

## Assessing ligand efficiencies using template-based molecular docking and Tabu-clustering on tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepin-2(1H)-one and -thione (TIBO) derivatives as HIV-1RT inhibitors

NITIN S SAPRE<sup>1,\*</sup>, SWAGATA GUPTA<sup>2</sup> and NEELIMA SAPRE<sup>3</sup>

<sup>1</sup>Department of Applied Chemistry, Shri GS Institute of Technology and Sciences, Indore 452 001

<sup>2</sup>Department of Chemistry, Govt. P.G. College, MHOW 453 441

<sup>3</sup>Department of Computer Science, SD Bansal College of Technology, Indore 453 331

e-mail: sukusap@yahoo.com; nsapre@sgsits.ac.in

MS received 28 February 2008; revised 12 April 2008

**Abstract.** A template-based flexible docking simulation followed by ‘Tabu-clustering’ was performed on a series of 38 TIBO derivatives as HIV-1 reverse transcriptase (HIV-1 RT) inhibitors. Four different templates of the Cl-TIBO (1-REV) were created and used as reference templates for docking and aligning. On the basis of the optimal conformation of the ligands, when fitting to the template, the respective scoring functions were obtained; different ligand efficiencies were evaluated and analysed. Statistical modelling using artificial neural network (ANN:  $r^2 = 0.922$ ) and multiple linear regression method (MLR:  $r^2 = 0.851$ ) showed good correlation between the biological activity, binding affinity, and different ligand efficiencies of the compounds, which suggest the robustness of the template-based binding conformations of these inhibitors. Our studies suggest that, template-based docking followed by ‘Tabu-clustering’ will give a better alignment of inhibitors with respect to the crystal coordinates and enhance the docking efficiency. Also, our study indicates that scoring functions based on 3D symmetry analysis along with heavy atoms count serve as a valuable tool for estimating the efficiency of the ligands. Thus, this is a novel method based on heavy atoms count predicting the binding affinity of the TIBO group of inhibitors, so that their therapeutic utility can be enhanced.

**Keywords.** TIBO; structure based drug design (SBDD); template; Tabu-clustering; HIV-1 reverse transcriptase (HIVRT); non-nucleoside reverse transcriptase inhibitor (NNRTI); pIC50; multiple linear regression (MLR); artificial neural network (ANN).

### 1. Introduction

The human immunodeficiency virus-1 (HIV-1), the causative agent for acquired immunodeficiency syndrome (AIDS), is the most interesting virus in the history of biomedical research.<sup>1–3</sup> At present, chemotherapy seems to be the main weapon in dealing with the dreaded disease caused by HIV-1 retro-virus. As a retrovirus, HIV has an envelop of lipid bi-layer membrane containing two copies of a single stranded RNA genome that codes for the structural proteins, surface glycoproteins, regulatory factors, and the enzymes reverse transcriptase (RT), protease, and integrase. Blocking the molecular mechanisms associated with HIV-1 pathogenesis has been the aim of

the researchers involved in AIDS eradication, but only limited success had been attained in this field.<sup>4,5</sup>

The results of Highly Active Antiretroviral Therapy (HAART) have received a setback owing to drug resistance as a result of incomplete suppression.<sup>6,7</sup> Thus, the issue of rapid emergence of NNRTI resistance has to be tackled by designing potent and efficient inhibitors in order to inhibit wild type HIV-1 as well as pre-existing resistant viral variants due to occurrence of mutations during ongoing viral replication.<sup>8–10</sup>

Non-nucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs) are an important group of structurally diverse compounds, which can act as highly effective inhibitors of the enzymatic activity of HIV-1 reverse transcriptase *in vitro* and of HIV-1 viral replication in cell culture and infected people. To

\*For correspondence

fulfill the criteria of an efficient NNRTI, a compound should bind specifically to the allosteric binding site, which is physically separated from the catalytic domain or the substrate binding site, of the HIV-1 RT at a concentration that is significantly lower than the concentration required to affect normal cell viability.<sup>11–14</sup> HIV-1 RT is a heterodimer composed of a p66 subunit carrying both DNA polymerase and RNase H-domain, and a proteolytically processed p51 subunit comprising only of the DNA polymerase domain. Both subunits are encoded in the same region of the viral genome, and a single mutation in the RT coding region will result in a heterodimer carrying the same amino acid substitution in both the subunits. The three-dimensional structure of each subunit in the heterodimer is different, and the amino acid substitution in each subunit cannot be considered structurally or functionally equivalent. All the NNRTIs, independent of their structure, bind in a hydrophobic pocket, located in the p66 subunit, approximately 10 Å from the polymerase binding site.<sup>15–17</sup>

