Quantitative structure–activity relationship analysis of pyridinone HIV-1 reverse transcriptase inhibitors using the k nearest neighbor method and QSAR-based database mining

Jose Luis Medina-Franco[a], Alexander Golbraikh[b], Scott Oloff[b], Rafael Castillo[a] &
Alexander Tropsha[b,*],
[a]*Departamento de Farmacia, Facultad de Química, Universidad Nacional Autónoma de México, 04510, Mexico City, Mexico;* [b]*Division of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina, CB# 7360, Chapel Hill, NC, 27599-7360, USA*

Summary

We have developed quantitative structure–activity relationship (QSAR) models for 44 non-nucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs) of the pyridinone derivative type. The k nearest neighbor (kNN) variable selection approach was used. This method utilizes multiple descriptors such as molecular connectivity indices, which are derived from two-dimensional molecular topology. The modeling process entailed extensive validation including the randomization of the target property (Y-randomization) test and the division of the dataset into multiple training and test sets to establish the external predictive power of the training set models. QSAR models with high internal and external accuracy were generated, with leave-one-out cross-validated R^2 (q^2) values ranging between 0.5 and 0.8 for the training sets and R^2 values exceeding 0.6 for the test sets. The best models with the highest internal and external predictive power were used to search the National Cancer Institute database. Derivatives of the pyrazolo[3,4-d]pyrimidine and phenothiazine type were identified as promising novel NNRTIs leads. Several candidates were docked into the binding pocket of nevirapine with the AutoDock (version 3.0) software. Docking results suggested that these types of compounds could be binding in the NNRTI binding site in a similar mode to a known non-nucleoside inhibitor nevirapine.

Abbreviations: AIDS – Acquired immunodeficiency syndrome; HIV – Human immunodeficiency virus; kNN – k Nearest neighbor; MZ – MolConnZ; NNRTIs – Non-nucleoside reverse transcriptase inhibitors; QSAR – Quantitative structure–activity relationships; RT – Reverse transcriptase

Introduction

The reverse transcriptase (RT) of human immuno-deficiency virus type 1 (HIV-1) is an attractive target to treat the acquired immune deficiency syndrome (AIDS) for which no universal successful chemotherapy is available to date [1, 2]. RT inhibitors are classified as nucleoside or non-nucleoside depending on their mechanism of action. One advantage of the latter group is that they lack toxic effects associated with the nucleosides. To date, three non-nucleoside RT inhibitors have been approved for the clinical use including nevirapine (VIRAMUNE), delavirdine (RESCRIPTOR) and efavirenz (SUSTIVA) [1].

*To whom correspondence should be addressed. Tel.: +1-919-966-2955; Fax: +1-919-966-0204; E-mail: alex_tropsha@unc.edu

However, all these and other approved drugs induce drug resistant variants of HIV-1.

Pyridinone derivatives [3] (cf., Table 1) are a class of non-nucleoside reverse transcriptase inhibitors (NNRTIs). Kinetic studies [3] and the analysis of resistance mutations [4] indicated that these derivatives should be binding at the same active site of HIV RT as other NNRTIs. Extensive structure–activity relationships (SAR) studies have been conducted for these compounds leading to potent inhibitors [5–8]. Compound **27** (cf., Table 1) showed good activity in clinical studies but resistant strains of the virus emerged [9, 10]. Several molecular modeling studies of the pyridinone derivatives have been reported including QSAR modeling using Hansch method [11, 12], 3D-QSAR approach [13], and docking [14].

Several crystal structures of HIV-1 RT in the apo form [15] and in complex with different NNRTIs were published [16–28] and successful structure-based design studies leading to the development of active NNRTIs have been reported [29]. However, no crystal structure for the RT-pyridinone complex is available. The great flexibility of the NNRTIs binding pocket is a major problem to apply this technique [17], and it is feasible that chemical structures that belong to different congeneric series may bind to the active site in different modes. Thus, ligand based approaches such as quantitative structure–activity relationships (QSAR) modeling may present a plausible alternative to the structure-based design of NNRTIs. In fact, recent studies in our group have established that thoroughly validated QSAR models could be used to mine chemical databases to discover novel compounds with the desired biological activity. For instance, the k-nearest-neighbor (kNN) variable selection QSAR approach [30, 31] has been applied successfully to the discovery of novel anticonvulsants [32, 33].

In this paper, we have applied the kNN QSAR approach to a dataset of 44 pyridinone derivatives (Table 1) [7, 8]. Our objective was to develop robust, validated QSAR models and search, with these models, the National Cancer Institute database NCISMA99 [34] for novel NNRTIs. Several compounds from this database have already been implicated as anti-AIDS agents [35], suggesting that NCISMA99 may continue to provide novel compounds with the anti-HIV activity. Screening the NCISMA99 database with our QSAR models identified several potential leads with high-predicted activity. These compounds were docked into the binding pocket of nevirapine in the HIV-1 RT, the structure with the highest resolution available to date, to assess the putative binding mode of these computational datamining hits within the HIV-1 RT binding site.

Computational methods

Dataset and biological data

The structures and *in vitro* activity of 44 pyridinone derivatives used in this study are listed in Table 1. Activities were converted into the corresponding $-\log_{10}IC_{50}$ values (pIC_{50}), were IC_{50} is the effective concentration of compound required to achieve 50% of inhibition of RT. The biological data was taken from the literature [7, 8].

