

Comprehensive model of wild-type and mutant HIV-1 reverse transcriptases

Flavio Ballante · Ira Musmuca · Garland R. Marshall ·
Rino Ragno

Received: 13 December 2011 / Accepted: 28 June 2012 / Published online: 26 July 2012
© Springer Science+Business Media B.V. 2012

Abstract An enhanced version of COMBINE that uses both ligand-based and structure-based alignment of ligands has been used to build a comprehensive 3-D QSAR model of wild-type HIV-1 reverse transcriptase and drug-resistant mutants. The COMBINER model focused on 7 different RT enzymes complexed with just two HIV-RT inhibitors, niverapine (NVP) and efavirenz (EFV); therefore, 14 inhibitor/enzyme complexes comprised the training set. An external test set of chiral 2-(alkyl/aryl)amino-6-benzylpyrimidin-4(3H)-ones (DABOs) was used to test predictability. The COMBINER model MC4, although developed using only two inhibitors, predicted the experimental activities of the test set with an acceptable average absolute error of prediction (0.89 pK_i). Most notably, the model was able to correctly predict the right eudismic ratio for two *R/S* pairs of DABO derivatives. The enhanced COMBINER approach was developed using only software freely available to academics.

Keywords 3-D-QSAR · HIV-1 reverse transcriptase · Drug resistance · NNRTI · RT mutants · Molecular modeling · COMBINER · DABO inhibitors · PLS

Introduction

When faced with an abundance of data from diverse sources of structure-activity studies, for example, SAR data on multiple drug-resistant mutants of a therapeutically important enzyme, choosing the best paradigm for an integrative analysis is difficult. An example common to most HIV antivirals is an abundance of crystal structures of diverse inhibitor complexes with both wild-type and mutant enzymes. The design of the next generation of inhibitors with enhanced ability to withstand the impact of known mutations should emerge from a 3-D QSAR analysis that integrates common features found in the diverse structure-activity studies within the structural 3-D framework of the enzyme. HIV-1 reverse transcriptase (HIV-RT) continues to be of therapeutic interest in the ongoing effort to provide AIDS therapeutics that have improved efficacy against multiple drug-resistant mutants, and provides a case in point for application of modern 3-D QSAR techniques.

The paradigm chosen for modeling of HIV-RT and several drug-resistant mutants was an enhanced 3-D QSAR version of the original COMBINE approaches of Gago et al. [1–3]. This Roman reincarnation, referred to in this manuscript hereafter as COMBINER to avoid confusion, used both ligand-based and structure-based alignment of inhibitors as pioneered by Musmuca et al. [4] in a previous study of hepatitis C virus NS5B polymerase inhibitors. The basic premise is that all experimental information is relevant if an appropriate frame of reference can be found, in this case, the 3-D-structure of the parent enzyme. An additional goal was the exclusive use of software readily available to the academic community to allow testing and validation of this approach by others. In a study by the Mai group, Rotili et al. [5] generated data for novel HIV-RT

F. Ballante · I. Musmuca · G. R. Marshall · R. Ragno
Rome Center for Molecular Design, Dipartimento di Chimica e
Tecnologie del Farmaco, Sapienza Università di Roma, P. le A.
Moro 5, 00185 Rome, Italy
e-mail: rino.ragno@uniroma1.it

G. R. Marshall (✉)
Department of Biochemistry and Molecular Biophysics,
Washington University School of Medicine, St. Louis, MO
63110, USA
e-mail: garlandm@gmail.com

inhibitors that was used as an external test set for the current iteration of the COMBINER model.

Methodology

Molecular modeling, COMBINER, and docking calculations

All molecular modeling calculations were performed on a 6-blade (8 Intel-Xeon E5520 2.27 GHz CPU with 24 GB DDR3 RAM each) cluster (48 CPU total) running the Debian GNU/Linux 5.03 operating system. The experimental activities of EFV and NVP reported by Rotili et al. [5] were performed as previously described [6, 7]. To build the non-experimental complexes, a cross-docking procedure previously described [4] was used with the AutodockVina program. Docking assessments were checked for either Autodock 4.2 or AutodockVina 1.1, root-mean-square-deviation (RMSD) errors are reported in Table 3.

All complexes were arbitrary superimposed using a reference template 1vrt, chosen for its superior crystallographic resolution ($R = 2.2 \text{ \AA}$). Superimpositions of the HIV-RT complexes were made with Chimera [8] using the command-line implementation of MatchMaker [9]. Prior to any minimization, all crystallographic waters were discarded by a procedure previously described [10–12]. Hydrogen atoms were added using the tleap module of AMBER [13]. Protonation states at pH 7.4 were selected, i.e., lysines, arginines, aspartates, and glutamates were assumed to be in the ionized form and parameters were calculated by means of the Antechamber module of AMBER. The complexes were solvated (SOLVATEOCT command) in a box extending 10 \AA with water molecules (TIP3 model) and neutralized with Na^+ and Cl^- ions. The solvated complexes were then refined by a single-point minimization using the Sander module of AMBER. The minimized complexes were realigned with MatchMaker using the same reference complex while maintaining the coordinates (experimental alignments) into ligands (key) and proteins (lock) and were used to obtain the energy-deconvolution matrix to develop the COMBINER models. Using Autogrid4 [14], three contributing energetic fields were calculated: the electrostatic (ELE), the steric (STE) and the desolvation (DRY). As HIV-RT is composed of 1,000 residues, 1,000 COMBINE descriptors were calculated for each field. Seven combination of the field were examined (ELE, STE, DRY, ELE + STE, ELE + DRY, STE + DRY and ELE + STE + DRY). The PLS algorithm as implemented in the R [15] environment, was used with an in-house script to compute all statistical calculations and cross-validations (Table 2).

Results and discussion

COMBINER model

To build the COMBINER model, training set selection was driven by both the availability of co-crystal structures and homogeneous inhibition data from the Mai lab. From a literature search, 14 complexes (characterized by 7 different HIV-RT wild-type and mutant enzymes) were selected as a training set using complexes with only two HIV-RT inhibitors, NVP and EFV, for which inhibition constants were available as previously tested by our collaborators [4].

As reported in Table 1, the training set was composed of NVP and EFV in complex with seven different HIV-RT enzymes (WT, L100I, K103N, V106A, V179D, Y181I, Y188L). Of the 14 complexes, structural data were experimentally available from the PDB for only five (WT/EFV: 1fk9 [16], K103N/EFV: 1fko [16], WT/NVP: 1vrt [17], L100I/NVP: 1slu [18], and K103N/NVP: 1fkp [16]). The other nine complexes (L100I/EFV, V106A/NVP, V106A/EFV, V179D/NVP, V179D/EFV, Y181I/NVP, Y181I/EFV, Y188L/NVP and Y188L/EFV) were directly modeled using side-chain structural information retrieved from other complexes present in the PDB and using the BUILD module of UCSF Chimera.

Different from the original COMBINE protocol, COMBINER used the Autogrid module of the AutoDock 4 suite [14] to compute the energy interactions between the inhibitors and each amino-acid residue of the enzyme in a complex. The ligand/residues/energy deconvolution matrix was directly obtained by the sum of the interaction energies between all ligand atoms and those composing each amino acid residue in HIV-RT. The complexes were optimized by a short energy minimization followed by docking experiments conducted with AutoDockVina [19]. From the Autogrid application, three kinds of interaction contributions were calculated: the STE, the ELE and the desolvation (DRY) ones. HIV-1 RT is a heterodimer with a subunit of 560 residues (p66) and a second subunit (p51) of 440 residues. Therefore, for each contribution, a total of 1,000 interactions were computed, and modeled using the PLS algorithm implemented in the R [20] environment. Considering all possible combination of contributions, seven different COMBINER models were independently derived (CM1–CM7, Table 2). From data reported in Table 2, all seven COMBINER models were highly robust and endowed with good predictive power. Among the seven models, CM1 and CM4 (Fig. 1) exhibited the best statistical-value profiles (compare r^2 , q^2 and SDEP values in Table 2).

