European Journal of Medicinal Chemistry 44 (2009) 2059–2070

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Thiocarbamates as non-nucleoside HIV-1 reverse transcriptase inhibitors: Docking-based CoMFA and CoMSIA analyses

Elena Cichero*, Sara Cesarini, Paola Fossa, Andrea Spallarossa, Luisa Mosti

Dipartimento di Scienze Farmaceutiche, Università degli Studi di Genova, Viale Benedetto XV n.3, 16132 – Genova. Italy

ARTICLE INFO

Article history:
Received 27 June 2008
Received in revised form
10 October 2008
Accepted 14 October 2008
Available online 1 November 2008

Keywords:
HIV-1
Reverse transcriptase
3D-QSAR
CoMFA
CoMSIA
Thiocarbamates

ABSTRACT

Thiocarbamates (TCs) have been recently identified as a new class of potent non-nucleoside HIV-1 Reverse Transcriptase (RT) inhibitors. A computational strategy based on molecular docking studies, followed by CoMFA and CoMSIA analyses, has been used to elucidate the atomic details of the RT/TC interactions and to identify the most important features impacting the TC antiretroviral activity. The CoMFA model resulted to be the more predictive, and gave $r^2 = 0.93$, $r^2_{cv} = 0.53$, SEE = 0.292, $F = 180$, and $r^2_{test} = 0.70$. The 3D-QSAR field contributions and the structural features of the RT binding site showed a good correlation. These studies will be useful to design new TCs with improved potency also against clinically relevant resistant mutants.

1. Introduction

Reverse transcriptase (RT) is a key enzyme in the HIV replication cycle and is one of the main targets in the development of drugs for treating HIV-infection and AIDS [1–5]. RT catalyzes the conversion of viral RNA into double stranded DNA that is then integrated in the host genome. Non-nucleoside RT inhibitors (NNRTIs) bind to an allosteric hydrophobic pocket (NNRTI binding site: NNIBS), located at about 10 Å far from the polymerase active site and created upon inhibitor interaction, and lock the enzyme into an inactive form by affecting the geometry of the polymerase active site aspartyl residues [6]. In the past fifteen years, more than fifty structurally diverse NNRTIs have been described [6–12]. The fact that cross-resistance extends to the whole NNRTI class calls for development of new agents capable of inhibiting clinically relevant NNRTI-resistant mutants [13,14].

Thiocarbamates (TCs), and in particular *O*-(2-phthalimidoethyl)–*N*-arylthiocarbamates (C-TCs) [15,16], their corresponding ring-opened analogues [15,16] and their non-phthalimidic congeners [17,18], have been recently described as a novel class of potent NNRTIs.

We performed a computational study with the aim to elucidate the most important RT/TC interactions and to identify the features really impacting the TC antiretroviral activity. Our aim was also to elaborate a quantitative structure–activity relationship (QSAR) model for enabling the TC antiretroviral activity prediction prior to synthesis, and to obtain useful suggestions for the design of new TCs with improved potency also against clinically relevant resistant mutants.

On the basis of the crystal structure of RT in complex with *O*-[2-(phthalimido)ethyl]-*N*-(4-chlorophenyl)thiocarbamate [19], a flexible docking simulation was performed on a series of 111 TCs [15–18]. The most probable docking poses were selected and submitted to 3D-QSAR studies involving Comparative Molecular Fields Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA).

2. Computational details

2.1. Dataset

A dataset of 111 TCs, screened according to the same pharmacological protocol, was selected from four sample data sets [15–18] and submitted to QSAR analysis. The molecular structures of TCs **1**–**111** (Tables 1–4) were built, parameterized (Gasteiger–Huckel method) and energy minimized within MOE using MMFF94 force field [20].

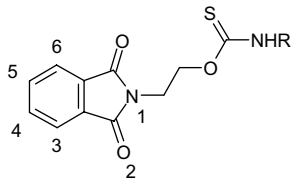
2.2. Docking-based ligand alignment

To locate the appropriate binding orientations and conformations of TCs within the NNIBS, a computational searching method

* Corresponding author. Tel.: +39 010 3538361; fax: +39 010 3538358.
E-mail address: cichero@unige.it (E. Cichero).

0223-5234/$ – see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.10.014

Table 1
Molecular structure of TC **1**–**44** (group 1)



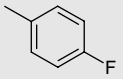
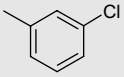
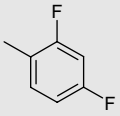
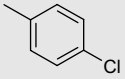
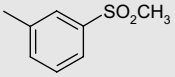
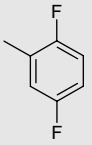
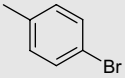
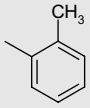
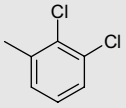
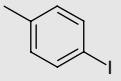
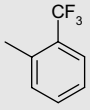
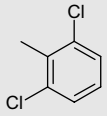
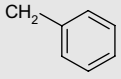
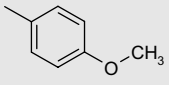
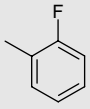
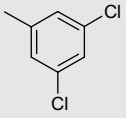
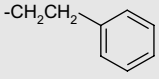
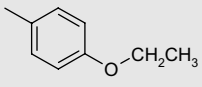
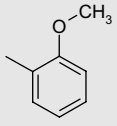
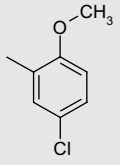
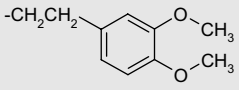
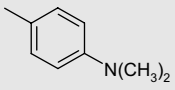
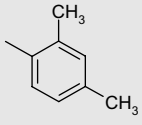
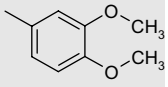
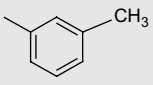
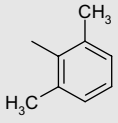
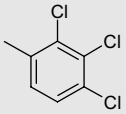
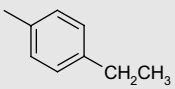
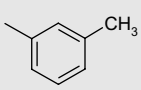
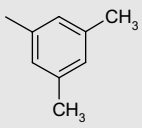
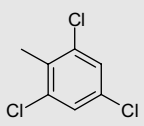
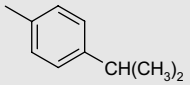
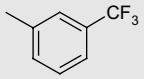
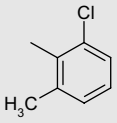
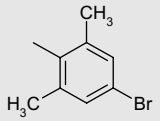
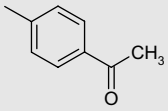
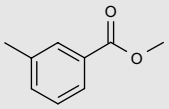
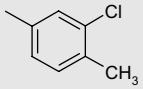
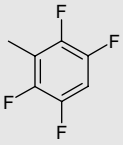
Compound	R	Compound	R	Compound	R	Compound	R
1	$-CH_2(CH_2)_2CH_3$	**12**		**23**		**34**	
2		**13**		**24**		**35**	
3		**14**		**25**		**36**	
4		**15**		**26**		**37**	
5		**16**		**27**		**38**	
6		**17**		**28**		**39**	
7		**18**		**29**		**40**	
8		**19**		**30**		**41**	
9		**20**		**31**		**42**	