Generation of molecular descriptors

All chemical structures were generated using SYBYL software [36]. Molecular topological indices [37, 38] were generated with the MolConnZ program (MZ descriptors) [39]. Overall, MolconnZ generates over 400 different descriptors. Most of these descriptors characterize chemical structure, but several depend on the arbitrary numbering of atoms in a molecule and are introduced solely for bookkeeping purposes. In our study, only 193 chemical descriptors were eventually used (after deleting these bookkeeping descriptors as well as those with zero value or zero variance). The descriptors were ranged-scaled prior to subsequent calculations (see [30] for details), since the absolute scales for MZ descriptors can differ by orders of magnitude. Accordingly, our use of range scaling avoided giving descriptors with significantly higher ranges a disproportional weight upon distance calculations as part of kNN procedure in multidimensional MZ descriptor space. All calculations were performed on an SGI Octane at the University of North Carolina's Laboratory for Molecular Modeling.
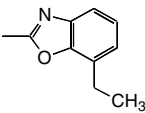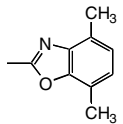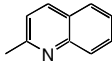
kNN QSAR method
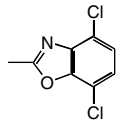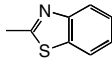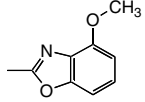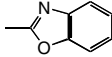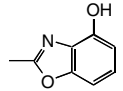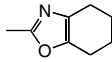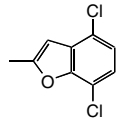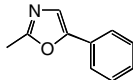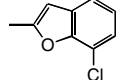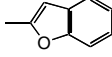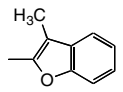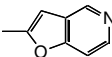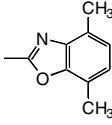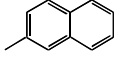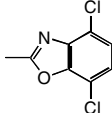
The kNN QSAR method [30] employs the kNN classification principle [40] and the variable selection procedure. Briefly, a subset of *nvar* (number of

Table 1. Structures, HIV-1 RT inhibitory activity, experimental and calculated by kNN QSAR models, of 44 pyridinone derivatives used in this study.
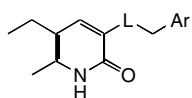
Compd.	L	Ar	Experimental pIC$_{50}$	Calculated pIC$_{50}$	Compd.	L	Ar	Experimental pIC$_{50}$	Calculated pIC$_{50}$
1	NH		6.68	6.49	13	NH		5.57	5.57
2	NH		6.48	6.39	14	NH		5.36	5.33
3	NH		6.55	6.70	15	NH		5.27	4.91
4	NH		6.47	6.11	16	NH		5.12	5.54
5	NH		6.46	6.54	17	NH		4.82	4.44
6	NH		6.34	6.04	18	NH		4.65	5.07
7	NH		6.30	6.31	19	NH		4.49	4.54
8	NH		6.28	6.10	20	NH		3.98	4.54
9	NH		5.96	5.78	21	NH		6.92	6.80
10	NH		5.72	5.51	22	NH		5.90	5.83
11	NH		5.63	5.79	23	NH		5.78	5.81
12	NH		5.60	5.68	24	NH		7.26	7.21

Table 1. Continued.

Compd.	L	Ar	Experimental pIC$_{50}$	Calculated pIC$_{50}$	Compd.	L	Ar	Experimental pIC$_{50}$	Calculated pIC$_{50}$
25	NH		6.59	6.43	**35**	CH$_2$		3.59	4.10
26	NH		7.70	7.65	**36**	CH$_2$		5.61	5.82
27	NH		7.72	7.61	**37**	CH$_2$		6.43	6.43
28	NH		6.74	6.56	**38**	CH$_2$		7.64	7.18
29	NH		6.36	6.43	**39**	CH$_2$		7.24	7.02
30	NH		7.24	7.22	**40**	CH$_2$		6.41	6.12
31	NH		6.47	6.51	**41**	CH$_2$		6.89	6.72
32	NH		5.71	6.51	**42**	CH$_2$		4.87	5.19
33	CH$_2$		4.30	4.64	**43**	CH$_2$		7.55	7.65
34	CH$_2$		5.08	5.43	**44**	CH$_2$		7.85	7.67

selected variables) descriptors is selected randomly as a hypothetical descriptor pharmacophore (HDP). The *nvar* is set to different values to obtain the best q^2 possible. The HDP is validated by leave-one-out cross-validation, where each compound is eliminated form the training set and its biological activity is predicted as the average activity of k most similar molecules ($k = 1$–5). The similarity is characterized by the Euclidean distance between compounds in multidimensional descriptor space. A method of simulated annealing with the Metropolis-like acceptance criteria is used to optimize the variable selection. Further details of the kNN method implementation, including the description of the simulated annealing procedure used for stochastic sampling of the descriptor space, are given elsewhere [30].

Robustness of QSAR models

The q^2 values for the models for experimental training sets were compared to those derived for so-called random datasets, which are generated by random shuffling of compounds' activities. The statistical significance of QSAR models for training sets was evaluated with the standard hypothesis testing method [41].

In this approach, two alternative hypothesis are formulated: (1) for H_0, $h = \mu$; (2) for H_1, h (μ, where μ is the average value of q^2 for random datasets and h is the q^2 value for the actual dataset. Thus, the null hypothesis, H_0, states that the QSAR model for the actual dataset is not significantly better than random models, whereas the alternative hypothesis, H_1, assumes the opposite (i.e., that the actual model is significantly better than random models). The decision-making is based on a standard one-tail test, which involves the following procedure.

(1) Determine the average value of q^2 (μ) and its standard deviation (σ) for random datasets.
(2) Calculate the Z score that corresponds to the q^2 value for the actual dataset: $Z = (h-\mu)/\sigma$
(3) Compare this Z score with the tabular critical values of Z_c at different levels of significance (α) [41] to determine the level at which H_0 should be rejected. If the Z score is higher than tabular values of Z_c (cf., Table 2), one concludes that at the level of significance that corresponds to that Z_c, H_0 should be rejected and,

Table 2. Frequently used α values and the corresponding critical values of Z_c for the one-tail test [41].

α	Z_c
0.10	1.28
0.05	1.64
0.01	2.33

therefore, H_1 should be accepted. In this case, it is concluded that the result obtained for the actual dataset is statistically better than those obtained for random datasets at the given level of significance.