As discussed by Gago et al. [2, 3] and common to other 3-D QSAR studies [21, 22], COMBINE-like models have to be analyzed by means of PLS coefficients and activity contribution (interaction energies multiplied by the PLS

Table 1 Structures, anti-HIV-RT activities (μM) of nevirapine (NVP) and efavirenz (EFV) used to build the COMBINER models

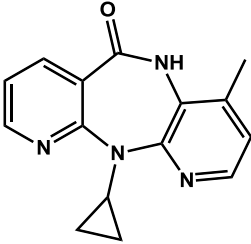
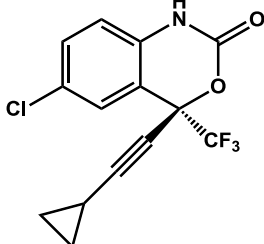
<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>Nevirapine (NVP)</p> </div> <div style="text-align: center;">  <p>Efavirenz (EFV)</p> </div> </div>		
RT	NVP	EFV
WT	0.4	0.03
L100I	9.0	0.12
K103N	7.0	0.16
V106A	10.0	0.04
V179D	2.0	0.10
Y181I	36.0	0.15
Y188L	18.0	0.38

Table 2 Statistical coefficients of the COMBINER models

CM	Model	r^2	SDEC	q^2_{LOO}	SDEP_{LOO}	q^2_{LSO5}	$\text{SDEP}_{\text{LSO5}}$	q^2_{LSO2}	$\text{SDEP}_{\text{LSO2}}$
1	DRY	0.91	0.31	0.82	0.43	0.79	0.46	0.63	0.58
2	ELE	0.80	0.45	0.51	0.71	0.49	0.72	0.37	0.79
3	STE	0.81	0.44	0.69	0.57	0.65	0.60	0.52	0.68
4	DRY_STE	0.88	0.35	0.78	0.48	0.75	0.50	0.61	0.61
5	ELE_STE	0.82	0.43	0.58	0.66	0.53	0.69	0.44	0.75
6	DRY_ELE	0.89	0.34	0.66	0.59	0.63	0.62	0.48	0.70
7	DRY_ELE_STE	0.86	0.38	0.66	0.59	0.62	0.62	0.50	0.70

CM COMBINER model number, r^2 conventional squared-correlation coefficient, SDEC standard error of calculation, q^2 cross-validation coefficient, LOO leave-one-out, SDEP standard error of prediction, LSO5 and LSO2 leave-some-out using 5 and 2 groups respectively

coefficients) plots. While PLS coefficients indicated which residues contributed most to the COMBINE relationships (general indication), the activity contributions provided the real pK_i contribution for each inhibitor/residue pair to the enhancement or decrease of the given inhibitor activity starting from a constant threshold value (intercept). Further indications of significance can be inferred from the PLS coefficients weighted by the standard deviation values ($\text{PLS} \times \text{SD}$) to give the overall importance of each amino-acid residue in the COMBINER model. In Figs. 2 and 3 are reported the PLS coefficients, the $\text{PLS} \times \text{SD}$ and activity-contribution histograms for CM1 and CM4 models, respectively.

Regarding the desolvation energy (DRY), from Figs. 2A and 3A, residues Leu100 (Ile100), Lys101, Lys103 (Asn103), Val106 (Ala106), Val179 (Asp179), Tyr181 (Ile181), Tyr188 (Leu188), Trp229, Leu234 and Tyr318

are mainly involved in defining either model CM1 or model CM4. As suggested by Wesson and Eisenberg [23], DRY is proportional to the change in the surface area that is available to water, therefore, the DRY energies are an estimation of the hydrophobic effect similar to the DRY probe in the Goodford GRID program [24]. The DRY interactions have only positive values; therefore, multiplication of the PLS value by the standard deviation of a certain residue can be interpreted in the same way as the 3-D-QSAR CoMFA [25] plots in which positive PLS Coeff \times SD values are directly correlated with enhanced activity and negative values correlate with decreased biological affinities (Fig. 2B). In Fig. 2B, residues Leu100 (Ile100), Lys101 and Tyr188 (Leu188) have the highest PLS Coeff \times SD values and, therefore, interaction with these residues are desirable, while low negative PLS Coeff \times SD values are associated with residues Trp229 and

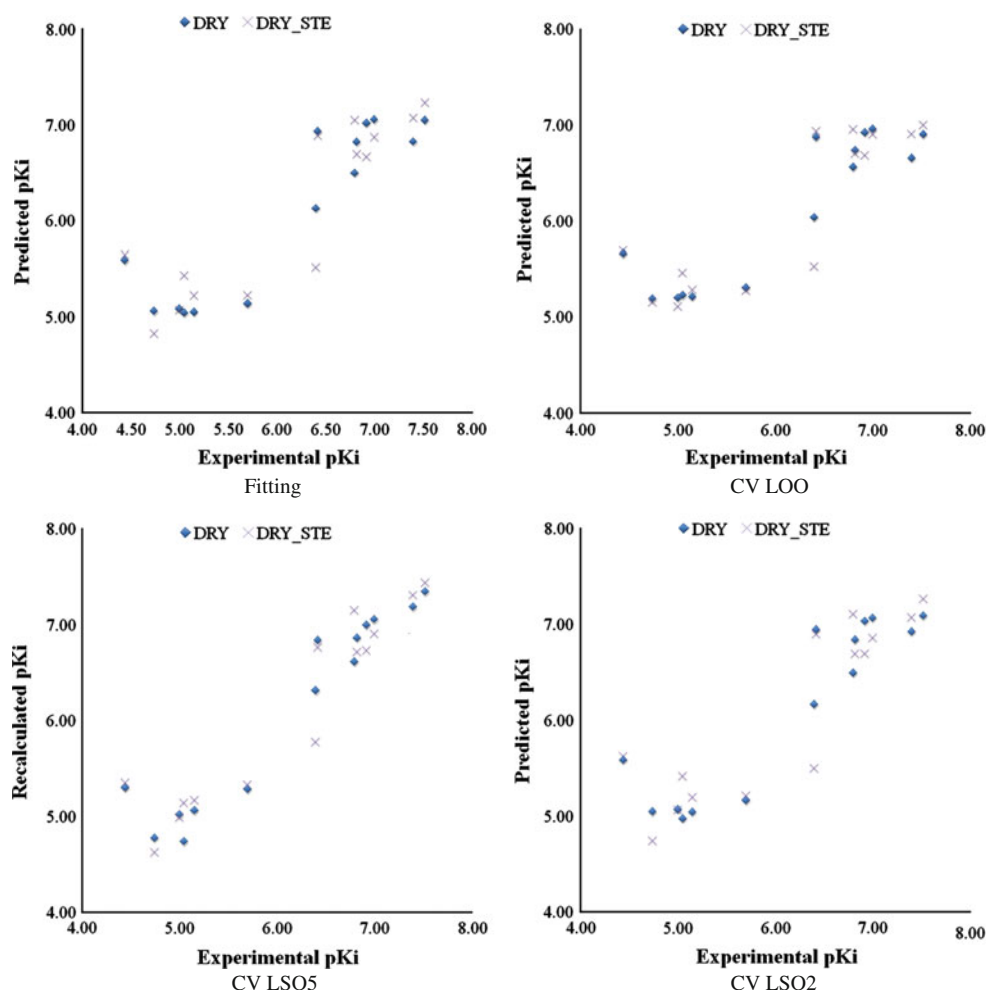


Fig. 1 Fitting and cross-validation plots (*LOO*, *LSO5* and *LSO2*) recalculate/experimental and predicted/experimental pK_i for COMBINER models CM1 and CM4