Table 1 (*continued*)

Compound	R	Compound	R	Compound	R	Compound	R
10		**21**		**32**		**43**	
11		**22**		**33**		**44**	

was applied. Starting from a database including all the 111 compounds, a docking procedure was performed. Thus, on the basis of the three-dimensional structure co-ordinates of the RT/**13** complex (PDB entry 2VG5) [19] (**13** in the bioactive conformation), each inhibitor was docked into the NNIBS using the flexible docking module implemented in MOE. Since for all compounds the best-docked geometries, evaluated in terms of "Affinity dG" (kcal/mol of total estimated binding energy), were in agreement with the crystallographic data of the RT/**13** complex (and thus already aligned), they were directly submitted to CoMFA [21] and CoMSIA [22] studies by Sybyl7.0 software [23].

2.3. 3D-QSAR analysis

2.3.1. Training set and test set

All the compounds were grouped into a training set, for model generation, and a test set, for model validation, containing 94 and 17 compounds, respectively. The molecules of the test set represent 20% (estimated as a good percentage to validate a molecular model) of the training set. 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_{50} values have been transformed into pEC_{50} values and then used as response variables. Compounds enzyme inhibitory activity covered 4 log orders.

2.3.2. CoMFA and CoMSIA analyses

CoMFA method is a widely used 3D-QSAR technique to relate the biological activity of a series of molecules with their steric and electrostatic fields, which are calculated 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, were calculated using the standard Tripos force field method. The column-filtering threshold value was set to 2.0 kcal/mol to improve the signal–noise ratio. A methyl probe with +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.

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

2.3.3. Partial least square analysis and models validation

Partial least-squares (PLS) method 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 the PLS analysis, CoMFA and CoMSIA columns with a smaller variance than 2.0 kcal mol^{-1} were filtered by using column filtering to improve the signal-to-noise ratio.

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-validated r^2 (r^2_{cv}). In the LOO methodology, one molecule is omitted from the dataset and a model is derived involving the rest of the compounds. Then, employing this model, the activity of the omitted molecule is predicted.

The ONC obtained from cross-validation methodology was used in the subsequent regression model. Final CoMFA and CoMSIA models were generated using noncross-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 some others might be included more then once. The mean correlation coefficient is represented as bootstrap r^2(r^2_{boot}). To validate the CoMFA and CoMSIA derived models, the predictive ability for the test set of compounds (expressed as r^2_{pred}) was determined by 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.

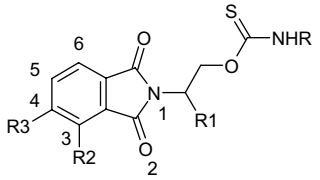
All the calculations were carried out using a PC with windows XP operative system and a PC with Linux Red Hat operative system.

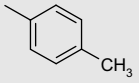
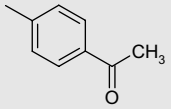
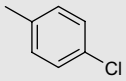
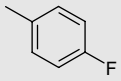
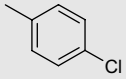
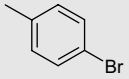
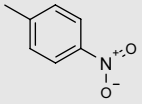
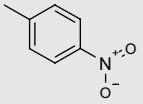
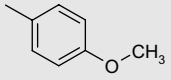
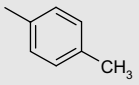
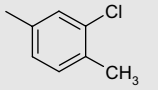
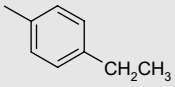
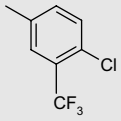
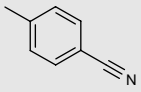
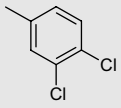
3. Results and discussion

3.1. Docking-based ligand alignment

Computational studies had already been performed to construct a docking model of TCs into the HIV-1 RT NNIBS [15–18]: low energy conformers of the TCs had been docked in the NNIBS using the X-ray co-ordinates of the crystal complex of the NNRTI PETT-1 with RT, after erasing PETT-1 from the complex. In this work, for the molecular modelling analysis we used the X-ray co-ordinates of the crystal structure of TC **13**/RT complex. The results of our docking calculations, performed by a different procedure (see Section 2) in

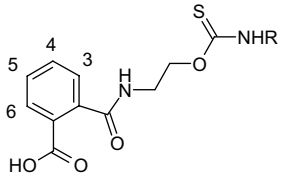
Table 2
Molecular structure of TC **45**–**60** (group 1)



Compound	R	R1	R2	R3	Compound	R	R1	R2	R3
45 (*R*)		CH_3	H	H	**53**		H	H	CH_3
46 (*S*)		CH_3	H	H	**54**		H	H	CH_3
47		H	CH_3	H	**55**		H	H	CH_3
48		H	CH_3	H	**56**		H	H	CH_3
49		H	H	CH_3	**57**		H	H	CH_3
50		H	H	CH_3	**58**		H	H	CH_3
51		H	H	CH_3	**59**		H	H	CH_3
52		H	H	CH_3	**60**		H	H	CH_3

comparison with the precedent studies, are in agreement with the results previously obtained. Briefly, according to our calculations, all inhibitors display an H-bond between the K101 main chain carbonyl and the thiocarbamic NH group, while several lipophilic interactions with two hydrophobic pockets (P_1 and P_2) are detected (Fig. 1). Region P_1 includes P95, Y181, Y188 and W229 residues, while P_2 consists of K103, V106, F227, L234 and P236 residues. The phthalimide moiety of compounds **1**–**60** (group 1, Tables 1 and 2), which include the most active molecules of the dataset, is located in the P_1 region establishing π–π stacking with Y181, Y188 and W229. Similarly, the benzoic phenyl ring of compounds **61**–**79** (group 2, Table 3) and part of the (hetero)aromatic ring of compounds **80**–**111** (group 3, Table 4) occupy the P_1 pocket. On the other side, the *N*-phenyl ring of all the compounds is oriented towards the P_2 pocket, establishing hydrophobic contacts with the side chains of V106, F227, L234 and P236. It appears evident that the difference in the activity of the three groups of TCs is mainly due to the different occupancy of pocket P_1. In order to validate the docking procedure, we compared the most probable binding pose, selected for each compound, with the X-ray structure of **13** in complex with RT, obtaining ligand conformations perfectly superposed on the available crystallographic data. Thus, they resulted to be aligned together inside the NNIBS for the following 3D-QSAR analysis.