Model validation: training and test set compound selection

To obtain training and test sets, we used a sphere exclusion algorithm described in detail elsewhere [42]. This algorithm allows construction of training and test sets of various sizes that cover the entire descriptor space of all compounds. To estimate the predictive power of a QSAR model, the following parameters were used [43]. (i) Correlation coefficient R^2 between the predicted and observed activities for the test set; (ii) Coefficients of determination [43] (predicted versus observed activities R^2 and observed versus predicted activities $R_0'^2$); (iii) slopes k and k' of regression lines (predicted versus observed activities, and observed versus predicted activities) through the origin. We suggest that a QSAR model has an acceptable predictive power if the following conditions are satisfied [42]:

$$q^2 > 0.5; \tag{1}$$

$$R^2 > 0.6; \tag{2}$$

$$|R_0^2 - R_0'^2| < 0.3 \tag{3}$$

$$0.85 \le k \le 1.15 \text{ or } 0.85 \le k' \le 1.15. \tag{4}$$

NCI database mining

Our database-mining strategy makes use of validated QSAR models developed for the available

series of biologically active compounds. The general flowchart of the database-mining procedure is shown in Figure 1 and includes the following major steps; a similar approach was successfully used recently for the discovery of novel anticonvulsant agents [33].

1. Develop validated variable selection QSAR models for a dataset of compounds with known structures and activities; define descriptor pharmacophores [44] and applicability domains [45] for all models.
2. Compute chemical descriptors used in QSAR model development for all compounds in the available chemical database(s); in our studies we have used molecular topological indices calculated with the MolConnZ program [39].
3. Calculate chemical similarity values (we use the Euclidean distance in the descriptor pharmacophore space) between all active probes (i.e., molecules used for QSAR model development) and every structure in the database.
4. Rank all database structures by their similarity to a probe(s), and select *M* structures within certain similarity threshold.
5. Predict biological activity values for these *M* structures based on preconstructed QSAR models using applicability domain.
6. Select structures to have high values of biological activity as *computational hits*.

Flexible docking

The geometry of chemical structures selected from the database mining were optimized in Sybyl with the Powell method [46] using the Tripos force field [47]. A gradient convergence criterion with a value of 0.05 kcal/mol was used. The docking experiments were performed with AutoDock 3.0 [48]. In short, AutoDock performs an automated docking of the ligand with user-specified dihedral flexibility within a protein rigid binding site. The program performs several runs in each docking experiment. Each run provides one predicted binding mode with the highest interaction energy.

The crystal structure of RT in complex with nevirapine [16] (cf. Figure 5) was obtained from the Protein Data Bank (PDB code 1vrt) [49]. For the validation of the docking protocol, the nevirapine coordinates were removed from the crystal structure and the bond orders were checked. Gasteiger charges [50] were assigned to the structure and the non-polar hydrogens were treated implicitly. The cyclopropyl group was allowed to rotate during docking. For the subsequent docking studies using compounds identified from database mining, nevirapine, magnesium ion and water molecules were removed from the original structure. With the AutoDock Tools [51], polar hydrogen atoms were added and Kollman charges
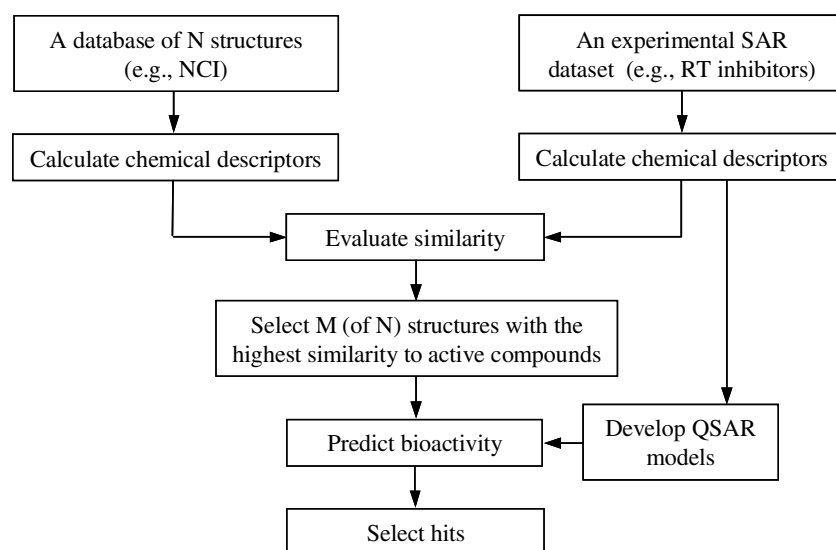

Figure 1. Flowchart of database mining that employs predictive QSAR models.
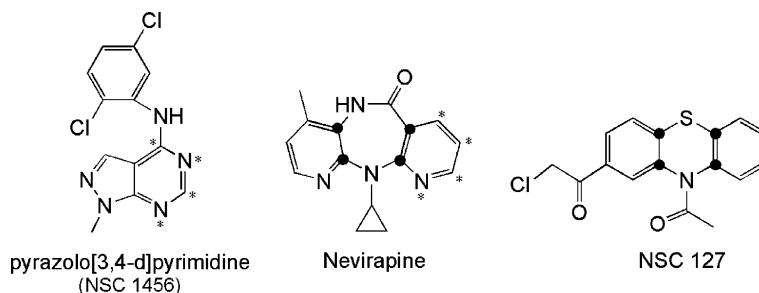
535

236

The statistical examination of the results was performed with one-tail hypothesis testing as described in the Computational methods. The average q^2 value for 30-descriptor models obtained for 10 different randomized datasets was 0.06 with a standard deviation of 0.18. The Z score for the most significant 30-descriptor model for the actual dataset was 2.71. This indicates that the probability that the kNN QSAR model constructed for the real dataset is random is less than 0.01.

kNN QSAR model validation

To obtain reliable and truly predictive QSAR models, it is necessary to demonstrate that the training set models can accurately predict activities of compounds in external datasets. We accept models with q^2 values for the training set greater than 0.5 and R^2 values for predicted versus actual activities of the test set compounds greater than 0.6 [42, 54]. Table 3 presents the 10 best models obtained from multiple kNN analyses. Figure 4 shows actual versus calculated activity values for the training and test sets based on the best model. The number of descriptors for the 10 best models varied between 10 and 26. Even for a test set as large as 19 compounds a good model with $q^2 = 0.52$ and $R^2 = 0.75$ was still obtained.

