Pergamon

Bioorganic & Medicinal Chemistry 7 (1999) 2283–2292

BIOORGANIC & MEDICINAL CHEMISTRY

Comparative Molecular Field Analysis (CoMFA) and Docking Studies of Non-nucleoside HIV-1 RT Inhibitors (NNIs)

M. L. Barreca,[a] A. Carotti,[b] A. Carrieri,[b] A. Chimirri,[a,*] A. M. Monforte,[c] M. Pellegrini Calace[b] and A. Rao[a]

[a]*Dipartimento Farmaco-Chimico, Università di Messina, Viale Annunziata, 98168 Messina, Italy*
[b]*Dipartimento Farmaco-Chimico, Università di Bari, Via E. Orabona 4, 70125 Bari, Italy*
[c]*Facoltà di Farmacia, Università di Catanzaro, 88021 Roccelletta di Borgia, Catanzaro, Italy*

Received 5 February 1999; accepted 1 April 1999

Abstract—A set of TIBO derivatives endowed with reverse transcriptase (RT) inhibitory activity were analyzed by comparative molecular field analysis (CoMFA). Besides conventional steric and electrostatic fields, molecular lipophilicity potential (MLP) was also used as a third field in CoMFA. An informative and statistically significant model ($q^2 = 0.70$, $r^2 = 0.90$, $s = 0.46$) was obtained by taking into account the three field types together. The key molecular determinants governing the RT inhibition by TIBO congeners were detected at the 3-D level by a careful analysis of the CoMFA isocontour maps. To challenge the predictive ability of the CoMFA model, an external set of thiazolobenzimidazole (TBZ) derivatives were examined. Good predictions, suggesting a similar binding mode for TIBO and TBZ derivatives, emerged. Flexible docking experiments on TBZ, TIBO and other NNIs confirmed common binding characteristics, as found out also by CoMFA, and moreover a good correlation between calculated binding energies and inhibitory potency was found.

Introduction

In the life of human immunodeficiency virus (HIV), the reverse transcriptase (RT) is a key multifunctional enzyme that constitutes an important target for the development of new antiviral agents. RT is essential for HIV replication and moreover contains multiple sites where inhibitors can efficiently bind.[1]

In previous publications,[2–5] the synthesis of a series of 1*H*,3*H*-thiazolo[3,4-*a*]benzimidazoles exhibiting significant anti-HIV activity led to the discovery of TBZ (**53**) (NSC 625487)[2] as a potent and selective non-nucleoside HIV-1 RT inhibitor.[6,7] TBZ (**53**) was active against a panel of biologically diverse strains, including the AZT resistant strain G910-6, but it was inactive against HIV-2. Combination of TBZ with AZT and ddI synergistically inhibited HIV-1-induced cell killing.[6]

The biological activity of TBZ (**53**) seems to be associated to its capability to assume a 'butterfly-like' conformation[6] as already observed for other NNIs such as Nevirapine (**1**)[8] and TIBO [**31**] (Chart 1).

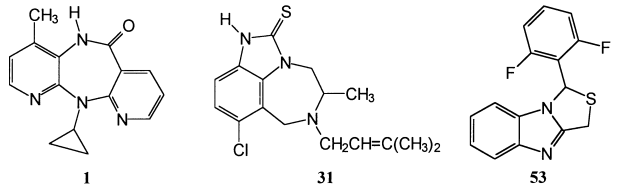
The aim of the present research was to obtain further insight into the relationships between the structure and inhibitory potency of TIBO RTIs, to locate at the 3-D level the main intermolecular interactions involved in the RT inhibition and finally to use these results to derive informative SAR on a series of TBZ derivatives prepared in our laboratories.[2,3,5]

A further important goal of our study was the demonstration of a common binding mode of a series of structurally diverse NNIs. To reach our research objectives two modeling techniques have been used, namely comparative molecular field analysis (CoMFA) and flexible docking, implemented, respectively, in the Sybyl 6.4[9] and QXP[10] molecular modeling packages.

CoMFA[11] is a widely-used tool for the study of quantitative structure–activity relationships (QSAR) at the 3-D level. Unlike the traditional Hansch analysis, which relies on substituent parameters, CoMFA relates the biological activity of a series of molecules with their steric and electrostatic fields sampled at grid points defining a large 3-D box around the molecule. CoMFA columns (descriptors) consist of steric (Lennard–Jones) and electrostatic (Coulomb) potentials computed for each molecule, at each grid point, by means of a suitable

* Corresponding author.

0968-0896/99/$ - see front matter
PII: S0968-0896(99)00181-9



1 31 53

Chart 1.

probe, usually an sp^3 carbon atom with a charge of +1. Partial least squares (PLS)[12] is the regression method used to develop the relationship between independent variables (steric and electrostatic potentials) and biological activity (dependent variable). PLS analysis produces model equations which explain the variance in the target property in terms of the independent variables. The optimum number of components (latent variables) is determined by cross-validation and the model predictive ability is assessed by cross-validated r^2 (r^2cv, q^2). The graphical representation of CoMFA results (isocontour maps) indicated the regions where the variation in steric and electrostatic properties of different molecules in a data set is correlated with the variation of biological activity.

CoMFA is not able to appropriately describe all binding forces, being based only on standard steric and electrostatic molecular fields to model receptor–ligand interactions. Furthermore, CoMFA describes only the enthalpic component of the ligand–receptor interactions. Introducing the molecular lipophilicity potential (MLP)[13] as an additional field has been shown to significantly improve the descriptive, predictive and interpretative powers of CoMFA in many cases. Indeed the MLP encodes hydrogen bonds and hydrophobic interactions not sufficiently described by the steric and electrostatic fields and includes also an entropy component.[14]

For these reasons, the CoMFA methodology, with the inclusion of the MLP, was selected as an appropriate tool to study the SAR of TIBO congeners. Regrettably, for TBZ derivatives the lack of parametrization of some molecular fragments prevented the use of MLP as a third field in CoMFA.