Leu234 meaning that the interaction with these residues should be minimized. Observing Fig. 3a, in model CM4, residues Leu100 (Ile100), Lys101 and Tyr188 (Leu188) are more sensitive to STE interactions, in agreement with the above. On the other hand, investigation on the energy of interaction on the STE field revealed that almost only negative values are present, in agreement with the fact that the 14 complexes were generated by means of docking experiments with van der Waals and hydrogen-bonding optimized. Thus the significance of the PLS Coef*StDev bars of histogram in Fig. 3B relative to the STE fields have inverse signification to those of the DRY fields. Although some redundancy occurs in the Autogrid-field calculations, the fact that the charge of the atom is incorporated in the calculation of desolvation interactions and that the STE fields is the sum of the interactions of the residue atoms, thus containing also hydrogen-bonding calculations, the DRY and the STE field together contain most of the ELE interactions. Similar analyses were also done for the ELE

(CM2), STE (CM3), ELE_STE (CM5), DRY_ELE (CM6) and the triple field containing COMBINER model CM7. In all COMBINER models containing the ELE field merged with other fields, its contribution to the description was almost negligible. As a matter of fact, the CM2 models (only ELE) had lower statistical coefficients, thus, indicating a lower correlation between the biological activities and ELE interactions. In the multifield models (CM4–CM7), therefore, the PLS code correctly recognized this low correlation and contribution of the ELE field was essentially eliminated. Since the models were obtained using single point RT-mutated forms, interesting sources of data are the activity-contribution plot of Figs. 2C and 3C. These plots reported the product of each residue field by the respective PLS coefficients. The sum of all these products and the intercept values for each complex returns the fitted values of the COMBINER models (Fig. 8). Due to the similar profile of the DRY field in both CM1 and CM4 models, only the DRY_STE double-field model is

Fig. 2 PLS coefficients

A, PLS \times SD **B** and activity-contribution histograms **C** for the DRY model CM1. Only bars with values higher than 0.001 and lower than -0.001 are shown

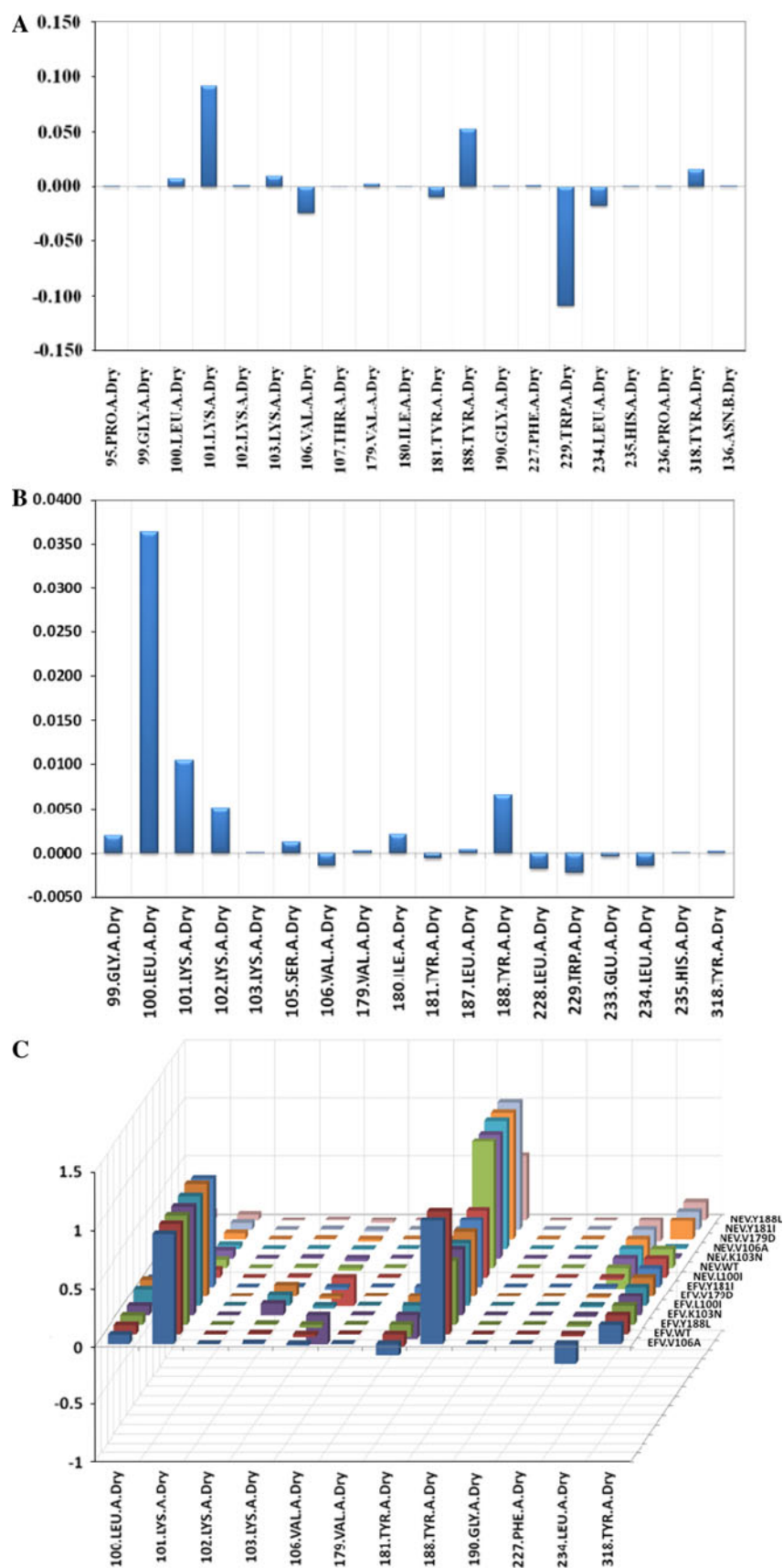
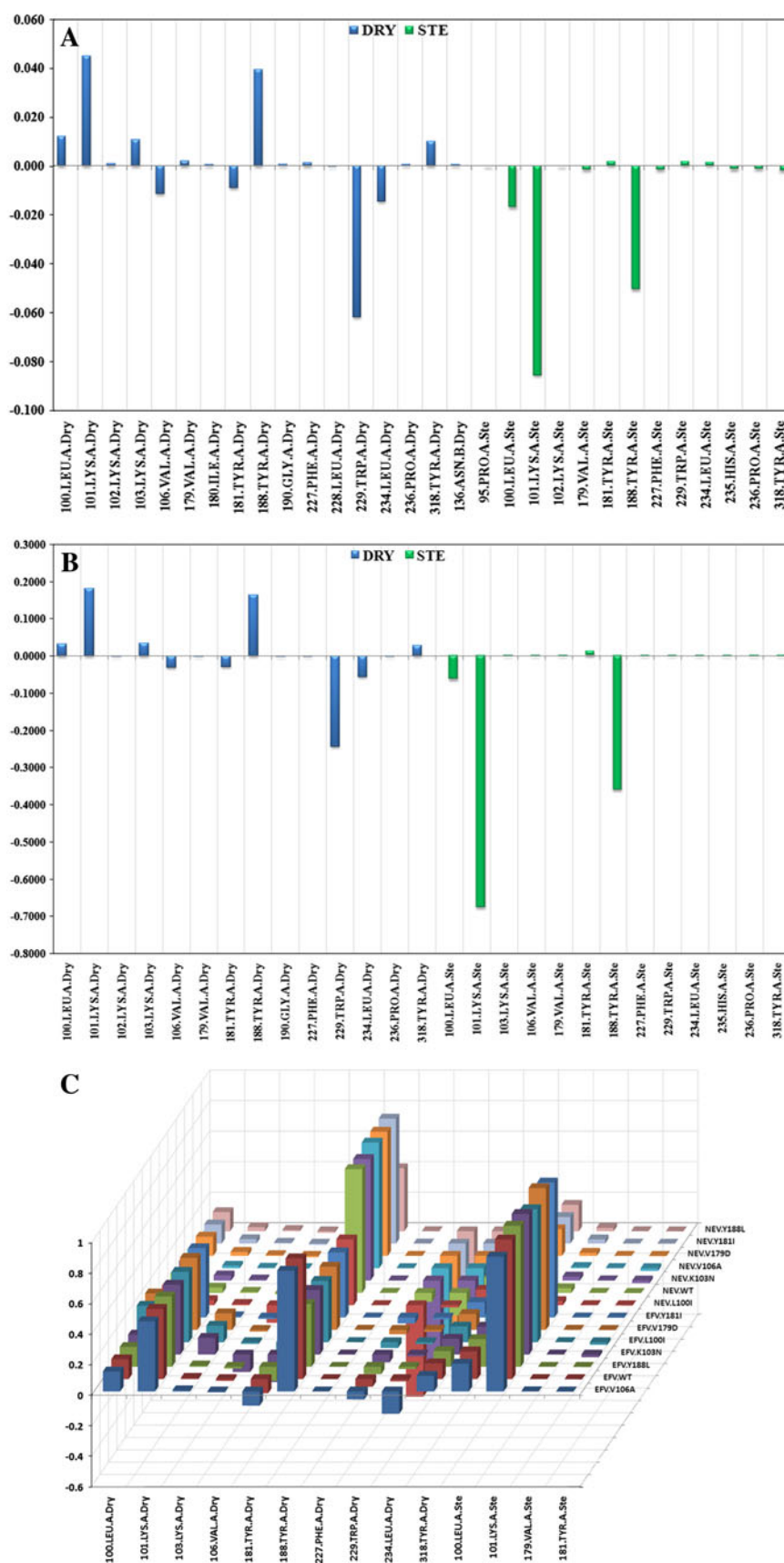