Table 3
Molecular structure of TC **61**–**79** (group 2)



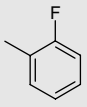
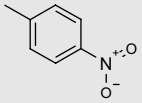
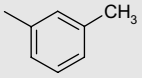
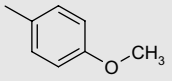
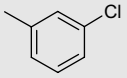
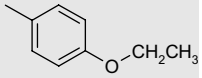
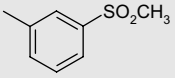
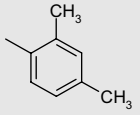
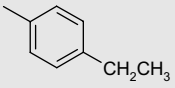
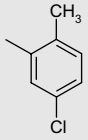
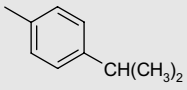
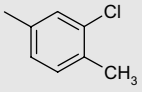
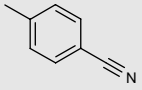
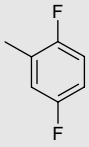
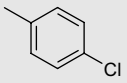
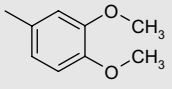
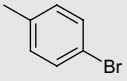
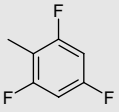
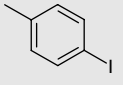
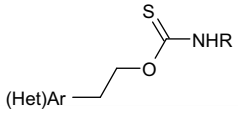
Compound	R	Compound	R
61		**71**	
62		**72**	
63		**73**	
64		**74**	
65		**75**	
66		**76**	
67		**77**	
68		**78**	
69		**79**	
70			

Table 4
Molecular structure of TC **80**–**111** (group 3)



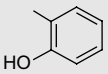
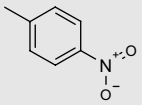
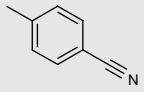
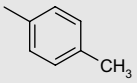
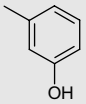
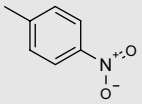
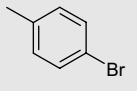
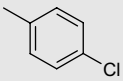
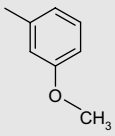
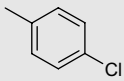
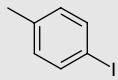
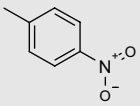
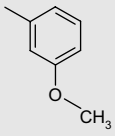
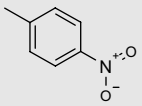
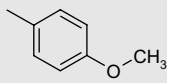
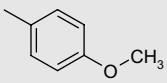
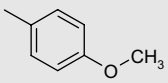
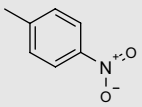
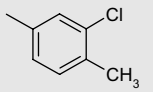
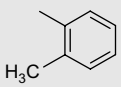
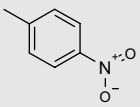
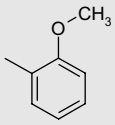
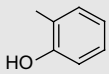
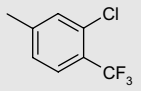
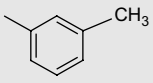
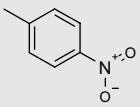
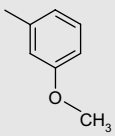
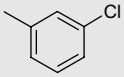
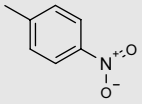
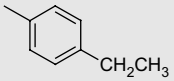
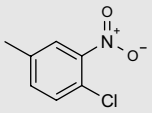
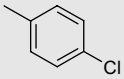
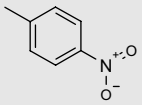
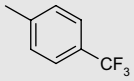
Compound	(Het) Ar	R	Compound	(Het) Ar–	R	Compound	(Het) Ar	R
80			**91**			**102**		
81			**92**			**103**		
82			**93**			**104**		
83			**94**			**105**		
84			**95**			**106**		
85			**96**			**107**		
86			**97**			**108**		
87			**98**			**109**		
88			**99**			**110**		

Table 4 (*continued*)

Compound	(Het) Ar	R	Compound	(Het) Ar–	R	Compound	(Het) Ar	R
89			**100**			**111**		
90			**101**					

3.2. CoMFA and CoMSIA analyses

The docking-based CoMFA analysis was performed dividing compounds **1**–**111** into a training set (**2**–**10**, **12**–**26**, **29**–**34**, **36**–**40**, **43**–**50**, **52**–**57**, **60**, **61**, **63**–**65**, **67**–**76**, **78**–**89**, **91**–**98**, **100**–**104**, **106**–**108**, **110**, **111**), for model generation, and into a test set (**1**, **11**, **27**, **28**, **35**, **41**, **42**, **51**, **58**, **59**, **62**, **66**, **77**, **90**, **99**, **105**, **109**), for model validation. CoMFA and CoMSIA studies were developed using, respectively, CoMFA steric and electrostatic fields, and CoMSIA hydrophobic, H-bond acceptor and H-bond donor properties, as independent variables, and the ligand pEC_{50} as dependent variable.

The final CoMFA model was generated employing noncross-validated PLS analysis with the optimum number of components (ONC) = 6 to give a noncross-validated $r^2(r^2_{ncv}) = 0.93$, standard error of estimate (SEE) = 0.292, steric contribution = 0.428 and

