

# Acylthiocarbamates as non-nucleoside HIV-1 reverse transcriptase inhibitors: docking studies and ligand-based CoMFA and CoMSIA analyses

Elena Cichero · Sara Cesarini · Andrea Spallarossa ·  
Luisa Mosti · Paola Fossa

Received: 13 October 2008 / Accepted: 26 November 2008 / Published online: 20 January 2009  
© Springer-Verlag 2009

**Abstract** Acylthiocarbamates (ATCs) have been identified as a class of potent non-nucleoside HIV-1 reverse transcriptase (RT) inhibitors. A computational strategy based on molecular docking studies followed by comparative molecular fields analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) was used to identify the most important features impacting ATC antiretroviral activity. The CoMSIA model proved to be the more predictive, with  $r^2_{ncv}=0.89$ ,  $r_{cv}^2=0.38$ , standard error of estimate (SEE)=0.494,  $F=84$ , and  $r^2_{pred}=0.81$ . The results of these studies will be useful in designing new ATCs with improved potency, also against clinically relevant resistant mutants.

**Keywords** HIV-1 · Reverse transcriptase · Docking · CoMFA · CoMSIA · Acylthiocarbamates

## Introduction

HIV-1 reverse transcriptase (RT) catalyses the conversion of a single-stranded RNA into a double stranded DNA that is then integrated into the host cell genome. This key role of RT in the HIV-1 life cycle singled out this enzyme as a preferred target for “Highly Active Antiretroviral Therapy” [1–5]. Non-nucleoside RT inhibitors are chemically diverse and selective RT targeting agents that bind to an allosteric hydrophobic pocket (non-nucleoside RT inhibitor binding site), located about 10 Å from the polymerase active site.

Upon inhibitor interaction, the non-nucleoside RT inhibitor binding site is created and the enzyme is locked into an inactive form due to the effect on the geometry of the polymerase active site aspartyl residues [6]. In the past 15 years more than 50 structurally diverse non-nucleoside RT inhibitors have been described [6–12]. The therapeutic efficacy of non-nucleoside RT inhibitors is severely limited by the emergence of HIV-1 drug-resistant mutants [13, 14]. Therefore, the search for new, selective and potent drugs able to inhibit also these mutant HIV forms remains a challenge.

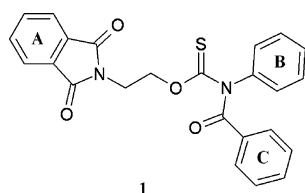
In previous studies, we identified *O*-(2-phthalimidoethyl)-*N*-aryl-*N*-acylthiocarbamates (ATCs) [15, 16] (Fig. 1 shows the lead ATC **1**) as potent non-nucleoside RT inhibitors.

To elucidate the molecular basis for RT/ATC interactions (in the absence of crystallographic data from RT/ATC complexes), and to identify features significantly impacting ATC antiretroviral activity, we performed docking studies on a series of 78 ATCs and elaborated two three-dimensional quantitative structure-activity relationship (3D-QSAR) models. These models should enable us to predict ATC antiretroviral activity prior to synthesis, and provide useful suggestions for the design of new ATCs with improved potency, also against clinically relevant resistant mutants.

The computational studies used to construct the docking model of ATCs bound into the non-nucleoside RT inhibitor binding site exploited the X-ray coordinates [17] of the complex between RT and *O*-[2-(phthalimido)ethyl]-*N*-(4-chlorophenyl)thiocarbamate (**I**, Fig. 2). **I** belongs to the class of thiocarbamates, potent non-nucleoside RT inhibitors recently identified by some of us [18, 19], and is a structurally simplified ATC deacylated derivative. Before the crystallographic structure of the RT/**I** complex became available, the template structure used by some of us

E. Cichero · S. Cesarini · A. Spallarossa · L. Mosti · P. Fossa (✉)  
Dipartimento di Scienze Farmaceutiche,  
Università degli Studi di Genova,  
Viale Benedetto XV n.3,  
16132 Genoa, Italy  
e-mail: fossap@unige.it

**Fig. 1** Chemical structure of **1** [lead acylthiocarbamate (ATC) compound]



for RT/ATC docking studies [15, 16] had been the crystal complex of RT with PETT-1 (Fig. 2), belonging to the non-nucleoside RT inhibitor class of PETT (phenethylthiazolylthiourea) derivatives [20], which has, however, fewer structural features in common with ATCs.

Even though a structure-based approach was possible using the constructed RT/ATC model, we employed a ligand-based approach based on comparative molecular fields analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) to obtain a complementary tool for ATC design.

## Materials and methods

### Data set

A dataset of 78 ATCs (**1–78**), screened according to the same pharmacological protocol, were selected from two samples datasets [15, 16] and submitted to QSAR analysis. The molecular structures of **1–78** (Table 1) were built, parameterised (Gasteiger-Huckel method) and energy minimised within MOE software using MMFF94 forcefield [21].

### Docking protocol

Since complexes between RT and ATCs were not available, we used the 3D structure co-ordinates of RT in complex with **1** (Fig. 2) (PDB entry 2VG5) [17] as the starting point for a preliminary manual docking simulation. Thiocarbamate **1** was chosen because it shares the *O*-[2-(phthalimido)ethyl]-*N*-phenylthiocarbamate moiety with ATCs. Compound **31**, showing the highest pEC<sub>50</sub> value (EC<sub>50</sub>=compound concentration [μM] required to achieve 50% protection of MT-4 cells from HIV-1 induced

cytopathogenicity, as determined by the MTT method) was selected for the preliminary docking simulation. Starting from its best energy minimised conformation, the corresponding RT/**31** complex model was derived by superimposition of **31** on the RT/**1** crystallographic structure (see Manual docking). A procedure of automated docking of the energy minimised conformations of **1–78** was successively performed using the previously derived putative RT/**31** complex as the ligand/receptor model.

### Manual docking

The *O*-[2-(phthalimido)ethyl]-*N*-(4-chlorophenyl)thiocarbamate moiety of ATC **31** was manually superimposed on **1** {i.e. *O*-[2-(phthalimido)ethyl]-*N*-(4-chlorophenyl)thiocarbamate} in the crystallographic structure of the RT/**1** complex. Then, **1** was erased and the RT/**31** complex was generated. The complex was then minimised within LigX, a module of MOE software. Briefly, receptor atoms far from the ligand were held fixed (constrained not to move at all), while residues within a certain distance (8 Å) could move so as ligand atoms were not fixed. A forcefield (MMFF94) energy minimisation was performed, which terminated when the root mean square (RMS) gradient of potential energy had fallen below a certain threshold, set to 0.05 kcal/molÅ.

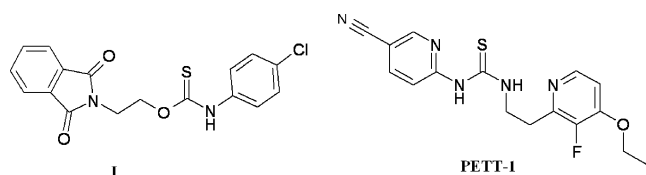
### Automated docking

Compounds **1–78** were docked into the non-nucleoside RT inhibitor binding site using the flexible docking module implemented in MOE software. For all compounds, the best-docked geometries, evaluated in terms of “London dG”, were refined by energy minimisation (MMFF94) and rescored according to “Affinity dG” (kcal/mol of total estimated binding energy). Following this procedure, on the basis of the final docking scoring function (S), we identified the most probable ATC binding conformation interacting with RT (lowest mean S value).