Fig. 3 PLS coefficients

A, PLS \times SD **B** and activity-contribution histograms **C** for DRY_STE model CM4. Only bars with values higher than 0.001 and lower than -0.001 are shown



considered for future comments. It could be argued that all statistics of the DRY model are slightly better or comparable to those of the DRY_STE model. It was decided, nevertheless, to focus only on the DRY_STE model so to have a more complete description of the ligand/enzyme interactions. Analyses of activity-contribution plots confirmed that the amino-acids mutations were directly and indirectly responsible for the different activity profiles of EFV and NVP. Any description of the detailed interaction network is far too complicated; after analysis of the CM1 and CM4 models plots reported in Figs. 2 and 3, a

schematic view (Fig. 4) on the direct influence to the NVP and EFV anti-RT activities by their surrounding residues (and their mutations) is presented.

COMBINER predictions

The reported COMBINER model CM4 was used to rationalize the role of mutation on the activity profile of (*R*)- and (*S*)-MC1501, and of (*R*)- and (*S*)-MC2082 reported by Rotili et al. [5]. The binding modes of the four DABO derivatives (Fig. 5) were analyzed by the means of the

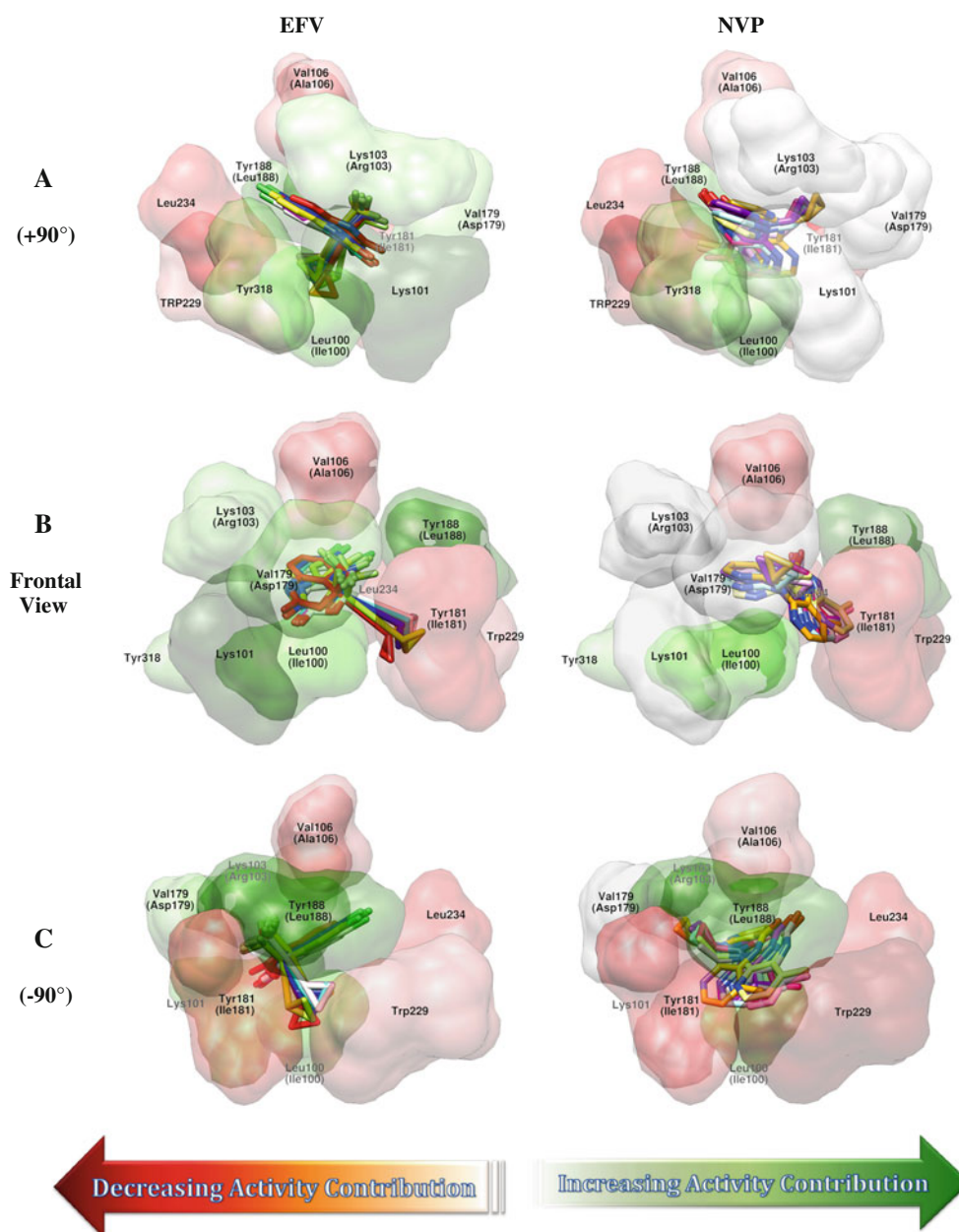


Fig. 4 Efavirenz (left column) and nevirapine (right column) with the surrounding residues surfaces as in the experimental complexes. The surfaces are colored by activity contribution. Here are three orthogonal views of the complexes (rotated along the X axes by $\pm 90^\circ$)