Database mining

Several compounds from the NCI database whose anti-AIDS activity has not been yet documented
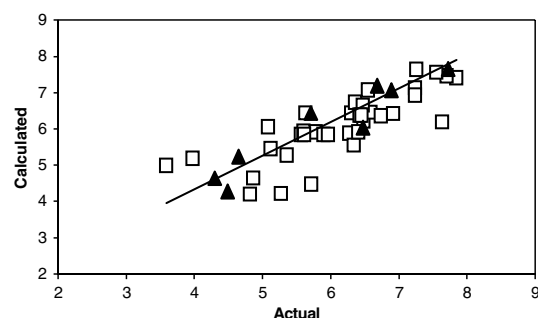
Figure 4. Actual versus calculated activity values for the training (squares) and test (triangles) set using kNN QSAR model 1 ($q^2 = 0.62$, $R^2 = 0.89$).

were predicted to be active. The structures of some of them are shown in Figure 5. It was especially interesting to observe the structural similarity of the compound NSC 127 and nevirapine. The core of both structures is a three-ring system involving two six-member aromatic rings. Both compounds have a small substituent (methylketone or cyclopropyl group) attached to a nitrogen located at the central ring (Figure 5).

Flexible docking. Validation of the docking protocol

The docking protocol was validated by removing nevirapine from the crystal structure and then predicting the binding mode with the AutoDock. All 100 docked energy conformations formed a single cluster within 1.0 Å as described in Computational methods. The RMSD between the lowest docked energy conformation found (docked energy = −9.67 kcal/mol; average docked energy

Table 3. Ten best kNN QSAR models.

Model number	Size of the training set (number of cpds.)	Size of the test set (number of cpds.)	Number of descriptors	q^2 (training set)	R^2 (test set)	R_0^2	$R_0'^2$
1	36	8	14	0.62	0.89	0.88	0.89
2	36	8	14	0.60	0.87	0.85	0.87
3	36	8	12	0.61	0.75	0.70	0.75
4	36	8	26	0.55	0.74	0.48	0.72
5	36	8	10	0.58	0.72	0.68	0.71
6	31	13	10	0.59	0.72	0.72	0.68
7	31	13	20	0.63	0.71	0.60	0.71
8	28	16	16	0.63	0.75	0.46	0.72
9	27	17	24	0.53	0.71	0.48	0.70
10	25	19	10	0.52	0.75	0.72	0.42
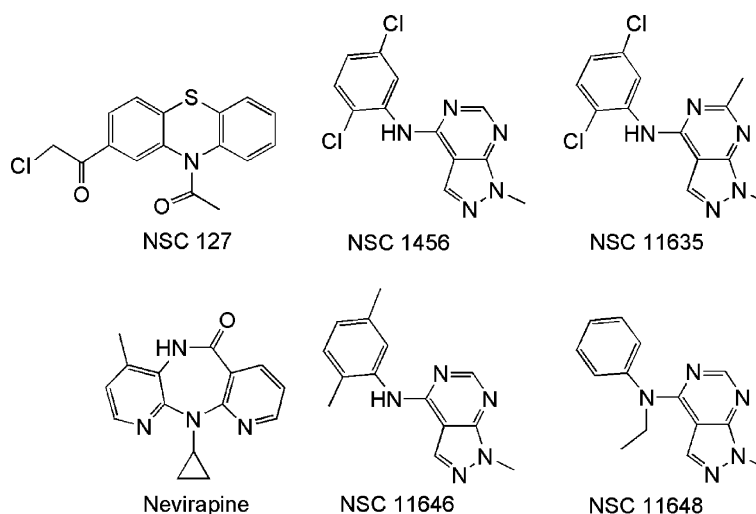
Figure 5. Chemical structures of the compounds found in the database mining and nevirapine.

for the cluster = −9.66 kcal/mol) and the actual coordinates was 0.70. This demonstrated that the protocol successfully predicted the binding position of nevirapine.

Docking of selected molecules resulting from the database mining

The compounds predicted most active from the database mining were visually inspected. Based on some common features with the known NNRTIs, e.g., the possibility to adopt a butterfly-like conformation within the NNRTIs binding pocket [16–20], the selected compounds were docked into the binding site of nevirapine. The results are summarized in Table 4.

The pyrazo[3, 4-*d*]pyrimidines NSC 1456, 11635, 11646 and 11648 [55] were found to adopt

a butterfly-like geometry inside the nevirapine binding pocket. Moreover, the predicted binding mode is similar to the binding mode of nevirapine. The binding pocket of most pyrazo[3,4-*d*]pyrimidines is formed by amino acids Pro95, Leu100, Lys101, Lys102, Lys103, Val106, Tyr181, Tyr188, Phe227, Trp229, Leu234, His235, Pro236, Asp237, Tyr318 of the p66 subunit and Glu138 of the p51 subunit [16]. Most of these amino acids are common to the binding pocket of other NNRTIs [16-28]. The binding mode of the predicted most active compound, NSC 11635, into the NNRTIs binding pocket is depicted in Figure 6. The binding mode of nevirapine is shown for comparison. The methyl group at position 6 of the pyrazolo[3,4-*d*]pyrimidine ring of NSC 11635 showed additional contacts with Pro225 of the p66 subunit. These additional contacts can explain the

Table 4. The results of docking of several selected compounds found by the means of database mining into the binding pocket of nevirapine.

