

Data mining using template-based molecular docking on tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepinone (TIBO) derivatives as HIV-1RT inhibitors

Nitin S. Sapre · Swagata Gupta · Nilanjana Pancholi ·
Neelima Sapre

Received: 22 February 2008 / Accepted: 9 June 2008 / Published online: 19 July 2008
© Springer-Verlag 2008

Abstract TIBO (Tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepinone) compounds are potent non-nucleoside reverse transcriptase inhibitors (NNRTIs) that show a great promise for the treatment of AIDS. A structure-based molecular modeling approach based on template-based flexible docking simulation followed by ‘Tabu clustering’ was performed on a series of 46 TIBO derivatives considered as training set of HIV-1 NNRTIs. Four different templates of the highest active ligand ($pIC_{50}=8.52$) of the series were used. The results were reasonably satisfactory. A good correlation was observed between the biological activity and binding affinity of the compounds, which suggest that identified binding conformations of these inhibitors are reliable. Statistical modeling yielded satisfactory results ($r^2=0.878$). Our studies suggest that template-based docking followed by ‘Tabu clustering’ enhances the docking efficiency. Also, cross-validation with a test-set containing 16 compounds gave satisfactory results ($r^2=0.836$). Data mining of PubChem database yielded a total of 31 hits (25 novel TIBO like compounds, as well as, 6 novel scaffolds) with enhanced binding efficacy as hits.

These hits may, be targeted toward potent lead-optimization and, help in designing and synthesizing novel compounds with enhanced therapeutic efficacy.

Keywords TIBO · Structure based drug design (SBDD) · Tabu-clustering · HIV-1 reverse transcriptase (HIV-1 RT) · Non-nucleoside reverse transcriptase inhibitor (NNRTI) · pIC_{50} · Data mining · Non-nucleoside inhibitor-binding pocket (NNIBP)

Introduction

The human immunodeficiency virus (HIV) is the cause of the acquired immunodeficiency syndrome (AIDS) and was first identified in 1981 [1]. Since then, AIDS has developed into a worldwide pandemic at an alarming proportion. In the last two decades, progress has been made to understand and control viral replication [2, 3]. The introduction of highly active antiretroviral therapy (HAART) has been specifically effective in reducing the mortality rate [4]. However, HAART has of yet not been able to eradicate HIV completely from patients who have undergone this therapy [5–7]. The emergence of resistant HIV viral strains is a limitation for all the concerned therapeutic classes. The emerging cross-resistance among the approved drugs such as nevirapine, delavirdine etc. has particularly limited the use of the non-nucleoside reverse transcriptase inhibitors (NNRTI) class of compounds. The development of anti-HIV compounds with more favorable side effect profiles and improved characteristics, continues to be an active field of pharmaceutical research, and many lead compounds still emerge from initial antiviral screens [8, 9].

N. S. Sapre (✉) · N. Pancholi
Department of Applied Chemistry,
Shri GS Institute of Technology and Sciences,
Indore, MP Pin 452001, India
e-mail: sukusap@yahoo.com

S. Gupta
Department of Chemistry, Govt. P.G. College,
MHOW MP, India

N. Sapre
Department of Computer Applications,
SD Bansal College of Technology,
Indore MP, India

It is a well established fact, that reverse transcriptase (RT) inhibitors are important drugs in the presently available therapy [10]. These are divided into the following two classes: nucleoside reverse transcriptase inhibitors (NRTIs), such as zidovudine (AZT, Retrovir), and non-nucleoside reverse transcriptase inhibitors (NNRTIs), such as nevirapine (Viramune) [11, 12]. The NNRTIs selectively inhibit HIV-1 reverse transcriptase by binding to an allosteric site located in the p66 subunit, approximately 10 Å from the polymerase binding site in a noncompetitive manner with respect to the substrate [13, 14]. Unusually, the allosteric hydrophobic pocket appears to be formed only on binding of the NNRTIs to the viral enzyme. The inner hydrophobic core is mainly stacked with TYR181, TYR188 amino acid residues. The other amino acids involved in the hydrophobic interactions with NNRTIs are VAL106, VAL108, PHE227, TRP229 and GLU138. Amino acid residues VAL179 and LYS101 are associated with hydrogen-bonding interactions [15, 16]. The flexibility or to say, the ‘plasticity’ of the NNRTI binding site has led to the discovery of many structurally diverse classes of allosteric enzyme inhibitors, providing the medicinal chemists with a broad platform to design improved anti-retroviral drugs [17–21]. NNRTIs have proven to be susceptible to rapid drug resistance [22]. The dynamics of HIV replication and the viral population are the huge reservoir of genetic variants in each HIV infected system. Also, numerous conformational variants are generated owing to the torsional flexibility, repositioning and reorientation of the inhibitors within the pocket [23]. Therefore, though an important constituent of the cocktail therapy, the immediate thrust is on designing of broad spectrum improved NNRTIs, which do not give rise to cross-resistance and are effective against all clinically relevant mutant strains [24].

Among the various computer-aided drug design (CADD) approaches, structure based drug design (SBDD) approach provides a better understanding of the receptor-binding site structure and interactions associated with it. Docking methods, typically use an energy-based scoring function to identify the most favorable ligand conformation when bound to the target. The general hypothesis is that lower energy scores represent better protein-ligand bindings as compared to higher energy values. Therefore, molecular docking can be formulated as an optimization problem, where the main task is to find a favorable ligand-binding mode with the lowest energy. The molecular docking studies on receptor structures deduced from X-ray crystallography and rationally designed NNRTIs have been quite successful in studying the ‘quasi-species’ nature of the virus and changes in the binding pocket shape, residue constituent, size and the mutational effects causing resistance [25, 26].

Drug designing is an iterative process, where computational skills are used to generate computer models of new

chemical entities to define their activity profiles, geometries and reactivity. Virtual screening is an efficient way of hastening the drug design process by screening large libraries of chemical compounds in a limited time. In the present study, we have used data mining, as a tool for virtual screening from a large database of compounds [27–30].

TIBO (Tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepinone) analogues, discovered in 1987 by screening a subset of the Janssen compound library of pharmacologically “inactive” compounds in a cell-based anti-HIV test, are one of the initially synthesized prominent NNRTIs [31]. Despite not fitting into a two-hinged-ring model like most of the NNRTIs, they show remarkable shape complementarity in geometrical arrangement on binding to reverse transcriptase with respect to other NNIs. Tivirapine, a TIBO compound in the clinical trials has been a promising candidate as an NNRTI [32]. The crystal structures of TIBO/RT complexes and their comparative studies reveal the probable energetic interactions and the binding orientations, which may have a breakthrough in discovering modified TIBO analogues against a variety of clinically relevant mutant RT enzymes combined with a favorable pharmacokinetic profile and low toxicity [33–38]. In our recently published work, simple docking studies on TIBO derivatives were performed [39]. Guided by the encouraging results obtained in the alignment of the redocked molecule, we have extended our docking studies to template-based search followed by ‘Tabu-clustering’. We were able to derive a better predictive model using regression analysis, which was subsequently used to cross-validate a test set and then to data mine PubChem compound database [40], to derive putative hit compounds. Here it is worth mentioning, that the database search was based on the same substructure search as mentioned in our earlier paper, but the number of hits obtained were more (31 as compared to 20) within the similar energetic profile.

Computational methods

Molecular structures

TIBO or tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepinone and its derivatives developed by Pauwels et al. [31] and their biological activities were taken for template-based docking studies. The molecular structures were drawn and geometrically optimized using ChemDraw Ultra 7.0.0 [41]. Though the docking engine itself prepares the molecules to be docked, we preferred prior minimization of the energy of molecules using MM2 Force Field in ChemDraw (as better docking values could be obtained) and then exported to Molegro Virtual Docker [42], where they were further prepared along with the proteins (charges and protonation

states were assigned) by the docking engine. The structure of HIV-1 RT protein (PDB code 1REV) was obtained from Protein Data Bank [43] [Research Collaboratory for Structural Bioinformatics (RCSB)]. PubChem compound database [40] was used for data mining to find probable hit candidates.