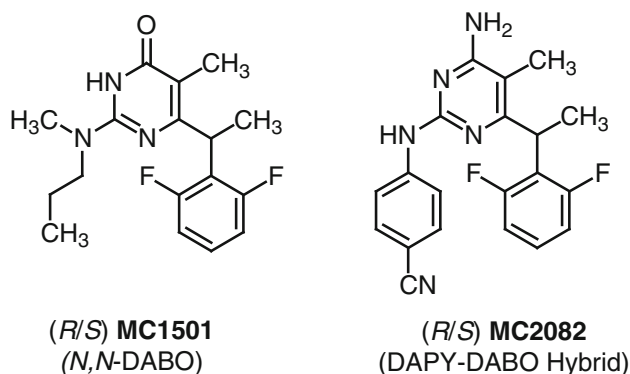


Fig. 5 Structures of racemic HIV-RT inhibitors resolved by Rotili et al. [5] used to validate CM4

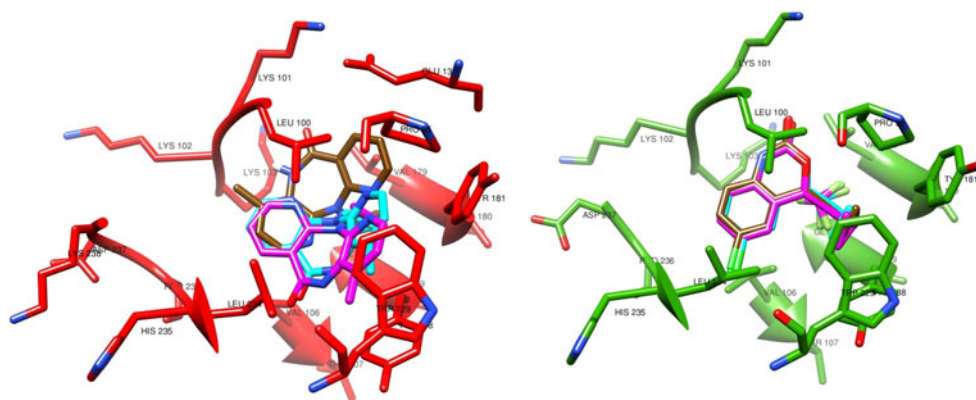
Vina program [19] which proved more reliable, as shown in Table 3 and Fig. 6, than Autodock [14] in reproducing the EFV- and NVP-experimental binding modes. In redocking Vina was more reliable than Autodock in reproducing the binding mode of both NVP and EFV starting from the experimental conformation of the ligands [4]. In view of these results and the fact that Vina was 10-times faster than Autodock, Vina was selected for docking experiments.

Figure 7 shows the binding modes of the DABO derivatives (structure of test set compounds are available in Appendix 1 and in the publication by Rotili et al. [5]) with the WT and the mutated HIV-RTs used in this study. Similarly to

Table 3 Docking assessment: root-mean-square deviations (RMSDs) displayed by the Vina and Autodock docking programs

PDB code	Mutation	Ligand	Vina		Autodock
			Exp	Mod	
1fk9	WT	EFV	0.33	0.41	0.29
1fko	K103N	EFV	0.35	0.43	0.59
1fkp	K103N	NVP	0.53	0.81	3.41
1slu	L100I	NVP	0.26	0.48	3.52
1vrt	WT	NVP	0.51	0.86	3.53

Fig. 6 Docking assessments: comparison of redocking by Vina and Autodock. In cyan are reported the experimental conformations in the 1vrt and 1fko complexes; in magenta are those redocked with Vina and in brown those obtained with Autodock. In red is shown HIV-RT in the 1vrt (nevirapine) complex and in green HIV-RT for 1fko (efavirenz)



previous studies [10–12], the *R*-conformations display an overall binding profile similar for either MC1501 or MC2082. In the *S*-configurations, the methyl at the C6-benzylic position (highlighted in red) prevented similar interactions [12]. The (*R*)-MC2082 binding mode is comparable with that of TMC278 (rilpivirine) [26], a recently reported DAPY derivative now undergoing clinical trials [27].

Figure 8 displays the (*R*)-MC2082 binding modes overlapped with the experimental complexes of efavirenz and TMC278 in wild-type and mutated HIV-RTs.

Once the binding modes of MC compounds were calculated, the COMBINER model CM4 was readily applied. As reported in Table 4, the COMBINER model, although developed with only two different HIV-RT inhibitors, predicted the experimental MC activities with a surprisingly acceptable average absolute-error-of-prediction (0.89 pK_i). The CM4 model percentage of prediction error ranged between 61.6 and 0.9 % with an average error of 14.3 % comparable to those experimentally reported by Rotili et al. [5] (37.5, 1.5 and 16.2 %, respectively).

Most notably, the COMBINER CM4-model correctly predicted the right eudismic ratio for the two *R/S* pairs of MC derivatives.

Application of the COMBINER CM4-model to the test set of Rotili et al. [5] (MC compounds) helped the interpretation of the calculated activity contributions (Fig. 9) directly highlighting difference between MC1501 and MC2082 upon binding to HIV-RTs. Figure 9 shows that the activity contributions associated to the interactions of the most active MC enantiomers (stereoisomers *R*) with residues Lys101 were mainly responsible for the higher activities of (*R*)-MC2082 versus the (*R*)-MC1501 with an average increase of activities of about 0.29 and 0.19 of pK_i units for the hydrophobic and STE fields, respectively.

Comparing the activity contributions of *R*- and *S*-enantiomers of MC1501, the hydrophobic effect of residue Lys101 became negligible, while that from Trp229 increased with an average contribution of 0.24 pK_i units. In comparison, the Lys101-related STE contribution was