In the present scenario, chemotherapy seems to be the main weapon in dealing with the dreaded disease caused by HIV-1 retrovirus. Together with nucleoside reverse transcriptase inhibitors (NRTIs) and protease inhibitors (PIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs) have gained a crucial place in the treatment of HIV-1 infections, and at present are in rapid progress. Virtually, all the drugs that have been licensed for clinical use for the treatment of HIV infections fall into one of the following categories: (i) nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs), that, following two phosphorylation or three phosphorylation steps act as chain terminators, at the substrate binding site of the reverse transcriptase, (ii) nonnucleoside reverse transcriptase inhibitors (NNRTIs), that interact with the reverse transcriptase at an allosteric, nonsubstrate binding site, (iii) protease inhibitors (PIs), that specifically inhibit the virus-associated protease or (iv) viral fusion inhibitor such as Fuzeon.<sup>18</sup> Numerous other stages of the viral replication cycle are also being investigated as possible chemotherapeutical targets such as viral attachment to the host cell, integration of the provirus into the host genome, packaging and assembly of virions. The strength of NNRTIs as anti-AIDS drug lies in their antiviral potency, high specificity and low toxicity as they do not directly affect the activities of the cellular polymerases.<sup>19</sup>

Computational methods have developed into useful tools in facilitating new drug discovery. These meth-

ods are simple and non-expensive and speed up the process of designing novel and potent molecules with desired biological activity. Quantitative structure-activity relationship (QSAR) and docking methods are two mostly used computational methods in structure-based drug design (SBDD).<sup>20,21</sup> In QSAR methodologies, a mathematical relationship, relating the biological activity to some molecular descriptors is obtained.<sup>22–24</sup> In docking studies, different search algorithms such as simulated annealing and genetic algorithm in combination with scoring function such as molecular mechanic calculations are being used to study the binding of the candidate ligands to an enzyme with known structure. Through docking procedures, not only new biological active compound is introduced, but also the chemistry of the ligand-enzyme interaction is well recognized. Docking studies involving NNI's and HIV1 RT have been performed previously.<sup>25</sup> Rationally designed NNRTIs deduced from computational assessment of changes in binding pocket size, shape and residue character that result from clinically observed NNRTI resistance-associated mutations exhibit high binding affinity for HIV-1 RT and a better pharmacokinetic profile.<sup>26</sup>

TIBO (Tetrahydroimidazo-[4,5,1-jk][1,4]-Benzodiazepin-2(1H)-one and -thione) derivatives,<sup>27</sup> are one of the important classes of non-nucleoside reverse transcriptase inhibitors which inhibit the replication of HIV-1.<sup>28,29</sup> TIBO derivatives, like most of the other non-nucleoside inhibitors, on binding to the binding pocket, adapt conformation which resemble the wings of a butterfly, and hence called as butterfly like conformation. The specific conformation of the 7-membered ring of the TIBO derivatives is responsible for producing their butterfly like arrangement. Comparison of the different RT-NNI complexes suggests modifications to the TIBO group of inhibitors which might enhance their binding and hence, potentially, their therapeutic efficacy.<sup>30–32</sup> A TIBO derivative, Tivirapine has already entered into the clinical trials.<sup>33</sup> Analysis of the crystalline structures of the NNRTI/RT complexes in both presence and absence of mutations is crucial for understanding the interaction and stereo-chemical changes associated with the receptor and ligand on binding in the non-nucleoside inhibitor-binding pocket.<sup>33–37</sup>

Recent reports have shown that docking scores are not much more accurate predictors of binding than the molecular weight,<sup>38</sup> or that raw atom counts are almost as good as '2D fingerprints'<sup>39</sup> have motivated us to perform the present studies wherein we have taken heavy atoms count and correlated it to various

scores as ligand efficiencies and utilized it in ascertaining the various binding efficacies, which seems to be a novel approach.

## 2. Computational methods

### 2.1 Molecular structures

TIBO or Tetrahydroimidazo-[4,5,1-jk][1,4]-Benzodiazepin-2(1H)-one and -thione and its derivatives developed by Pauwels *et al*<sup>27</sup> along with their biological activities are taken for template-based docking studies. The molecular structures were drawn and optimized using ChemDraw ultra 7.0<sup>40</sup> and exported to Molegro Virtual Docker<sup>41</sup> 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<sup>42</sup> [Research Collaboratory for Structural Bioinformatics (RCSB) (<http://www.rcsb.org/pdb>)].