Docking simulations

MolDock specifically uses guided differential evolution algorithm [44, 45], wherein cavity prediction algorithm [46] is incorporated to constrain the predicted poses during the search process. The docking scoring function of MolDock is the modified version of piecewise linear potential (PLP) scoring functions [47, 48] and GEMDOCK [49]. As mentioned in our previous work, [39] the scoring function takes care of van der Waals term as well as hydrogen bond directionality. The fitness of a candidate solution is evaluated in terms of intermolecular interaction energy between the ligand and the protein, and the intramolecular interaction energy of the ligand. The robustness of the docking process is mitigated by a more stringent re-ranking of the top scoring poses.

Docking templates have been used (in order to focus the search) in the present study and are implemented as scoring functions, rewarding poses similar to the specific pattern. A template is a collection of groups, where each group represents a chemical feature for an atom (e.g., ‘hydrogen acceptor atoms’). Each template group contains a number of centers, which are optimal 3D positions for the group feature.

If an atom matches a group definition, it will be rewarded depending on its distance to the group centers by using the following (Gaussian) formula for each center:

$$e = \omega * \exp\left(\frac{-d^2}{r_0^2}\right) \quad (1)$$

where d is the distance from the position of the atom to the center in the group. ω is a weight factor for the template group and r_0 is a distance parameter.

Hardware and software

Molegro Virtual Docker 2007.2.2 [36] was run on a Windows XP based pentium IV 2.66 GHz core 2 duo processor PC (with 512 MB of memory).

Results and discussion

In our earlier work, we have presented simple docking studies with respect to TIBO derivatives and 9-Cl-TIBO/

HIV-RT complex (PDB CODE-1REV) [39]. In the present work, we have extended our docking studies to template-based docking followed by Tabu clustering. We have divided the molecular set into a training set and a test set. The specific aim behind the present study, is to obtain enhanced docking efficacy in terms of energetic relationship and to obtain a robust model, which could be further used to datamine newer and novel probable hit compounds from vast external database/databases.

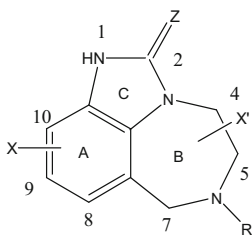
The chemical structures of 46 TIBO derivatives of the training set are given in Table 1 along with their respective biological activity, expressed in terms of pIC₅₀ (where IC₅₀ is the effective concentration of a compound required to activate 50% protection of MT-4 cell against the cytopathic effect of HIV-1) and the substituents as X, Z, R and X'. X is the substitution on the aromatic ring ‘A’, X' is substitution on the 7-member ring ‘B’, Z indicates the presence of =O or =S attached to the five member ring ‘C’ and R is the substitution attached to nitrogen in the ring ‘B’.

Validation of the docking method

To obtain valid binding modes of the inhibitors used, the MolDock docking parameters had to be validated first to ensure proper ligand orientation and positioning for the crystal structure (1REV) as imported from the Protein Data Bank. Thus the ligand 9Cl-TIBO, in the conformation found in the crystal structure was extracted and docked back into the corresponding binding pocket to determine the ability of MolDock to reproduce the alignment of the inhibitor observed in the crystal structure. We have used the same docking protocol as mentioned in our earlier work [39], as it could reproduce the best alignment in terms of root mean square deviation (RMSD) not only in our simulation study, but also reported in other docking studies [50–52]. The RMSD of 0.269 Å showed a satisfactory agreement between the alignments of the redocked and crystal ligand coordinates. Figure 1 shows the best fit redocked coordinates with respect to the crystal structure. The hydrogen bond interactions involved the five member ring of ligand and the backbone of LYS101 residue of the non-nucleoside inhibitor-binding pocket (NNIBP), hydrophobic interactions involved TYR181, PRO95, PHE227, LEU234, VAL106, TYR188 and the DMA fragment of the ligand at 6-N position of the diazepine ring.

Docking of the molecule training set

Following the validation of the docking method using 9Cl-TIBO, a training dataset of 46 molecules belonging to TIBO derivatives with varied activity range (4.0–8.52) were docked into the same coordinates of the crystal structure. As compared to our previous work [39], we have made the

Table 1 Anti-HIV-1 activity (inhibitory concentration pIC_{50}), reranking scores and binding affinity (E_{binding} in kJ/mol) of TIBO derivatives (training set)

S.No	X	Z	R	X'	pIC_{50}^{α}	Binding Affinity*	Score-1 ^a	Score-2 ^b	Score-3 ^c	Score-4 ^d
1	H	S	DMA [‡]	5-Me(S)	7.36	-22.77	-101.40	-101.56	-64.12	-2.71E-187
2	9-Cl	S	DMA	5-Me(S)	7.47	-24.81	-115.21	-114.81	-86.98	-2.09E-190
3	8-Cl	S	DMA	5-Me(S)	8.37	-25.70	-118.17	-117.31	-89.30	-9.90E-188
4	8-F	S	DMA	5-Me(S)	8.24	-25.16	-121.47	-121.19	-96.09	-1.28E-189
5	8-SMe	S	DMA	5-Me(S)	8.30	-23.97	-123.83	-123.85	-86.24	-7.92E-191
6	8-OMe	S	DMA	5-Me(S)	7.47	-24.01	-125.47	-125.40	-97.80	-1.05E-190
7	8-OC ₂ H ₅	S	DMA	5-Me(S)	7.02	-24.77	-134.48	-135.96	-91.38	-2.32E-191
8	8-CONH ₂	O	DMA	5-Me(S)	5.20	-21.54	-127.11	-129.05	-89.45	-4.94E-189
9	8-Br	O	DMA	5-Me(S)	7.33	-24.30	-117.01	-118.79	-87.54	-8.33E-188
10	8-Br	S	DMA	5-Me(S)	8.52	-25.31	-121.17	-120.45	-94.48	-6.75E-188
11	8-I	O	DMA	5-Me(S)	7.06	-24.62	-112.83	-113.51	-80.82	-2.42E-187
12	8C=CH	O	DMA	5-Me(S)	6.36	-22.49	-134.03	-132.41	-106.46	-1.37E-187
13	8C=CH	S	DMA	5-Me(S)	7.53	-23.27	-121.23	-120.65	-81.51	-7.37E-186
14	8-CH ₃	O	DMA	5-Me(S)	6.00	-21.96	-121.25	-120.88	-95.10	-1.46E-188
15	8-N(CH ₃) ₂	O	CPM [§]	5-Me(S)	5.18	-20.68	-127.24	-130.91	-94.82	-9.72E-188
16	9-NH ₂	O	CPM	5-Me(S)	4.22	-20.87	-104.37	-104.99	-81.59	-1.45E-189
17	9-F	S	DMA	5-Me(S)	7.60	-25.02	-122.56	-122.69	-94.72	-1.65E-191
18	9-Me	O	DEA [#]	5-Me(S)	6.50	-21.68	-113.01	-112.46	-72.95	-1.51E-176
19	H	O	2-MA [¥]	5-Me(S)	4.33	-20.66	-98.90	-96.72	-80.34	-3.67E-194
20	H	O	CPM	5-Me(S)	4.36	-19.95	-102.23	-98.19	-81.56	-2.67E-193
21	H	O	CH ₂ CH ₂ CH ₂ CH ₃	5-Me(S)	4.00	-20.38	-96.22	-94.77	-15.01	-3.81E-190
22	H	O	DMA	5-Me(S)	4.90	-21.49	-109.03	-109.54	-83.69	-4.75E-187
23	H	O	DMA[R(+)]	5-Me(S)	4.66	-20.80	-109.37	-112.14	-82.90	-2.35E-188
24	8-Cl	S	DMA	H	7.34	-24.78	-116.78	-115.41	-84.29	-9.22E-189
25	H	O	2-MA	4-Me	4.50	-19.95	-104.08	-101.70	-83.19	-2.03E-188
26	H	O	C ₃ H ₇	4-CHMe ₂	4.13	-19.66	-124.68	-124.91	-101.69	-6.86E-190
27	H	O	2-MA	4-CHMe ₂	4.90	-19.95	-115.17	-112.75	-82.96	-5.91E-193
28	H	O	2-MA	4-C ₃ H ₇	4.32	-20.22	-121.58	-119.23	-44.34	-1.50E-189
29	H	O	DMA	7-Me	4.92	-20.95	-117.31	-117.24	-85.18	-6.61E-190
30	8-Cl	O	DMA	7-Me	6.84	-23.97	-123.14	-120.92	-78.66	-5.67E-189
31	9-Cl	O	DMA	7-Me	6.80	-23.50	-110.87	-109.88	-84.06	-3.60E-188
32	H	S	C ₃ H ₇	7-Me	5.61	-21.31	-102.61	-102.65	-82.44	-1.24E-192
33	8-Cl	S	DMA	7-Me	7.92	-24.52	-120.19	-122.16	-83.99	-3.00E-190
34	9-Cl	S	DMA	7-Me	7.64	-24.51	-116.35	-116.33	-41.22	-1.25E-193
35	H	O	DMA	4,5-di-Me(Cis)	4.25	-20.63	-119.25	-120.12	-91.37	-1.11E-187
36	H	S	DMA	4,5-di-Me(Cis)	5.65	-21.52	-116.62	-116.55	-80.38	-5.10E-192
37	H	S	CPM	4,5-di-Me(trans)	4.87	-20.17	-102.59	-99.64	-76.03	-2.98E-193
38	H	S	DMA	4,5-di-Me(trans)	4.84	-20.96	-118.17	-118.30	-86.13	-1.16E-192
39	H	S	DMA	5,7-di-Me(trans)	7.38 ^γ	-20.63	-130.55	-131.37	-105.62	-6.59E-186
40	H	S	DMA	5,7-di-Me(Cis)	5.94	-22.84	-109.27	-109.75	-79.02	-1.05E-187
41	9-Cl	O	DMA	5,7-di-Me(R,R-trans)	6.64	-23.74	-139.80	-138.32	-108.82	-5.81E-186
42	9-Cl	S	DMA	5,7-di-Me(R,R-trans)	6.32	-23.18	-115.87	-112.91	11.42	-2.92E-186
43	H	S	DMA	4,7-di-Me(trans)	4.59	-20.90	-133.86	-135.21	-108.16	-8.28E-190