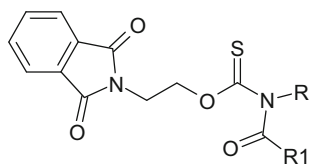
### CoMFA and CoMSIA analyses

### Training- and test-set

The compounds were grouped into a training set, for model generation, and a test set, for model validation, containing 67 and 11 compounds, respectively. Both the training and the test set were divided manually according to a representative range of biological activities and structural variations. For QSAR analysis, EC<sub>50</sub> values were transformed into pEC<sub>50</sub> values and then used as response variables. The enzyme inhibitory activity of the compounds covered 4 log orders of magnitude.



**Fig. 2** Chemical structure of *O*-[2-(phthalimido)ethyl]-*N*-(4-chlorophenyl) thiocarbamate (**1**)

**Table 1** Molecular structure of acylthiocarbamates (ATCs) 1–78

Comp	R	R1	Comp	R	R1
1			12		
2			13		
3			14		
4			15		
5			16		
6			17		
7			18		
8			19		
9			20		
10			21		
11					

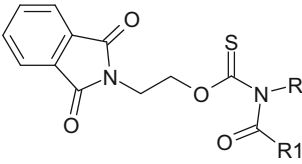
### Molecular alignment

Molecular alignment is the most sensitive parameter in 3D-QSAR analysis, and is becoming one of the most determining factors in computing robust and meaningful statistical models. In the present study, the geometry optimised structures of 1–78 were aligned together on the basis of the common acylthiocarbamic moiety, by the align database command in Sybyl7.0 [22].

### CoMFA and CoMSIA interaction energies

The CoMFA method [23] is a 3D-QSAR technique widely used to relate the biological activity of a series of molecules to their steric and electrostatic fields, which are calculated by placing the aligned molecules, one by one, in a 3D cubic lattice with a 2 Å grid spacing. The van der Waals potential and Coulombic terms, which represent steric and electrostatic fields, respectively, are then

Table 1 (continued)

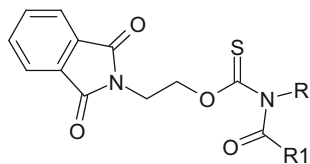


Comp	R	R1	Comp	R	R1
22			33		
23			34		
24			35		
25			36		
26			37		
27			38		
28			39		
29			40		
30			41		
31			42		
32			43		

calculated using the standard Tripos force field method. The column-filtering threshold value was set to 2.0 kcal/mol to improve the signal-to-noise ratio. A methyl probe with a +1 charge was used to calculate the CoMFA steric and electrostatic fields. A 30 kcal/mol energy cut-off was applied to avoid infinity of energy values inside the molecule. The CoMSIA method [24] calculates five

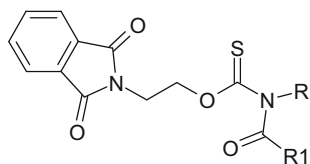
descriptors, namely the steric, electrostatic, and hydrophobic parameters, and the H-bond donor and H-bond acceptor properties. The similarity index descriptors were calculated using the same lattice box employed for CoMFA calculations, and an  $sp^3$  carbon as probe atom with +1 charge, +1 hydrophobicity and +1 H-bond donor and +1 H-bond acceptor properties.

Table 1 (continued)



Comp	R	R1	Comp	R	R1
44			55		
45			56		
46			57		
47			58		
48			59		
49			60		
50			61		
51			62		
52			63		
53			64		
54			65		

Table 1 (continued)



Comp	R	R1	Comp	R	R1
66			73		
67			74		
68			75		
69			76		
70			77		
71			78		
72					

#### Partial least square analysis and model validation

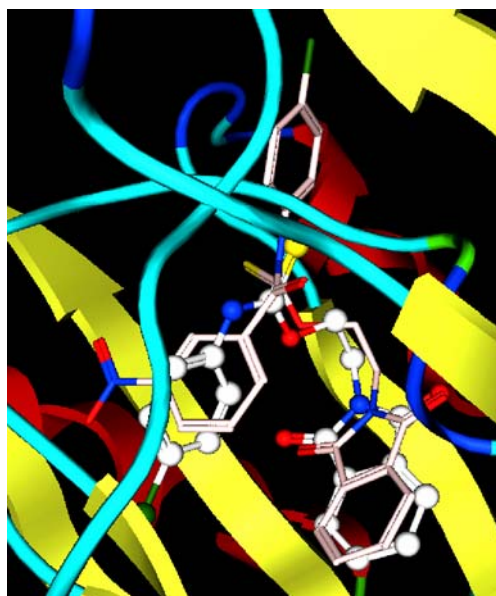
The partial least-squares (PLS) approach—an extension of the multiple regression analysis—was used to derive the 3D-QSAR models, in which the CoMFA and CoMSIA descriptors were used as independent variables and  $pEC_{50}$  values were used as dependent variables. Prior to PLS analysis, CoMFA and CoMSIA columns with a variance smaller than  $2.0 \text{ kcal mol}^{-1}$  were filtered using column filtering to improve the signal-to-noise ratio.

The leave one out (LOO) cross-validation method was used to check the predictivity of the derived model and to identify the optimal number of components (ONC) leading to the highest cross-validation  $r^2$  ( $r^2_{cv}$ ). In the LOO cross-validation methodology, one molecule is omitted from the

dataset and a model involving the rest of the compounds is derived. By employing this model, the activity of the omitted molecule is then predicted.

The optimal number of components obtained from cross-validation methodology was used in the subsequent regression model. Final CoMFA and CoMSIA models were generated using non-cross-validated PLS analysis. To further assess the statistical confidence and robustness of the derived models, a 100-cycle bootstrap analysis was performed. This is a procedure in which  $n$  random selections out of the original set of  $n$  objects are performed several times (100 times were required to obtain a good statistical information). In each run, some objects may not be included in the PLS analysis, whereas others might be included more than once. The mean correlation coefficient is represented as bootstrap  $r^2$ .