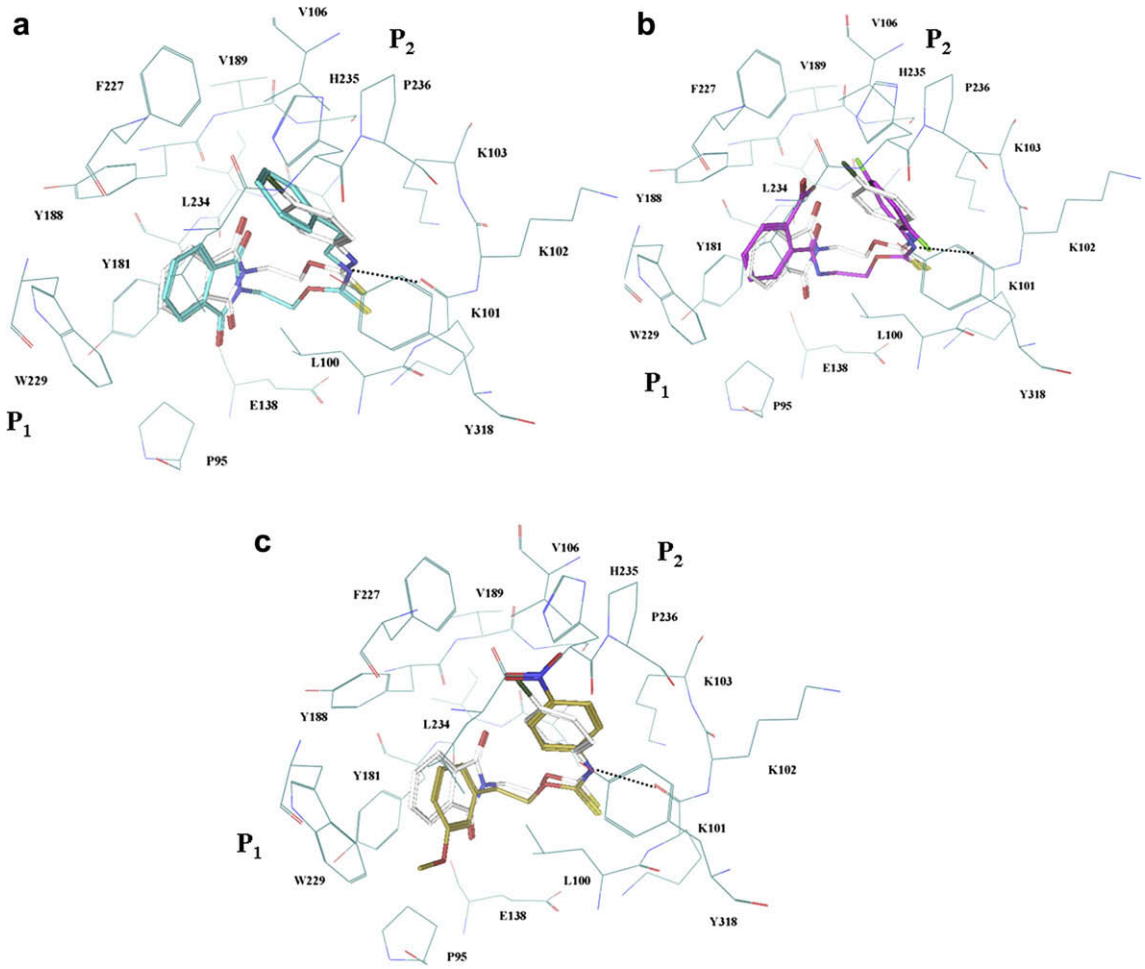


Fig. 1. Docking analysis into HIV-1 reverse transcriptase X-ray structure: compound **13** X-ray pose is reported in stick coloured by atom type. (a) Docking pose of **6**, depicted in stick (colour code: C. green; N. blue; O. red; S. yellow); (b) docking pose of **77**, reported in stick (colour code: C. magenta; N. blue; O. red; S. yellow), (c) docking pose of **94**, depicted in stick (colour code: C. yellow; N. blue; O. red; S. yellow). The most important residues are labelled. Hydrogen bonds are coloured in black. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Summary of CoMFA results

No. compounds	94
Optimal number of components (ONC)	6
Leave one out $r^2(r^2_{loo})$	0.504
Cross-validated $r^2(r^2_{cv})$	0.531
Standard error of estimate (SEE)	0.292
Noncross-validated $r^2(r^2_{ncv})$	0.93
F value	180.561
Steric contribution	0.428
Electrostatic contribution	0.572
Bootstrap $r^2(r^2_{boot})$	0.948
Standard error of estimate r^2_{boot}	0.243
Test set $r^2(r^2_{pred})$	0.70

electrostatic contribution = 0.572. The model reliability thus generated was supported by bootstrapping results. All statistical parameters supporting CoMFA model are reported in Table 5.

A CoMSIA model consisting of hydrophobic, H-bond donor and H-bond acceptor fields with a $r^2_{ncv} = 0.83$, SEE = 0.446, hydrophobic contribution = 0.465, H-bond donor contribution = 0.151 and H-bond acceptor contribution = 0.383 was derived. All statistical parameters supporting CoMSIA model are reported in Table 6.

Experimental and predicted binding affinities' values for the training set and test set are reported in Tables 7 and 8 respectively, while distribution of experimental and predicted pEC_{50} values for the training set and test set according to CoMFA and CoMSIA models are represented in Fig. 2. The CoMFA model resulted to be the more predictive.

Since CoMFA and CoMSIA field effects on the target properties can be viewed as 3D coefficient contour plots (identifying important regions where any change in these fields may affect the biological activity), they could be helpful to optimize TCs as NNRTIS. The 3D-QSAR analysis maps are described and discussed in the following sections.

3.2.1. CoMFA steric and electrostatic regions

As shown in Fig. 3, the steric contour map predicts favourable interaction polyhedra for the 3 and 4 positions of the phthalimide moiety of TCs **1**–**60** (group 1), for the corresponding 3 and 5 positions of the ring opened TCs **61**–**79** (group 2) and for the *para*-position of the (hetero)aryl portion of TCs **80**–**111** (group 3). These results are in agreement with the highest pEC_{50} values of TCs **47**–**60**, bearing a methyl at phthalimide positions 3 (**47**, **48**) or 4 (**49**–**60**), among TCs of group 1. For all the compounds, the *N*-phenyl ring, and in particular its *para*-substituent, is oriented toward a sterically favoured green region, while its *meta*-position is surrounded by sterically disfavoured yellow polyhedra. The reliability of the steric map calculation is underlined by the activity trend of the *N*-*para*-phenyl

Table 6
Summary of CoMSIA results

No. compounds	94
Optimal number of components (ONC)	7
Leave one out $r^2(r^2_{loo})$	0.413
Cross-validated $r^2(r^2_{cv})$	0.502
Standard error of estimate (SEE)	0.446
Noncross-validated $r^2(r^2_{ncv})$	0.830
F value	59.532
Hydrophobic contribution	0.465
H-bond donor contribution	0.151
H-bond acceptor contribution	0.383
Bootstrap $r^2(r^2_{boot})$	0.863
Standard error of estimate (r^2boot)	0.398
Test set $r^2(r^2_{pred})$	0.52

Table 7
Experimental and predicted pEC_{50} values of training set compounds (**2**–**10**, **12**–**26**, **29**–**34**, **36**–**40**, **43**–**50**, **52**–**57**, **60**, **61**, **63**–**65**, **67**–**76**, **78**–**89**, **91**–**98**, **100**–**104**, **106**–**108**, **110**, **111**)

