90}$  mutant/ $IC_{90}$  WT) for mutant viruses [14]

De Lucca et al. [14] also examined a panel of mutant viruses including mutant strains with single amino acid substitutions (I84V), double amino acid substitutions (I84V/V82F), and multiple amino acid changes corresponding to mutations observed in clinical isolates of patients treated with Ritonavir<sup>®</sup>. They reported [14] the resistance data as a ratio of  $IC_{90}$  mutant/ $IC_{90}$  WT. We analyzed this resistance data and obtained significant correlations (QSAR 3–5) with inverted parabolic CMR term of whole molecule as well as 3rd position P2 benzyl



substituents (Table 2). Although QSAR 3–5 are derived for the resistance data ( $IC_{90}$  mutant/ $IC_{90}$  WT) whereas QSAR 2 is for antiviral activity data ( $IC_{90}$ ), they both represent viral inhibition. Also the QSAR models (2–5) are quite parallel and a comparison of these models and their inversion point provide useful insight about the interactions at the binding site of WT as well as mutant virus.

Resistance data ( $IC_{90}$  mutant/ $IC_{90}$  WT) of single amino acid substituted mutant virus (I84V) [14] (Table 2)

$$\begin{aligned} \text{Log}(IC_{90} \text{ mutant}/IC_{90} \text{ WT}) &= -5.991(\pm 3.318) \text{ CMR} \\ &+ 0.153(\pm 0.088) \text{ CMR}^2 \\ &+ 58.733(\pm 31.125) \\ n &= 8, r = 0.948, r^2 = 0.899, q^2 = 0.691, s = 0.176 \end{aligned} \quad (3a)$$

For CMR, inversion point = 19.575 (19.139–20.764)

$$\begin{aligned} \text{Log}(IC_{90} \text{ mutant}/IC_{90} \text{ WT}) &= -0.766(\pm 0.342) \text{ CMR}_{R,3} \\ &+ 0.125(\pm 0.072) \text{ CMR}_{R,3}^2 \\ &+ 0.342(\pm 0.345) \\ n &= 8, r = 0.945, r^2 = 0.894, q^2 = 0.587, s = 0.171 \end{aligned} \quad (3b)$$

For  $CMR_{R,3}$ , inversion point = 3.060 (2.582–4.320)

The key difference observed between the wild-type and single amino acid substituted I84V mutant models (QSAR 3) were the significant change in the value of ‘inversion point’ of CMR. The mutant viral QSAR models showed a higher inversion point (minimum value) of CMR. An increase in the volume-dependent-polarizability term refers to the larger size of the binding pocket and increase in polar characteristics [24]. For the single amino acid substituted virus I84V, the ‘inversion point’ of CMR seems to increase almost by 1.5 log units (19.575 (QSAR 3a) vs. 17.955 (QSAR 2b)). Similarly, the inversion points for  $CMR_{R,3}$  was also found higher (3.060 (QSAR 3b) vs. 1.579 (QSAR 2b)). These observations indicate that the single mutant I84V pocket may be more spacious and would allow bigger substituents (especially 3rd position P2 benzyl substituents) to fit in the cavity. These observations seem to agree well with the other experimental studies where a bigger pocket was reported for HIVPR mutant [28]. The replacement of Isoleucine (I) by Valine (V) indicates a decrease in the carbon chain length of the amino acid by unit carbon atom (Fig. 1). Thus, the receptor pocket can accommodate voluminous congeners with higher CMR value giving higher ‘inversion point’.