Results and Discussion

High resolution crystallographic coordinates of HIV-1 RT[15] and its complexes with some NNIs were used as good starting points for the 3-D-QSAR CoMFA study. Those data were particularly useful to select reliable inhibitor conformations both for the alignment in CoMFA and for the subsequent docking studies. In particular CoMFA was carried out starting from a TIBO conformation recovered from the crystallographic data of RT/TIBO R86183 (**31**) complex.[8]

TIBO, tetrahydroimidazo(4,5,1-JK)(1,4)-benzodiazepin-2(1*H*)-ones and thiones,[16–18] are potent and highly selective anti-HIV-1 inhibitors whose mechanism of action has been ascribed to an allosteric inhibition of viral RT, through a tight binding in a region close to the polymerase site.[19]

Thirty-seven TIBO derivatives (Table 1) carrying different substituents at N-6 and at the benzodiazepine ring and an ureidic or thioureidic function at the condensed five-membered ring, were selected as a training set (TS) for the CoMFA study. The selection was made so as to have a good variation of steric, electrostatic and lipophilic properties. Moreover the biological data were homogeneous and regularly distributed in a relatively wide activity range (pIC_{50} from 3.80 to 8.52).

Also the TIBO molecules used as a prediction set (PS, Table 1) showed a good spread and distribution of data points. The inhibitory potency was expressed as $-\log IC_{50}$ where IC_{50} is the concentration of compound required to achieve 50% protection of MT-4 cells against the cytopathic effect of HIV-1.

Table 1 lists the chemical structures and biological activities of TIBO derivatives forming the TS, used to derive the CoMFA models, and the PS used to challenge their predictive capability.

CoMFA studies on TIBO congeners

The statistical results obtained in the CoMFA study of TIBO inhibitors are summarized in Table 2.

With the exception of model 1, PLS models based on single field are not very significant in terms of predictive (q^2) and fitting (r^2, s) values. The pair-wise combination of the three fields yielded only marginal statistical improvements. The most satisfactory statistical model, both in fitting and predictive terms, was model 7 combining all the three fields.

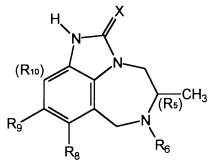
The relative field contributions were 49% steric, 30% electrostatic, 21% lipophilic suggesting a balanced model with prevalent steric effects. Its fitting power can easily be judged from the plot of measured versus predicted pIC_{50} values shown in Figure 1.

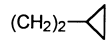
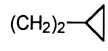
The CoMFA steric, electrostatic and lipophilic field signals for model 7 are shown as single field contour maps and as merged fields in Figure 2.

Figure 2 (upper left) shows the steric contour map from model 7. Compounds **3** ($pIC_{50}=3.80$), **8** ($pIC_{50}=5.18$) and **26** ($pIC_{50}=8.52$) were added to aid interpretation. The green and red polyhedra represent areas of steric bulk tolerance and steric hindrance, respectively. The red contours surround groups like $NHCOCH_3$ of compound **3** and OCH_3 of compound **8**, while the green contours surround the side chain at N-6. Compound **26** (the most active one) occupies only sterically allowed areas.

Figure 2 (upper right) shows compounds **8** ($pIC_{50}=5.18$) and **34** ($pIC_{50}=8.48$) embedded in the CoMFA electrostatic contour maps. The magenta and white polyhedra define, respectively, areas of favorable or

Table 1. Chemical structures of TIBO derivatives **2**–**47**



Compounds	X	R_6	R_5	R_9	R_x[a]	pIC_{50}	Set_x[b]
2	O		H	H		5.00	PS
3	O		H	$NHCOCH_3$		3.80	TS
4	O		H	NO_2		4.48	TS
5	O		H	NH_2		4.22	TS
6	O		$N(CH_3)_2$	H		5.18	TS
7	O	$CH_2CH=C(CH_3)_2$	H	H		5.40	TS
8	O	$CH_2CH=C(CH_3)_2$	H	H	R_{10}=OCH_3	5.18	TS
9	O	$CH_2CH=C(CH_3)_2$	C≡CH	H		6.36	PS
10	O	$CH_2CH=C(CH_3)_2$	CN	H		5.94	PS
11	O	$CH_2CH=C(CH_3)_2$	CH_3	H		6.00	PS
12	O	$CH_2CH=C(CH_3)_2$	I	H		7.06	TS
13	O	$CH_2CH=C(CH_3)_2$	Br	H		7.33	TS
14	O	$CH_2CH=C(CH_3)_2$	$CONH_2$	H		5.20	TS
15	O	$CH_2CH=C(CH_3)_2$	H	CF_3		5.23	TS
16	O	$CH_2CH=C(CH_3)_2$	H	Cl		6.74	TS
17	O		H	Cl		5.66	TS
18	O	$CH_2C(CH_3)=CH_2$	H	H		4.82	TS
19	O	$CH_2CH=C(Et)_2$	H	CH_3		6.50	TS
20	O	$CH_2CH=C(Et)_2$	H	Cl		7.54	TS
21	O	$CH_2CH_2CH_3$	H	H		4.22	TS
22	O		H	Cl		6.42	PS
23	S	$CH_2CH=C(CH_3)_2$	H	H		7.36	TS
24	S	$CH_2CH=C(CH_3)_2$	H	H	R_{10}=OCH_3	5.33	TS
25	S	$CH_2CH=C(CH_3)_2$	OCH_2CH_3	H		7.02	TS
26	S	$CH_2CH=C(CH_3)_2$	Br	H		8.52	TS
27	S	$CH_2CH=C(CH_3)_2$	Cl	H		7.34	TS
28	S	$CH_2CH=C(CH_3)_2$	F	H		8.24	TS
29	S	$CH_2CH=C(CH_3)_2$	CN	H		7.25	TS
30	S	$CH_2CH=C(CH_3)_2$	CH_3	H		7.85	TS
31	S	$CH_2CH=C(CH_3)_2$	Cl	H		8.34	TS
32	S	$CH_2CH=C(CH_3)_2$	I	H		7.32	TS
33	S	$CH_2CH=C(CH_3)_2$	C≡CH	H		7.53	PS
34	S	$CH_2CH=C(CH_3)_2$	H	Cl		8.48	TS
35	S	$CH_2CH=C(CH_3)_2$	H	Cl	R_5=H	6.80	TS
36	S	$CH_2CH=C(CH_3)_2$	H	F		7.60	TS
37	S	$CH_2CH=C(CH_3)_2$	H	CF_3		6.31	TS
38	S		H	Cl		6.38	TS
39	S	$CH_2C(CH_3)=CH_2$	H	H		7.85	TS
40	S	$CH_2CH=C(Et)_2$	H	Cl		7.92	TS
41	S	$CH_2CH=CH_2$	H	H		4.17	TS
42	S		H	H		7.22	TS
43	S		H	Cl		7.47	TS
44	S		H	NO_2		5.61	PS
45	S		H	Cl		7.88	TS
46	S	$CH_2CH_2CH_3$	H	H		5.78	TS

[a] Additional structural features.
[b] Used in the prediction set (PS) or training set (TS).