Compound	Experimental pEC_{50}	CoMFA Model		CoMSIA Model	
		Predicted pEC_{50}	Residual	Predicted pEC_{50}	Residual
2	4.95	5.18	−0.23	5.26	−0.31
3	6.52	6.40	0.12	5.93	0.60
4	4.22	4.10	0.12	4.46	−0.24
5	4.58	4.11	0.47	4.13	0.46
6	5.04	5.48	−0.44	5.33	−0.29
7	5.09	4.73	0.36	5.33	−0.24
8	5.92	6.49	−0.57	6.00	−0.08
9	7.10	7.02	0.08	6.49	0.61
10	7.40	6.76	0.64	6.62	0.79
12	7.00	7.06	−0.06	6.54	0.46
13	7.40	6.87	0.54	6.84	0.56
14	7.52	7.09	0.43	6.84	0.68
15	7.70	7.66	0.04	7.62	0.08
16	7.52	7.60	−0.08	6.25	1.27
17	5.15	5.80	−0.65	5.59	−0.44
18	6.22	6.56	−0.34	6.29	−0.07
19	5.75	6.06	−0.31	6.20	−0.45
20	5.82	5.63	0.19	5.65	0.17
21	4.87	4.70	0.17	5.08	−0.21
22	4.68	4.71	−0.03	5.84	−1.16
23	5.70	5.81	−0.11	6.33	−0.63
24	5.15	5.29	−0.14	5.03	0.12
25	5.43	5.13	0.31	4.97	0.46
26	4.19	4.58	−0.39	4.94	−0.75
29	5.22	5.15	0.07	5.85	−0.63
30	4.12	4.30	−0.18	4.45	−0.33
31	4.92	4.93	−0.01	4.56	0.36
32	4.64	4.70	−0.06	4.49	0.15
33	6.10	6.34	−0.24	6.69	−0.59
34	6.30	6.17	0.13	5.78	0.52
36	4.52	4.81	−0.29	4.86	−0.34
37	4.68	5.03	−0.35	4.24	0.44
38	5.40	5.70	−0.30	5.16	0.24
39	4.32	4.11	0.21	4.55	−0.23
40	5.45	5.05	0.40	5.79	−0.34
43	5.05	4.90	0.15	5.39	−0.34
44	5.00	5.35	−0.35	5.44	−0.44
45	7.15	6.66	0.49	6.72	0.43
46	6.70	7.04	−0.34	6.49	0.22
47	7.30	7.25	0.05	6.97	0.33
48	7.15	6.78	0.37	7.22	−0.07
49	6.52	6.88	−0.36	6.54	−0.02
50	7.70	7.67	0.03	7.44	0.26
52	8.00	7.78	0.22	8.08	−0.08
53	7.15	7.29	−0.14	8.19	−1.04
54	7.52	7.43	0.09	7.15	0.37
55	8.10	7.81	0.30	7.83	0.27
56	8.00	7.91	0.09	8.32	−0.32
57	7.52	7.69	−0.17	6.95	0.57
59	6.30	5.75	0.55	6.44	−0.14
60	6.70	7.00	−0.30	7.24	−0.54
63	5.52	5.57	−0.05	5.29	0.23
64	4.96	4.63	0.33	4.68	0.28
65	6.52	6.37	0.15	6.36	0.16
67	7.00	7.07	−0.07	7.43	−0.43
68	7.00	7.10	−0.09	7.07	−0.07
69	7.00	6.98	0.02	6.60	0.40
70	6.89	6.94	−0.05	6.37	0.52
71	6.52	6.79	−0.27	6.63	−0.11
72	6.69	6.87	−0.18	6.79	−0.10
73	5.70	5.15	0.55	5.54	0.17
74	4.30	4.05	0.25	4.41	−0.11
75	4.53	4.46	0.07	4.69	−0.16
76	5.37	5.58	−0.21	5.48	−0.11
78	4.79	5.25	−0.46	5.05	−0.26
79	5.14	5.25	−0.11	4.95	0.19
80	5.22	5.63	−0.41	5.45	−0.23
81	5.37	5.58	−0.21	5.34	0.03
82	6.30	5.96	0.34	5.99	0.31
83	6.05	6.09	−0.04	6.12	−0.07
84	4.77	4.97	−0.20	5.37	−0.60

Table 7 (*continued*)

Compound	Experimental pEC_{50}	CoMFA Model		CoMSIA Model	
		Predicted pEC_{50}	Residual	Predicted pEC_{50}	Residual
85	6.15	6.12	0.03	5.93	0.22
86	5.77	5.89	−0.12	5.22	0.55
87	6.52	6.25	0.27	6.40	0.12
88	6.40	6.70	−0.30	6.49	−0.09
89	5.82	6.15	−0.33	6.37	−0.55
91	5.00	5.57	−0.57	5.32	−0.32
92	5.43	5.32	0.11	5.44	−0.01
93	7.09	6.98	0.11	5.98	1.11
94	6.70	6.31	0.39	6.29	0.41
95	5.85	5.82	0.03	6.35	−0.50
96	6.52	6.34	0.18	6.73	−0.21
97	6.00	6.12	−0.12	6.31	−0.31
98	5.95	5.64	0.31	5.69	0.26
100	5.00	5.02	−0.02	5.44	−0.44
101	4.89	4.65	0.24	4.94	−0.05
102	6.70	6.60	0.10	6.88	−0.18
103	6.70	6.50	0.20	6.46	0.24
104	6.30	6.36	−0.06	6.87	−0.57
106	6.15	5.92	0.23	6.13	0.03
107	5.26	5.62	−0.36	5.73	−0.47
108	5.05	4.96	0.09	4.79	0.26
110	5.53	5.78	−0.25	5.40	0.13
111	7.00	6.70	0.30	6.74	0.26

derivatives: [isopropyl **10** ($pEC_{50} = 7.40$) > ethyl **9** ($pEC_{50} = 7.10$) > unsubstituted **8** ($pEC_{50} = 5.92$), and iodo **15** ($pEC_{50} = 7.70$) > bromo **14** ($pEC_{50} = 7.52$) > chloro **13** ($pEC_{50} = 7.40$) > fluoro **12** ($pEC_{50} = 7.00$)], as well as by the difference in pEC_{50} between TCs **13** (*para*-chloro, $pEC_{50} = 7.40$) and **23** (*meta*-chloro, $pEC_{50} = 5.70$) and between TCs **68** (*para*-chloro, $pEC_{50} = 7.00$) and **63** (*meta*-chloro, $pEC_{50} = 5.52$).

Table 8
Experimental and predicted pEC_{50} values of test set compounds (**1**, **11**, **27**, **28**, **35**, **41**, **42**, **51**, **58**, **59**, **62**, **66**, **77**, **90**, **99**, **105**, **109**)

Compound	Experimental pEC_{50}	CoMFA Model		CoMSIA Model	
		Predicted pEC_{50}	Residual	Predicted pEC_{50}	Residual
1	5.14	4.44	0.70	5.96	−0.82
11	6.85	6.73	0.12	7.31	−0.46
27	5.52	5.61	−0.09	5.60	−0.08
28	5.25	4.84	0.41	5.81	−0.56
35	5.10	5.96	−0.86	6.04	−0.94
41	4.96	4.30	0.66	4.35	0.61
42	5.00	5.63	−0.63	4.82	0.19
51	7.70	7.69	0.01	7.68	0.02
58	6.40	7.29	−0.89	7.19	−0.79
59	5.62	6.31	−0.69	6.16	−0.54
62	5.30	5.96	−0.66	6.04	−0.74
66	5.43	5.56	−0.13	5.66	−0.23
77	4.29	4.43	−0.14	4.83	−0.54
90	4.89	4.90	−0.01	5.22	−0.33
99	6.30	5.96	0.34	4.78	1.52
105	5.25	5.03	0.22	5.65	−0.40
109	5.05	4.43	0.62	4.76	0.29

According to the electrostatic fields contour map of the CoMFA analysis, plotted in Fig. 4, all the compounds correctly position their polar moieties according to the distribution of CoMFA electrostatic favourable (blue) and unfavourable (red) regions. Thus, more electropositive substituents are predicted to be favoured (blue region) around the *meta*- and *para*-positions of the *N*-phenyl ring, while less positive moieties are predicted to be favoured (red region) around the ring plane.