Compd.	Cluster[a]	Members	Lowest energy (kcal/mol)[b]	Average energy (kcal/mol)[c]
NSC 11635	1	93	−10.03	−10.02
NSC 11646	1	100	−9.37	−9.36
NSC 1456	1	62	−9.34	−9.29
NSC 11648	1	9	−8.61	−8.57
NSC 127	2	43	−10.28	−10.16
	4	40	−9.74	−9.71

[a]The best clusters of a total 100 runs are shown.
[b]Value for the optimal structure in the cluster. Docking energies are as determined by AutoDock.
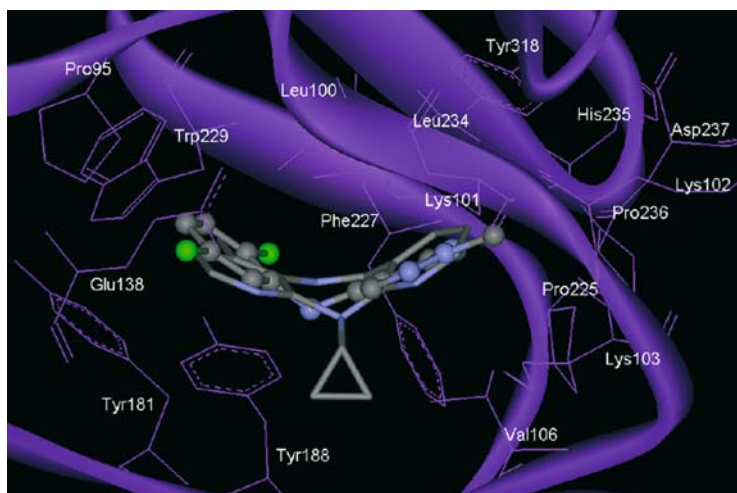[c]Average value of docking energies of all members in the cluster.

238

Figure 6. Predicted binding mode of NSC 11635 (balls and sticks) in the binding pocket of HIV-1 RT. Nevirapine (sticks) is displayed for comparison. Hydrogens are omitted for clarity.

lowest energy interaction calculated by AutoDock for NSC 11635 compared with the energy interactions calculated for NSC 11646 and NSC 1456 (Table 4). This methyl group, that is missing in the structures of NSC 11646 and NSC 1456 (Figure 5), could be an important structural feature in the putative activity of NSC 11635. NSC 11648 presented additional interactions between the ethyl group and the amino acids Val179 and Gly190 of the p66 subunit (cf., Figure 7d). However, for this compound contacts with Asp237 and Glu138 were not observed.

In order to compare the binding mode predicted for the pyrazo[3,4-*d*]pyrimidines with the predicted for the pyridinone derivatives, the binding mode of the former compounds was superimposed with the binding mode of the potent inhibitor **27** previously reported [14]. We have shown that the predicted binding mode of **27** is representative of most potent pyridinone derivatives [13, 14]. The comparisons are shown in Figure 7. The predicted binding mode of all compounds is similar to the binding mode of the pyridinone derivative. The phenyl ring of NSC 11635, NSC 11646 and NSC 1456, all with a 2,5-dichloro or 2,5-dimethyl substitution, may be roughly overlapped with the benzoxazol ring of **27**. Noteworthy, both chlorine or methyl groups of the phenyl ring of these pyrazo[3,4-*d*]pyrimidines occupies almost the same binding position that the chlorine groups of **27** (Figure 7a–c) interacting

mainly with amino acids Pro95, Tyr181, Tyr188 and Trp229. Interesting to note, it has been shown that pyridinone derivatives di-substituted with methyl or chlorine groups at the positions 4 and 7 of the benzoxazol ring are more active than the corresponding mono or unsubstituted compounds [7, 8, 11]. These observations suggest the putative biological importance of the chloro or methyl groups at the phenyl ring of compounds NSC 11635, NSC 11646 and NSC 1456.

For NSC 11648, which lacks of substituents at the phenyl ring, a rotated conformation was predicted if compared to the binding orientation of the former three pyrazo[3,4-*d*]pyrimidines. The pyrazolo[3,4-*d*]pyrimidine ring occupies a similar binding region that the benzoxazol ring of **27** (Figure 7d). The ethyl group was predicted to occupy the same pocket region that the aminomethylene group of the pyridinone derivative.