### 2.2 Docking simulations

Docking simulations have been performed using MolDock software, which combines differential evolution with a cavity prediction algorithm.<sup>43</sup> The guided differential evolution algorithm combines the differential evolution optimization technique with a cavity prediction algorithm. Differential evolution (DE)<sup>44</sup> has been successfully applied to molecular docking.<sup>45</sup> Fast and accurate identification of potential binding modes during the search process is made by the use of predicted cavities. The docking scoring function of MolDock make use of piece-wise linear potential (PLP)<sup>46,47</sup> and is further extended in GEMDOCK.<sup>48</sup> The docking scoring function takes hydrogen bond directionality into account. The highest ranked poses are again reranked to increase the docking accuracy further. Only the ligand properties were represented in the individuals, as the protein has limited side chain flexibility during the docking process. The fitness of a candidate solution is the sum of the intermolecular interaction energy between the ligand and the protein, and the intramolecular interaction energy of the ligand. The docking scoring function,  $E_{score}$ , is defined by the following energy terms

$$E_{score} = E_{inter} + E_{intra}, \quad (1)$$

where  $E_{inter}$  is the ligand–protein interaction energy:

$$E_{inter} = \sum_{i \in ligand} \sum_{j \in protein} \left[ E_{PLP}(r_{ij}) + 332 \cdot 0 \frac{q_i q_j}{4 r_{ij}^2} \right]. \quad (2)$$

The summation runs over all heavy atoms in the ligand and all heavy atoms in the protein, including cofactor atoms and water molecule atoms that might be present. The second term of the (2) describes the electrostatic interactions between charged atoms.

$E_{intra}$  is the internal energy of the ligand:

$$E_{intra} = \sum_{i \in ligand} \sum_{j \in ligand} E_{PLP}(r_{ij}) + \sum_{flexiblebonds} A[1 - \cos(m^* \theta - \theta_0)] + E_{clash}. \quad (3)$$

The double summation is between all atom pairs in the ligand, excluding atom pairs that are connected by two bonds or less. The second term is a torsional energy term, parameterized according to the hybridization types of the bonded atoms,  $\theta$  is the torsional angle of the bond. The last term,  $E_{clash}$ , assigns a penalty of 1000 if the distance between two atoms (more than two bonds apart) is less than 2.0 Å. Thus, the  $E_{clash}$  term punishes infeasible ligand conformations.

Our present study makes use of docking templates so as to focus the search. Templates 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 such as hydrogen acceptor atoms or donor atoms. Each template group contains a number of centers: optimal 3D positions for the group feature.

Following Gaussian formula is used for rewarding each group center:

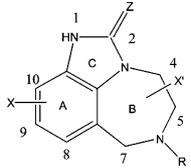
$$e = \omega^* \exp\left(\frac{-d^2}{r_0^2}\right), \quad (4)$$

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

## 3. Hardware and software

Molegro Virtual Docker 2007.2.2<sup>41</sup> was run on a Windows XP based Pentium IV 2.66 GHz PC (with 512 MB of memory).

**Table 1.** Anti-HIV-1 activity (inhibitory concentration pIC<sub>50</sub>), binding affinity ( $E_{\text{binding}}$  in kJ/mol) and ligand efficiencies (LE 1, LE 2, LE 3, LE 4 and LE 5) of tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepin-2(1H)-one and -thione (TIBO derivatives).

No.	X	Z	R	X'	pIC <sub>50</sub> <sup>α</sup>	Binding affinity <sup>δ</sup>	LE 1 <sup>a</sup>	LE 2 <sup>b</sup>	LE 3 <sup>c</sup>	LE 4 <sup>d</sup>	LE 5 <sup>e</sup>
											