Moreover, red polyhedra occupy the positions 5 and 6 and one of the carbonyl oxygen of the phthalimide moiety of TCs **1**–**60** (group 1); besides, they occupy the oxygen of the carboxylic function of TCs **61**–**79** (group 2) and the positions *ortho* and *meta* of

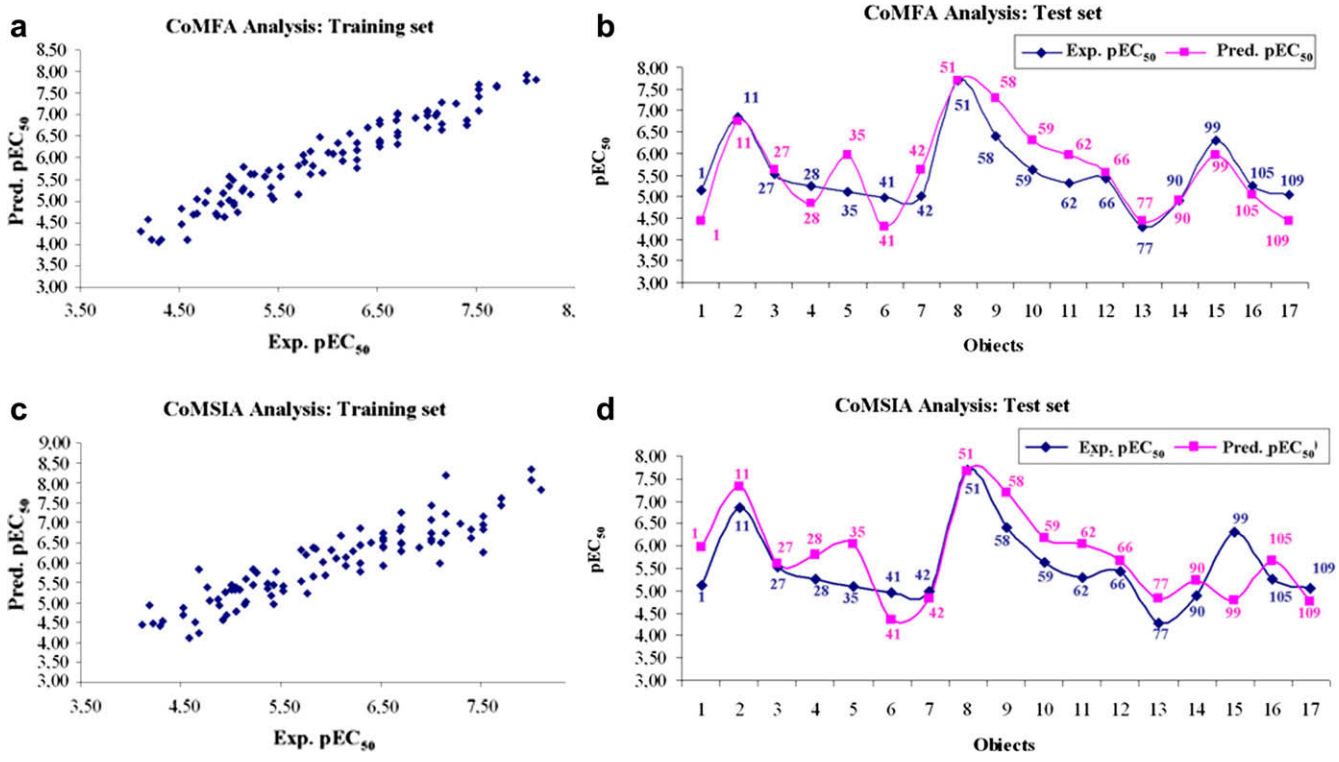


Fig. 2. Distribution of experimental and predicted pEC_{50} values for training set compounds according to CoMFA analysis (a), for test set compounds according to CoMFA analysis (b), for training set compounds according to CoMSIA analysis (c), for test set compounds according to CoMSIA analysis (d).

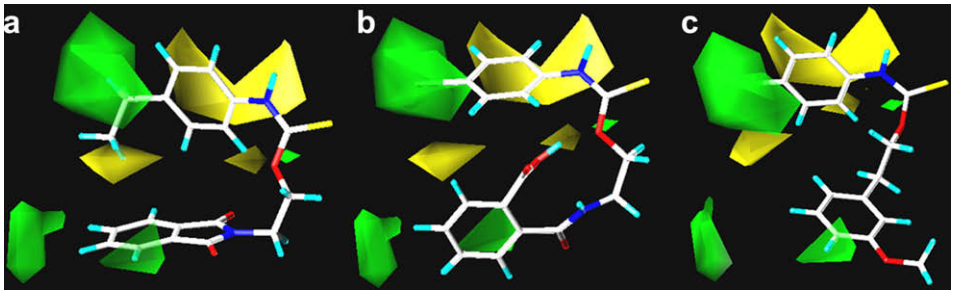


Fig. 3. Contour maps of CoMFA steric regions (green, favoured; yellow, disfavoured) are shown around compounds **9** (A), **69** (B) and **93** (C). The compounds are depicted in stick and coloured by atom type. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

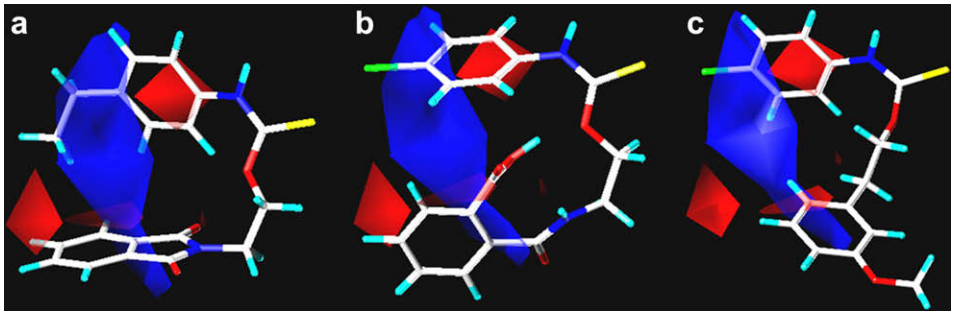


Fig. 4. Contour maps of CoMFA electrostatic regions are shown around compounds **9** (A), **69** (B) and **93** (C), depicted in stick and coloured by atom type. Blue regions are favourable for more positively charged groups and red regions are favourable for less positively charged groups. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