Table 1 (continued)

S.No	X	Z	R	X'	pIC ₅₀ ^α	Binding Affinity*	Score-1 ^a	Score-2 ^b	Score-3 ^c	Score-4 ^d
44	9-Cl	S	CPM	5-Me	7.47 ^γ	-20.43	-111.36	-111.29	-90.29	-2.92E-189
45	H	O	2-MA	5-Me(S)	4.46	-21.05	-105.33	-105.17	-82.65	-1.52E-190
46	H	S	2-MA	5-Me(S)	5.48	-22.12	-117.75	-116.51	-95.76	-9.38E-188

[‡] DMA=3,3-Dimethylallyl, [¥] 2-MA=2-methylallyl, [#] DEA=3,3-Diethylallyl, [§] CPM = cyclopropylmethyl.

^α pIC₅₀=-logIC₅₀ (where IC₅₀ is the effective concentration of a compound required to activate 50% protection of MT-4 cell against the cytopathic effect of HIV-1).

*E_{binding} in kJ/mol, ^a Score 1 = MolDock score, ^b Score 2 = docking score (pose energy), ^c Score 3 = the reranking score, (all in arbitrary energy units). ^d Score-4 = similarity score(it is the similarity contribution from the docking template with respect to the ligand).

^γ Data points not used in deriving equation.

docking protocol more efficient by incorporating templates and using simplex evolution algorithm (MolDock SE) along with Grid scoring function (MolDock Grid). In our earlier work, we had used only MolDock Grid. Pose clustering is enhanced using 'Tabu clustering' technique, which ensures greater diversity of the returned poses. Prior to docking, templates (Fig. 2a,b,c,d) based on steric (shape matching taking into consideration all the atoms), hydrogen acceptor, hydrogen donor and ring (aromatic as well aliphatic) were derived from the ligand with highest activity (pIC₅₀=8.52). Similarity groups were set up with following parameters (overall strength=-500.00, resolution of energy grid=0.40). The simplex evolution algorithm performs a combined local/global search on the poses generated by the pose generator and the following parameters were set (No. of runs=10, max. iterations=2000, max. population size=50) for SE. Poses were generated at default energy threshold with SE max. steps at 380, neighbor distance factor at 1.00 and an extra penalty term. For pose clustering, 'Tabu clustering' technique has been used, wherein poses similar to solutions from earlier runs were penalized. RMSD calculation was done by automorphism (intrinsic ligand symmetries) with a threshold of 2 Å. The docked 3D-structures of TIBO derivatives were scored, reranked and then compared with the X-ray crystallographic structure of 9-Cl TIBO. All the compounds of this series share a basic common pattern, therefore these inhibitors dock back into the non-nucleoside inhibitor binding pocket (NNIBP) extremely well as expected. The binding conformations were observed to occupy slightly separate, but overlapping binding regions. Most of the top ranking compounds, with 'good fit' (with different activity range) were found to prefer a single binding position in the allosteric pocket (Fig. 3). A flipped arrangement with respect to the imidazolone ring in the binding site was also observed in some of the compounds.

TIBO compounds show a similar basic structure, thus, as expected, similar interactions were observed. Hydrogen

bonding interactions were observed between the imidazolone ring of ligand and the LYS101 residue. Hydrophobic interactions seem to stabilize the TIBO compounds. The major contributor to the hydrophobicity appears to be the diazepine ring and the groups attached to it at 6-N. The high activity of the compounds containing DMA moiety at the diazepine ring may be attributed to the strong hydrophobic nature of DMA. Favorable electrostatic interactions are stabilized by aromatic moieties of the receptor (TYR181 and TYR188) as well as that of the TIBOs.

Correlation between binding affinity and activity

An important aspect of a docking engine is to predict the binding efficacy in terms of energy as well as conformation of an inhibitor within the receptor NNIBP. The binding affinity is the measure of the receptor-ligand interaction in a given dataset. The binding affinity (kJ/mol) of a pose is

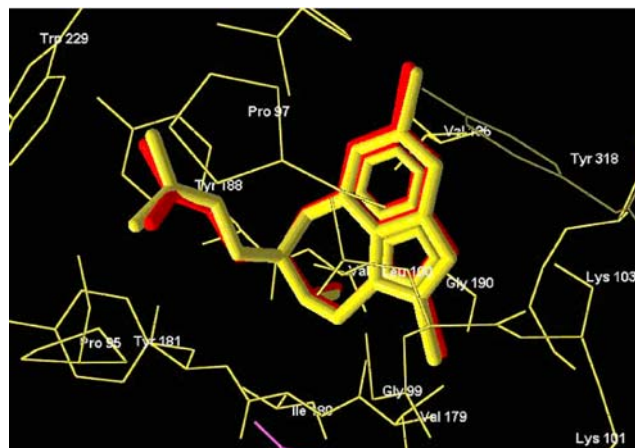


Fig. 1 Conformation of 9-Cl TIBO crystal structure [PDB code: 1REV] (yellow) as compared to redocked conformation of 9-Cl TIBO (red)