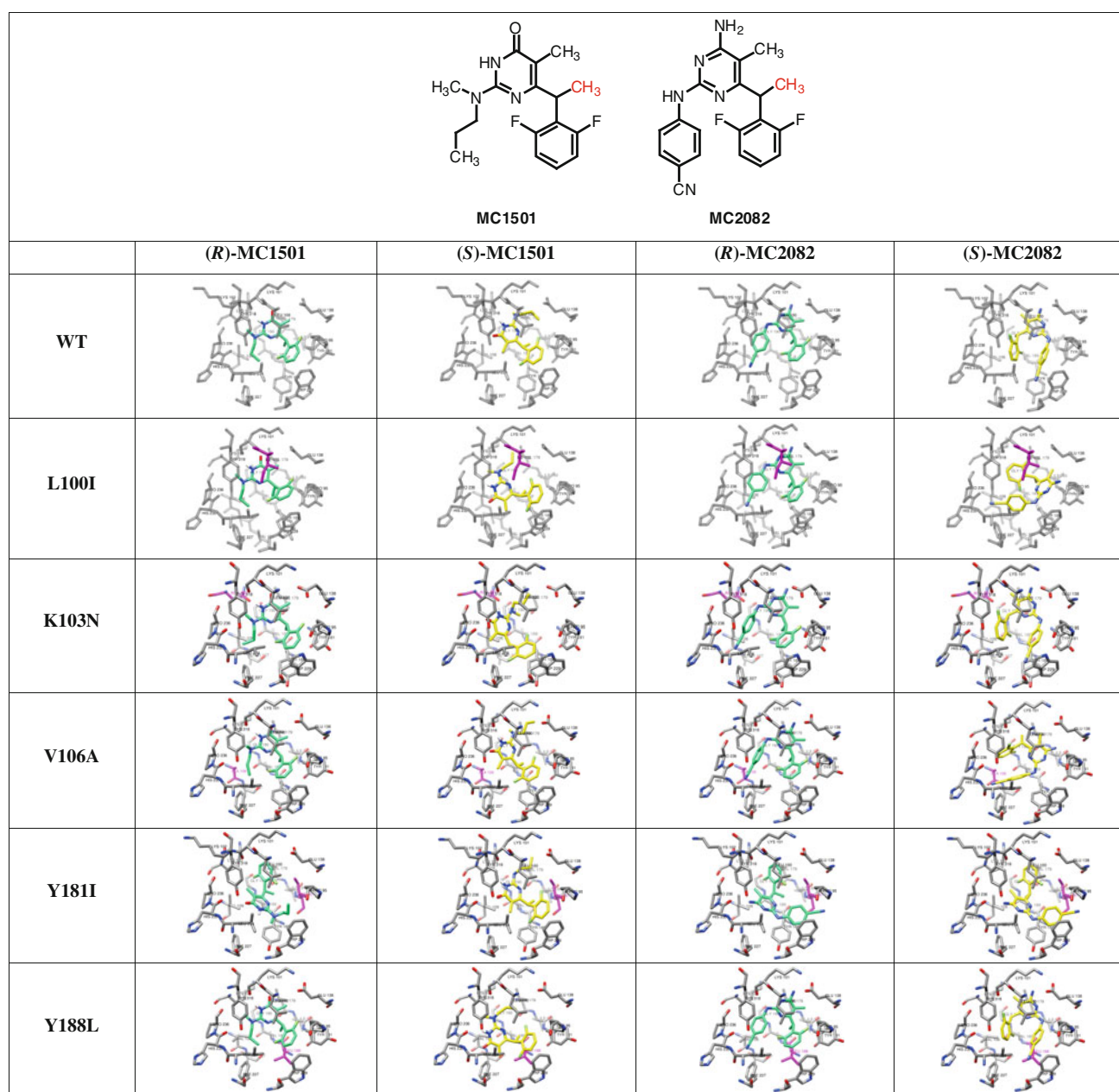


Fig. 7 Vina-proposed binding modes for the MC1501 and MC2082 enantiomers in six different HIV-RT proteins. The molecular structures are shown with the C6-methyl group highlighted in red at the top of the figure

more than doubled (see Tables 5, 6). In the case of MC2082 *R*- and *S*-enantiomers, the activity contribution of Lys101 was only reduced by 32 % (0.17), Trp229 increased to 0.16, and the Lys101 STE contribution increased more than 5 times (1.05).

Single-point mutations within the COMBINer CM4 model showed that residue 188 demonstrated a key role in modulating the interactions of the ligands both in its wild-type (Tyr188) and in the Leu188 mutation. Interestingly, for another mutating residue, residue 188 seemed to offset loss of interactions as a result of the other residue mutation, most

remarkably in the case of the more active compounds (*R*)-MC1501 and (*R*)-MC2082. Comparing the activity-contribution profile of (*R*)-MC2082 docked into wild-type HIV-RT and in the V106A-mutated protein, the only values changed drastically were those associated with Tyr188. A possible explanation for this might be that incoming missing interactions for the (*R*)-MC2082/Val106 → (*R*)-MC2082/Ala106 replacement are filled readily by the augmented (*R*)-MC2082/Tyr188 interactions (compare Tyr188 positions in Fig. 3), but this remains speculative without further analyses.

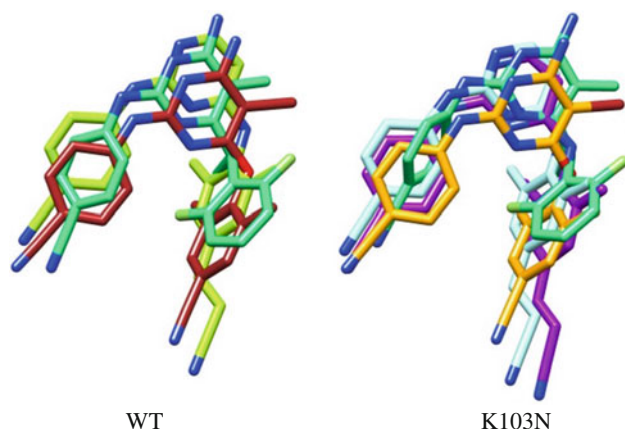


Fig. 8 Binding modes of (*R*)-MC2082 overlapped with etravirine and TMC278. On the *left* side are shown (*R*)-MC2082 in green, etravirine (3mec) in *brown* and TMC278 (2zd1) in *light green*, all bound to wild-type HIV-RT. On the *right* side are shown (*R*)-MC2082 (*green*) binding mode in K103N-mutated RT overlapped with etravirine (*orange*) that was co-crystallized with K103N HIV-RT, TMC278 (*light blue*) in the K103N–Y181C double mutant (3bgr) and in the L100I–K103 double mutant (*purple*, 2ze2)

Finally, Tables 5 and 6 clearly demonstrated that most mutations force the ligands to primarily re-adapt their interaction network around the two non-mutating Lys101 and Trp229 residues, supplying alternate ways of hydrogen-bond and hydrophobic anchor points with which ligands interact upon complex formation.

Fig. 9 Activity-contribution histograms calculated for the test MC compounds. Only bars with values higher than 0.001 and lower than −0.001 are shown

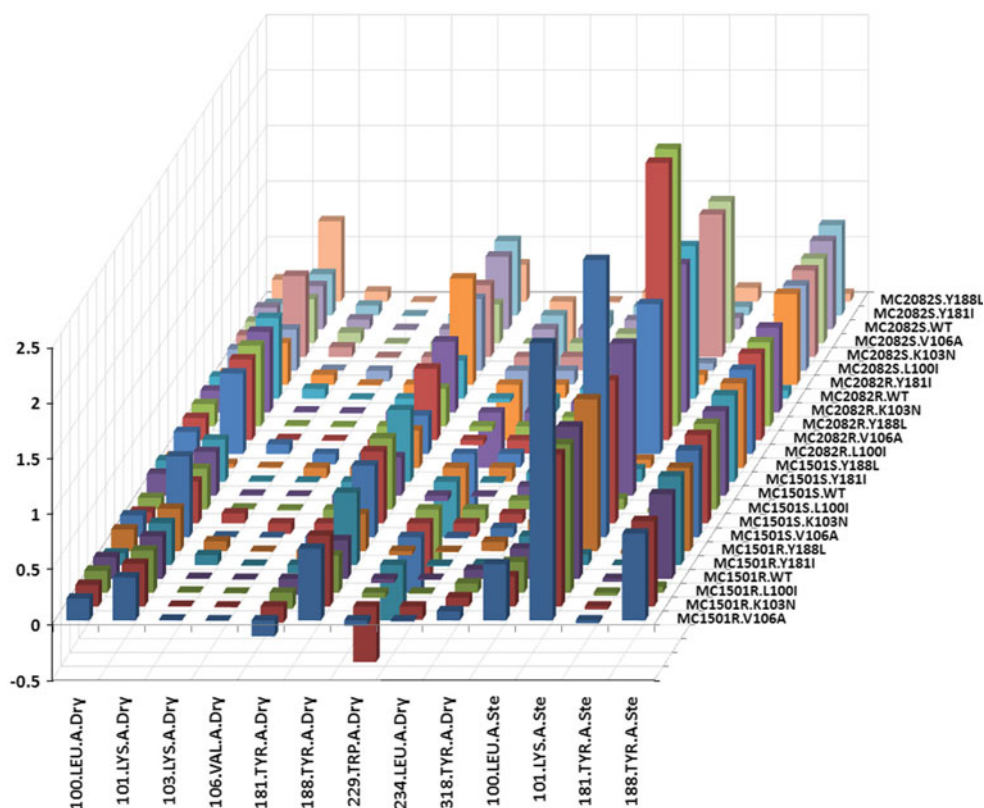


Table 4 Experimental and COMBINER CM4-model predicted activities of MC compounds of Rotili et al. [5]