Table 5 presents a comparison between the activity predicted by the *k*NN QSAR models, the docked energy and the free energy of binding as estimated by AutoDock for four compounds. Obviously, there is a good agreement between these predictions, with similar relative ranking of all compounds by all computational methods. Noteworthy, there is a good agreement between the activity predicted by *k*NN models for all the pyridinone derivatives in the dataset, and the experimental activities (Table 1). In docking studies for several pyridinone derivatives, we have
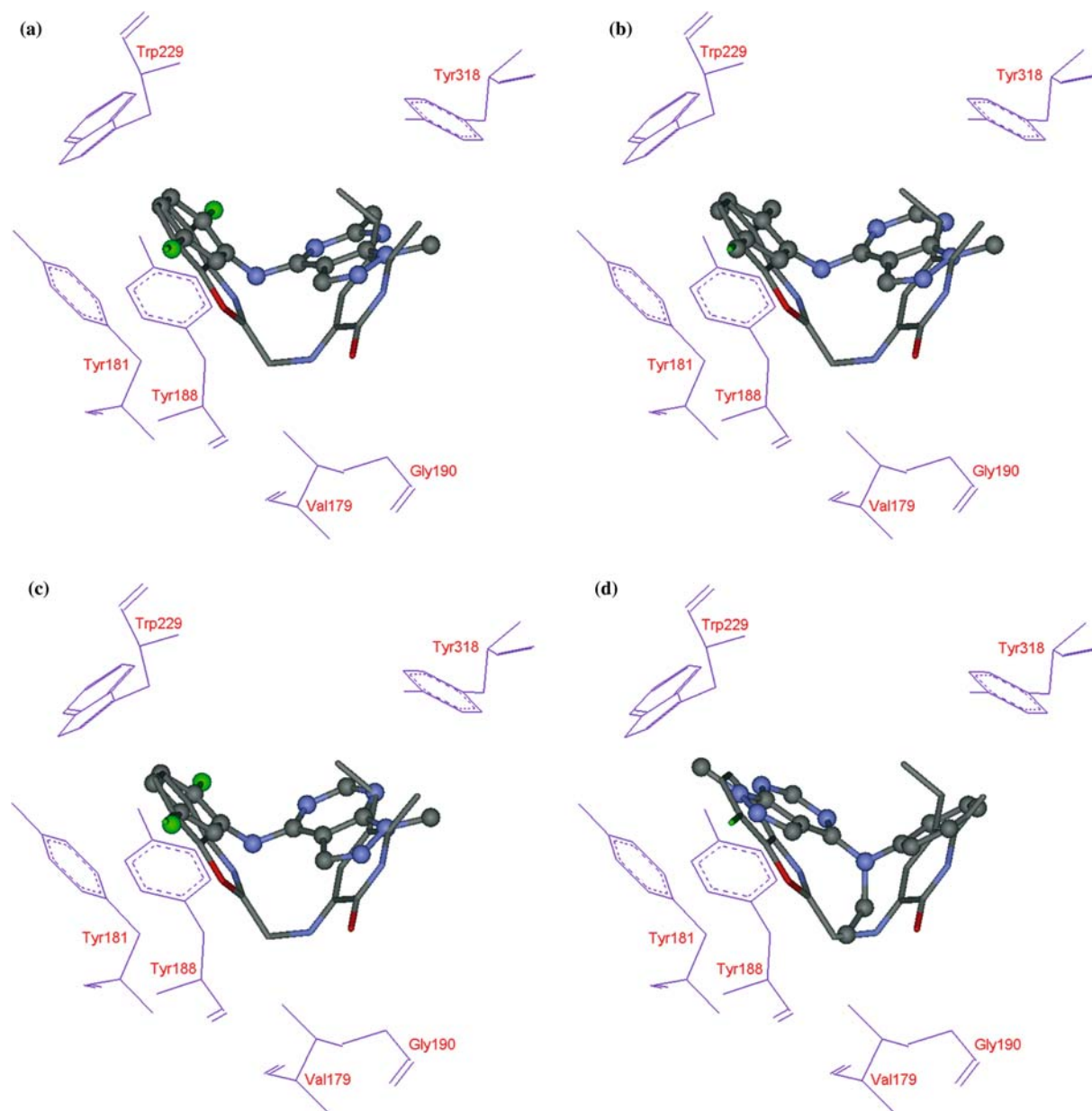
Figure 7. Comparison of the binding conformations of the pyrazo[3,4-*d*]pyrimidines (balls and sticks) with **27** (sticks) [13]. (a) NSC 11635. (b) NSC 11646. (c) NSC 1456. (d) NSC 11648. The position of Val179, Tyr181, Tyr188, Gly190, Trp229 and Tyr318 is displayed for reference. Hydrogens are omitted for clarity.

reported an acceptable quantitative correlation between the predicted binding free energy and the biological activity especially for structural similar compounds, i.e., all compounds with a benzoxazol ring [13]. Although a quantitative correlation between experimental activities and predictions were better with QSAR, all the above observations suggest that the pyrazo[3,4-*d*]pyrimidines represent promising novel NNRTIs leads.

Compound NSC 127 [56] was also found to adopt a butterfly-like conformation in the nevirapine binding site. The two predicted binding modes, represented by the most populated clusters 2 and 4 (Table 4), are similar to the binding mode

240

Table 5. Predicted activity by *k*NN best models, docking energy and free energy of binding as calculated by AutoDock for selected molecules.

Compd.	Predicted activity by *k*NN models (pIC$_{50}$)[a]	Docking energy (kcal/mol)[b]	Free energy of binding (kcal/mol)[b]
NSC 11635	6.44	−10.03	−9.40
NSC 11646	6.34	−9.37	−8.72
NSC 1456	6.12	−9.34	−8.65
NSC 11648	5.89	−8.61	−7.98

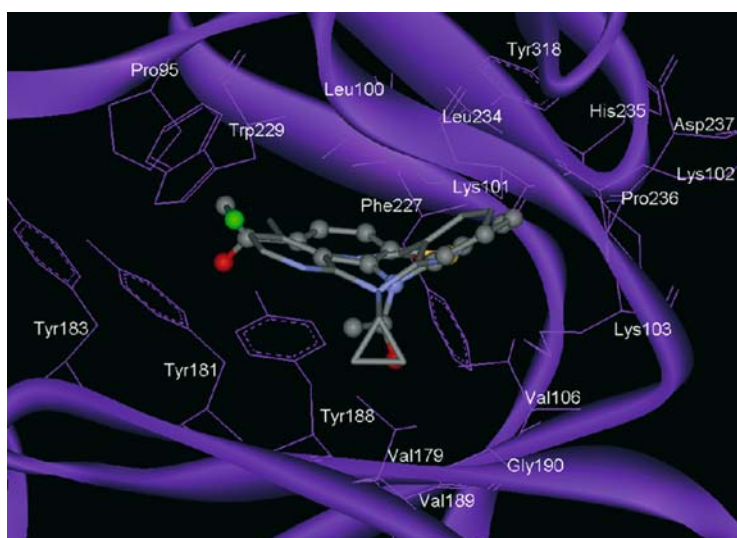[a]Average of predictions.
[b]Value for the optimal structure.