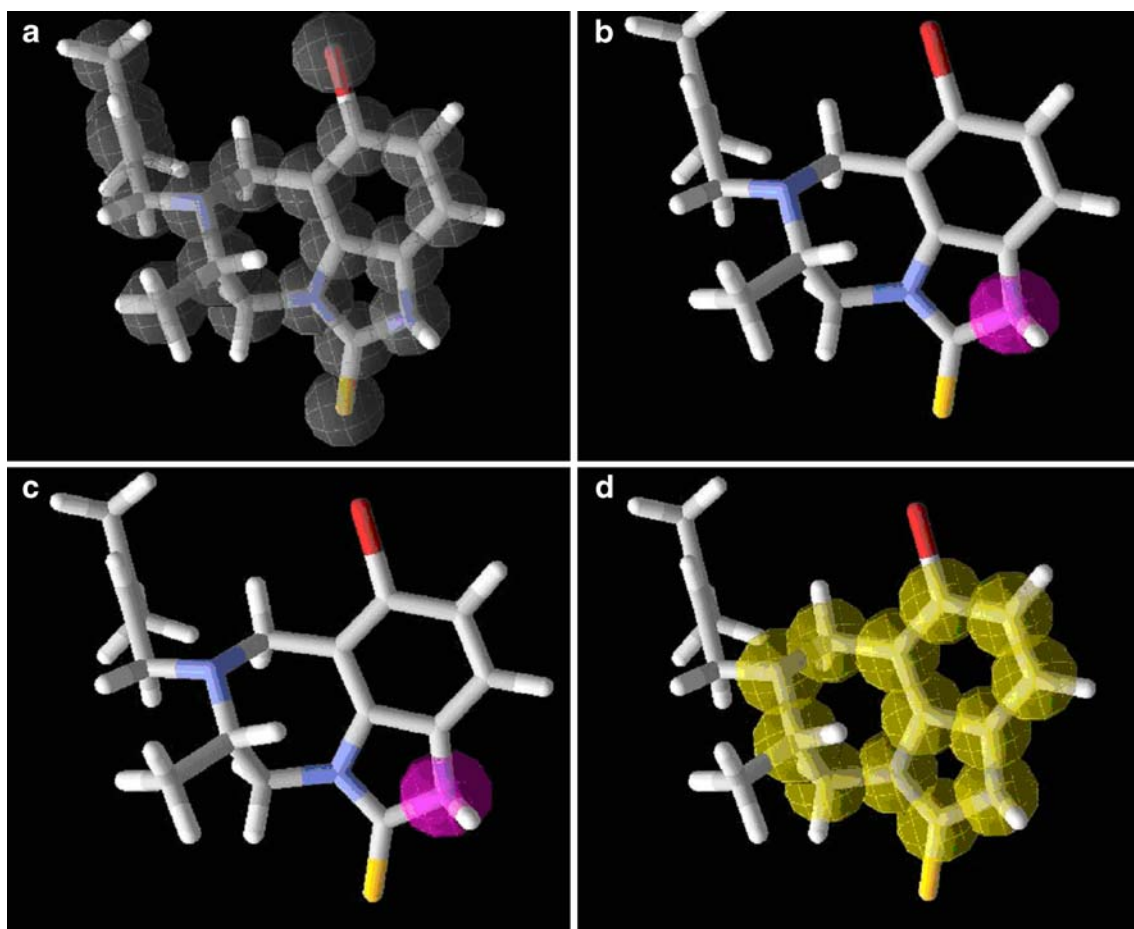


Fig. 2 Templates showing different groups for the most active ligand (cpd. 10, $pIC_{50}=8.52$): (a) steric, (b) hydrogen acceptor, (c) hydrogen donor, (d) ring

given by: $E_{binding} = -5.68 * pKi$ (The numerical factor corresponds to a temperature of 297 K) [42]. The reranked scores predicting the binding affinities spanned the range of -19.63 to -25.7 kJ/mol, which is again better than our previous study using simple docking. The scoring robustness of the docking engine was ascertained by correlating binding affinities with the biological activity pIC_{50} of the TIBO derivatives. The models generated used binding affinities as a sole descriptor variable via linear regression technique (LR). Leave-one-out (LOO) procedure and N-cross validated (N-CV) method were adopted for validation of the results. Biological activity (pIC_{50}) was taken as the dependent variable. The auto generated random seed used in the model training was 1255830047. In these equations, n is the number of compounds, r is the correlation coefficient, q^2 is the cross-validated r^2 from the (LOO) or (N-CV) procedure, Spearman rank correlation coefficient (ρ), MSE is the mean squared error and $PRESS$ is the predictive sum of squares.

The binding affinity ($E_{binding}$) in kJ/mol, MolDock score, similarity score, docking score and a more stringent rerank score (all in arbitrary energy units) are also recorded in

Table 1. The similarity score is not a measure of scoring the pose efficiency, but it shows the similarity contribution of the docking template with respect to the concerned ligand and is used for evaluating the docking score.

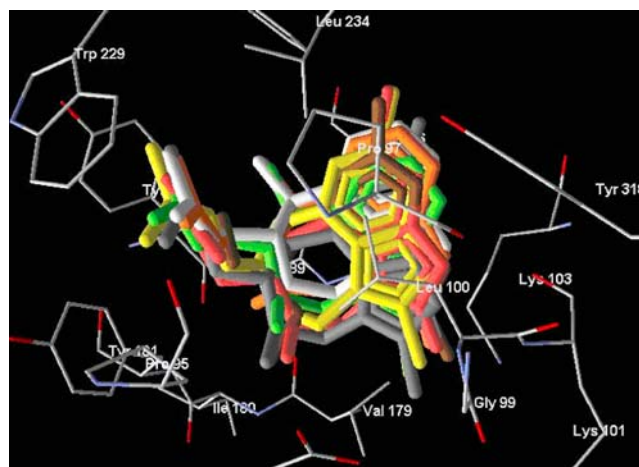


Fig. 3 Docked conformation of 9-Cl TIBO crystal structure (yellow) with six best fit TIBO derivatives. Compound no. 39 (orange), 40 (green), 41 (pink), 42 (grey), 43 (white), 44 (brown)

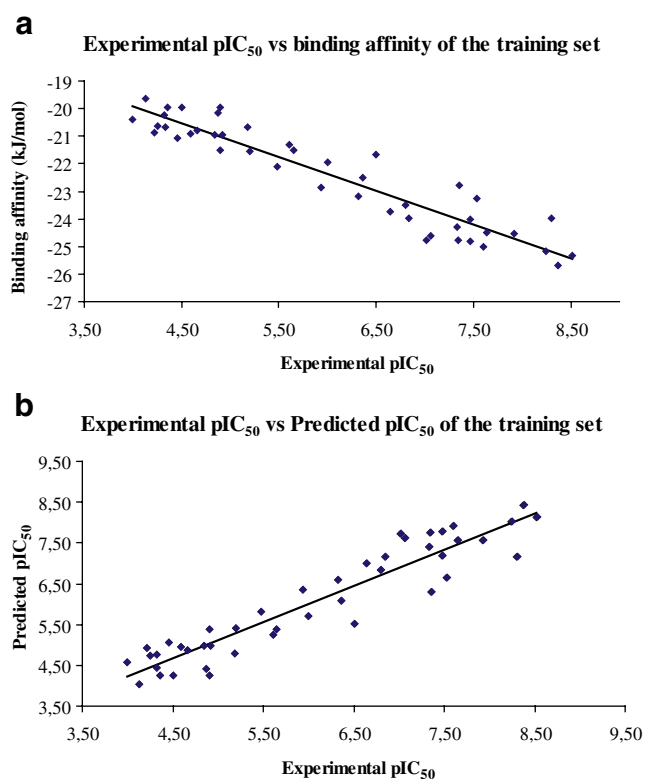


Fig. 4 **a** Graph between experimental activity (pIC₅₀) and binding affinity (E_{binding} in kJ/mol) of the training set of TIBO compounds. **b** Graph between experimental pIC₅₀ and predicted pIC₅₀ of the training set of TIBO compounds

Linear regression technique (LR)

The best model relating biological activity with the binding affinity derived using LR (N-CV) with $N = 10$ is presented below:

$$\text{Activity (pIC}_{50}) = -0.721932(0.942)E_{\text{binding}} - 10.1443 \quad (2)$$

$$(n = 44 \quad r = 0.937 \quad r^2 = 0.878 \quad r_{\text{adj}}^2 = 0.875)$$

$$\text{Spearman}(\rho) = 0.922 \quad q^2 = 0.877 \quad \text{MSE} = 0.239$$

$$\text{PRESS} = 10.50)$$

The results derived from (N-CV) with $N = 5$ showed a near similar result ($r^2=0.875$, $\text{MSE}=0.246$). The results derived from LOO also reaffirmed the robustness of the model ($r^2=0.878$). The results thus obtained were validated, as shown in Fig. 4 a, by correlating biological activity with the binding affinity.