	MC1501				MC2082			
	<i>R</i>		<i>S</i>		<i>R</i>		<i>S</i>	
	Exp	Pred	Exp	Pred	Exp	Pred	Exp	Pred
WT	8.70	7.46	6.93	7.20	6.81	7.21	4.52	5.77
V106A	8.52	9.19	6.45	5.78	9.52	9.43	6.62	7.51
K103N	7.02	7.17	6.01	7.52	8.52	9.11	7.19	7.52
L100I	7.02	6.69	4.40	7.11	8.10	7.49	6.74	6.03
Y188L	6.71	7.51	4.40	5.11	8.10	7.09	4.40	5.95
Y181I	6.35	6.05	4.40	6.12	6.12	6.25	6.29	5.48

Conclusions

The COMBINER approach integrates multiple sources of SAR information to build a self-consistent model of the amino-acid residues in both wild-type and mutant enzymes responsible for molecular recognition and discrimination. As with all such underdetermined 3-D QSAR models, predictability is the only real means of selection one model over another. This study on HIV-RT used a minimal set of inhibitor complexes to extract possible models for HIV-RT variants that rationalize the experimentally observed inhibitory activity of a novel set of compounds described

Table 5 COMBINER CM4-model predicted activity contributions of MC1501 with average values higher than 0.01 absolute pK_i values

Field	Dry									Ste			
Residue number	100	K101	103	106	181	188	T229	L234	Y318	100	K101	181	188
(R)-MC1501.WT	0.19	0.38	0.01	−0.01	−0.15	0.35	−0.04	−0.01	0.08	0.27	1.37	−0.03	0.76
(R)-MC1501.L100I	0.20	0.38	0.01	−0.01	−0.15	0.33	−0.05	−0.01	0.08	0.28	1.34	−0.03	0.05
(R)-MC1501.K103N	0.19	0.38	0.01	−0.01	−0.15	0.64	−0.51	−0.12	0.08	0.27	1.37	−0.03	0.77
(R)-MC1501.V106A	0.20	0.39	0.01	0.00	−0.14	0.65	−0.05	−0.01	0.08	0.51	2.55	−0.03	0.78
(R)-MC1501.Y181I	0.10	0.38	0.09	−0.01	−0.07	0.65	−0.51	−0.01	0.01	0.26	0.10	0.00	0.80
(R)-MC1501.Y188L	0.19	0.38	0.09	−0.01	−0.15	0.33	−0.04	−0.01	0.08	0.27	1.37	−0.03	0.75
Average	0.18	0.38	0.03	−0.01	−0.13	0.49	−0.20	−0.03	0.07	0.31	1.35	−0.02	0.65
SD	0.04	0.00	0.04	0.00	0.03	0.17	0.24	0.04	0.03	0.10	0.78	0.01	0.29
Max	0.20	0.39	0.09	0.00	−0.07	0.65	−0.04	−0.01	0.08	0.51	2.55	0.00	0.80
Min	0.10	0.38	0.01	−0.01	−0.15	0.33	−0.51	−0.12	0.01	0.26	0.10	−0.03	0.05
Range	0.09	0.01	0.08	0.00	0.07	0.32	0.47	0.11	0.08	0.25	2.45	0.02	0.74
(S)-MC1501.WT	0.20	0.39	0.01	0.00	−0.08	0.35	−0.05	−0.01	0.08	0.28	1.37	−0.03	0.76
(S)-MC1501.L100I	0.10	0.37	0.01	−0.01	−0.15	0.64	−0.52	−0.12	0.08	0.26	0.10	−0.03	0.77
(S)-MC1501.K103N	0.10	0.38	0.09	−0.09	−0.08	0.65	−0.51	−0.12	0.08	0.26	1.29	0.00	0.80
(S)-MC1501.V106A	0.20	0.73	0.01	0.00	−0.08	0.65	−0.50	−0.01	0.08	0.51	2.58	0.00	0.78
(S)-MC1501.Y181I	0.19	0.38	0.01	−0.01	−0.08	0.65	−0.52	−0.01	0.08	0.27	0.11	0.00	0.78
(S)-MC1501.Y188L	0.10	0.03	0.01	−0.09	−0.08	0.34	−0.52	−0.12	0.08	0.26	0.07	0.00	0.76
Average	0.15	0.38	0.02	−0.03	−0.09	0.54	−0.44	−0.07	0.08	0.31	0.92	−0.01	0.78
SD	0.05	0.22	0.03	0.05	0.03	0.16	0.19	0.06	0.00	0.10	1.01	0.01	0.01
Max	0.20	0.73	0.09	0.00	−0.08	0.65	−0.05	−0.01	0.08	0.51	2.58	0.00	0.80
Min	0.10	0.03	0.01	−0.09	−0.15	0.34	−0.52	−0.12	0.08	0.26	0.07	−0.03	0.76
Range	0.09	0.69	0.08	0.09	0.07	0.31	0.47	0.11	0.00	0.25	2.50	0.02	0.04
RvsS ^a	0.03	0.00	0.01	0.03	−0.05	−0.05	0.24	0.04	−0.01	0.00	0.43	−0.01	−0.12

^a Differences between (R)-MC1501 and (S)-MC1501 activity-contribution averages. In bold are highlighted values cited in the interpretations of prediction reported in the text

Table 6 COMBINER CM4-model predicted activity contributions of MC2082 with average values higher than 0.01 absolute pK_i values

Field	Dry									Ste			
Residue number	100	K101	103	106	181	188	T229	L234	Y318	100	K101	181	188
(R)-MC2082.WT	0.20	0.73	0.09	−0.01	−0.15	0.34	−0.03	−0.12	0.09	0.28	1.37	−0.03	0.07
(R)-MC2082.L100I	0.20	0.73	0.09	−0.09	−0.15	0.35	−0.51	−0.12	0.08	0.28	1.35	−0.03	0.76
(R)-MC2082.K103N	0.20	0.73	0.01	−0.01	−0.15	0.64	−0.51	−0.12	0.08	0.28	1.34	−0.03	0.76
(R)-MC2082.V106A	0.20	0.73	0.01	−0.01	−0.15	0.64	−0.05	−0.12	0.08	0.51	2.54	−0.03	0.78
(R)-MC2082.Y181I	0.10	0.37	0.09	−0.01	−0.08	0.96	−0.54	−0.12	0.08	0.26	0.09	−0.03	0.82
(R)-MC2082.Y188L	0.20	0.73	0.01	−0.01	−0.15	0.34	−0.04	−0.12	0.08	0.51	2.55	−0.03	0.76
Average	0.18	0.67	0.05	−0.02	−0.14	0.54	−0.28	−0.12	0.08	0.35	1.54	−0.03	0.66
SD	0.04	0.14	0.04	0.03	0.03	0.25	0.26	0.00	0.00	0.12	0.92	0.00	0.29
Max	0.20	0.73	0.09	−0.01	−0.08	0.96	−0.03	−0.12	0.09	0.51	2.55	−0.03	0.82
Min	0.10	0.37	0.01	−0.09	−0.15	0.34	−0.54	−0.12	0.08	0.26	0.09	−0.03	0.07
Range	0.10	0.35	0.08	0.09	0.07	0.62	0.51	0.00	0.00	0.25	2.46	0.00	0.74
(S)-MC2082.WT	0.20	0.39	0.09	−0.01	−0.15	0.65	−0.52	−0.12	0.08	0.27	0.10	−0.03	0.79
(S)-MC2082.L100I	0.19	0.37	0.01	−0.09	−0.15	0.65	−0.51	−0.12	0.08	0.26	0.07	−0.03	0.77
(S)-MC2082.K103N	0.20	0.73	0.09	−0.01	−0.15	0.65	−0.52	−0.12	0.08	0.27	1.28	−0.03	0.78
(S)-MC2082.V106A	0.19	0.40	0.09	−0.01	−0.08	0.35	−0.01	−0.12	0.09	0.27	1.28	0.00	0.76
(S)-MC2082.Y181I	0.10	0.37	0.09	−0.01	−0.08	0.67	−0.55	−0.12	0.08	0.26	0.08	0.00	0.81