Table 2. CoMFA statistical results

Model	Fields	NOC	$q^{2\,a}$	$r^{2\,a}$	$r^2_{pred}{}^a$	$s_{pred}{}^a$
1	ste	2	0.70	0.81	0.78	0.46
2	ele	1	0.52	0.86		
3	lipo	3	0.55	0.73		
4	ste, ele	2	0.65	0.78	0.69	0.59
5	ste, lipo	2	0.71	0.81		
6	ele, lipo	1	0.58	0.67		
7	ste, ele, lipo	4	0.70	0.90	0.96	0.15

[a] q, r, NOC are the cross-validated correlation coefficient, the correlation coefficient and the optimal number of components, respectively; r_{pred} and s_{pred} are the correlation coefficient and the standard deviation referred to the linear correlation between predicted and observed values for the prediction set.

unfavorable interactions for negative charges. The magenta area surrounds the ureidic or thioureidic groups, while the white area is occupied by electron-rich groups in the case of less active compounds.

In Figure 2 (lower left) the lipophilic contour map is shown with compounds **3** ($pIC_{50} = 3.80$), **34** ($pIC_{50} = 8.48$) and **26** ($pIC_{50} = 8.52$). The yellow color refers to a large region of favorable lipophilic binding occupied by halogen substituents at position 9 occurring in very active compounds. It is worth noting that for this class of inhibitors no favorable hydrophilic interactions were found by MLP.

In Figure 2 (lower right) the isocontour graphs coming from the merging of the three single field maps are shown together with ligands **3**, **8**, **26** and **34**: highly active molecules (**34** and **26**) do not occupy forbidden steric regions whereas positive electrostatic contributions can be mainly located close to electron-rich groups in the magenta regions. Lipophilic substituents such as Cl and Br in compounds **34** and **26** respectively, lie on a lipophilic yellow region. Low activity molecules such as **3** and **8** show the OCH_3 and $NHCOCH_3$ substituents within unfavorable steric (red) and electrostatic (white) zones.

The salient features of model 7 emerging from the previous analysis of the contour maps furnish significant insight into the physicochemical characteristics and location of the key interactions governing the inhibitory binding process of TIBO congeners.

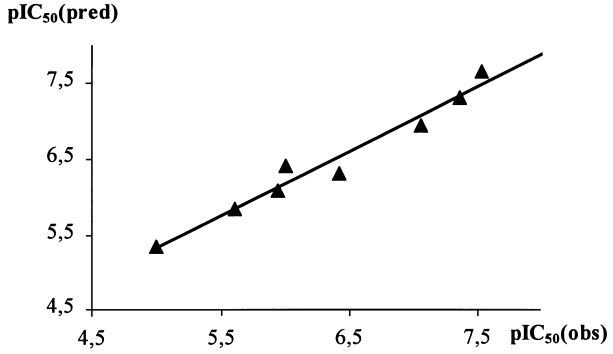


Figure 1. Plot of measured versus predicted pIC_{50} values for the TIBO prediction set.

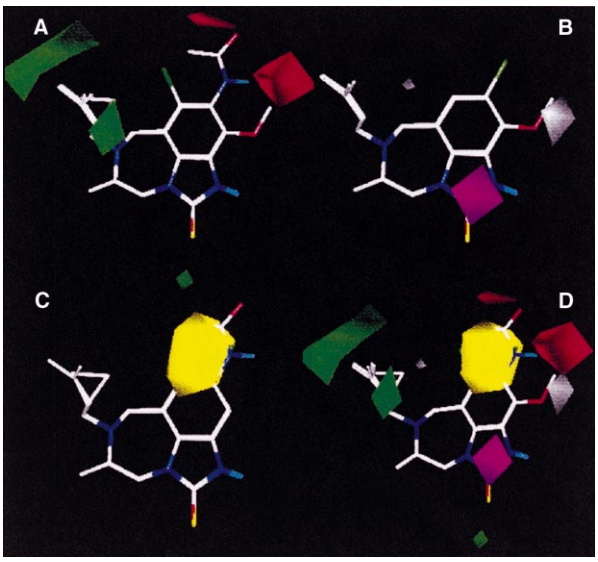


Figure 2. Isocontour maps for PLS model 7 expressed as STEV×COEFF. (A) Steric map (contour levels: 0.052 green; −0.048 red. Color codes: green increased affinity, red decreased affinity), compounds **3**, **8** and **26** are shown to help interpretation. (B) Electrostatic map (contour levels: 0.043 white; −0.095 magenta. Color codes: white decreased affinity for negative charge; magenta increased affinity for negative charge), compounds **8** and **34** are shown to help interpretation. (C) Lipophilic map (contour levels: 0.023 yellow. Color codes: yellow increased affinity), compounds **3**, **34** and **26** are shown to help interpretation. (D) CoMFA three-field isocontour map, compounds **3**, **8**, **34** and **26** are shown to help interpretation.

Interestingly, our 3-D-QSAR analysis of TIBO congeners is in full agreement with a recent QSAR study[20] carried out by the traditional Hansch approach on the same set of compounds. Our results permit, however, a better definition of the favorable lipophilic interaction of substituents at positions 8 and 9 and a detection of a steric effect for substituents at position 10.

To further evaluate the validity of model 7, especially in predictive terms, compounds forming the PS were examined. The statistical parameters used for this evaluation were the r^2_{pred} and s_{pred}, which represent, respectively, the squared correlation coefficient resulting from the regression equation of the observed versus predicted activities and the standard deviation as a measure of the prediction error.

The r^2_{pred} and s_{pred} from model 7 were much better than those from model 1, assuming the values of 0.96 and 0.15, respectively.

These figures reinforce the high statistical value of the PLS model 7. From Table 3, it can be seen that the activities of all the compounds in the PS were predicted within 0.40 log units of their actual pIC_{50} with an average absolute error of 0.19 log units along a range of 2.53 units. Also these findings were quite gratifying.