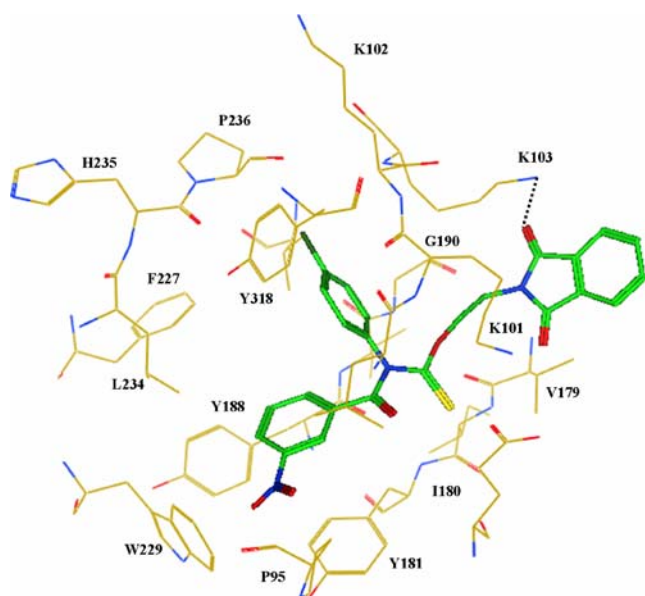




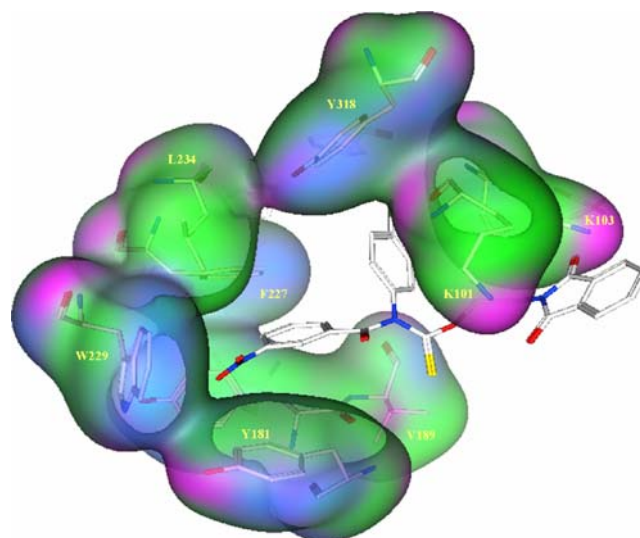
**Fig. 3** Manual docking by superimposition of **31** on the X-ray structure of **I** in complex with RT. **I** is depicted in ball and stick, coloured by atom type. Compound **31** is shown in stick (pink C, blue N, red O, yellow S)

#### Predictive correlation coefficient

To further validate the CoMFA and CoMSIA derived model, the predictive ability for the test set of compounds



**Fig. 4** Automated docking analysis of the RT/ATC **31** complex obtained by a manual docking procedure followed by minimisation of the energy of the complex. Compound **31** is reported in stick (green C, blue N, red O, yellow S). Only residues at 5 Å distance from ligands are reported—the most important are labelled. Black dotted lines Hydrogen bonds



**Fig. 5** Selected docking pose of **31** into the non-nucleoside RT inhibitor binding site. The RT surface electrostatic distribution (Connolly surface) is shown. Green Hydrophobic regions, magenta H-bond regions, blue mildly polar regions

(expressed as  $r^2_{pred}$ ) was determined using the following equation:

$$r^2_{pred} = (SD - PRESS) / SD$$

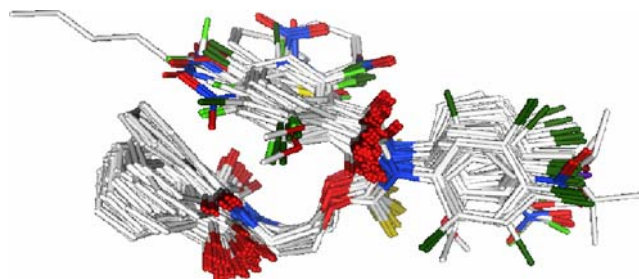
SD is the sum of the squared deviations between the biological activities of the test set molecules and the mean activity of the training set compounds, and PRESS is the sum of the squared deviation between the observed and the predicted activities of the test set compounds.

All calculations were carried out using a PC, under a Windows XP operative system and SGI O2 Silicon Graphics.

## Results and discussion

### Docking simulations

In the absence of crystallographic data on RT/ATC complexes, and before that the X-ray coordinates of the



**Fig. 6** Alignment of compounds **1–78** employed for comparative molecular fields analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA)

**Table 2** Summary of comparative molecular fields analysis (CoMFA) results

Parameter	
No. of compounds	67
ONC	6
$r^2_{\text{LOO}}$	0.490
$r^2_{\text{cv}}$	0.489
SEE	0.347
$r^2_{\text{ncv}}$	0.920
<i>F</i> value	115.53
Steric contribution	0.523
Electrostatic contribution	0.477
$r^2_{\text{boot}}$	0.951
SEE $r^2_{\text{boot}}$	0.268
Test set $r^2$ ( $r^2_{\text{pred}}$ )	0.77

ONC Optimal number of components, LOO leave one out, cv cross-validated, SEE standard error of estimate, ncv non cross-validated, boot bootstrap

**Table 3** CoMFA analysis: experimental (*Exp.*) and predicted (*Pred.*) pEC<sub>50</sub> values of training set compounds (1–3, 5–9, 11–14, 16, 17, 19–36, 38–41, 43–45, 48–58, 60, 62–69, 71–78)

Compound	Exp. pEC <sub>50</sub>	Pred. pEC <sub>50</sub>	Residual	Compound	Exp. pEC <sub>50</sub>	Pred. pEC <sub>50</sub>	Residual
1	6.40	6.42	-0.02	40	7.70	8.30	-0.60
2	4.96	4.99	-0.03	41	7.70	7.11	0.59
3	4.96	4.96	0.00	43	8.10	7.68	0.42
5	4.96	4.82	0.14	44	8.00	7.78	0.22
6	5.00	5.10	-0.09	45	7.70	8.00	-0.30
7	4.96	5.55	-0.59	48	8.15	8.00	0.15
8	5.22	5.47	-0.25	49	7.00	7.49	-0.49
9	4.96	5.40	-0.44	50	7.52	7.28	0.24
11	5.92	5.14	0.78	51	5.25	5.44	-0.19
12	6.42	5.65	0.78	52	8.05	7.84	0.21
13	5.46	5.51	-0.04	53	8.10	7.84	0.26
14	7.52	7.53	-0.01	54	6.40	5.86	0.54
16	7.00	7.77	-0.77	55	8.10	8.08	0.02
17	7.60	7.39	0.21	56	7.15	7.48	-0.33
19	8.10	7.58	0.52	57	7.70	7.65	0.05
20	8.30	8.46	-0.16	58	8.15	8.15	0.00
21	8.22	7.96	0.26	60	7.52	7.92	-0.40
22	7.46	7.53	-0.07	62	6.00	5.96	0.04
23	8.00	8.01	-0.01	63	7.30	7.34	-0.04
24	8.10	8.43	-0.33	64	6.70	6.63	0.07
25	8.00	8.21	-0.21	65	6.52	6.53	-0.01
26	6.30	6.63	-0.33	66	5.05	5.44	-0.39
27	5.07	4.88	0.19	67	7.30	7.04	0.26
28	7.40	7.24	0.16	68	4.66	4.83	-0.17
29	7.15	7.29	-0.14	69	7.05	6.90	0.15
30	8.00	7.49	0.51	71	7.30	7.28	0.02
31	8.82	8.70	0.12	72	6.30	6.36	-0.06
32	7.30	7.40	-0.10	73	6.46	6.43	0.03
33	8.30	8.20	0.10	74	7.52	7.69	-0.17
34	7.22	7.05	0.18	75	7.70	7.13	0.57
35	7.46	7.60	-0.14	76	5.59	5.71	-0.12
36	6.70	6.43	0.27	77	4.77	4.83	-0.06
38	6.40	6.63	-0.23	78	6.16	5.91	0.25
39	7.40	7.41	-0.01				

RT/I complex became available, the template structure used by some of us for RT/ATC docking studies [15, 16] had been the crystal complex of RT with PETT-1 (Fig. 2).