Table 6 continued

Field	Dry									Ste			
	100	K101	103	106	181	188	T229	L234	Y318	100	K101	181	188
(S)-MC2082.Y188L	0.20	0.72	0.09	−0.01	−0.15	0.33	−0.50	−0.01	0.08	0.27	0.12	−0.03	0.07
Average	0.18	0.50	0.08	−0.02	−0.12	0.55	−0.44	−0.10	0.08	0.27	0.49	−0.02	0.66
SD	0.04	0.18	0.03	0.04	0.04	0.16	0.21	0.04	0.00	0.01	0.61	0.01	0.29
Max	0.20	0.73	0.09	−0.01	−0.08	0.67	−0.01	−0.01	0.09	0.27	1.28	0.00	0.81
Min	0.10	0.37	0.01	−0.09	−0.15	0.33	−0.55	−0.12	0.08	0.26	0.07	−0.03	0.07
Range	0.09	0.36	0.09	0.09	0.07	0.34	0.54	0.11	0.00	0.01	1.22	0.02	0.74
RvsS ^a	0.00	0.17	−0.03	0.00	−0.01	−0.01	0.16	−0.02	0.00	0.08	1.05	−0.01	−0.01

^a Differences between (R)-MC2082 and (S)-MC2082 activity-contribution averages. In bold are highlighted the values cited in the interpretations of predictions reported in the text

by Rotili et al. including the relative activity of two different sets of stereoisomers. Obviously, prediction of novel inhibitors and their activities against HIV-RT is a logical next step to validate the utility of the COMBINER approach. Extension to similar problems, such as prediction of isoform selectivity of the eleven zinc-based histone deacetylase inhibitors, is underway.

Acknowledgments The authors thank the research group (Rotili et al. [5]) of Prof. Antonello Mai for sharing their data about the separation and activity of their DABO derivatives prior to publication. In addition, Garland R. Marshall acknowledges financial support from the Dipartimento di Chimica e Tecnologie del Farmaco, Facoltà di Farmacia e Medicina, Sapienza Università di Roma, which made his visiting professorship in Rome feasible.

References

- Lozano JJ, Pastor M, Cruciani G, Gaedt K, Centeno NB, Gago F, Sanz F (2000) 3-D-QSAR methods on the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. *J Comput Aided Mol Des* 14:341–353
- Perez C, Pastor M, Ortiz AR, Gago F (1998) Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. *J Med Chem* 41:836–852
- Rodriguez-Barrios F, Gago F (2004) Chemometrical identification of mutations in HIV-1 reverse transcriptase conferring resistance or enhanced sensitivity to arylsulfonylbenzonitriles. *J Am Chem Soc* 126:2718–2719
- Musmuca I, Caroli A, Mai A, Kaushik-Basu N, Arora P, Ragno R (2010) Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. *J Chem Inform Model* 50:662–676
- Rotili D, Samuele A, Tarantino D, Ragno R, Musmuca I, Ballante F, Botta G, Morera L, Pierini M, Cirilli R, Nawrozki MB, Gonzalez E, Clotet B, Artico M, Este JA, Maga G, Mai A (2012) 2-(Alkyl/aryl)amino-6-benzylpyrimidin-4(3H)-ones as inhibitors of wild-type and mutant HIV-1: enantioselectivity studies. *J Med Chem* 55:3558–3562
- Cancio R, Mai A, Rotili D, Artico M, Sbardella G, Clotet-Codina I, Este JA, Crespan E, Zanolli S, Hubscher U, Spadari S, Maga G (2007) Slow-, tight-binding HIV-1 reverse transcriptase non-nucleoside inhibitors highly active against drug-resistant mutants. *ChemMedChem* 2:445–448
- Samuele A, Facchini M, Rotili D, Mai A, Artico M, Armand-Ugon M, Este JA, Maga G (2008) Substrate-induced stable enzyme-inhibitor complex formation allows tight binding of novel 2-aminopyrimidin-4(3H)-ones to drug-resistant HIV-1 reverse transcriptase mutants. *ChemMedChem* 3:1412–1418
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
- Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* 7:339
- Mai A, Sbardella G, Artico M, Ragno R, Massa S, Novellino E, Greco G, Lavecchia A, Musiu C, La Colla M, Murgioni C, La Colla P, Loddo R (2001) Structure-based design, synthesis, and biological evaluation of conformationally restricted novel 2-alkylthio-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J Med Chem* 44:2544–2554
- Quaglia M, Mai A, Sbardella G, Artico M, Ragno R, Massa S, del Piano D, Setzu G, Doratiotto S, Cotichini V (2001) Chiral resolution and molecular modeling investigation of rac-2-cyclopentylthio-6-[1-(2,6-difluorophenyl)ethyl]-3,4-dihydro-5-methylpyrimidin-4(3H)-one (MC-1047), a potent anti-HIV-1 reverse transcriptase agent of the DABO class. *Chirality* 13:75–80
- Ragno R, Mai A, Sbardella G, Artico M, Massa S, Musiu C, Mura M, Marturana F, Cadeddu A, La Colla P (2004) Computer-aided design, synthesis, and anti-HIV-1 activity in vitro of 2-alkylamino-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as novel potent non-nucleoside reverse transcriptase inhibitors, also active against the Y181C variant. *J Med Chem* 47:928–934
- Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668–1688
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock and AutoDockTools: automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791

15. Mevik B-H, Wehrens R (2007) The pls package: principal component and partial least squares regression in R. *J Stat Softw* 18(2):1–24
16. Ren J, Milton J, Weaver KL, Short SA, Stuart DI, Stammers DK (2000) Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. *Structure* 8:1089–1094
17. Ren J, Esnouf R, Garman E, Somers D, Ross C, Kirby I, Keeling J, Darby G, Jones Y, Stuart D et al (1995) High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat Struct Biol* 2:293–302
18. Ren J, Nichols CE, Chamberlain PP, Weaver KL, Short SA, Stammers DK (2004) Crystal structures of HIV-1 reverse transcriptases mutated at codons 100, 106 and 108 and mechanisms of resistance to non-nucleoside inhibitors. *J Mol Biol* 336:569–578
19. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461
20. R-Development-Core-Team R: a language and environment for statistical computing. <http://www.R-project.org>
21. Ballante F, Ragno R (2012) 3-D QSAutogrid/R: an alternative procedure to build 3-D QSAR models. *Methodologies and applications. J Chem Inf Model* 52:1674–1685
22. Baroni M, Costantino G, Cruciani G, Riganelli D, Valigi R, Clementi S (1993) Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3-D-QSAR problems. *Quant Struct Activ Relatsh* 12:9–20
23. Wesson L, Eisenberg D (1992) Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci* 1:227–235
24. Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* 28:849–857
25. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J Am Chem Soc* 110:5959–5967
26. Azijn H, Tirry I, Vingerhoets J, de Bethune MP, Kraus G, Boven K, Jochmans D, Van Craenenbroeck E, Picchio G, Rimsky LT (2010) TMC278, a next-generation nonnucleoside reverse transcriptase inhibitor (NNRTI), active against wild-type and NNRTI-resistant HIV-1. *Antimicrob Agents Chemother* 54:718–727
27. Macarthur RD (2011) Clinical trial report: TMC278 (rilpivirine) versus efavirenz as initial therapy in treatment-naive, HIV-1-infected patients. *Curr Infect Dis Rep* 13:1–3