Table 2 records the observed and the calculated values of pIC₅₀ for the training set of TIBO derivatives. The quality of correlation is demonstrated by their residual values, i.e., the difference between their experimental and predicted pIC₅₀.

A good fit was observed between experimental and predicted activity, as shown in Fig. 4b, reconfirming the

robustness of the docking procedure adopted herein. Thus template-based molecular docking studies using different templates of the ligand with highest biological activity followed by ‘Tabu clustering’ (computationally more expensive) have provided a reasonably satisfactory model and were instrumental in providing better docking results.

A very interesting exception was observed in the case of two TIBO compounds [Fig. 3, Cpd. no. 39 (orange) and 44

Table 2 Experimental, predicted and residuals values of pIC₅₀ and binding affinity (kJ/mol) of TIBO derivatives (training set)

S.No	Exp-pIC ₅₀	E_{binding}	Pred-pIC ₅₀	Residual
1	7.36	-22.768	6.29	1.07
2	7.47	-24.809	7.77	-0.30
3	8.37	-25.702	8.41	-0.04
4	8.24	-25.165	8.02	0.22
5	8.30	-23.969	7.16	1.14
6	7.47	-24.005	7.19	0.28
7	7.02	-24.769	7.74	-0.72
8	5.20	-21.541	5.41	-0.21
9	7.33	-24.301	7.40	-0.07
10	8.52	-25.311	8.13	0.39
11	7.06	-24.618	7.63	-0.57
12	6.36	-22.494	6.09	0.27
13	7.53	-23.266	6.65	0.88
14	6.00	-21.962	5.71	0.29
15	5.18	-20.679	4.78	0.40
16	4.22	-20.866	4.92	-0.70
17	7.60	-25.020	7.92	-0.32
18	6.50	-21.678	5.51	0.99
19	4.33	-20.660	4.77	-0.44
20	4.36	-19.949	4.26	0.10
21	4.00	-20.381	4.57	-0.57
22	4.90	-21.491	5.37	-0.47
23	4.66	-20.802	4.87	-0.21
24	7.34	-24.784	7.75	-0.41
25	4.50	-19.951	4.26	0.24
26	4.13	-19.655	4.05	0.08
27	4.90	-19.951	4.26	0.64
28	4.32	-20.220	4.45	-0.13
29	4.92	-20.954	4.98	-0.06
30	6.84	-23.967	7.16	-0.32
31	6.80	-23.505	6.82	-0.02
32	5.61	-21.311	5.24	0.37
33	7.92	-24.521	7.56	0.36
34	7.64	-24.512	7.55	0.09
35	4.25	-20.625	4.75	-0.50
36	5.65	-21.517	5.39	0.26
37	4.87	-20.165	4.41	0.46
38	4.84	-20.964	4.99	-0.15
39	5.94	-22.844	6.35	-0.41
40	6.64	-23.736	6.99	-0.35
41	6.32	-23.181	6.59	-0.27
42	4.59	-20.900	4.94	-0.35
43	4.46	-21.053	5.05	-0.59
44	5.48	-22.123	5.83	-0.35

Table 3 Experimental anti-HIV-1 activity, predicted anti-HIV-1 activity (inhibitory concentration pIC_{50}), their residual values and binding affinity (E_{binding} in kJ/mol) of TIBO derivatives (Test set)

S.No	X	Z	R	X'	E_{binding} (kJ/mol)	Exp- pIC_{50}	Pred- pIC_{50}	Residual
47	8-I	S	DMA	5-Me(S)	-25.186	7.32	8.04	-0.72
48	9,10-di-Cl	S	DMA	5-Me(S)	-28.410	7.60	10.37	-2.77
49	H	O	$CH_2CH=CH_2$	5-Me(S)	-21.478	4.15	5.36	-1.21
50	H	O	$CH_2CH_2CH=CH_2$	5-Me(S)	-21.109	4.30	5.09	-0.79
51	H	O	$CH_2CH_2CH_3$	5-Me(S)	-20.174	4.05	4.42	-0.37
52	H	O	2-MA[S(+)]	5-Me(S)	-20.760	4.72	4.84	-0.12
53	H	O	$CH_2CH=CHMe(E)$	5-Me(S)	-21.937	4.24	5.69	-1.45
54	H	O	$CH_2CH=CHMe(Z)$	5-Me(S)	-21.243	4.46	5.19	-0.73
55	H	O	$CH_2C(Me)=CHMe(E)$	5-Me(S)	-21.482	4.54	5.36	-0.82
56	H	O	DMA[S(+)]	5-Me(S)	-20.797	5.40	4.87	0.53
57	H	O	$CH_2C(H=CH_2)=CH_2$	5-Me(S)	-21.573	4.15	5.43	-1.28
58	9-Cl	S	DMA	H	-24.946	6.80	7.87	-1.07
59	H	O	2-MA	5,5-di-Me	-21.864	4.64	5.64	-1.00
60	9-Cl	S	2-MA	4-Me(S)	-24.482	6.17	7.53	-1.36
51	9-Cl	S	CPM	4-Me(R)	-23.343	5.66	6.71	-1.05
62	H	O	C_3H_7	5-Me	-20.150	4.22	4.40	-0.18

(brown) with activities 7.38 and 7.47 respectively]. Although they showed a good geometric fit, they yielded a lower value of binding affinity as compared to other higher actives. This suggests some complexity in interaction in the putative binding site.

Cross-validation with the test-set

A test-set consisting of 16 TIBO compounds (activities in the range of 4.12 – 7.60) was docked to assess the consistency of the aforementioned results obtained. The test-set has been made, based on two criteria: Firstly, the whole activity range has been spanned and secondly, emphasis has been laid on choosing the compounds with varied substituents attached to 6-N of ring 'B', as this region is instrumental in having major hydrophobic interactions with the surrounding residues (DMA being the best). We have also tried not to include too many DMA substituted compounds, so as to refrain from having a biased test-set. Table 3 presents the 16 compounds with their experimental and predicted activity. The table also shows the respective residual values indicating a good quality of relatedness. A satisfactory agreement was observed between the experimental and predicted activity ($r^2=0.836$, $Spearman(\rho)=0.692$) and the binding affinity of the compounds as shown in Fig. 5 a and b.

Data mining

We have used the same PubChem compound database as reported earlier [39], containing more than 10 million compounds, for data mining. A similarity based search and further screening using Lipinski's rule yielded 162 compounds.