In this work, for docking simulations the template used was the RT/I complex, as I is structurally more similar than PETT-1 to ATCs.

#### Manual docking

According to our calculations, ATC 31, the most active ATC so far synthesised, superimposed on the crystallographic structure of RT/I (Fig. 3), displays a H-bond interaction between the K101 backbone NH and the oxygen atom of the acyl carbonyl function. The RT/31 complex was minimised and employed for the following automated docking procedure.



**Table 4** CoMFA analysis: experimental and predicted pEC<sub>50</sub> values of test set compounds (**4**, **10**, **15**, **18**, **37**, **42**, **46**, **47**, **59**, **61**, **70**)

Compound	Exp. pEC <sub>50</sub>	Pred. pEC <sub>50</sub>	Residual
<b>4</b>	4.96	5.30	-0.34
<b>10</b>	5.22	5.25	-0.03
<b>15</b>	7.00	6.12	0.88
<b>18</b>	8.15	7.36	0.79
<b>37</b>	6.70	7.24	-0.54
<b>42</b>	7.52	7.23	0.30
<b>46</b>	7.70	7.80	-0.10
<b>47</b>	8.10	7.95	0.15
<b>59</b>	7.70	6.97	0.73
<b>61</b>	7.40	6.94	0.46
<b>70</b>	7.22	7.11	0.11

### Automated docking

**ATC 1-78 docking poses** As shown for compound **31** in Fig. 4, all inhibitors display a H-bond between the K103 ε-amino group and one oxygen of the phthalimide moiety (positioned at the entrance of the non-nucleoside RT inhibitor binding site), and several lipophilic interactions with two hydrophobic pockets (P<sub>1</sub> and P<sub>2</sub>) are detected. Region P<sub>1</sub> includes residues P95, Y181, Y188 and W229, whereas P<sub>2</sub> consists of residues V106, F227, L234, P236 and Y318. The ATC acyl moiety is located in the P<sub>1</sub> region and establishes π–π stacking with Y181, Y188 and W229, while the *N*-phenyl ring is oriented towards the P<sub>2</sub> pocket, and is involved in π–π interactions with Y318. The position of the phthalimide moiety at the entrance of the non-nucleoside RT inhibitor binding site is in agreement with the RT/ATC docking model previously elaborated by some of us [15, 16]. However, in the preceding model, the *N*-phenyl ring established Van der Waals interactions with amino acids of the P<sub>1</sub> pocket, while the phenyl ring of the benzoyl group made hydrophobic contacts with V106 and K103.

To verify the reliability of the derived RT/ATC complex, the selected docking pose of ATC **31** was compared with

the RT surface electrostatic distribution (Connolly surface), as reported in Fig. 5. A good correlation between the hydrophilic and hydrophobic features of the external phthalimide moiety (ring A) of **31** with those of the RT residues (in particular K101 and K103) confirmed that the selected ATC conformation is the most probable one involved in the interaction with RT.

### CoMFA and CoMSIA analyses

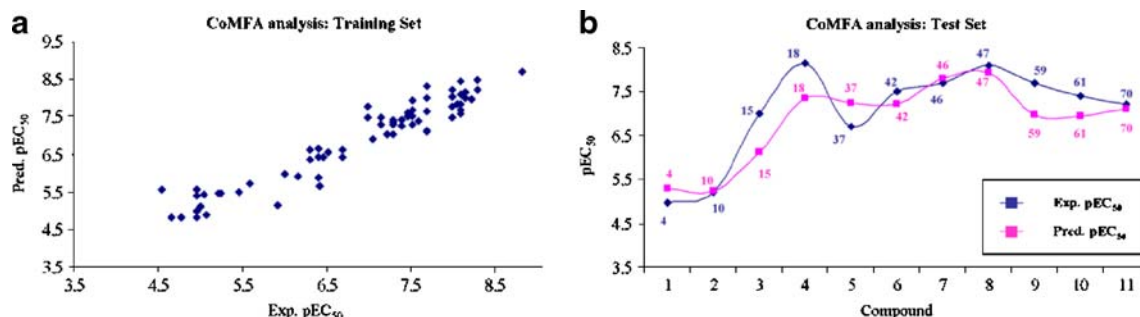
To develop the 3D-QSAR analyses, **1-78** were aligned together on the basis of the common acylthiocarbamic moiety, as reported in Fig. 6.

Ligand-based CoMFA and CoMSIA analyses were performed by dividing compounds **1-78** into a training set (**1-3**, **5-9**, **11-14**, **16**, **17**, **19-36**, **38-41**, **43-45**, **48-58**, **60**, **62-69**, **71-78**) for model generation, and a test set (**4**, **10**, **15**, **18**, **37**, **42**, **46**, **47**, **59**, **61**, **70**) for model validation. CoMFA and CoMSIA studies were developed using, respectively, CoMFA steric and electrostatic fields and CoMSIA steric, electrostatic, hydrophobic and H-bond acceptor properties, as independent variables, and the ligand pEC<sub>50</sub> as a dependent variable. For CoMSIA analysis, the H-bond donor descriptor was not taken into consideration because the ligands display no H-bond donor groups.

The final CoMFA model was generated employing non-cross-validated PLS analysis with the optimum number of components (ONC)=6 to give a non-cross validated  $r^2$  ( $r^2_{ncv}$ )=0.92, standard error of estimate (SEE)=0.347, steric contribution=0.523 and electrostatic contribution=0.477. All statistical parameters supporting the CoMFA model are reported in Table 2. The model reliability thus generated was supported by bootstrapping results (see Table 2).

Experimental and predicted binding affinity values for the training set and test set are reported in Tables 3 and 4, respectively, while distribution of experimental and predicted pEC<sub>50</sub> values for training set and test set according to the CoMFA model are represented in Fig. 7.

A CoMSIA model consisting of steric, electrostatic, hydrophobic and H-bond acceptor fields with a non-cross



**Fig. 7** Distribution of experimental and predicted pEC<sub>50</sub> values for training set compounds (**a**) and test set compounds (**b**) according to CoMFA analysis

**Table 5** Summary of comparative molecular similarity indices analysis (CoMSIA) results

Parameter	
No. of compounds	67
ONC	6
$r^2_{\text{LOO}}$	0.359
$r^2_{\text{cv}}$	0.380
SEE	0.494
$r^2_{\text{ncv}}$	0.89
<i>F</i> value	83.807
Steric contribution	0.269
Electrostatic contribution	0.160
Hydrophobic contribution	0.331
H-bond acceptor contribution	0.240
$r^2_{\text{boot}}$	0.90
SEE $r^2_{\text{boot}}$	0.386
Test set $r^2$ ( $r^2_{\text{pred}}$ )	0.81

validated  $r^2_{\text{ncv}}=0.89$ ,  $\text{SEE}=0.494$ , steric contribution=0.269, electrostatic contribution=0.160, hydrophobic contribution=0.331, and H-bond acceptor contribution=0.240 was derived. All statistical parameters supporting the CoMSIA model are reported in Table 5.