**Table 2** The resistance data and physicochemical parameters of mutant viruses used for deriving QSAR 3, 4 and 5

S. no.	R	CMR	$CMR_{R,3}$	Log ( $IC_{90}$ mutant/ $IC_{90}$ wild-type)								
				Single mutant (I84V)		Double mutant (I84V/V82F)		Multiple mutant (Ritonavir®)				
				Obsd. <sup>14</sup>	Calcd.	Obsd. <sup>14</sup>	Calcd.	Obsd. <sup>14</sup>	Calcd.			
					QSAR 3a	QSAR 3b	QSAR 4a	QSAR 4b	QSAR 5a	QSAR 5b		
1 <sup>a</sup>	3-Amino	16.57	0.37	0.83	1.48	1.08	2.30	2.31	2.14	1.88	2.02	1.75
2	4-Amino	16.57	0.00	1.49	1.48	1.34	2.38	2.31	2.47	2.23	2.02	2.30
3 <sup>c,d</sup>	3-(Pyrazol-1-yl)-4-amino	18.30	1.73	0.52	0.35	0.39	1.70	1.19	1.28	1.85	0.60	0.64
4 <sup>a,b</sup>	3-(1,2,4-Triazol-1-yl)-4-amino	18.09	1.52	0.00	0.43	0.47	1.30	1.28	1.38	0.95	0.64	0.72
5	3-(1,2,4-Triazol-3-yl)	17.63	1.52	0.52	0.67	0.47	1.26	1.51	1.38	0.58	0.84	0.72
6	3-(Pyrazol-3-yl)-4-amino	18.30	1.73	0.34	0.35	0.39	1.30	1.19	1.28	0.32	0.60	0.64
7	3-(N-thiazol-2-yl)carboxamido	18.80	2.59	0.30	0.19	0.20	1.34	1.04	1.01	0.95	0.68	0.68
8 <sup>d</sup>	3-(Pyrid-2-yl)acetyl	19.47	3.26	0.00	0.10	0.18	0.70	0.97	0.96	1.00	1.12	1.10
9 <sup>b,c</sup>	3-N-(benzimidazol-2-yl)methylamino	20.45	4.25	0.00	0.22	0.35	0.48	1.12	1.12	–	2.47	2.36
10	3-N-(benzthiazol-2-yl)methylamino	20.83	4.63	0.52	0.34	0.48	1.30	1.26	1.26	1.40	3.22	3.04

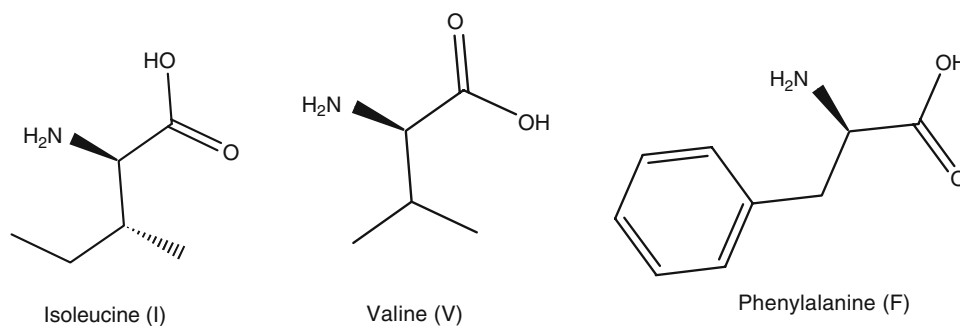
<sup>a</sup> Not included in deriving QSAR 3a

<sup>b</sup> Not included in deriving QSAR 3b

<sup>c</sup> Not included in deriving QSAR 4a and 4b

<sup>d</sup> Not included in deriving QSAR 5a and 5b

**Fig. 1** Amino-acids observed in I84V/V82F mutation with different side-chains (R-group)



Resistance data ( $IC_{90}$  mutant/ $IC_{90}$  WT) of double amino acid substituted mutant virus (I84V/V82F) [14] (Table 2)

$$\begin{aligned} \text{Log } (IC_{90} \text{ mutant}/IC_{90} \text{ WT}) &= -6.200(\pm 3.820) \text{ CMR} \\ &+ 0.159(\pm 0.103) \text{ CMR}^2 \\ &+ 61.355(\pm 35.346) \\ n &= 8, r = 0.944, r^2 = 0.891, q^2 = 0.708, s = 0.222 \end{aligned} \quad (4a)$$

For CMR, inversion point = 19.478 (18.946–21.323)

$$\begin{aligned} \text{Log } (IC_{90} \text{ mutant}/IC_{90} \text{ WT}) &= -0.946(\pm 0.437) \text{ CMR}_{R,3} \\ &+ 0.148(\pm 0.091) \text{ CMR}_{R,3}^2 \\ &+ 2.473(\pm 0.434) \\ n &= 8, r = 0.947, r^2 = 0.897, q^2 = 0.711, s = 0.215 \end{aligned} \quad (4b)$$

For  $\text{CMR}_{R,3}$ , inversion point = 3.203 (2.672–4.897)

For the double amino acid substituted mutant virus I84V/V82F also the resistance data gave QSAR 4a with inversion point higher than the WT (19.478 (QSAR 4a) vs. 17.955 (QSAR 2a)). This model also hints toward a larger binding pocket in this double mutant that can adjust bigger substituents. Similar to other models (QSAR 1b, 2b, 3b), the double mutant resistance data also correlated well with the CMR value of 3rd position P2 benzyl substituent (QSAR 4b) designed to fit S2/S3 pocket. Here also it was noticed that the inversion point for double mutant is higher than the WT (3.203 (QSAR 4b) vs. 1.579 (QSAR 2b)).