1	9-Cl	S	DMA <sup>ε</sup>	5-Me(S)	7.47	-25.02	-3.01	-1.19	8.79	-21.83	-21.36
2	8-Cl	S	DMA	5-Me(S)	8.37	-25.04	-4.34	-1.19	0.52	-21.46	-21.36
3	8-F	S	DMA	5-Me(S)	8.24	-24.98	-4.32	-1.19	0.47	-21.55	-21.41
4	8-SMe	S	DMA	5-Me(S)	8.30	-23.61	-3.97	-1.07	3.26	-20.90	-20.69
5	8-OMe	S	DMA	5-Me(S)	7.47	-23.45	-3.85	-1.07	2.62	-20.62	-20.53
6	8-OC <sub>2</sub> H <sub>5</sub>	S	DMA	5-Me(S)	7.02	-24.34	-3.21	-1.06	5.24	-20.32	-20.06
7	8-CN	O	DMA	5-Me(S)	5.94	-18.51	-4.04	-0.84	1.11	-20.77	-20.64
8	8-CHO	S	DMA	5-Me(S)	6.73	-20.65	-3.80	-0.94	1.27	-20.60	-20.56
9	8C=-CH	S	DMA	5-Me(S)	7.53	-21.47	1.09	-0.98	29.12	-19.53	-19.60
10	8-CH <sub>3</sub>	O	DMA	5-Me(S)	6.00	-20.59	-4.00	-0.98	1.33	-21.51	-21.28
11	8-N(CH <sub>3</sub> ) <sub>2</sub>	O	CPM <sup>§</sup>	5-Me(S)	5.18	-20.82	-2.60	-0.95	10.86	-20.94	-20.71
12	9-NH <sub>2</sub>	O	CPM	5-Me(S)	4.22	-22.06	-3.67	-1.10	3.63	-22.98	-22.56
13	9-NHCOMe	O	CPM	5-Me(S)	3.80	-19.17	-3.61	-0.83	0.93	-19.93	-19.56
14	9-F	S	DMA	5-Me(S)	7.60	-24.98	-3.05	-1.19	9.28	-21.86	-21.42
15	9-Me	O	DEA <sup>#</sup>	5-Me(S)	6.50	-19.61	-3.76	-0.85	3.06	-19.85	-19.88
16	H	O	CH <sub>2</sub> CO <sub>2</sub> Me	5-Me(S)	3.07	-23.17	-1.71	-1.16	10.23	-22.62	-22.03
17	H	O	CH <sub>2</sub> C=-CH	5-Me(S)	3.24	-21.42	-1.13	-1.19	18.38	-23.44	-23.09
18	H	O	CH <sub>2</sub> CH <sub>2</sub> CH=CH <sub>2</sub>	5-Me(S)	4.30	-20.78	-0.17	-1.09	21.11	-22.29	-22.03
19	H	O	CH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	5-Me(S)	4.05	-20.36	-4.45	-1.13	-0.68	-24.06	-23.63
20	H	O	CH <sub>2</sub> CH=CHMe(Z)	5-Me(S)	4.46	-21.55	-4.37	-1.13	-0.15	-23.22	-22.81
21	H	O	DMA	5-Me(S)	4.90	-20.93	-4.04	-1.05	0.83	-22.37	-22.11
22	H	O	CH <sub>2</sub> C(Me)=CHMe(E)	5-Me(S)	4.54	-21.68	-4.70	-1.08	-0.86	-22.37	-21.90
23	H	O	DMA[S(+)]	5-Me(S)	5.40	-20.91	-4.66	-1.05	-0.60	-22.39	-22.04
24	H	O	CH <sub>2</sub> C(CH=CH <sub>2</sub> )=CH <sub>2</sub>	5-Me(S)	4.15	-21.85	-2.41	-1.09	9.81	-22.63	-22.19
25	8-Cl	S	DMA	H	7.34	-25.21	-4.97	-1.26	-3.26	-22.37	-22.11
26	9-Cl	S	DMA	H	6.80	-24.51	2.70	-1.23	37.99	-21.34	-21.27
27	H	O	2-MA <sup>¶</sup>	5,5-di-Me	4.64	-20.77	-1.79	-1.04	14.88	-22.32	-22.03
28	H	O	2-MA	4-Me	4.50	-20.29	-3.61	-1.07	7.39	-23.95	-23.86
29	H	O	C <sub>3</sub> H <sub>7</sub>	4-CHMe <sub>2</sub>	4.13	-18.63	-5.07	-0.93	-1.06	-23.14	-23.11
30	H	O	DMA	7-Me	4.92	-20.46	-4.64	-1.02	-1.25	-22.32	-22.05
31	8-Cl	O	DMA	7-Me	6.84	-23.97	-3.88	-1.14	3.73	-21.90	-21.74
32	9-Cl	O	DMA	7-Me	6.80	-24.08	-4.58	-1.15	-0.25	-22.15	-21.96
33	8-Cl	S	DMA	7-Me	7.92	-24.48	-3.93	-1.17	3.97	-21.81	-21.73
34	9-Cl	S	DMA	7-Me	7.64	-24.53	-2.79	-1.17	13.39	-21.83	-21.74
35	H	O	DMA	4,5-di-Me(cis)	4.25	-20.44	-4.26	-0.97	-0.17	-22.07	-21.63
36	H	S	DMA	4,5-di-Me(cis)	5.65	-21.44	-4.28	-1.02	0.54	-21.44	-21.19
37	H	S	DMA	4,5-di-Me(trans)	4.84	-20.99	-4.65	-1.00	-2.21	-21.49	-21.29
38	H	S	DMA	4,7-di-Me(trans)	4.59	-20.18	-5.76	-0.96	-4.31	-22.95	-22.79