Molecular modeling of the inhibitory potency of TBZ congeners

The predictive ability of PLS model 7 was tested also on an external set of structurally diverse inhibitors

Table 3. Observed and predicted values from model 7 for the TIBO prediction set

Compounds observed	pIC_{50} (obs)	pIC_{50} (pred)	pIC_{50} (obs)−pIC_{50} (pred)
2	5.00	5.34	−0.34
44	5.61	5.84	−0.23
10	5.94	6.07	−0.13
11	6.00	6.40	−0.40
22	6.42	6.31	0.11
12	7.06	6.94	0.12
23	7.36	7.29	0.07
33	7.53	7.64	−0.11

constituted by the 1-aryl-1*H*,3*H*-thiazolo[3,4-*a*]benzimidazoles **47**–**53** (TBZ) reported in Table 4.

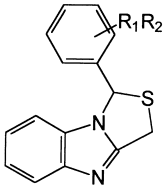
The 3-D structures of TBZ were compared with those of TIBO derivatives, aiming at the maximum overlap of the most plausible pharmacophoric elements. The enantiomers with R absolute configuration were selected since they give a better fit in a butterfly-like geometry.

The molecular model of the most active compound TBZ (**53**), was constructed using available crystallographic data[6] and its geometry optimized by the semi-empirical molecular orbital method AMI.[21] Then, a conformational search was carried out on the phenyl torsion angle N(2)-C(1)-C(10)-C(11) looking for the minimum energy conformer. All the other TBZ derivatives were constructed from the structure of compound **53** and optimized with the same method.

The best geometric fit between the two classes of compounds TIBO and TBZ resulted from the alignment shown in Figure 3 (rms = 0.18) in which the fitted elements were numbered and labeled with an asterisk.

The orientation of compound **53** resulting from this fit was used as starting conformation for docking studies into the non-nucleosidic binding site of RT.

Table 4. Chemical structures and inhibition data of TBZ derivatives **47**–**53**



Compounds	R_1R_2	pIC_{50} (obs)
47	2Cl,5NO_2	5.12
48	3CN	5.46
49	3F	5.54
50	2NO_2	5.65
51	2Cl,6Cl	6.30
52	2Cl,6F	6.39
53	2F,6F	6.41

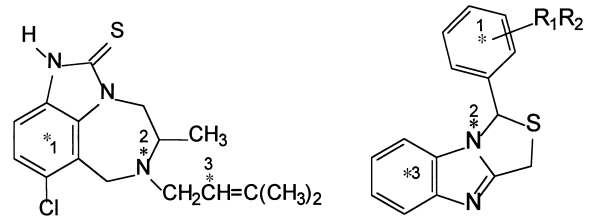


Figure 3. Superimposition criteria of TIBO and TBZ compounds: fitted elements are labeled with an asterisk and numbered.

Figure 4 shows compounds **47**–**53**, overlaid according to the previously indicated fitting elements, surrounded by CoMFA steric, electrostatic and lipophilic contour maps. In other terms, the isocontour maps, as resulting from the CoMFA analysis of TIBO, were used as field template to test, at a qualitative level, their capability to predict the rank order of activity of another different class of RT inhibitors. It was interesting to observe in Figure 4 that the lipophilic region surrounds the halogen of the substituted aromatic rings in the most active TBZ compounds, while the nitro group of compound **47** ($pIC_{50} = 5.12$) is embedded in the red and white polyhedra which represent areas of steric hindrance and unfavorable interactions for negative charges.

These findings indicated that model 7 can be used to correctly estimate, at least at a qualitative level, the inhibitory activity of TBZ derivatives. As previously stated, the lack of parametrization of some molecular fragments prevented the possibility to calculate the MLP for TBZ molecules and therefore model 7 could not be used for a quantitative prediction of their inhibitory potency. However, the relatively good statistics of the steric and electrostatic two-fields model 4 (54 and 46% contribution, respectively) prompted us to use it for a tentative quantitative prediction of the TBZ inhibitory activity: model 4 revealed a certain predictive ability with an average absolute error of 0.3 log units.

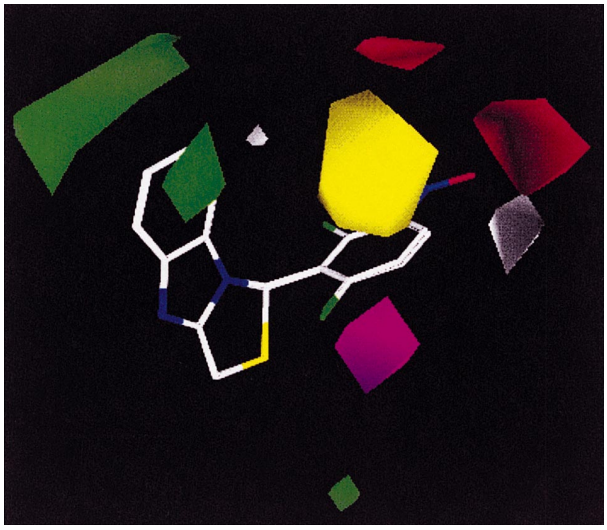


Figure 4. Merging of TBZ congeners **47**–**53** in the CoMFA three-field isocontour map from model 7.

Docking studies of NNIs into RT

RT is a dimeric protein having a polymerase site (catalytic aminoacids are ASP110, ASP185 and ASP186), and an allosteric binding site close to the active site and partially included in it. The most important aminoacids involved in hydrophobic interactions with NNIs are VAL106, VAL108, TYR181, TYR188, PHE227, TRP229 and GLU696 while VAL179 and LYS101 may be responsible for hydrogen bonding interactions.[22,23]

To further prove a common binding mode with RT of some selected NNIs and our TBZ derivatives, an intermolecular docking study was performed by means of the QXP (Quick eXPlore) molecular modeling package[10] and the interaction energies calculated from the minimized structures of the RT-inhibitor complexes.

QXP has been recently proposed as an original and rapid method for docking, template (pharmacophore) fitting and pseudo-receptor building. Its algorithms take advantage of the classical Monte Carlo[24] procedure to perform efficient conformational searches on flexible cyclic and acyclic molecules being the energy calculations made by a mixed AMBER/MM2 force-field.[25]

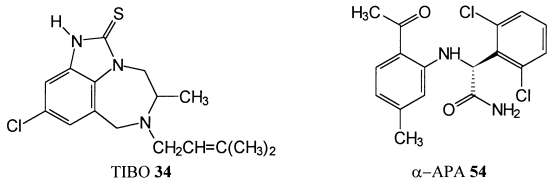
For the molecular fitting procedure, QXP uses a superposition force-field which automatically assigns short-range attractive forces to similar atoms in different molecules according to the following equation $E_{sup} = K_{sup}\ (dist^2\text{-}d^2_{cut})/d^4_{cut}$ where E_{sup} is the superimposition energy, *dist* is the interatomic distance and d_{cut} is the cutoff distance. K_{sup} is the energy constant for a perfect superimposition. Fast Monte Carlo searches are used to match proposed molecules to a template or a pharmacophore and for flexible docking to a binding site. The Polak–Ribiere conjugate gradient (PRCG) minimization[26] is then used to optimize the fitting or binding energies.