The change of Isoleucine (I) to Valine (V) in double mutant increases the pocket size but the change of Valine (V) to Phenylalanine (F) reduces the pocket size (Fig. 1). Since, the Phenylalanine (F) is larger with its bulky phenyl ring compared with two-methyl groups of Valine (V); it in turn make the pocket slightly smaller in size as compare to I84V mutant. This is also reflected in a slightly lower inversion point of double mutant as compared to single mutant (19.478 (QSAR 4a) vs. 19.575 (QSAR 3a)). We did not observe the same pattern for 3rd position P2 benzyl substituents (3.203 (QSAR 4b) vs. 3.060 (QSAR 3a)). However, these are very small differences and our results are based on a small dataset. We need some more similar models for comparative study before coming to final

conclusion. More such studies using bigger datasets will help in laterally validating these results.

Resistance data ( $IC_{90}$  mutant/ $IC_{90}$  WT) of multiple amino acids substituted mutant (Ritonavir<sup>®</sup>) virus (M46I/I63P/A71V/V82F/I84V) [14] (Table 2)

$$\begin{aligned} \text{Log } (IC_{90} \text{ mutant}/IC_{90} \text{ WT}) &= -15.993(\pm 13.106) \text{ CMR} \\ &+ 0.435(\pm 0.366) \text{ CMR}^2 \\ &+ 147.574(\pm 116.960) \\ n &= 7, r = 0.925, r^2 = 0.856, q^2 = 0.417, s = 0.318 \end{aligned} \quad (5a)$$

For CMR, inversion point = 18.380 (17.989–21.168)

$$\begin{aligned} \text{Log } (IC_{90} \text{ mutant}/IC_{90} \text{ WT}) &= -1.621(\pm 0.841) \text{ CMR}_{R,3} \\ &+ 0.385(\pm 0.251) \text{ CMR}_{R,3}^2 \\ &+ 2.295(\pm 0.612) \\ n &= 7, r = 0.948, r^2 = 0.899, q^2 = 0.658, s = 0.266 \end{aligned} \quad (5b)$$

For  $\text{CMR}_{R,3}$ , inversion point = 2.106 (1.748–3.226)

For the multiple substitution of amino acid corresponding to mutations observed in clinical isolates of patients treated with Ritonavir<sup>®</sup>, our analysis gave similar result with an ‘inversion point’ for CMR. Although, the constant term is rather large in QSAR 5a but QSAR 5b shows robust correlation with  $\text{CMR}_{R,3}$ . Similar to the QSAR derived for single mutant and double mutant here also we see a higher inversion point. In this mutant there are various changes in number of carbon atoms at different position and the loss of sulfur; a polar atom, at 46th position of the protease enzyme. In total, the viral pocket loses 2-carbon atoms and 1-sulfur atom. This causes an increase in the average value of CMR compared to wild type but not to the same extent as it does for single mutant (18.380 (QSAR 5a) vs. 17.955 (QSAR 2a) vs. 19.575 (QSAR 3a)). The loss of sulfur atom reduces the polar interaction of amino acid with the ligands.

Although statistically a robust model, this QSAR is based only on seven data points and two parameters. It has been recommended that minimum five data points per parameter should be used for a valid meaningful QSAR [21]. We have reported this QSAR because of its similarity

**Table 3** Comparison of inversion point in QSAR 1–5

Model number	Viral type	Biological end points	Range in biological activity	CMR range	Inversion point (CMR parabola)	CMR <sub>R,3</sub> range	Inversion point (CMR <sub>R,3</sub> parabola)
QSAR 1	Wild-type	Log $1/K_i$	9.64–10.80	16.36–20.83	17.812	0.0–4.63	1.515
QSAR 2	Wild-type	Log $1/IC_{90}$	6.12–8.52	16.36–20.83	17.955	0.0–4.63	1.579
QSAR 3	I84V <sup>a</sup>	Log (IC <sub>90</sub> mutant/ IC <sub>90</sub> WT)	0.0–1.49	16.57–20.83	19.575	0.0–4.63	3.060
QSAR 4	V84F <sup>b</sup>	Log (IC <sub>90</sub> mutant/ IC <sub>90</sub> WT)	0.48–2.38	16.57–20.83	19.478	0.0–4.63	3.203
QSAR 5	Ritonavir mutant virus <sup>c</sup>	Log (IC <sub>90</sub> mutant/ IC <sub>90</sub> WT)	0.32–2.23	16.57–20.83	18.380	0.0–4.63	2.106