<sup>ε</sup>DMA = 3,3-Dimethylallyl, <sup>¶</sup>2-MA = 2-methylallyl, <sup>#</sup>DEA = 3,3-Diethylallyl, <sup>§</sup>CPM = cyclopropylmethyl

<sup>α</sup>pIC<sub>50</sub> = -logIC<sub>50</sub> (where IC<sub>50</sub> is the effective concentration of a compound required to activate 50% protection of MT-4 cell against the cytopathic effect of HIV-1)

<sup>δ</sup>E<sub>binding</sub> in kJ/mol, <sup>a</sup>LE 1 = MolDock score/heavy atom count, <sup>b</sup>LE 2 = binding affinity/heavy atom count, <sup>c</sup>LE 3 = reranking score/heavy atom count, <sup>d</sup>LE 4 = similarity score based on templates/heavy atom count, <sup>e</sup>LE 5 = docking score (pose energy)/heavy atom count. (all in arbitrary energy units)

#### 4. Results and discussion

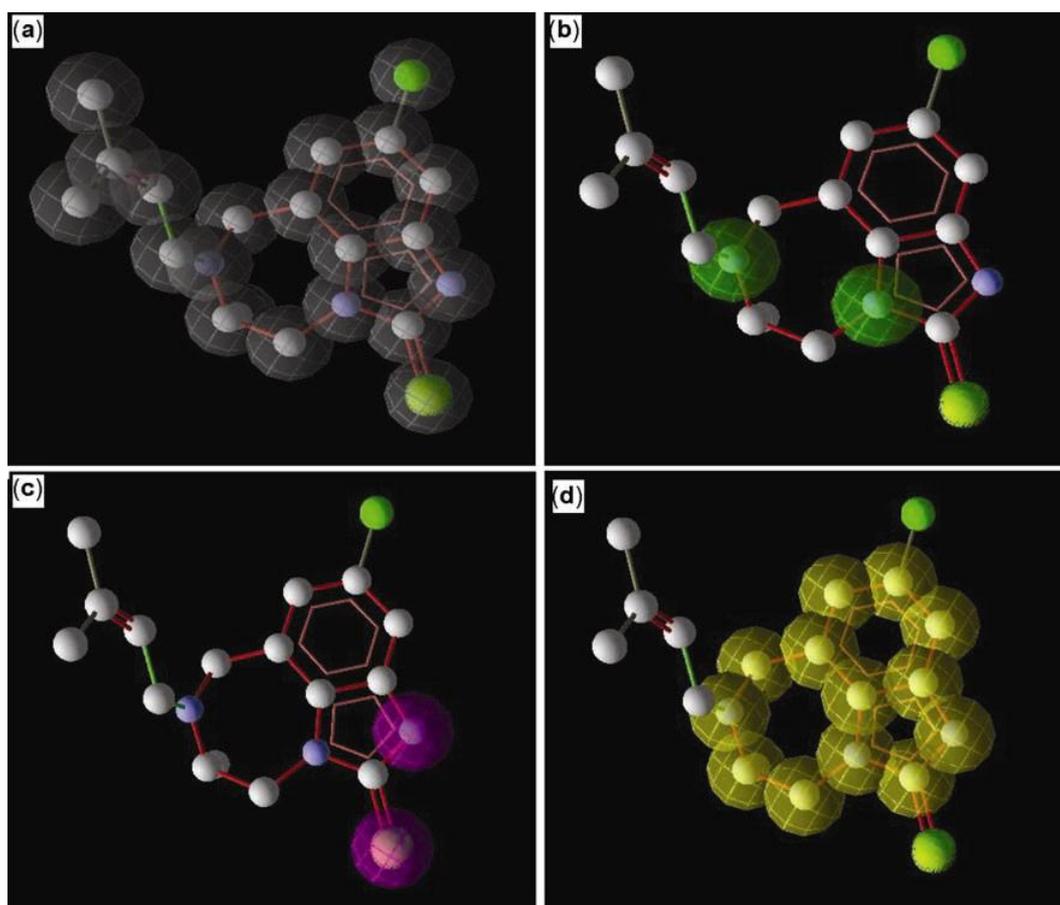
Template docking can be used when knowledge about the 3D conformation of a ligand is available. From the known conformations, it is possible to create a template with features expected to be relevant for the binding. Our specific aims for the present study are three-fold. Firstly, to derive the templates of the compound in the crystal structure, secondly, to assess the efficacy of various ligand-efficiencies along with the binding affinities, based on the fitness evaluation by template-assisted molecular docking, and thirdly, to understand their relationship with the antiviral activity through various statistical analyses. During docking, for each atom in the ligand, score contributions from all centers in all matching groups are taken into account. A single atom may contribute to several centers in several groups, where an atom is not restricted to the closest matching center or a single group.