Docking is carried out using a molecular mechanics force-field without superposition energy. Guide atoms are automatically generated inside the binding site before the search starts. A dielectric constant of $\varepsilon = 4r$ is used in the calculation.

The docking study of NNIs was limited to TIBO (**34**), TBZ (**53**), *Nevirapine* (**1**) and the α-anilinophenylacetamide derivative (α-APA) (**54**) (Chart 2).

Preliminary superimposition studies of the diverse NNIs were performed before docking simulations, aimed at evaluating possible binding similarities of NNIs ligands. A qualitative analysis of the binding mode found in RT–NNIs crystal complexes showed that compounds **34**, **1** and **54** could bind to RT assuming a superimposable butterfly-like conformation (Fig. 5, upper part).

A flexible fitting procedure (see Experimental) was applied to the selected NNIs molecules. In the best developed alignment, two unsaturated zones corresponding to the wings of the butterfly and a hydrophobic alkyl region on the plane dissecting the two wings are visible (Fig. 5, lower part).



TIBO **34** α–APA **54**

Chart 2.

Moreover, to validate the TIBO/TBZ alignment chosen in our CoMFA studies, a flexible fitting of compounds **53** and **34** was performed by QXP. The resulting alignment was quite similar to the one previously used in CoMFA, as shown in Figure 6.

Finally, flexible docking simulations (see Experimental) of the ligands into the RT allosteric binding site were executed. The resulting TIBO (**34**)–RT complex shows the ligand NH forming a hydrogen bond with the backbone carbonyl oxygen of LYS101, the allylic group interacting with the phenolic rings of TYRI81 and TYR188, the first one being probably involved in a *edge-to-face* π–π stacking.[27] In fact, a measured distance

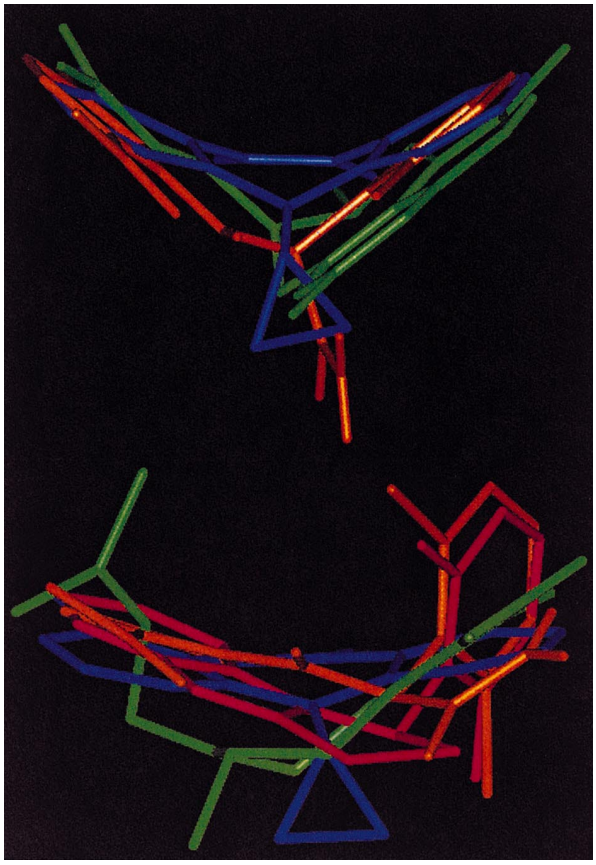


Figure 5. Upper part: superimposition of NNIs **34** (green), **1** (blue), and **54** (orange) as extracted from their crystal complexes. Lower part: alignment of compounds **34** (green), **53** (red), **1** (blue) and **54** (orange) resulting from QXP fitting.

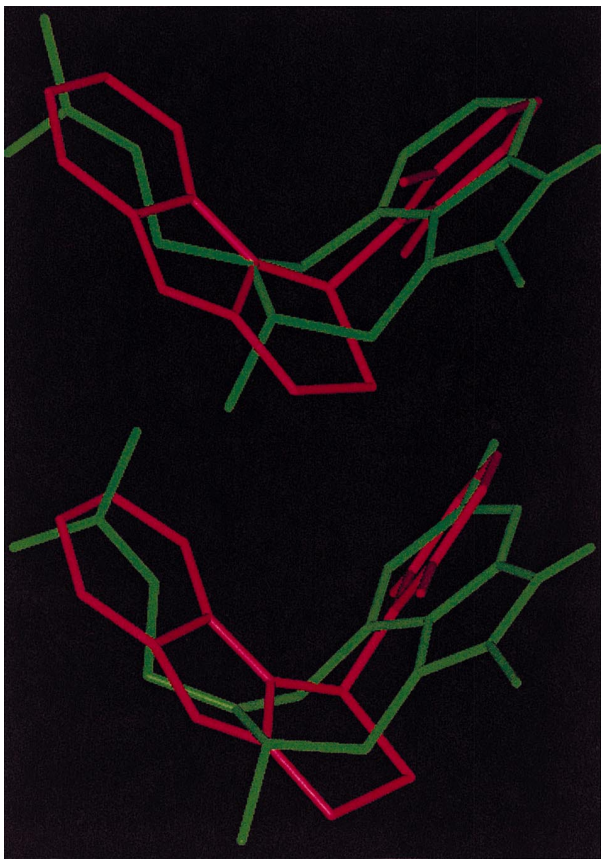


Figure 6. Alignments of compounds **34** (green) and **53** (red) resulting from QXP (upper part) and CoMFA (lower part).

of 2.64 Å between the centers of mass of the two groups and a nearly orthogonality (87.8°) of the planes passing through them was observed. Moreover, the methyl group is trapped in a hydrophobic pocket defined by TYR181, LEU100 and GLU696 (Fig. 7, upper part).