<sup>a</sup> Single amino acid substituted mutant virus

<sup>b</sup> Double amino acid substituted mutant virus

<sup>c</sup> Multiple amino acid substituted mutant (Ritonavir) virus

with other models. A smaller set of total nine molecules and insufficient spread in the parameter value precluded a more detailed analysis. These observations can be better explained in large datasets with a wide range.

To summarize, our analyses gave similar results not only for wild-type but also for mutant variants of HIV. The subtle but significant difference in the optimum value of the physico-chemical parameter under study indicates change in the shape, size and various characteristics of the active site. The mutation in enzyme is associated with the change in different amino acid substituent at the active site as described earlier. This may result in a greater occlusion at the sub-sites in the enzyme due to bulky amino acid substitution and vice-versa or a change in hydrophobic and/or electronic characteristics. The mutant QSAR models were found with higher value of inversion point. An increase in the volume-dependent-polarizability term refers to the larger size of the pocket and decrease in polar characteristics. Designing a substituent with higher value of CMR than the inversion point observed may help in achieving better potency against these mutants. We have included the inversion point for WT as well as each mutant with each QSAR and a brief summary is given in Table 3.

QSAR model often points toward the presence of outliers (compounds that do not fit in the final QSAR) that seem to be ‘congeners’ but in fact are not. This is usually due to including too many more or less similar compounds in a data set. The calculated value of these molecules was either too high or too low than the corresponding observed value (see corresponding tables for the data). This problem of “misfit” of the congeners in the final QSAR could be associated with any one of the following reasons:

1. Outliers due to what seem to be ‘congeners’ but are not.
2. Mathematical form of the QSAR may be off the mark.

3. Different rates of metabolism of the members of a set.
4. The quality of the experimental data.
5. Finally, the parameters used may not be the best. Sometimes, experimentally obtained parameters are better than those calculated and vice versa.

Thus we believe that outliers are the leads to new understanding, and covering them up by including them in a QSAR, at the cost of lower  $r^2$ , can be more confusing than helpful.

## Conclusion

In principle, practically any potential drug binding to the protein surface can alter the conformational redistribution. Therefore, appropriate ligands, point mutations, or any other external conditions may facilitate a population shift. It may lead a presumably non-allosteric protein to behave allosterically [32]. Our QSAR analysis, which results in the correlation of biological activity with an inversion point, is an indicative of the ‘allosteric binding’. The allostery, which we observe, seems to be ligand induced and may be true only for the type of substituents analyzed here if not for all protease inhibitors. Further investigations focusing on the possible significance of allosteric binding behavior of HIVPR could extend the clinical and therapeutic applications of anti-HIV drugs.

**Acknowledgements** This study was supported by grant 2 R15 GM 069323-02 from NIH/NIGMS. The authors are indebted to Biobyte Corp., Claremont, CA for the use of CQSAR program.