The chemical structures of 38 TIBO derivatives are given in table 1 along with their biological acti-

vities, expressed in terms of  $pIC_{50}$  (where  $IC_{50}$  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 types of substituent as X, Z, R and X'. X represents the substitution on the aromatic ring A, X' represents substitution on the 7-member ring B, Z represents presence of O or S attached to the five member ring C and R is the substitution attached to N in the ring B.

##### 4.1 Validation of the docking method

For our present studies, we have selected the 9-Cl-TIBO/HIV-RT complex (PDB CODE-1REV). 9Cl-TIBO is extracted from the complex (1REV) and redocked using flexible docking simulations into its original structure of RT. The starting coordinates of the HIV-1 RT/TIBO complex (1REV) were imported from the Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)) and bond orders, charges, flexible torsions were assigned to them. Bond flexibility of the ligands was checked and the protein was protonated and side



**Figure 1.** Templates showing different groups for the 9-Cl TIBO crystal structure [PDB code: 1REV]: (a) steric, (b) hydrogen acceptor, (c) hydrogen donor, (d) ring.

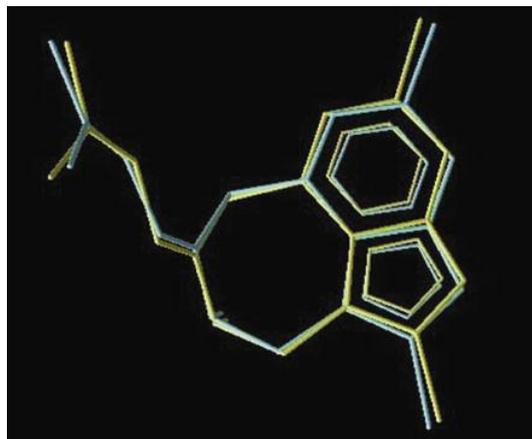
chains were minimized using Nedler–Mead simplex algorithm. Prior to docking, templates (figures 1a–d) based on steric (21 groups, shape matching taking into consideration all the atoms), hydrogen acceptor (2 groups), hydrogen donor (2 groups) and ring (13 groups, aromatic as well aliphatic) were derived for the crystal-fit ligand. Similarity groups were set-up with the following parameters (overall strength =  $-500.00$ , resolution energy grid =  $0.40$ ) and Ligand Evaluator with resolution  $0.40$  Å was used for scoring. MolDock Optimizer, based on an evolutionary algorithm<sup>49</sup> has been used as docking search algorithm. Following parameters were set: No. of runs = 10, max. iterations = 2000, max. population size = 50, scaling factor =  $0.50$ , cross over rate =  $0.90$ , offspring scheme = scheme 1, termination step = variance based. It was followed by pose clustering so that the best-scoring pose was not missed out. For pose clustering, ‘Tabu clustering’ technique has been applied to ensure greater diversity of the returned poses, wherein poses similar to solutions from earlier runs were penalized. RMSD calculation was done by atom ID with a threshold of  $2$  Å. Fitness evaluations were done by Ligand Evaluator scoring function with a RMSD threshold of  $2.00$  Å and an energy penalty term. The poses were reranked to increase the docking accuracy. The root mean square deviation (RMSD) of  $0.186$  Å between the redocked and crystal ligand coordinates indicates a good alignment of the experimental and calculated positions. The RMSD obtained was better than that obtained by simple docking ( $0.269$  Å) using the same docking engine as well as the other docking algorithms.<sup>48</sup> Figure 2 shows the best-fit redocked coordinates with respect to the crystal structure. In general, it is observed that binding affinity solely cannot be a measure of higher biological activity. Sometimes compounds with higher biological activity show lower binding affinity values which may be attributed to certain complexity in protein flexibility and interactions.<sup>50</sup>

#### 4.2 Docking of the molecule set

After the validation of the docking method using 9Cl-TIBO, a dataset of 38 molecules belonging to TIBO derivatives with varied activity range ( $3.0$ – $8.52$ ) were docked into the same coordinates of the crystal structure. The docked 3D-structures of TIBO derivatives were scored, reranked and then compared with the X-ray crystallographic structure of 9-

Cl-TIBO. The basic backbone of the TIBO derivatives has a common pattern with a diazepine ring, imidazolone ring and phenyl ring, therefore as expected, similar alignment and interactions are obtained. The result demonstrates that, based on template docking, the TIBO inhibitors can be docked and aligned into the NNRTI allosteric binding site extremely well. A good alignment (figure 3) was observed between the dataset ligands and the crystal ligand. The ligand efficiencies based on different scoring functions and heavy atoms count were evaluated and relative binding affinities were calculated.