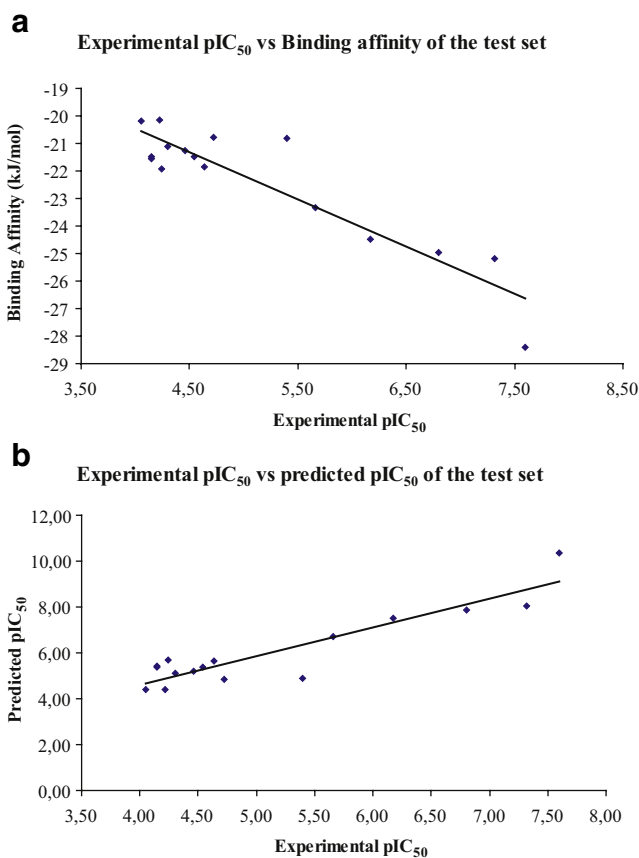


Fig. 5 a Graph between experimental activity (pIC_{50}) and binding affinity (E_{binding} in kJ/mol) of the test set of TIBO compounds. b Graph showing comparison between experimental pIC_{50} and predicted pIC_{50} for the test-set

Table 4 Binding affinities and predicted values of pIC₅₀ of 31 compounds from PubChem compound Database

S.No	Comp ID	E _{binding}	Pred- pIC ₅₀
63	10265094	-26.131	6.00
64	10376208	-25.148	8.01
65	460857	-25.194	8.04
66	504495	-27.611	9.79
67	10061731	-31.077	5.18
68	504501	-24.655	7.65
69	465251	-27.279	9.55
70	465289	-26.566	9.03
71	6451104	-28.005	10.07
72	10365317	-25.310	8.13
73	452838	-25.503	8.27
74	460859	-25.355	8.16
75	10517380	-29.607	11.23
76	3010438	-27.808	9.93
77	465250	-25.840	8.51
78	465287	-25.388	8.18
79	6451106	-26.419	8.93
80	10380506	-27.302	9.57
81	3010459	-26.873	9.26
82	465249	-25.245	8.08
83	10040548	-27.688	9.84
84	10315128	-34.043	14.43
85	1038250	-33.259	13.87
86	10473297	-27.677	9.84
87	10589822	-27.405	9.64
88	452844	-31.399	12.52
89	465288	-28.342	10.32
90	6451105	-28.918	10.73
91	10444087	-24.139	7.28
92	452826	-24.442	7.50
93	465243	-24.127	7.27

Most of the compounds had the molecular weight in the range of 250–400, Hydrogen bond donor (1 or 2), Hydrogen bond acceptor (1–5) and log P (≤ 5.5). Neglecting the compounds with unfavorable functional groups such as –CHO, –NO₂ and the compounds already existing in training and the test-set, finally a dataset of 136 compounds was selected. These compounds were docked to evaluate their respective binding affinities and predict their anti-viral activities. We also docked tivrapipe (R-86183), a potent TIBO drug already in clinical trials, into the same binding pocket ($E_{\text{binding}}=-25.05$ kJ/mol). It is worth mentioning here, that we were able to retrieve more number of hits (31) as compared to (20) the one in our earlier study [39]. This signifies the robustness of our present model. On the basis of the binding energies obtained, finally 31 compounds were selected ($E_{\text{binding}}=-24$ to -34 kJ/mol and pIC₅₀=7 to 14.5), which could serve as probable hit candidate toward a potent lead-optimization. Table 4 and Fig. 6 present the binding affinities and predicted values of pIC₅₀ of compounds from the PubChem database. Structures of selected hit compounds from the PubChem compound database are shown in Fig. 7. Fifteen compounds were found to be same as retrieved in our earlier work (In both the cases, we have considered a cut off of minimum binding affinity of -24 kJ/mol). The rest of the compounds have the basic TIBO skeleton. This shows that, the template-based similarity measure could contribute to a better docking protocol and could retrieve more compounds with similar structure as compared to simple docking. Thus, our present model was able to evaluate the binding affinities in a more precise way.

It is interesting to note that 27 of the 31 hits show better binding affinity as compared to tivrapipe. Most of the compounds have the basic TIBO skeleton. The datamined

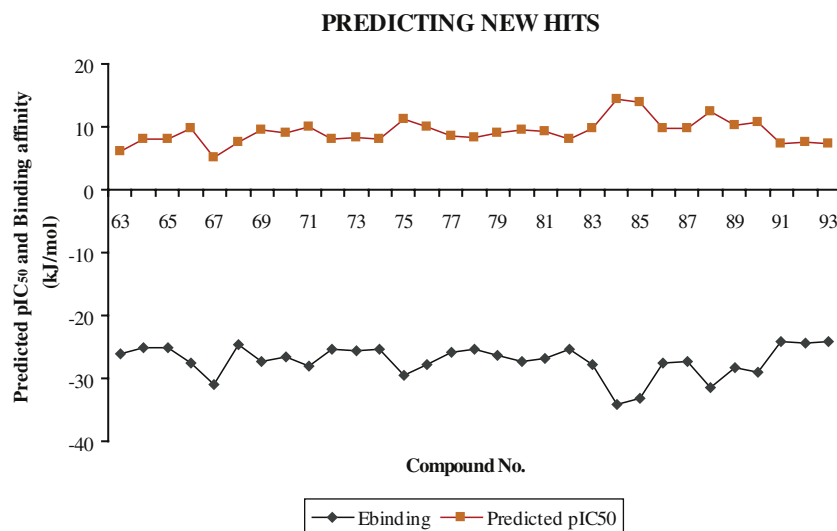
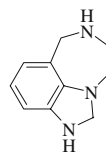
Fig. 6 Graph showing binding affinity (E_{binding} in kJ/mol) and their respective predicted pIC₅₀ of compounds from PubChem compound database

Fig. 7 Structure of selected 31 compounds from the PubChem compound database. Compound ID are given below



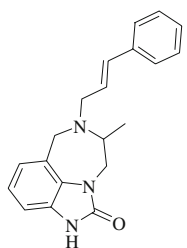
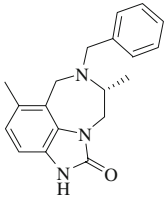
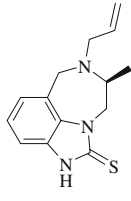
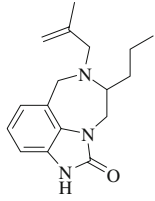
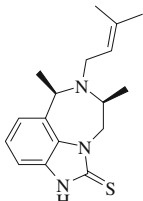
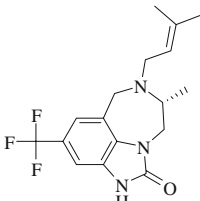
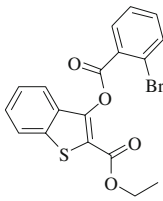
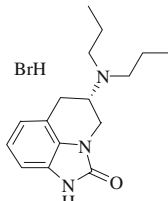
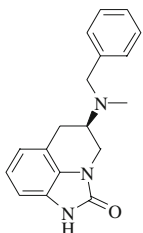
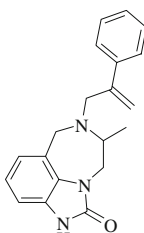
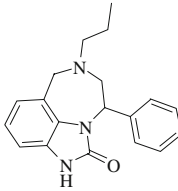
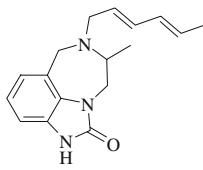
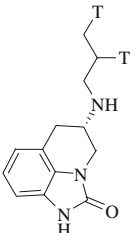
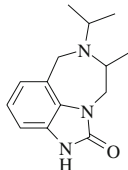
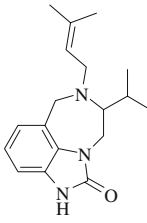
Template structure used for sub-structure search with compound similarity $\geq 90\%$.