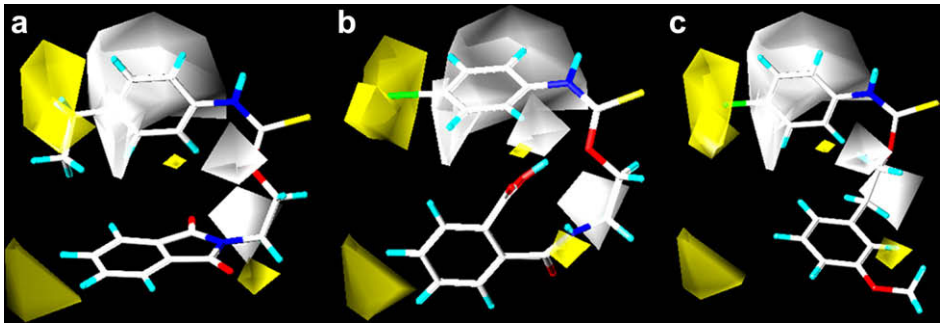


Fig. 5. Contour maps of the CoMSIA hydrophobic regions (yellow, favoured; white, disfavoured) are shown around compounds **9** (A), **69** (B) and **93** (C). All the compounds are depicted in stick and coloured by atom type. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the (hetero)aryl portion of TCs **80**–**111** (group 3). The activity trends of the *N*-*para*-methoxyphenyl TCs supports these results [compare the phthalimide derivative **16** ($pEC_{50} = 7.52$), the corresponding ring-opened derivative **72** ($pEC_{50} = 7.69$) and the 2-pyridyl TC **105** ($pEC_{50} = 5.25$) with the phenyl TC **84** ($pEC_{50} = 4.77$)].

3.2.2. CoMSIA hydrophobic H-bond acceptor and H-bond donor regions

The calculated CoMSIA hydrophobic contours (Fig. 5) are in agreement with the compound binding modes proposed by the molecular modelling as well as with the ligand steric requirements

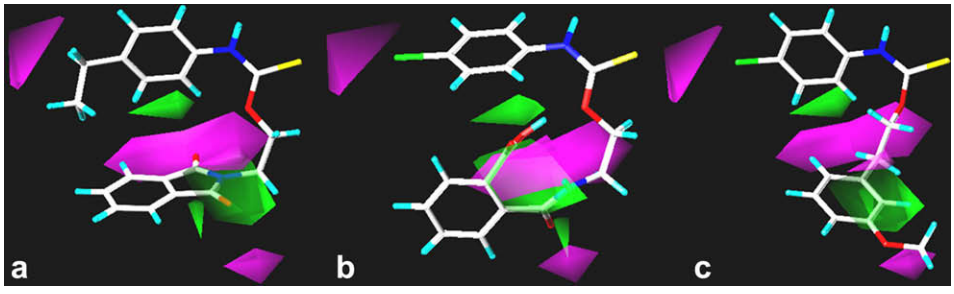


Fig. 6. CoMSIA hydrogen bond acceptor polyhedra are reported around compounds **9** (A), **69** (B) and **93** (C), depicted in stick and coloured by atom type. H-bond acceptor groups: magenta, favoured; green, disfavoured. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

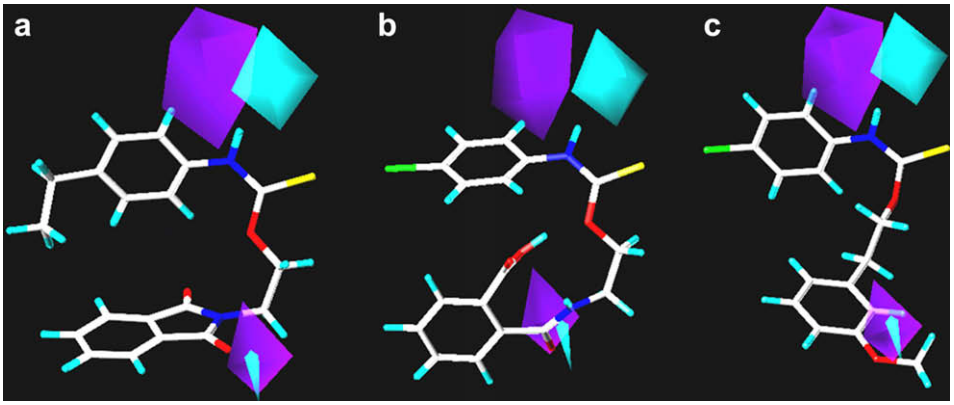


Fig. 7. CoMSIA hydrogen bond donor polyhedra are reported around compounds **9** (A), **69** (B) and **93** (C), depicted in stick and coloured by atom type. H-bond donor groups: cyano, favoured; purple, disfavoured. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

pointed out by CoMFA analysis, as exemplified by the following activity trend of the *N*-phenyl derivatives: isopropyl **10** (pEC_{50} = 7.40) > ethyl **9** (pEC_{50} = 7.10) > unsubstituted **8** (pEC_{50} = 5.92) and iodo **15** (pEC_{50} = 7.70) > bromo **14** (pEC_{50} = 7.52) > chloro **13** (pEC_{50} = 7.40) > fluoro **12** (pEC_{50} = 7.00) > unsubstituted **8** (pEC_{50} = 5.92).

To take into account the role of H-bond acceptor and H-bond donor groups for the antiretroviral activity, the corresponding CoMSIA contours were calculated (Figs. 6 and 7). The CoMSIA H-bond acceptor map corresponds to the H-bond donating groups of the receptor and, similarly, the CoMSIA H-bond donor region describes the areas where the H-bond acceptor groups of the receptor should be located.

As shown in Fig. 6, H-bond acceptor groups are predicted to be favoured (magenta region) around one of the phthalimide oxygen atom of TCs **1**–**60** (group 1), the carboxylic function of TCs **61**–**89** (group 2) and the positions 2 and 5 of the (hetero)aryl moiety of TCs **90**–**111** (group 3). These hypotheses are supported by the following experimental data: 2-pyridyl **105** (pEC_{50} = 5.25) > phenyl **84** (pEC_{50} = 4.77); 3-methoxyphenyl **93** (pEC_{50} = 7.09) > phenyl **82** (pEC_{50} = 6.30); 3-methoxyphenyl **94** (pEC_{50} = 6.70) > phenyl **83** (pEC_{50} = 6.05).

On the contrary, H-bond acceptor groups are predicted to be unfavourable (green region) in proximity of the *N*-phenyl *meta*- and *ortho*-positions and also around the carboxylic OH of TCs **61**–**89**. Moreover, H-bond donor groups (Fig. 7) are predicted to be favoured (cyan region) around the thiocarbamic NH group.

Finally, according to all the five descriptors, the substitution at the *ortho*-positions of the *N*-phenyl ring results to be disfavoured.