The interactions observed were similar to those mentioned in our earlier publication.<sup>50</sup> The amino acid residue LYS101 shows hydrogen bond interactions with imidazolone ring of ligand. Extensive hydrophobic interactions are instrumental in stabilizing the TIBO compounds. The diazepine ring and the group attached to it at 6-N are strongly hydrophobic. The high activity of DMA (dimethyl allyl) group attached to diazepine ring is attributed to its hydrophobicity and perhaps is the reason of high activity of compounds with it. The residues involved in hydrophobic interactions with the diazepine ring and 6-N are TYR181, TYR185, TYR229, LEU234, VAL106, GLY190, VAL189 and PRO95. The 6-member and the 5-member ring are lesser hydrophobic in nature as compared to the 7-member ring. The hydrophobic interactions involved with the phenyl ring are PHE227, TYR318 and LEU234. The imidazolone ring interacts hydrophobically with LYS101, LYS103. Favourable electrostatic interac-



**Figure 2.** Conformation of 9-Cl TIBO crystal structure [PDB code: 1 REV] (yellow) as compared to redocked conformation of 9-Cl TIBO (blue).

tions are observed between the ligand and the binding pocket. Electrostatics of molecules provide a highly informative means of characterizing the essential electronic features of inhibitors and their stereoelectronic complementarities with the receptor site on the basis of ionic and polar interactions between the host and the guest. As expected, the major role is played by the aromatic moiety present in the structure.  $\pi$ - $\pi$  interactions are observed between the aromatic moieties of the amino acid residues (TYR181 and TYR188) that converge at the inner side of the binding cavity and the aromatic moieties of the ligand. The electrostatic interaction is stabilized by these stacking type interactions.

### 4.3 Correlation between binding affinity, ligand efficiencies and activity

Table 1 also records binding affinity ( $E_{\text{binding}}$ ) in kJ/mol and ligand efficiencies (LE 1, LE 2, LE 3, LE 4 and LE 5 based on MolDock score, binding affinity, rerank score, similarity score, docking score and heavy atoms count).

The primitive job of any docking engine is in predicting the binding affinity and interaction energies along with the binding conformation of an inhibitor with the receptor. The binding affinity is the measure of favourable interaction between the receptor and the ligand. The binding affinity (kJ/mol) of a particular pose is given by:  $E_{\text{binding}} = -5.68 * pKi$  (The numerical factor corresponds to a temperature of 297 K). The reranked scores predicted the binding affinities in the range of -18.63 to -25.04 kJ/mol. The reliability of the ligand efficiencies and binding affinities as evaluated by the docking engine were ascertained by correlating them with the biological activity of the TIBO derivatives. Various models were generated using multiple linear regression technique (MLR) and artificial neural networks (ANN) using in-built data analyser. Leave-one-out (LOO) procedures as well as N-cross validated (N-CV) method were used for validation of results. Biological activity ( $pIC_{50}$ ) was taken as the dependent variable.

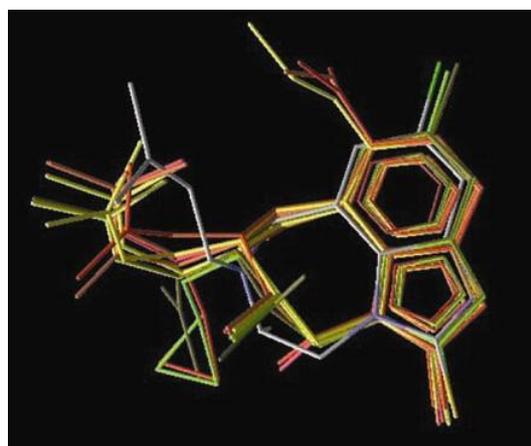
The auto generated random seed used in the model training was 416619787. In these equations,  $n$  is the number of compounds,  $r$  is the correlation coefficient,  $q^2$  is the cross-validated  $r^2$  from the (LOO) or (NCV) procedure,  $\rho$  is the Spearman rank correlation coefficient, MSE is the mean squared error and PRESS is the predictive sum of squares.