63. CID 10265094	64. CID 10376208	65. CID 460857	66. CID 504495
67. CID 10061731	68. CID 504501	69. CID 465251	70. CID 465289
71. CID 6451104	72. CID 10365317	73. CID 452838	74. CID 460859
75. CID 10517380	76. CID 3010438	77. CID 465250	78. CID 465287

compounds belonging to TIBO family, specifically have halogens attached at 8 or 9 position and phenyl group or a bulky chain either directly attached to 6-N or to DMA chain. Among the halogens, fluorine containing compounds give better binding affinities. Compound numbers 63, 75, 86, 87, and 91 have different scaffolds from that of the TIBO. The scaffolds, where the diazepine ring has been replaced by a six member ring showed better binding affinity and hence better predicted activity. Compound no. 85 (CID 1038250) shows an altogether different scaffold. This compound is significantly different from all other

compounds because of the absence of a 7-member ring. The 5-member ring has a sulphur atom in place of N-H. The most astonishing thing about this compound is the absence of nitrogen. Almost, all of the NNRTIs in the present drug discovery pipeline have one or more nitrogen. Significantly, the very same compound was also reported in our previous report [39]. The bromo-benzoyloxy moiety and the benzothiophene ring position themselves in different planes. The compound occupies more space of the available volume of the NNIBP as compared to 1-REV. While comparing the interactions of 1038250 with 1-REV, the bromo-phenyl ring

Fig. 7 (continued)

			
79. CID 6451106	80. CID 10380506	81. CID 3010459	82. CID 465249
			
83. CID 10040548	84. CID 10315128	85. CID 1038250	86. CID 10473297
			
87. CID 10589822	88. CID 452844	89. CID 465288	90. CID 6451105
			
91. CID 10444087	92. CID452826	93. CID 465243	

seems to be involved in the H-bonding with LYS101 (Fig. 8a), whereas in the case of 1-REV, the interaction was observed with the 5-member ring. Favorable π - π interactions are expected between the benzothiazepine ring and TYR181 and TYR188 amino acid residues (seems to be more sandwiched, $\sim 1.3\text{\AA}$ closer) as compared to 1-REV. The benzothiazepine moiety shows stronger hydrophobic interactions with TYR181, TYR188, VAL108, TYR318, PRO95, PRO97, TRP229 and LEU234 in a similar manner as that of the DMA moiety in 1-REV. The closer contact observed here with TYR181 and TYR188 is a significant phenomenon, as it forms an important part of the hydrophobic core that interacts with most of the butterfly-like NNRTIs. Closer interactions of relatively large carboxylic acid ethyl ester moiety attached to the thiophene ring and the amino acid residue PHE227 (the carboxylic acid ethyl ester moiety is $\sim 2.31\text{\AA}$ closer as compared to the diazepine moiety) seems to apparently influence the binding mode

while hydrophobic interactions are observed with VAL106, VAL108 and VAL179. Bromo-benzoyloxy moiety interacts hydrophobically with LYS101, LYS103, PRO236 and LEU100 (Fig. 8b). The bromo-benzoyloxy moiety as well as the carboxylic acid ethyl ester moiety are deeply embedded and seem to reorient in the NNIBP in such a fashion, so as to resemble a butterfly arrangement, as observed in most of the NNRTIs. Favorable electrostatic interactions are also observed between the receptor and the ligand (Fig. 8c). The electrostatic interactions provide a highly informative means of characterizing the essential electronic features of drugs and their stereoelectronic complementarity with the receptor site. A closer stacking interaction of TYR181 and TYR188 with the aromatic moieties of the ligand seems to have a decisive effect on their mutual orientation.

The torsional bending in the compound helps itself to adjust within the NNIBP in a better way. More torsion angle

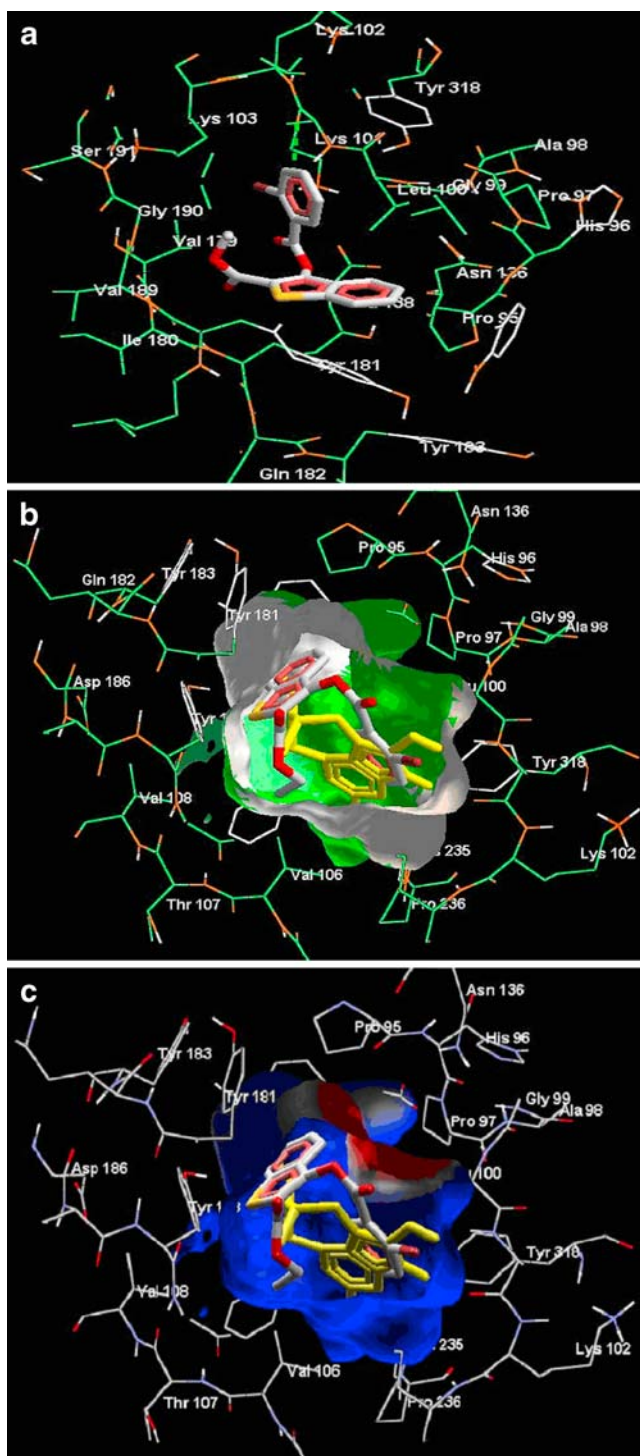
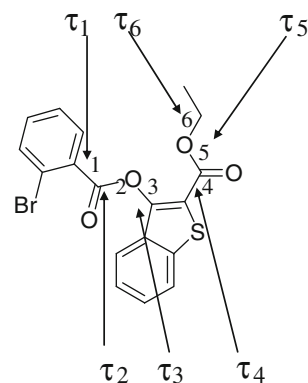


Fig. 8 Figures showing various interactions of compound number 85 (CID 1038250, 3-(2-Bromo-benzoyloxy)-benzo[b]thiophene-2-carboxylic acid ethyl ester). (a) Hydrogen bonding interactions, (b) Hydrophobic interactions (as compared to 1-REV), (c) Electrostatic interactions (as compared to 1-REV)

Fig. 9 The structure of 3-(2-Bromo-benzoyloxy)-benzo[b]thiophene-2-carboxylic acid ethyl ester (CID 1038250) depicting various torsional bonds



variations were observed in the compound as compared to its unbounded form as well as 1-REV (Fig. 9 and Table 5). All the six torsional bending seems to be instrumental in their ability to undergo conformational changes and repositioning within the pocket. The torsional freedom thus facilitates the conformational changes of the compound.

Though the potency of compound 1038250 (mol. Wt.= 405, hydrogen bond donor=0, hydrogen bond acceptor=4, log P=5.4) is not a sufficient parameter for it to be considered as an effective drug candidate, but the torsional freedom of the compound can be exploited to facilitate the conformational changes within the compound, and might be instrumental in compensating the effects of a resistance mutation. This particular scaffold (with very good predicted activity) along with other aforementioned scaffolds can serve as precursors for design and synthesis of newer and novel NNRTIs against HIV-1 RT.