## References

1. Chrusciel RA, Strohbach JW (2004) *Curr Top Med Chem* 4:1097
2. Gulick RM (2003) *Clin Microbiol Infect* 9:186



3. van Heeswijk RP, Veldkamp A, Mulder JW, Meenhorst PL, Lange JM, Beijnen JH, Hoetelmans RM (2001) *Antivir Ther* 6:201
4. Turner D, Schapiro JM, Brenner BG, Wainberg MA (2004) *Antivir Ther* 9:301
5. Mehandru S, Markowitz M (2003) *Expert Opin Investig Drugs* 12:1821
6. (a) Fujita T (1997) *Quant Struct-Act Relat* 16:107; (b) Yang GF, Huang X (2006) *Curr Pharm Des* 12:4601
7. Walters WP, Murcko MA (2002) *Adv Drug Deliv Rev* 54:255
8. (a) Jorgensen WL, Duffy EM (2002) *Adv Drug Deliv Rev* 54:355; (b) Abdel-Rahman HM, Al-karamany GS, El-Koussi NA et al (2002) *Curr Med Chem* 9:1905
9. (a) Hansch C, Leo A, Mekapati SB, Kurup A (2004) *Bioorg Med Chem* 12:3391; (b) Luco JM, Marchevsky E (1999) *J Chem Inf Comput Sci* 39:396
10. (a) Garg R, Gupta SP, Gao H, Mekapati SB, Debnath AK, Hansch C (1999) *Chem Rev* 99:3525; (b) Kurup A, Garg R, Hansch C (2001) *Chem Rev* 101:2573
11. (a) Kurup A, Mekapati SB, Garg R, Hansch C (2003) *Curr Med Chem* 10:1819; (b) Debnath AK (2005) *Curr Pharm Des* 11:3091; (c) Gayathri P, Pande V, Sivakumar R, Gupta SP (2001) *Bioorg Med Chem* 9:3059
12. Bhatarai B, Garg R (2005) *Bioorg Med Chem* 13:4078
13. Patel D, Garg R (2005) *Bioorg Med Chem Lett* 15:3767
14. De Lucca GV, Kim UT, Liang J, Cordova B, Klabe RM, Garber S, Bachelier LT, Lam GN, Wright MR, Logue KA, Erickson-Viitanen S, Ko SS, Trainer GL (1998) *J Med Chem* 41:2411
15. (a) Hansch C, Garg R, Kurup A, Mekapati SB (2003) *Bioorg Med Chem* 11:2075; (b) Hansch C, Garg R, Kurup A (2001) *Bioorg Med Chem* 9:283
16. (a) Mekapati SB, Hansch C (2001) *Bioorg Med Chem* 9:2885; (b) Garg R, Kurup A, Hansch C (2001) *Bioorg Med Chem* 9:3161; (c) Christopoulos A (2002) *Nat Rev Drug Discov* 1:198
17. Lam PYS, Ru Y, Jadhav PK, Aldrich PE, De Lucca GV, Eyer-mann CJ, Chang C-H, Emmett G, Holler ER, Daneker WF, Li L, Confalone PN, McHugh RJ, Han Q, Markwalder JA, Seitz SP, Bachelier LT, Rayner MM, Klabe RM, Shum L, Winslow DL, Kornhauser DM, Jackson DA, Erickson-Viitanen S, Sharpe TR, Hodge CN (1996) *J Med Chem* 39:3514
18. Cramer RD III, Bunce JD, Patterson DE, Frank IE (1988) *Quant Struct-Act Relat* 7:18
19. CQSAR program, Biobyte Corp., 201 W. 4th street, Claremont, CA 91711
20. Leo A (1993) *Chem Rev* 93:1281
21. (a) Hansch C, Leo A (eds) (1995) *Exploring QSAR: fundamentals and applications in chemistry and biology*. American Chemical Society, Washington; (b) Hansch C, Leo A, Hoekman D (eds) *Exploring QSAR: hydrophobic, electronic and steric constants*. American Chemical Society, Washington
22. Garg R, Bhatarai B (2004) *Bioorg Med Chem* 12:5819
23. (a) Koshland DE Jr (1958) *Proc Natl Acad Sci USA* 44:98; (b) Koshland DE Jr, Boyer PD (eds) (1970) *The enzymes*, vol 1. Academic Press, New York, p 342
24. (a) Hansch C, Klein T (1986) *Acc Chem Res* 19:392; (b) Kubinyi H (ed) (1993) *QSAR: Hansch analysis and related approaches*. VCH, New York
25. (a) Monod J, Changeux J-P, Jacob F (1963) *J Mol Biol* 6:306; (b) Monod J, Wyman J, Changeux J-P (1965) *J Mol Biol* 12:88
26. (a) Toh H, Ono M, Saigo K, Miyata T (1985) *Nature* 315:691; (b) Miller M, Jaskolski M, Rao JKM, Leis J, Wlodawer A (1989) *Nature* 337:576
27. McCammon JA (2005) *Biochim Biophys Acta* 1754:221
28. (a) Perryman AL, Lin J-H, McCammon JA (2004) *Protein Sci* 13:1108; (b) Perryman AL, Lin J-H, McCammon JA (2006) *Chem Biol Drug Des* 67:336; (c) Perryman AL, Lin J-H, McCammon JA (2006) *Biopolymers* 82:272
29. Hornak V, Simmerling C (2007) *Drug Discov Today* 12:132
30. Babine RE, Bender SL (1997) *Chem Rev* 97:1359
31. Appelt K (1993) *Perspect Drug Discov Des* 1:23
32. Gunasekaran K, Ma B, Nussinov R (2004) *Proteins* 57:433