Experimental and predicted binding affinity values for the training set and test set are reported in Tables 6 and 7, respectively, while distribution of experimental and predicted pEC<sub>50</sub> values for training set and test set according to the CoMSIA model are represented in Fig. 8.

Since CoMFA and CoMSIA field effects on the target properties can be viewed as 3D coefficient contour plots, thus identifying important regions where any change in these fields may affect the biological activity, they could be helpful in optimising ATCs as non-nucleoside RT inhibitors. The 3D-QSAR analysis maps are described and discussed in the following sections.

**Table 6** CoMSIA analysis: experimental and predicted pEC<sub>50</sub> values of training set compounds (1–3, 5–9, 11–14, 16, 17, 19–36, 38–41, 43–45, 48–58, 60, 62–69, 71–78)

Compound	Exp. pEC <sub>50</sub>	Pred. pEC <sub>50</sub>	Residual	Compound	Exp. pEC <sub>50</sub>	Pred. pEC <sub>50</sub>	Residual
1	6.40	6.29	0.11	40	7.70	8.26	-0.56
2	4.96	5.33	-0.37	41	7.70	7.26	0.44
3	4.96	4.65	0.31	43	8.10	7.75	0.35
5	4.96	4.93	0.03	44	8.00	7.65	0.35
6	5.00	5.13	-0.13	45	7.70	7.67	0.03
7	4.96	4.69	0.27	48	8.15	8.14	0.01
8	5.22	5.19	0.03	49	7.00	7.24	-0.24
9	4.96	5.97	-1.01	50	7.52	7.13	0.39
11	5.92	5.72	0.20	51	5.25	5.78	-0.53
12	6.42	6.81	-0.39	52	8.05	8.12	-0.07
13	5.46	5.05	0.41	53	8.10	8.28	-0.18
14	7.52	6.76	0.76	54	6.40	6.04	0.37
16	7.00	7.14	-0.14	55	8.10	8.19	-0.09
17	7.60	7.30	0.30	56	7.15	7.24	-0.09
19	8.10	7.40	0.70	57	7.70	7.73	-0.03
20	8.30	7.84	0.46	58	8.15	7.86	0.29
21	8.22	8.21	0.02	60	7.52	6.91	0.61
22	7.46	7.66	-0.20	62	6.00	6.40	-0.40
23	8.00	8.06	-0.06	63	7.30	7.52	-0.22
24	8.10	8.47	-0.37	64	6.70	6.08	0.62
25	8.00	7.97	0.03	65	6.52	6.39	0.13
26	6.30	6.77	-0.47	66	5.05	5.47	-0.42
27	5.07	5.35	-0.28	67	7.30	7.42	-0.12
28	7.40	7.10	0.30	68	4.66	4.97	-0.31
29	7.15	7.25	-0.10	69	7.05	6.48	0.57
30	8.00	7.86	0.14	71	7.30	7.28	0.02
31	8.82	8.48	0.34	72	6.30	7.05	-0.75
32	7.30	7.67	-0.37	73	6.46	6.12	0.34
33	8.30	8.40	-0.10	74	7.52	7.01	0.51
34	7.22	8.27	-1.05	75	7.70	7.28	0.42
35	7.46	7.62	-0.16	76	5.59	5.96	-0.37
36	6.70	6.48	0.22	77	4.77	4.87	-0.10
38	6.40	6.70	-0.30	78	6.16	5.94	0.22
39	7.40	7.31	0.09				

**Table 7** CoMSIA analysis: experimental and predicted pEC<sub>50</sub> values of test set compounds (**4**, **10**, **15**, **18**, **37**, **42**, **46**, **47**, **59**, **61**, **70**)

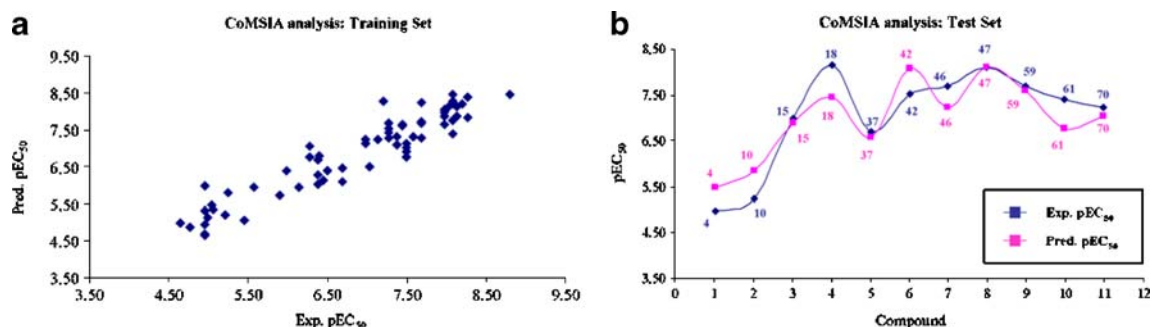
Compound	Exp. pEC <sub>50</sub>	Pred. pEC <sub>50</sub>	Residual
<b>4</b>	4.96	5.48	-0.52
<b>10</b>	5.22	5.84	-0.62
<b>15</b>	7.00	6.90	0.10
<b>18</b>	8.15	7.45	0.71
<b>37</b>	6.70	6.57	0.13
<b>42</b>	7.52	8.10	-0.58
<b>46</b>	7.70	7.24	0.46
<b>47</b>	8.10	8.10	0.00
<b>59</b>	7.70	7.59	0.11
<b>61</b>	7.40	6.77	0.63
<b>70</b>	7.22	7.03	0.19

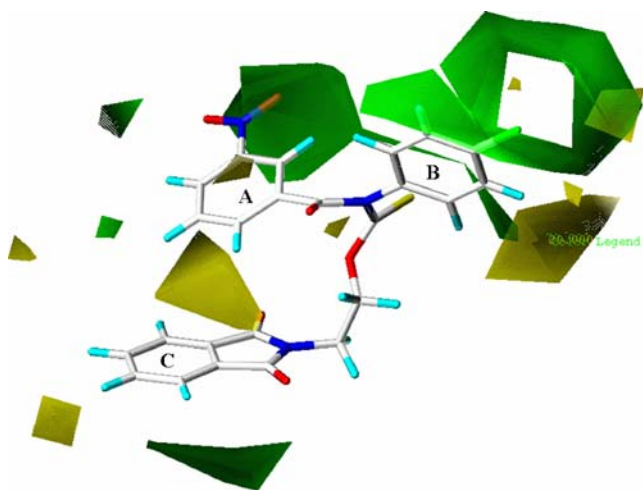
### CoMFA steric and electrostatic regions