The binding mode of the α-APA (**54**) (not shown in Fig. 7) presents the acetophenyl group parallel to the phenolic ring of TYRI81 (but perpendicular to the corresponding allylic portion of TIBO (**34**)), the dichlorophenyl group in the same place of the TIBO bicyclic region, and the methyl substituent into the hydrophobic pocket. Moreover, one of the amidic hydrogens forms a HB with the backbone carbonyl oxygen atom of VAL179.

In the binding of *Nevirapine* (**1**) (not shown in Fig. 7), several favorable lipophilic interactions seem to take place and the butterfly-like conformation is practically conserved (pyridine rings take up the wing zones).

TBZ derivative **53** shows the indolic region interacting hydrophobically with the side-chains of PHE227 and TYR188 and the difluorophenyl group close to the hydrophobic pocket (Fig. 7, lower part). Unlike TIBO (**34**) and α-APA (**54**), TBZ (**53**) forms no hydrogen bond. Thus, the formation of H-bonds does not seem essential for an efficient binding to RT. *Nevirapine* (**1**), in fact, strongly interacts with RT even if no HB is formed.

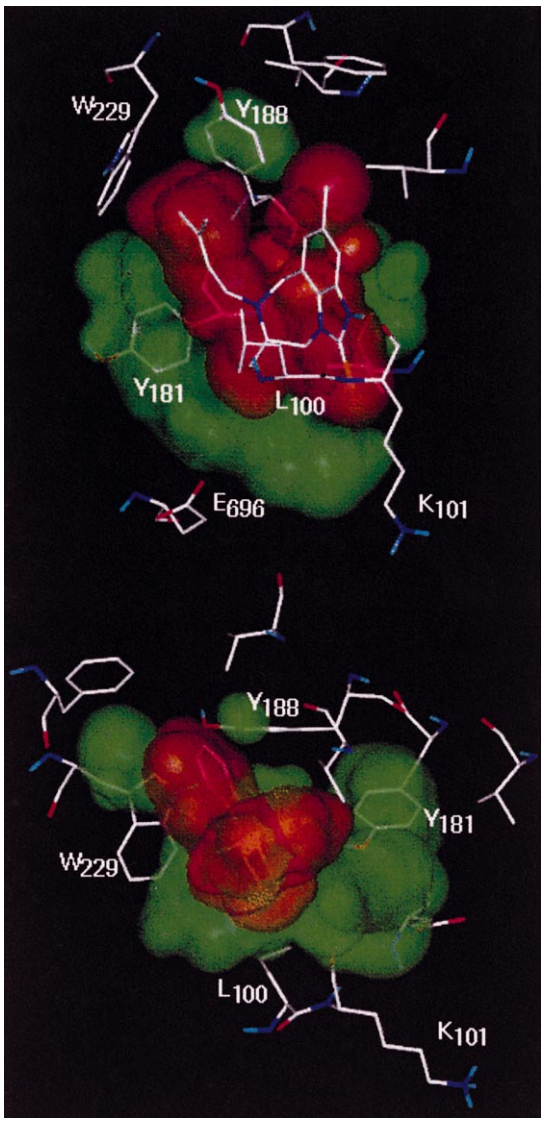


Figure 7. Upper part: minimum binding energy conformation of compound **34** into RT allosteric binding site. Red and green surfaces represent respectively the ligand and receptor cavity van der Waals surfaces as computed by the program Surfnet l.4.[37] Lower part: minimum binding energy conformation of compound **53** into RT allosteric binding site.

The interaction energies, calculated only for ligands for which homologous biological data were available, are shown in Table 5. The value obtained for TBZ (**53**)–RT complex is about 10 Kcal/mol higher than TIBO (**34**)–RT and this is in full agreement with its lower inhibitory potency. The higher stability of the TIBO (**34**)–RT complex could result both from the formation of one hydrogen bond with the LYS101 and from much more extended hydrophobic interactions of **34** compared to **53** as can be easily perceived by a simple visual comparison of the red ligand surfaces in the upper and lower parts of Figure 7.

Conclusions

A very informative and statistically significant CoMFA model was developed for a large series of TIBO derivatives

Table 5. Binding interaction energies of the listed ligands with RT

Ligand	pIC_{50} (obs)	E_i (kcal/mol)
TIBO (**34**)	8.30	−64.49
α–APA (**54**)	8.48	−73.82
TBZ (**53**)	6.30	−54.30

endowed with interesting inhibitory potency towards the HIV reverse transcriptase. The main molecular determinants responsible for high inhibitory activity were identified and proved valid also for a series of TBZ derivatives displaying the same type of biological activity. The butterfly-like conformation and a suitable spatial location of lipophilic and electron-rich groups are the key structural requirements for a potent enzyme inhibition by the two classes of NNIs, TIBO and TBZ, extensively examined in this study.

Comparative molecular field analysis and flexible docking experiments showed that TBZ and TIBO have very similar structural and binding properties. A plausible interpretation of the lower potency elicited by TBZ congeners compared to TIBO derivatives was also given.

As already stressed in recent publications,[28–30] the coordinated use of CoMFA and docking approaches, can yield significant and complementary insight into complex biological processes such as the enzyme–ligand binding.

Finally, it could be interesting to challenge the validity of the common binding mode proposed herein for TIBO and TBZ, by synthesizing new and original NN inhibitors designed on the basis of the key molecular requirements for high potency identified in the present study. Future work towards this important goal has been already planned.

Experimental

Molecular modeling

For our CoMFA studies of TIBO derivatives, compound **31** (Table 1), whose crystal structure complexed with HIV-1 RT is known, was chosen as the template to define the most likely binding conformation to be used for molecular alignment.

The molecule was isolated from its crystal complex RT–TIBO (Brookhaven Protein Data Bank, entry code 1HNV) and optimized by the semiempirical quantum mechanics method AM1 using the molecular modeling software Sybyl 6.4. The torsion angles of the side-chain, found essential to maintain a butterfly-like molecular shape, were frozen in the conformation seen in the X-ray crystallographic study of the complex.

Three-dimensional structures of the other molecules were constructed from compound **31**. A conformational search of the principal rotatable bonds was then executed and the best fitted butterfly-like conformation was selected (allowing an energy penalty up to 3 kcal/mol over the global minimum) and optimized by the same method used for the template.