Figure 8. Top ranked binding mode (cluster 2) of NSC 127 (balls and sticks) in the binding pocket of HIV-1 RT. Nevirapine (sticks) is displayed for comparison. Hydrogens are omitted for clarity.

of nevirapine. The binding pocket of NSC 127 is formed by amino acids Pro95, Leu100, Lys101, Lys102, Lys103, Val106, Val179, Tyr181, Tyr183, Tyr188, Val189, Gly190, Phe227, Trp229, Leu234, His235, Pro236, Asp237 and Tyr318 (Figure 8). As expected, the methylketone moiety of NSC 127 could be occupying the same pocket region that the cyclopropyl group of nevirapine making contacts with Val106, Val179, Tyr181, Tyr188, Val189 and Gly190.

Conclusions and future studies

We have developed and validated QSAR models for a series of NNRTIs. The models were further used to search the NCI database for novel inhibitors. Several classes of compounds were predicted to be active. The most promising novel leads are pyrazolo[3,4-*d*]pyrimidines and pheno-thiazine derivatives. Flexible docking studies of these types of compounds to HIV-1 RT have suggested that they could be binding to the NNRTIs binding site in a very similar mode to other NNRTIs, especially nevirapine.

Biological testing of these types of derivatives, e.g., NSC 11635 and NSC 127, is the next step in our effort to discover novel anti-HIV-1 compounds. The approaches employed in this paper for the rational discovery of HIV RT inhibitors could be adapted for similar studies of many biologically active series of compounds.

References

1. De Clercq, E., Int. J. Biochem. Cell Biol., 36 (2004) 1800.
2. Jacobo-Molina, A. and Arnold, E., Biochemistry, 30 (1991) 6351.
3. Goldman, M.E., Nunberg, J.H., O'Brien, J.A., Quintero, J.C., Schleif, W.A., Freund, K.F., Gaul, S.L., Saari, W.S., Wai, J.S., Hoffman, J.M., Anderson, P.S., Hupe, D.J., Emini, E.A. and Stern, A.M., Proc. Natl. Acad. Sci. USA, 88 (1991) 6863.
4. Nunberg, J.H., Schleif, W.A., Boots, E.J., O'Brien, J.A., Quintero, J.C., Hoffman, J.M., Emini, E.A. and Goldman, M.E., J. Virol., 65 (1991) 4887.
5. Saari, W.S., Hoffman, J.M., Wai, J.S., Fisher, T.E., Rooney, C.S., Smith, A.M., Thomas, C.M., Goldman, M.E., O'Brien, J.A., Nunberg, J.H., Quintero, J.C., Schleif, W.A., Emini, E.A., Stern, A.M. and Anderson, P.S., J. Med. Chem., 34 (1991) 2922.
6. Hoffman, J.M., Wai, J.S., Thomas, C.M., Levin, R.B., O'Brien, J.A. and Goldman, M.E., J. Med. Chem., 35 (1992) 3784.
7. Saari, W.S., Wai, J.S., Fisher, T.E., Thomas, C.M., Hoffman, J.M., Rooney, C.S., Smith, A.M., Jones, J.H., Bamberger, D.L., Goldman, M.E., O'Brien, J.A., Nunberg, J.H., Quintero, J.C., Schleif, W.A., Emini, E.A. and Anderson, P.S., J. Med. Chem., 35 (1992) 3792.
8. Hoffman, J.M., Smith, A.M., Rooney, C.S., Fisher, T.E., Wai, J.S., Thomas, C.M., Bamberger, D., Barnes, J.L., Williams, T.M., Jones, J.H., Olson, B.D., O'Brien, J.A., Goldman, M.E., Nunberg, J.H., Quintero, J.C., Schleif, W.A., Emini, E.A. and Anderson, P.S., J. Med. Chem., 36 (1993) 953.
9. Davey, R.T., Dewar, R.L., Reed, G.F., Vasudevachari, M.B., Polis, M.A., Kovacs, J.A., Fallon, J., Walker, R.E., Masur, H., Haneiwich, S.E., O'Neill, D.G., Decker, M.R., Metcalf, J.A., Deloria, M.A., Laskin, O.L., Salzman, N. and Lane, H.C., Proc. Natl. Acad. Sci. USA, 90 (1993) 5608.
10. Saag, M.S., Emini, E.A., Laskin, O.L., Douglas, J., Lapidus, W.I., Schleif, W.A., Whitley, R.J., Hildebrand, C., Byrnes, V.W., Kappes, J.C., Anderson, K.W., Massari, F.E., Shaw, G.M. and L-697,661 Working Group., N. Engl. J. Med., 329 (1993) 1065.
11. Garg, R., Gupta, S.P., Gao, H., Babu, M.S., Debnath, A.K. and Hansch, C., Chem. Rev., 99 (1999) 3525.
12. Gupta, S.P., In Jucker, E. (Ed.), Progress in Drug Research. Birkhäuser Verlag, Basel, Switzerland, 2002, vol. 58, pp. 252–253.
13. Medina-Franco, J.L., Rodríguez-Morales, S., Juárez-Gordiano, C., Hernández-Campos, A. and Castillo, R., J. Comput. Aid. Mol. Des., 18 (2004) 345.
14. Medina-Franco, J.L., Rodríguez-Morales, S., Juárez-Gordiano, C., Hernández-Campos, A., Jiménez-Barbero, J. and Castillo, R., Bioorg. Med. Chem., 12 (2004) 6085.
15. Esnouf, R., Ren, J., Ross, C., Jones, Y., Stammers, D. and Stuart, D., Nat. Struct. Biol., 2 (1995) 303.
16. Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D. and Stammers, D., Nat. Struct. Biol., 2 (1995) 293.
17. Hopkins, A.L., Ren, J., Esnouf, R.M., Willcox, B.E., Jones, E.Y., Ross, C., Miyasaka, T., Walker, R.T., Tanaka, H., Stammers, D.K. and Stuart, D.I., J. Med. Chem., 39 (1996) 1589.
18. Ren, J., Esnouf, R., Hopkins, A., Ross, C., Jones, Y., Stammers, D. and Stuart, D., Structure, 3 (1995) 915.
19. Ren, J., Milton, J., Weaver, K.L., Short, S.A., Stuart, D.I. and Stammers, D.K., Structure, 8 (2000) 1089.
20. Ren, J., Nichols, C., Bird, L.E., Fujiwara, T., Suginoto, H., Stuart, D.I. and Stammers, D.K., J. Biol. Chem., 275 (2000) 14316.
21. Esnouf, R.M., Ren, J., Hopkins, A.L., Ross, C.K., Jones, E.Y., Stammers, D.K. and Stuart, D.I., Proc. Natl. Acad. Sci. USA, 94 (1997) 3984.
22. Ren, J., Diprose, J., Warren, J., Esnouf, R.M., Bird, L.E., Ikemizu, S., Slater, M., Milton, J., Balzarini, J., Stuart, D.I. and Stammers, D.K., J. Biol. Chem., 275 (2000) 5633.
23. Chan, J.H., Hong, J.S., Hunter, III R.N., Orr, G.F., Cowan, J.R., Sherman, D.B., Sparks, S.M., Reitter, B.E., Andrews, III C.W., Hazen, R.J., St Clair, M., Boone, L.R., Ferris, R.G., Creech, K.L., Roberts, G.B., Short, S.A., Weaver, K., Ott, R.J., Ren, J., Hopkins, A., Stuart, D.I. and Stammers, D.K., J. Med. Chem., 44 (2001) 1866.
24. Hsiou, Y., Das, K., Ding, J., Clark, A.D. Jr., Kleim, J.P., Rosner, M., Winkler, I., Riess, G., Hughes, S.H. and Arnold, E., J. Mol. Biol., 284 (1998) 313.
25. Ren, J., Esnouf, R.M., Hopkins, A.L., Stuart, D.I. and Stammers, D.K., J. Med. Chem.,, 42 (1999) 3845.
26. Ren, J., Esnouf, R.M., Hopkins, A.L., Warren, J., Balzarini, J., Stuart, D.I. and Stammers, D.K., Biochemistry, 37 (1998) 14394.
27. Hogberg, M., Sahlberg, C., Engelhardt, P., Noreen, R., Kangasmetsa, J., Johansson, N. G., Oberg, B., Vrang, L., Zhang, H., Sahlberg, B.L., Unge, T., Lovgren, S., Fridborg, K. and Backbro, K., J. Med. Chem., 43 (2000) 304.
28. Das, K., Clark, A.D. Jr., Lewi, P.J., Heeres, J., Jonge, M.R.De, Koymans, L.M.H., Vinkers, H.M., Daeyaert, F., Ludovici, D.W., Kukla, M.J., Corte, B.De, Kavash, R.W., Ho, C.Y., Ye, H., Lichtenstein, M.A., Andries, K., Pauwels, R., Béthune, M.-P.De, Boyer, P.L., Clark, P., Hughes, S.H., Janssen, P.A.J. and Arnold, E., J. Med. Chem., 47 (2004) 2550.
29. Mao, C., Subeck, E.A., Venkatachalam, T.K. and Uckun, F.M., Biochem. Pharmacol., 60 (2000) 1251.
30. Zheng, W. and Tropsha, A., J. Chem. Inf. Comput. Sci., 40 (2000) 185.
31. Tropsha A., Cho S.J. and Zheng, W., In Parrill, A.L. and Reddy, M.R., (Eds.), ACS Symposium Series 719, American Chemical Society, Washington, DC, 1999, pp. 198–211.
32. Shen, M., LeTiran, A., Xiao, Y., Golbraikh, A., Kohn, H. and Tropsha, A., J. Med. Chem., 45 (2002) 2811.
33. Shen, M., Béguin, C., Golbraikh, A., Stables, J.P., Kohn, H. and Tropsha, A., J. Med. Chem., 47 (2004) 2356.
34. http://search.nci.nih.gov/search97/cgi/s97_cgi.
35. http://dtp.nci.nih.gov/docs/aids/aids_data.html.
36. The program Sybyl is available from Tripos Associates, St. Louis, MO.
37. Randic, M., J. Am. Chem. Soc., 97 (1975) 6609.