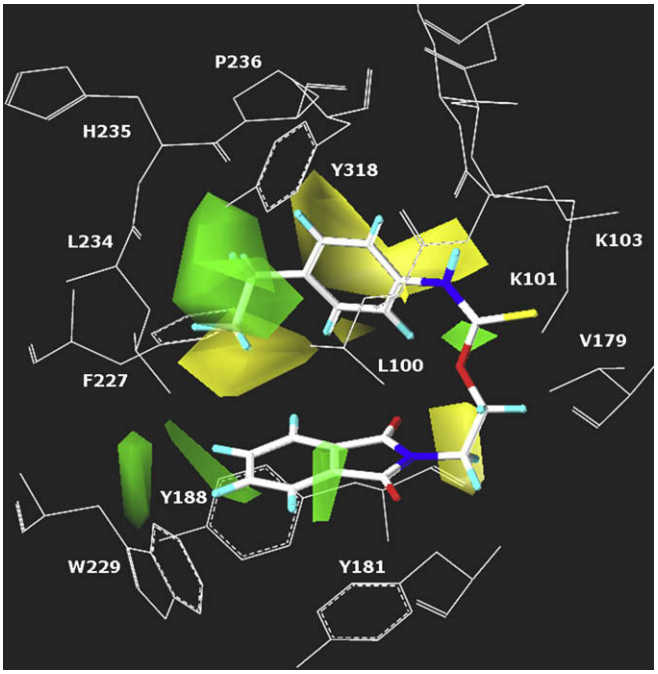


Fig. 8. CoMFA steric contour map (green, favoured; yellow, disfavoured) displayed around compound **9** and superimposition in the NNIBS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.3. A comparison between the CoMFA and CoMSIA analyses and the TC/RT docking model

In order to verify the reliability of the 3D-QSAR models, CoMFA and CoMSIA maps have been compared with the results of the docking analysis. For simplicity, in Fig. 8 we have reported only the CoMFA steric map superimposed to the docking model of TC **9** into the NNIBS.

The CoMFA steric model proves to match with the NNIBS 3D topology, suggesting bulky substitution in proximity of the two deep enzyme pockets P_1 and P_2. Thus, the first cavity surrounds the substituents at positions 3 and 4 of the phthalimide moiety, while the second pocket is occupied by the substituent at the *para*-position of the *N*-phenyl ring. The CoMFA electrostatic map highlights the importance of the π–π stacking interactions between the phthalimide moiety and Y181. The CoMSIA hydrophobic map points out the beneficial presence of substituents at the phthalimide positions 3 and 4, establishing hydrophobic contacts with P95 and W229, respectively, and at the *para*-position of the *N*-phenyl ring, displaying hydrophobic contacts with V106, F227, L234 and P236. Notably, W229, L234 and P236 are highly conserved aminoacids in the NNIBS and therefore they are recognized of strategic relevance for the design of new NNRTIs more resilient to the effects of RT mutations in this site [12]. With this purpose, the introduction of an alkyl group, such as ethyl or propyl at the phthalimide positions 3 and 4 could be particularly beneficial.

The CoMSIA H-bond donor map confirms the importance of the formation of the hydrogen bond between the thiocarbamic NH group and the K101 main chain carbonyl, essential interaction for the TC antiretroviral activity.

4. Conclusions

The 3D-QSAR field contributions showed good correlation with the 3D topology of the NNIBS. CoMFA and CoMSIA analyses highlight the importance of a good equilibrium between the lipophilic

and electrostatic properties of the ligand surface to obtain a proper occupation of the two enzyme pockets P_1 and P_2.

The computational studies here presented analyse the main interactions responsible for TC activity and give useful suggestions for the synthesis of new analogues with improved potency also against clinically relevant resistant mutants. In the future, the models elaborated will be exploited to design new TCs 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

[1] H. Jonckheere, J. Anne, E. De Clercq, Med. Res. Rev. 20 (2000) 129–154.
[2] E. De Clercq, Farmacognosia 56 (2001) 3–12.
[3] E. De Clercq, Expert. Opin. Emerg. Drugs 10 (2005) 241–273.
[4] E. De Clercq, J. Med. Chem. 48 (2005) 1297–1313.
[5] G. Barbaro, A. Scozzafava, A. Mastrolorenzo, C.T. Supuran, Curr. Pharm. Des. 11 (2005) 1805–1843.
[6] J. Balzarini, Curr. Top. Med. Chem. 4 (2004) 921–944.
[7] E. De Clercq, Antiviral. Res. 38 (1998) 153–179.
[8] O.S. Pedersen, E.B. Pedersen, Antivir. Chem. Chemother. 10 (1999) 285–314.
[9] O.S. Pedersen, E.B. Pedersen, Synthesis 4 (2000) 479–495.
[10] G. Campiani, A. Ramunno, G. Maga, V. Nacci, C. Fattorusso, B. Catalanotti, E. Morelli, E. Novellino, Curr. Pharm. Des. 8 (2002) 615–657.
[11] E. De Clercq, Chem. Biodiversity 1 (2004) 44–64.
[12] R. Pauwels, Curr. Opin. Pharmacol. 4 (2004) 437–446.
[13] A.J. Leigh Brown, S.D. Frost, W.C. Mathews, K. Dawson, N.S. Hellmann, E.S. Daar, D.D. Richman, S.J. Little, J. Infect. Dis. 187 (2003) 683–686.
[14] D.D. Richman, S.C. Morton, T. Wrin, N. Hellmann, S. Berry, M.F. Shapiro, S.A. Bozzette, AIDS 18 (2004) 1393–1401.
[15] A. Spallarossa, S. Cesarini, A. Ranise, O. Bruno, S. Schenone, P. La Colla, G. Collu, G. Sanna, B. Secci, R. Loddo. Eur. J. Med. Chem. (2008). doi:10.1016/j.ejmech.2008.09.024
[16] A. Ranise, A. Spallarossa, S. Cesarini, F. Bondavalli, S. Schenone, O. Bruno, G. Menozzi, P. Fossa, L. Mosti, M. La Colla, G. Sanna, M. Murreddu, G. Collu, B. Busonera, M.E. Marongiu, A. Pani, P. La Colla, R. Loddo, J. Med. Chem. 48 (2005) 3858–3873.
[17] S. Cesarini, A. Spallarossa, A. Ranise, O. Bruno, P. La Colla, B. Secci, G. Collu, R. Loddo, Bioorg. Med. Chem. 16 (2008) 4173–4185.
[18] S. Cesarini, A. Spallarossa, A. Ranise, P. Fossa, P. La Colla, G. Sanna, G. Collu, R. Loddo, Bioorg. Med. Chem. 16 (2008) 4160–4172.
[19] A. Spallarossa, S. Cesarini, A. Ranise, M. Ponassi, T. Unge, M. Bolognesi, Biochem. Biophys. Res. Commun. 365 (2008) 764–770.
[20] MOE: Chemical Computing Group Inc. Montreal. H3A 2R7 Canada, http://www. chemcomp. com.
[21] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Prog. Clin. Biol. Res. 291 (1989) 161–165.
[22] G. Klebe, U. Abraham, T. Mietzner, J. Med. Chem. 37 (1994) 4130–4146.
[23] Sybyl 7 0. Tripos Inc, 1699 South Hanley Road. St Louis, Missouri 63144. USA.