All molecules were superimposed by least-square fitting of some selected atoms, indicated by an asterisk in Figure 8, on the corresponding atoms of the template.

For molecular modeling studies of TBZ derivatives, the template molecule (**53**) was built from its crystallographic data (available for the S enantiomer) with the CRYSYN module of Sybyl 6.4 and fully optimized by the semiempirical quantum mechanics method AM1. The subsequent conformational search on the inter ring torsion angle was executed with a 30° increment in the range 0–180°. The geometry of the lowest energy conformer compared well to that found by X-ray analysis. The remaining TBZ derivatives were built in the ground of compound **53** and optimized with the same method.

CoMFA study

A 3-D cubic lattice, with a 2 Å grid spacing, was generated around the aligned compounds based on the molecular volume of the structures (grid beyond the molecules extended by 4 Å in all directions). Steric and electrostatic potentials were generated by a sp^3 carbon atom probe with a charge of +1. The option 'drop electrostatic' was set to NO.

The molecular lipophilic potentials were calculated by using the Broto and Moreau atomic fragment system[31] and used in CoMFA. The fields values were calculated at each grid point as the scalar product of the associated QSAR coefficients times the standard deviation of all values in the corresponding column of the data field table (STDEV×COEFF).

The CoMFA models were derived by the partial least squares (PLS) method, implemented in Sybyl 6.4, with cross-validation (leave-one-out procedure).

The predictivity ability of the model was quantified in terms of q^2 which is defined as:

$$q^2 = (SD - PRESS)/SD$$

where $PRESS = \Sigma\ (Ypred\text{-}Yactual)^2$ and $SD = \Sigma\ (Yactual\text{-}Ymean)^2$.

SD is the sum of squares of deviations of the observed values from their mean and PRESS is the prediction error sum of squares. The final models were developed by a conventional (non-cross-validated) regression analysis with the optimum number of components equal to

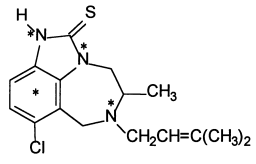


Figure 8. Fitting elements in TIBO derivatives (labeled by an asterisk).

that yielding the highest q^2 value. It can be observed in Table 2 that all the retained final CoMFA models had q^2 well above 0.3, which corresponds to a low probability of chance correlation ($p < 0.05$, that is $< 5\%$).[32] However, to be safer on the potential risk of chance correlations, the biological data were reassigned randomly to the compounds and the PLS analysis were repeated. In no case were good predictive PLS models obtained (q^2 always < 0.2). The final non-cross-validated equations are not very useful to represent efficiently the CoMFA models and therefore 3-D graphics or isocontour maps were developed. They represent areas in the space where steric, electrostatic and lipophilic interactions are responsible for the observed variations of the biological activity.

Docking study

X-ray structures of NNIs and reverse transcriptase were recovered from Brookhaven Protein Database. In particular, crystal structures of α–APA (**54**), TIBO (**34**) and *Nevirapine* (**1**) complexed with RT allosteric binding site can be retrieved from PDB (entry codes 1RVL, 1RVQ and 1RVO, respectively).[33] The whole RT crystal structure used for our study was recovered from RT–TIBO (**34**) complex (code 1TVR, resolution factor 3Å).[34] Homogeneous biological data were available for α–APA (**54**) ($pIC_{50} = 8.30$), TIBO (**34**) ($pIC_{50} = 8.48$) and TBZ (**53**) ($pIC_{50} = 6.41$, for the R enantiomer).[7]

Missing side-chains of the unbound RT 3-D structure were added using the fragment library of the BIOPOLYMER module in Sybyl 6.4 software package.

The obtained 3-D structure was checked by PROCHECK3.3 software.[35] The resulting Ramachandran plot shows unfavorable ϕ and Ψ values only for the 0.2% of aminoacids.

Prior to the docking study, the most likely alignment for NNIs was evaluated. The binding modes observed in 1RVL, 1RVM, 1RVO and 1RVQ complexes were qualitatively analyzed by rigid fitting of C-alpha atoms of the aminoacids in the allosteric binding site. TIBO (**34**)–RT complex was chosen as the template 3-D model. The diverse superimposition modes of NNIs on the template TIBO (**34**) extracted from its complex with the whole RT were studied by an automatic flexible fitting procedure. This calculation was performed with the TFIT algorithm of QXP software (1000 cycles of Monte Carlo conformational search). In the multiple alignment the lowest superimposition energy conformer was selected whereas in the TIBO/TBZ alignment a higher energy conformer (ranking fifth, energy penalty of about 2 kcal/mol over the global minimum) giving an alignment close to that used in CoMFA was chosen.

Subsequent flexible docking studies were performed by MCDOCK algorithm of QXP (1000 cycles of Monte Carlo search). To keep the cpu time at an acceptable level, only a portion of the target protein including all the important aminoacids involved in the ligand binding was considered; it was comprised within a sphere of 15 Å ray centered at the mass center of TIBO (**34**) in the TIBO–RT complex. Ligand molecule and side-chains of the RT aminoacids included within a sphere of 5 Å ray around the ligand were made flexible.

Minimum energy complexes resulting from each calculation were optimized by molecular mechanics (Merck Molecular force-field,[36] PRCG method for convergence).

The interaction energy E_i of the protein–ligand complexes was obtained as follows:

$$E_i = E_{pl} - [E_p + E_l]$$

where E_{pl} is the total energy of the complex, and E_p and E_l are the energies of the isolated allosteric binding site and ligand, respectively.

Acknowledgements

The authors wish to thank Dr. Colin McMartin for the free access to the QXP software package. Financial support from CNR and MURST is kindly acknowledged.