38. Kier, L.B. and Hall, L.H. Molecular Connectivity in Chemistry and Drug Research. Academic Press, New York, 1976.
39. Molconn-Z, version 4.05, Hall Associates Consulting, Quincy, MA.
40. Sharaf, M.A., Illman, D.L. and Kowalski, B.R. Chemometrics. John Wiley & Sons, New York, 1986.
41. Gilbert, N., Statistics, W.B. Saunders, Co., Philadelphia, PA, 1976.
42. Golbraikh, A. and Tropsha, A., J. Comput. Aid. Mol. Des.,, 16 (2002) 357.
43. Sachs, L., Applied Statistics, A Handbook of Techniques. Springer, Verlag, 1984, pp. 349.
44. Tropsha, A. and Zheng, W., Curr. Pharm. Design, 7 (2001) 599.
45. Tropsha, A., Gramatica, P. and Gombar, V.K., Quant. Struct. Act. Relat. Comb. Sci., 22 (2003) 69.
46. Powell, M.J.D., Math. Program., 12 (1977) 241.
47. Clark, M. and Cramer, R.D., N., J. Comp. Chem., 10 (1989) 982.
48. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., J. Comp. Chem., 19 (1998) 1639.
49. http://www.rcsb.org.
50. Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3219.
51. http://www.scripps.edu/pub/olson-web/doc/autodock/tools.html.
52. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
53. Mehler, E.L. and Solmajer, T., Protein Eng., 4 (1991) 903.
54. Golbraikh, A. and Tropsha, A., J. Mol. Graph. Model., 20 (2002) 269.
55. Cheng, C.C. and Robins, R.K., J. Org. Chem., 21 (1956) 1240.
56. Chan, C., Yin, H., Garforth, J., McKie, J.H., Jaouhari, R., Speers, P., Douglas, K.T., Rock, P.J., Yardley, V., Croft, S.L. and Fairlamb, A.H., J. Med. Chem., 41 (1998) 4910.