As shown in Fig. 9, the steric contour map predicts favourable (green) interaction polyhedra for the positions *para* of ring B, *meta* of ring C and position 3 of the phthalimide moiety, while unfavourable (yellow) interaction polyhedra surround the positions *meta* of ring B, *ortho* of ring C and position 4 of the phthalimide substructure. The reliability of this steric map calculation is underlined by the activity trend of the 2-thenoyl ATCs [cf. *N*-*para*-fluorophenyl **15**, *N*-*para*-chloro **19** and *N*-*para*-nitro **25** (pEC<sub>50</sub>=7.00–8.10) with *N*-*meta*-fluoro **9**, *N*-*meta*-chloro **10** and *N*-*meta*-nitro **12** (pEC<sub>50</sub>=4.96–6.42)], the benzoyl derivatives [*N*-*para*-chloro **16** (pEC<sub>50</sub>=7.00) and *N*-*para*-bromo **22** (pEC<sub>50</sub>=7.46) vs *N*-unsubstituted **1** (pEC<sub>50</sub>=6.40)], the 3,4-dichloro ATCs [cf. *N*-*para*-chloro **52** (pEC<sub>50</sub>=8.05) and *N*-*para*-methyl **67** (pEC<sub>50</sub>=7.30) with *N*-unsubstituted **72** (pEC<sub>50</sub>=6.30)] and the 4-chloro-3-nitro analogues [*N*-*para*-chloro **57** (pEC<sub>50</sub>=7.70) and *N*-*para*-methyl **70** (pEC<sub>50</sub>=7.22) vs *N*-unsubstituted **73** (pEC<sub>50</sub>=6.46)]. In addition, reliability is confirmed by the activity trend of the *N*-*para*-chloro-phenyl ATCs [cf. 3-methylbenzoyl **30** (pEC<sub>50</sub>=8.00) with 2-methylbenzoyl **28** (pEC<sub>50</sub>=7.40) and 3-nitrobenzoyl **31** (pEC<sub>50</sub>=8.82) with 2-nitrobenzoyl **29** (pEC<sub>50</sub>=7.15)] and the *N*-*para*-methyl-phenyl ana-

logues [cf. 3-methylbenzoyl **39** (pEC<sub>50</sub>=7.40) with 2-methylbenzoyl **37** (pEC<sub>50</sub>=6.70) and 3-nitrobenzoyl **40** (pEC<sub>50</sub>=7.70) with 2-nitrobenzoyl **38** (pEC<sub>50</sub>=6.40)]. The unfavourable substitution of the *ortho* position of ring C is also demonstrated by the lower pEC<sub>50</sub> values of the 2,6-dichloro- and 2,6-dimethoxy ATCs compared to those of the less bulky 2,6-difluoro congeners [cf. the *N*-*para*-chlorophenyl derivatives **51** (pEC<sub>50</sub>=5.25) and **54** (pEC<sub>50</sub>=6.40) with **46** (pEC<sub>50</sub>=7.70), and the *N*-*para*-methyl-phenyl analogues **66** (pEC<sub>50</sub>=5.05) and **68** (pEC<sub>50</sub>=4.66) with **62** (pEC<sub>50</sub>=6.00)].

According to the electrostatic fields contour map of the CoMFA analysis plotted in Fig. 10, more electropositive substituents are predicted to be favoured (blue region) around the *ortho* position of ring C, while less positive moieties are predicted to be favoured (red region) in the proximity of the positions *para* and *meta* of both rings B and C. These results are supported by the evidence of stronger activity of ATC **12** (*N*-*meta*-nitrophenyl, pEC<sub>50</sub>=6.42) compared to the other *N*-*meta*-substituted-phenyl ATCs (**6–11**, **13**: pEC<sub>50</sub>=4.96–5.92), and by the high pEC<sub>50</sub> values of the isonicotinoyl and 3- and 4-nitrobenzoyl ATCs (*N*-*para*-chlorophenyl **20**, **21**, **31**, **34**, **55–57**, pEC<sub>50</sub>=7.22–8.82; *N*-*para*-methylphenyl **40**, **42**, **69**, **70**, pEC<sub>50</sub>=7.05–7.70). The reliability of the electrostatic map calculation is also highlighted by the higher potency of the *N*-*para*-chlorophenyl ATCs compared to the corresponding *N*-*para*-methylphenyl congeners [**28** (pEC<sub>50</sub>=7.40) vs **37** (pEC<sub>50</sub>=6.70); **29** (pEC<sub>50</sub>=7.15) vs **38** (pEC<sub>50</sub>=6.40); **30** (pEC<sub>50</sub>=8.00) vs **39** (pEC<sub>50</sub>=7.40); **31** (pEC<sub>50</sub>=8.82) vs **40** (pEC<sub>50</sub>=7.70); **44** (pEC<sub>50</sub>=8.00) vs **60** (pEC<sub>50</sub>=7.52); **45** (pEC<sub>50</sub>=7.70) vs **61** (pEC<sub>50</sub>=7.40); **46** (pEC<sub>50</sub>=7.70) vs **62** (pEC<sub>50</sub>=6.00); **48** (pEC<sub>50</sub>=8.15) vs **63** (pEC<sub>50</sub>=7.30); **49** (pEC<sub>50</sub>=7.00) vs **64** (pEC<sub>50</sub>=6.70); **50** (pEC<sub>50</sub>=7.52) vs **65** (pEC<sub>50</sub>=6.52); **51** (pEC<sub>50</sub>=5.25) vs **66** (pEC<sub>50</sub>=5.05); **52** (pEC<sub>50</sub>=8.05) vs **67** (pEC<sub>50</sub>=7.30); **54** (pEC<sub>50</sub>=6.40) vs **68** (pEC<sub>50</sub>=4.66); **55** (pEC<sub>50</sub>=8.10) vs **69** (pEC<sub>50</sub>=7.05); **57** (pEC<sub>50</sub>=7.70) vs **70** (pEC<sub>50</sub>=7.22); **58** (pEC<sub>50</sub>=8.15) vs **71** (pEC<sub>50</sub>=7.30)].

**Fig. 8** Distribution of experimental and predicted pEC<sub>50</sub> values for training set compounds (a) and test set compounds (b) according to CoMSIA analysis

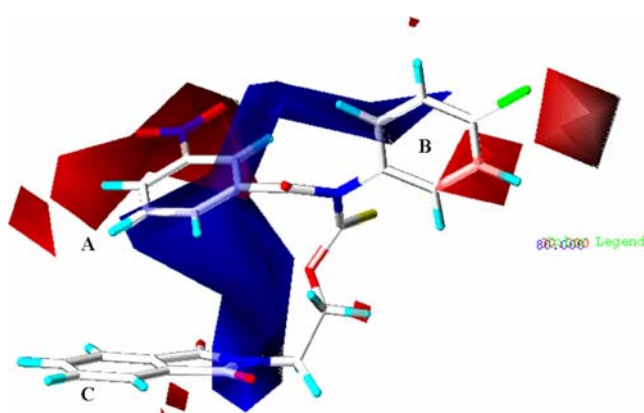


**Fig. 9** Contour maps of CoMFA steric regions (*green* favoured, *yellow* disfavoured) are shown around compound **31**. The compound is depicted in *stick* and coloured by atom type