References

1. De Clercq, E. *Med. Res. Rev*. **1996**, *16*, 125.
2. Chimirri, A.; Grasso, S.; Monforte, A. M.; Monforte, P.; Zappalà, M. *Il Farmaco* **1991**, *46*, 817.
3. Chimirri, A.; Grasso, S.; Monforte, A. M.; Monforte, P.; Zappalà, M. *Il Farmaco* **1991**, *46*, 925.
4. Schultz, R. J.; Bader, J. P.; Chimirri, A.; Covey, J. M.; Hill, D. L.; Haugwitz, R. D.; Guziec, F. S.; Narayanan, V. L. *Proc. Am. Assoc. Cancer Res*. **1992**, *33*, 519.
5. Monforte, P.; Monforte, A. M.; Zappalà, M.; Romeo, G.; Grasso, S.; Chimirri, A. *US Patent* **1993**, *5*, 217, 984.
6. Chimirri, A.; Grasso, S.; Molica, C.; Monforte, A. M.; Monforte, P.; Zappalà, M.; Bruno, G.; Nicolò, R; Witvrouw, M.; Jonckeere, H.; Balzarini, J.; De Clereq, E. *Antiviral Chem. Chemother*. **1997**, *8*, 363.
7. Chimirri, A.; Grasso, S.; Monforte, A. M.; Monforte, P.; Rao, A.; Zappalà, M.; Bruno, G.; Nicolò, F.; Pannecouque, C.; Witvrouw, M.; De Clercq, E. *Antiviral Chem. Chemother*. **1998**, *9*, 431.
8. Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A. J.; Hugles, S. H.; Arnold, E. *Struct. Biol*. **1995**, *2*, 407.
9. Tripos Inc., St. Louis, Missouri.
10. McMartin, C.; Bohacek, R. S. *J. Comp. Aid. Mol. Des*. **1997**, *11*, 333.
11. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc*. **1988**, *110*, 5959.
12. Dunn, W. J.; Mold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. *Quant. Struct. -Act. Relat*. **1984**, *3*, 131.
13. Gaillard, P.; Carrupt, P.-A.; Testa, B.; Boundon, A. *J. Comp. -Aid. Mol. Des*. **1994**, *8*, 83.
14. Testa, B.; Carrupt, P.-A.; Gaillard, P.; Billois, F.; Weber, P. *Pharm. Res*. **1996**, *13*, 126.
15. Esnouf, R.; Ren, J.; Ross, C.; Jones, Y.; Stamners, D.; Stuart, D. *Struct. Biol*. **1995**, *2*, 303.
16. Kukla, M. J.; Breslin, H.; Diamond, C. J.; Grous, P. P.; Ho, C. Y.; Miranda, M.; Rodgers, J. D.; Sherrill, R. G.; De Clercq, E.; Pauwels, R.; Andries, K.; Moens, L. J.; Janssen, M. A. C.; Janssen, P. A. J. *J. Med. Chem*. **1991**, *34*, 3187.

17. Breslin, H. J.; Kukla, M. J.; Ludovici, D. W.; Mohrbacher, R.; Ho, W.; Miranda, M.; Rodgers, J. D.; Hitchens, T. K.; Leo, G.; Gauthier, D. A.; Ho, C. Y.; Scott, M. K.; De Clercq, E.; Pauwels, R.; Andries, K.; Janssen, M. A. C.; Janssen, P. A. J. *J. Med. Chem*. **1995**, *38*, 771.
18. Ho, W.; Kukla, M. J.; Breslin, H. J.; Ludovici, D. W.; Grous, P. P.; Diamond, C. J.; Miranda, M.; Rodgers, J. D.; Ho, C. Y.; De Clercq, E.; Pauwels, R.; Andries, K.; Janssen, M. A. C.; Janssen, P. A. J. *J. Med. Chem*. **1995**, *38*, 794.
19. Spence, R. A.; Kati, W. M.; Anderson, K. S.; Johnson, K. A. *Science* **1995**, *267*, 988.
20. Gupta, S. P.; Garg, R. *J. Enz. Inhib*. **1996**, *11*, 23.
21. MOPAC6.0 is available from QCPE (506).
22. Mager, P. *P. Med. Res. Rev*. **1997**, *17*, 235.
23. Tantillo, C.; Ding, J.; Jacobo-Molina, A.; Nanni, R. G.; Boyer, P. L.; Hughes, S. H.; Pauwels, R.; Andries, K.; Janssen, P. A. J.; Arnold, E. *J. Mol. Biol*. **1994**, *243*, 369.
24. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys*. **1953**, *21*, 1087.
25. Weiner, S. J.; Kollmann, P. A.; Case, A. D.; Singh, U. C.; Ghio, C.; Algona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc*. **1984**, *106*, 765.
26. Polak, E.; Rebiere, G. *Rev. Franc. Inf. Rech. Oper*. **1969**, *16*, 35.
27. Nishio, M., Hirota, M., Umezawa, Y., *The CH-Pi Interaction: Evidence, Nature, and Consequences*; Paperback, 1998.
28. Carrieri, A.; Altomare, C.; Barreca, M. L.; Contento, A.; Carotti, A.; Hansch, C. *Il Farmaco* **1994**, *49*, 573.
29. Carrieri, A.; Brasili, L.; Leonetti, F.; Pigini, M.; Giannella, M.; Bousquet, P.; Carotti, A. *Bioorg. Med. Chem*. **1997**, *5*, 843.
30. Pigini, M.; Bousquet, P.; Brasili, L.; Carrieri, A.; Cavagna, R.; Dontwill, M.; Gentili, F.; Giannella, M.; Leonetti, F.; Piergentili, A.; Quaglia, W.; Carotti, A. *Bioorg. Med. Chem*. **1998**, *6*, 2245.
31. Broto, P.; Moreau, G.; Vandyeke, C. *Eur. J. Med. Chem*. **1984**, *19*, 61.
32. Agarwal, A.; Pearson, P. P.; Taylor, E. W.; Li, H. B.; Dahlgren, T.; Hersolf, M.; Yang, Y.; Lambert, G.; Nelson, D. L.; Regan, J. W.; Martin, A. R. *J. Med. Chem*. **1993**, *36*, 4006.
33. Gussio, R.; Pattarbiraman, N.; Zaharevitz, D. W.; Kellogg, G. E.; Topol, I. A.; Rice, W. G.; Schaeffer, C. A.; Erickson, J. W.; Burt, S. K. *J. Med. Chem*. **1996**, *39*, 1645.
34. Das, K.; Ding, J.; Hsiou, Y.; Clark, A. D., Jr.; Moereels, H.; Koymans, L.; Andries, K.; Pawels, R.; Janssen, P. A.; Boyer, P. L.; Clark, P.; Smith, R. H., Jr.; Kroeger Smith, M. B.; Michejda, C. J.; Hughes, S. H.; Arnold, E. *J. Mol. Biol*. **1996**, *264*, 1085.
35. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Cryst*. **1993**, *26*, 283.
36. Halgren, T.A. *J. Comp. Chem*., **1996**, *17*.
37. Laskowski, R. A. *J. Mol. Graph*. **1995**, *13*, 323.