Conclusions

The conclusions and significant accomplishments resulting from the present study may be summarized as follows:

In this work, template-based molecular docking studies were carried out to explore the efficacy of docking technique so as to obtain a better model as compared to that of a simple docking protocol, to facilitate design of new and effective NNRTIs. The docking simulations using templates

Table 5 Table showing the torsional bending in degrees responsible for reorientation and repositioning in the NNIBP for the PubChem compound 1038250

Torsional bending	Degrees (°)
τ_1	58.54
τ_2	138.84
τ_3	59.18
τ_4	120.87
τ_5	137.65
τ_6	107.38

of the highest active compound of the series ($pIC_{50}=8.52$) could generate a better model as compared to simple docking protocol used earlier [39]. Also more number of hits could be retrieved from the external database (PubChem compound database). The most significant accomplishment of our present work is the identification of a novel scaffold (CID 1038250), which could be a precursor to a newer and novel series of inhibitors of HIV-1 RT. The other newer scaffolds and novel modifications of TIBO too have better predicted activities. On the basis of correlations obtained between the biological activity and binding affinity and inferences obtained from the observed interactions, it can be inferred that the template-based molecular docking followed by ‘Tabu clustering’, can be a better alternative to simple docking protocol in evaluating more favorable binding modes of TIBO derivative’s top ranking compounds and can serve as a better tool for datamining large databases to yield novel hit candidates and lead optimization.

Acknowledgments The author wishes to acknowledge Dr. R.C. Saraswat, and Dr. P.K. Sen, S.G.S.I.T.S, Indore.

References

- Coffin JM (1995) *Science* 267:483–489
- Silvestri R, Maga G (2006) *Expert Opin Ther Pat* 16:939–962
- Matthis G, Torsten U, Danielson U (2006) *J Med Chem* 49:2375–2383
- Stephenson J (1997) *JAMA* 277:614–616
- Gazzard B (1999) *Int J Clin Pract Suppl* 103:45–48
- Lazzari S, de Felici A, Sobel H, Bertagnolio S (2004) *AIDS* 18: S49–53
- Palella FJ, Chmiel JS, Moorman AC, Holmberg SD (2002) *AIDS* 16:1617–1626
- De Clercq E (2005) *J Med Chem* 48:1297–1313
- Pauwels R (2004) *Curr Opin Pharmacol* 4:437–446
- De Clercq E (1990) *Trends Pharmacol Sci* 11:198–204
- Balzarini J (2004) *Curr Top Med Chem* 4:921–944
- Ren JS, Esnouf RM, Garman E, Somers D, Ross C (1995) *Nat Struct Biol* 2:293–302
- Esnouf R, Ren JS, Ross C, Jones Y, Stammers D (1995) *Nat Struct Biol* 2:303–308
- Hopkins AL, Ren J, Esnouf RM, Willcox BE, Jones EY (1996) *J Med Chem* 39:1589–1600
- Mager PP (1997) *Med Res Rev* 17:235–276
- Tantillo C, Ding J, Jacobo-Molina A, Nanni RG, Boyer PL, Hughes SH, Pauwels R, Andries K, Janssen PA, Arnold E (1994) *J Mol Biol* 243:369–384
- Esnouf RM, Ren J, Hopkins AL, Ross CK, Jones EY (1997) *Proc Natl Acad Sci USA* 94:3984–3989
- Campiani G, Ramunno A, Maga G, Nacci V, Fattorusso C (2002) *Curr Pharm Des* 8:615–657
- De Clercq E (1999) *IL Farmaco* 54:26–45
- De Clercq E (1998) *Antiviral Res* 38:153–179
- Ding J, Das K, Tantillo C, Zhang W, Clark ADJ, Pauwels R, Moereels H, Koymans L, Janssen PAJ, Smith RHJ, Kroeger Koepke R, Michejda CJ, Hughes SH, Arnold E (1995) *Structure* 3:365–379
- Das K, Levi PJ, Hughes SH, Arnold E (2005) *Prog Biophys Mol Biol* 88:209–231
- Zhou Z, Lin X, Madura JD (2006) *Infect Disord - Drug Targets* 6:391–413
- De Clercq E (2004) *Chem Biodiv* 1:44–64
- Barreca ML, Rao A, De Luca L, Zappala M, Monforte AM, Maga G, Panecouque C, Balzarini J, De Clercq E, Chimirri A, Monforte P (2005) *J Med Chem* 48:3433–3437
- D’Cruz OJ, Uckun FM (2006) *J Antimicrob Chemother* 57:411–423
- Batista J, Bajorath J (2007) *J Chem Inf Model* 47:59–68
- Bender A, Mussa Y, Glen RC (2004) *J Chem Inf Comput Sci* 44:1708–1718
- Batista J, Godden JW, Bajorath J (2006) *J Chem Inf Comput Sci* 46:1937–1944
- Bender A, Glen RC (2005) *J Chem Inf Model* 45:1369–1375
- Pauwels R, Andries K, Desmyter J, Schols D, Kukla MJ, Breslin HJ, Raeymaeckers A, Van Gelder J, Woestenborghs R, Heykants J (1990) *Nature* 343:470–474
- De Clercq E (1995) *Clin Microbiol Rev* 8:200–239
- Ren J, Esnouf R, Hopkins A, Ross C, Jones Y, Stammers D, Stuart D (1995) *Structure* 3:915–926
- Zhou Z, Madura JD (2004) *J Chem Inf Comput Sci* 44:2167–2178
- Das K, Ding J, Hsiou Y, Clark AJ, Moereels H, Koymans L, Andries K, Pauwels R, Janssen P, Boyer P, Clark P, Smith RJ, Kroeger SM, Michejda C, Hughes S, Arnold E (1996) *J Mol Biol* 264:1085–1100
- V Ding J, Das K, Moereels H, Koymans L, Andries K, Janssen PA, Hughes SH, Arnold E (1995) *Nature Struct Biol* 2:407–415
- Bahram H, Mohammad Hossein Tabaei S, Fatemeh N (2005) *J Mol Struct Theochem* 732:39–45
- Saen-oon S, Kuno M, Hannongbua S (2005) *Proteins* 61:859–863
- Sapre NS, Gupta S, Pancholi N, Sapre N (2007) *J. Comput. Aided Mol. Des.* doi:10.1007/S10822-007-9161-8
- <http://pubchem.ncbi.nlm.nih.gov>
- ChemDraw Ultra 7.0.0 (www.cambridgesoft.com)
- www.molegro.com (free trial version)
- <http://www.rcsb.org/pdb>
- Storn R, Price K (1995) *Differential Evolution - A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces*; Technical Report; International Computer Science Institute: Berkeley, CA
- Thomsen R (2003) *Flexible Ligand Docking Using Differential Evolution*, Proceedings of the 2003 Congress on Evol Comp 4:2354–2361
- Thomsen R, Christensen MH (2006) *J Med Chem* 49:3315–3321
- Gehlhaar DK, Verkhivker G, Rejto PA, Fogel DB, Fogel LJ, Freer ST (1995) *Docking conformationally flexible small molecules into a protein binding site through evolutionary programming*. Proceedings of the Fourth International Conference on Evolutionary Programming, 615–627.
- Gehlhaar DK, Bouzida D, Rejto PA (1998) *Fully automated and rapid flexible docking of inhibitors covalently bound to serine proteases*. Proc Seventh Int Conf Evol Prog 449–461
- Yang JM, Chen CC (2004) *Proteins* 55:288–304
- Rawal RK, Kumar A, Siddiqi IS, Katti SB (2007) *J Mol Model* 13:155–161
- Rizzo RC, Wang DP, Tirado-Rives J, Jorgensen WL (2000) *J Am Chem Soc* 122:12898–12900
- Zhou Z, Madura JD (2004) *Proteins* 57:493–498