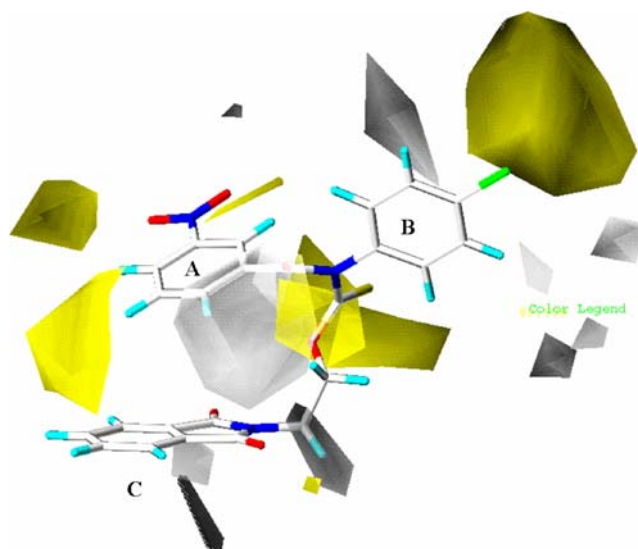
The CoMSIA steric and electrostatic regions are in agreement with the CoMFA steric and electrostatic areas.

#### CoMSIA hydrophobic and H-bond acceptor regions

The calculated CoMSIA hydrophobic contours (Fig. 11) predict favourable hydrophobic substituents (yellow areas) around the sulphur atom, the *para* position of ring B and C, and in the proximity of one *meta* position and one *ortho* position of ring C. On the contrary, lipophilic groups seem to be detrimental for activity (white areas) around the other *ortho* position of ring C and one *meta* position of ring B, in the proximity of the methylene adjacent to the phthalimide nitrogen and around one of the two phthalamide carbonyls. The reliability of this hydrophobic map calculation is underlined by the activity trend of the 2-thenoyl ATCs [cf. *N*-*para*-chlorophenyl **19** ( $pEC_{50}=8.10$ ) and *N*-*para*-iodo **23** ( $pEC_{50}=8.00$ ) with *N*-*para*-fluoro **15** ( $pEC_{50}=7.00$ )], the

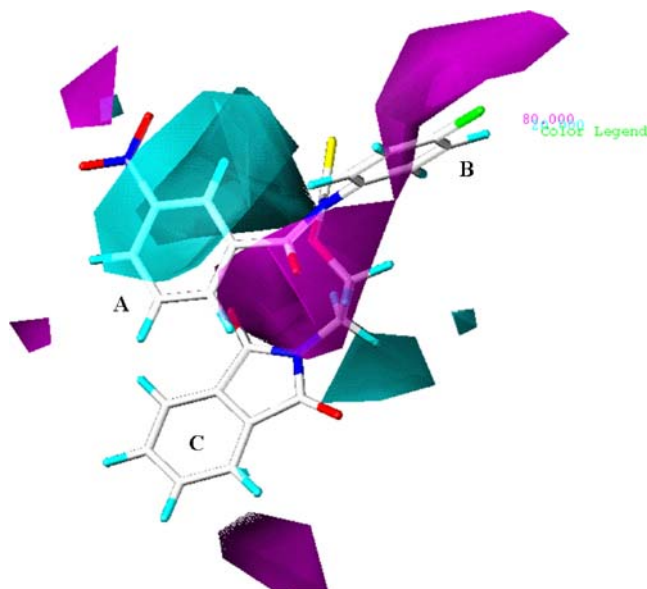


**Fig. 10** Contour maps of CoMFA electrostatic regions are shown around compound **31**, depicted in *stick* and coloured by atom type. *Blue* Favourable for more positively charged groups, *red* favourable for less positively charged groups



**Fig. 11** Contour maps of the CoMSIA hydrophobic regions (*yellow* favoured, *white* disfavoured) are shown around compounds **31**, reported in *stick* and coloured by atom type

benzoyl derivatives [*N*-*para*-chloro **16** ( $pEC_{50}=7.00$ ) and *N*-*para*-bromo **22** ( $pEC_{50}=7.46$ ) vs *N*-unsubstituted **1** ( $pEC_{50}=6.40$ )], the 3,4-dichloro ATCs [cf. *N*-*para*-chloro **52** ( $pEC_{50}=8.05$ ) and *N*-*para*-methyl **67** ( $pEC_{50}=7.30$ ) with *N*-unsubstituted **72** ( $pEC_{50}=6.30$ )] and the 4-chloro-3-nitro derivatives [*N*-*para*-chloro **57** ( $pEC_{50}=7.70$ ) and *N*-*para*-methyl **70** ( $pEC_{50}=7.22$ ) vs *N*-unsubstituted **73** ( $pEC_{50}=6.46$ )]. In addition, the results are in agreement with the high  $pEC_{50}$  values of the 2- and 6-chloro-isonicotinoyl ATCs **20** ( $pEC_{50}=8.30$ ) and **21** ( $pEC_{50}=8.22$ ), the 4-chlorobenzoyl derivative **17** ( $pEC_{50}=7.60$ ), the



**Fig. 12** CoMSIA hydrogen bond acceptor polyhedra are reported around compounds **31** depicted in *stick* and coloured by atom type. H-bond acceptor groups: *magenta* favoured, *cyan* disfavoured



2-, 3- and 4-methylbenzoyl ATCs **28**, **30**, **32**, **39**, **41** ( $pEC_{50}$ =7.30–8.00), the 4-trifluoromethylbenzoyl derivative **33** ( $pEC_{50}$ =8.30), the 2,4-, 3,4-, 3,5-dichlorobenzoyl analogues **50**, **52**, **53**, **67** ( $pEC_{50}$ =7.30–8.10), the 4-bromo-3-methylbenzoyl ATCs **58** ( $pEC_{50}$ =8.15) and **71** ( $pEC_{50}$ =7.30), the 2,5-dimethylfuroyl analogue **59** ( $pEC_{50}$ =7.70), the 1- and 2-naphthoyl ATCs **74** ( $pEC_{50}$ =7.52) and **75** ( $pEC_{50}$ =7.70).

To take into account the role of H-bond acceptor groups for antiretroviral activity, the corresponding CoMSIA contours were calculated (Fig. 12) (CoMSIA H-bond acceptor map corresponds to the H-bond donating groups of the receptor).

As shown in Fig. 12, H-bond acceptor groups are predicted to be favoured (magenta regions) around position 3 of the phthalamide substructure, the carbonyl of the acyl moiety and positions *para* of ring B and *meta* of ring C. Moreover, H-bond acceptor functions would be unfavourable (cyan polyhedra) around the methylene adjacent to the phthalamide nitrogen and one of the *ortho* positions of ring B. These results are supported by the evidence of the high potency of the *N*-*para*-nitrophenyl ATCs **24** ( $pEC_{50}$ =8.10) and **25** ( $pEC_{50}$ =8.00) and of the 3-nitrobenzoyl ATCs **31** ( $pEC_{50}$ =8.82) and **40** ( $pEC_{50}$ =7.70).

The information obtained by these modelling and 3D-QSAR studies provides useful suggestions for the synthesis of ATCs endowed with higher potency and improved resistance profiles. The introduction of a H-bond acceptor group on phthalamide position 3 (for example the introduction of a nitro group or isosteric replacement of the ring carbon atom with a nitrogen atom) might allow the establishment of a H-bond interaction with the  $\epsilon$ -amino group of K101 or K103. The *para* position of ring B might also be exploited to establish hydrophobic contacts with P236 by introducing bulky lipophilic substituents such as a cyclohexyl or phenyl ring. The introduction of an aliphatic chain at position 5 of the isonicotinoyl or 3-nitrobenzoyl acyl moiety (acyl groups bearing a favourable H-bond acceptor function) might allow hydrophobic contacts to be made with L234. Notably, L234 and P236 (along with W229) are highly conserved amino acids in the non-nucleoside RT inhibitor binding site and therefore are recognised to be of strategic relevance in the design of new non-nucleoside RT inhibitors more resilient to the effects of RT mutations in this site [12].

## Conclusions

The docking analysis and 3D-QSAR studies presented here highlight the main interactions responsible for ATC antiretroviral activity. Moreover, they provide useful suggestions for the synthesis of new analogues with improved

potency, also against clinically resistant mutants. In the future, the models elaborated will be exploited to design new ATCs and predict their activity prior to synthesis.

**Acknowledgements** This work was supported by University of Genova, Progetto Ateneo 2007. Fondazione Carige is gratefully acknowledged for financially supporting E.C. and S.C.

## References

- Jonckheere H, Anne J, De Clercq E (2000) The HIV-1 reverse transcription (RT) process as target for RT inhibitors. *Med Res Rev* 20:129–154
- De Clercq E (2001) New developments in anti-HIV chemotherapy. *Farmacologia* 56:3–12
- De Clercq E (2005) Emerging anti-HIV drugs. *Expert Opin Emerg Drugs* 10:241–273
- De Clercq E (2005) New approaches toward anti-HIV chemotherapy. *J Med Chem* 48:1297–1313
- Barbaro G, Scozzafava A, Mastrolorenzo A, Supuran CT (2005) Highly active antiretroviral therapy: current state of the art, new agents and their pharmacological interactions useful for improving therapeutic outcome. *Curr Pharm Des* 11:1805–1843
- Balzarini J (2004) Current status of the non-nucleoside reverse transcriptase inhibitors of human immunodeficiency virus type 1. *Curr Top Med Chem* 4:921–944
- De Clercq E (1998) The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. *Antiviral Res* 38:153–179
- Pedersen OS, Pedersen EB (1999) Non-nucleoside reverse transcriptase inhibitors, the NNRTI boom. *Antiviral Chem Chemother* 10:285–314
- Pedersen OS, Pedersen EB (2000) The flourishing syntheses of non-nucleoside reverse transcriptase inhibitors. *Synthesis* 4:479–495
- Campiani G, Ramunno A, Maga G, Nacci V, Fattorusso C, Catalanotti B, Morelli E, Novellino E (2002) Non-nucleoside HIV-1 reverse transcriptase (RT) inhibitors: past, present, and future perspectives. *Curr Pharm Des* 8:615–657
- De Clercq E (2004) Non-nucleoside reverse transcriptase inhibitors (NNRTIs): past, present, and future. *Chem Biodiversity* 1:44–64
- Pauwels R (2004) New non-nucleoside reverse transcriptase inhibitors (NNRTIs) in development for the treatment of HIV infections. *Curr Opin Pharmacol* 4:437–446
- Leigh Brown AJ, Frost SD, Mathews WC, Dawson K, Hellmann NS, Daar ES, Richman DD, Little SJ (2003) Transmission fitness of drug-resistant human immunodeficiency virus and the prevalence of resistance in the antiretroviral-treated population. *J Infect Dis* 187:683–686
- Richman DD, Morton SC, Wrin T, Hellmann N, Berry S, Shapiro MF, Bozzette SA (2004) The prevalence of antiretroviral drug resistance in the United States. *AIDS* 18:1393–1401
- Spallarossa A, Cesarini S, Ranise A, Schenone S, Bruno O, Borassi A, La Colla P, Pezzullo M, Sanna G, Collu G, Secci B, Loddio R (2008) Parallel synthesis, molecular modeling and further structure-activity relationship studies of new acylthiocarbamates as potent non-nucleoside HIV-1 reverse transcriptase inhibitors. *Eur J Med Chem*. doi:10.1016/j.ejmech.2008.10.032
- Ranise A, Spallarossa A, Schenone S, Bruno O, Bondavalli F, Vargiu L, Marceddu T, Mura M, La Colla P, Pani A (2003) Design, synthesis, SAR, and molecular modeling studies of acylthiocarbamates: a novel series of potent non-nucleoside HIV-

- 1 reverse transcriptase inhibitors structurally related to phenethylthiazolylthiourea derivatives. *J Med Chem* 46:768–781
17. Spallarossa A, Cesarini S, Ranise A, Ponassi M, Unge T, Bolognesi M (2008) Crystal structures of HIV-1 reverse transcriptase complexes with thiocarbamate non-nucleoside inhibitors. *Biochem Biophys Res Commun* 365:764–770
  18. Spallarossa A, Cesarini S, Ranise A, Bruno O, Schenone S, La Colla P, Collu G, Sanna G, Secci B, Loddo R (2008) Novel modifications in the series of *O*-(2-phthalimidoethyl)-*N*-substituted thiocarbamates and their ring-opened congeners as non-nucleoside HIV-1 reverse transcriptase inhibitors. *Eur J Med Chem*. doi:10.1016/j.ejmech.2008.09.024
  19. Ranise A, Spallarossa A, Cesarini S, Bondavalli F, Schenone S, Bruno O, Menozzi G, Fossa P, Mosti L, La Colla M, Sanna G, Murreddu M, Collu G, Busonera B, Marongiu ME, Pani A, La Colla P, Loddo R (2005) Structure-based design, parallel synthesis, structure–activity relationship, and molecular modeling studies of thiocarbamates, new potent non-nucleoside HIV-1 reverse transcriptase inhibitor isosteres of phenethylthiazolylthiourea derivatives. *J Med Chem* 48:3858–3873
  20. Ren J, Diprose J, Warren J, Esnouf RM, Bird LE, Ikemizu S, Slater M, Milton J, Balzarini J, Stuart DI, Stammers DK (2000) Phenylethylthiazolylthiourea (PETT) non-nucleoside inhibitors of HIV-1 and HIV-2 reverse transcriptases. Structural and biochemical analyses. *J Biol Chem* 275:5633–5639
  21. MOE: Chemical Computing Group Inc, Montreal, H3A 2R7 Canada. <http://www.chemcomp.com>
  22. Sybyl 7.0, Tripos Inc, St Louis, MO
  23. Cramer RD III, Patterson DE, Bunce JD (1989) Recent advances in comparative molecular field analysis (CoMFA). *Prog Clin Biol Res* 291:1–165
  24. Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J Med Chem* 37:4130–4146