4.3a Multiple linear regression technique (MLR): The best model relating biological activity with the binding affinity derived using MLR (LOO) is presented below:

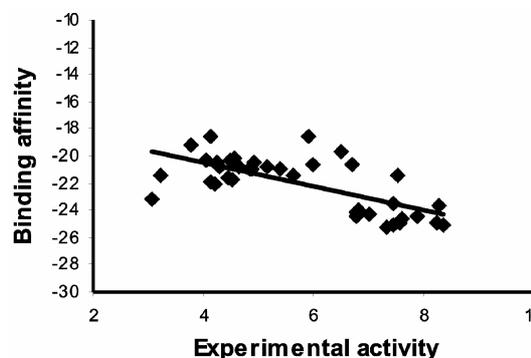
$$\text{Activity (pIC}_{50}) = -0.114 (\pm 0.145)E_{\text{binding}} - 1.406 (\pm 1.525)LE_1 - 8.56 (\pm 0.593)LE_2 + 0.212 (\pm 1.21)LE_3 + 4.06 (\pm 2.80)LE_4 - 2.94 (\pm 1.91)LE_5 - 13.58 \quad (5)$$

$$(n = 38, r = 0.922, r^2 = 0.851, r^2_{\text{adj}} = 0.822, \text{Spearman } (\rho) = 0.906, q^2 = 0.850, \text{MSE} = 0.358, \text{PRESS} = 13.63).$$

Also results derived from MLR (N-CV) using  $N = 10$ , provided similar results ( $r^2 = 0.847, q^2 = 0.846$ ).



**Figure 3.** Docked conformation of 9-Cl TIBO crystal structure (CPK) with eleven best aligned TIBO derivatives. (Compound no. 1, 4, 6, 11, 12, 14, 16, 27, 28, 30, 36).



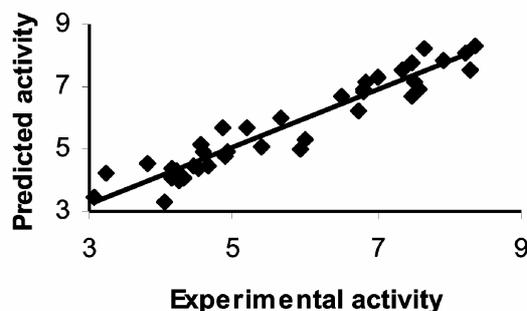
**Figure 4.** Graph between experimental activity ( $pIC_{50}$ ) and binding affinity ( $E_{\text{binding}}$  in kJ/mol) of TIBO derivatives.

**Table 2a.** Comparison of different statistical analyses.

	$r^2$	$r_{\text{adj}}^2$	Spearman (rho)	PRESS	MSE	$q^2$
MLR (LOO)	0.851	0.822	0.909	13.63	0.358	0.850
MLR (N-CV) N = 10	0.847	0.817	0.909	14.03	0.369	0.846
ANN (Back propagation)	0.902	–	0.923	–	0.200	–

**Table 2b.** The relevance scores of various descriptors as evaluated by artificial neural network (back propagation method).

Descriptor	Relevance score
LE 4	100
LE 1	50
Affinity	28
LE 3	14
LE 2	12
LE 5	6

**Figure 5.** Graph between experimental  $\text{pIC}_{50}$  and predicted  $\text{pIC}_{50}$  of TIBO derivatives.

4.3b *Artificial neural networks (ANN)*: The back-propagation method has been used for training the ANN model. Using the same random seed, following parameters were fixed: max training epoch = 2000, learning rate = 0.50, output learning rate = 0.30, momentum = 0.20, number of neurons in the first hidden layer = 20, number of neurons in the second hidden layer = 10, initial weight ( $\pm$ ) = 0.50. A better prediction was obtained ( $r^2 = 0.922$ , Spearman (rho) = 0.918 and MSE = 0.219). The above statistical analyses suggest the robustness of the method of evaluating ligand efficiencies based on template-assisted docking followed by ‘Tabu-clustering’ procedure adopted here. The statistical modelling results and relevance score of different descriptors are summarised in tables 2a and 2b, respectively. Both MLR and ANN methods suggest that the most relevant descriptor is  $\text{LE}_4$  (similarity score/heavy atom count), which strongly reaffirm that scoring function

based on similarity analysis of 3D structures are crucial in the docking studies as compared to scoring functions based on other parameters.

The results obtained, thus are validated by correlating biological activity with the binding affinity (figure 4). A good agreement was found between the affinity and anti-viral activity.

Table 3 records the observed and the calculated (5) values of  $\text{pIC}_{50}$  for the set of TIBO derivatives. The quality of correlation is demonstrated by their residual values i.e. the difference between their experimental and predicted  $\text{pIC}_{50}$ .

A good fit was observed between experimental and predicted activity as shown in figure 5, reconfirming the robustness of the template-based docking and ligand efficiency assessment procedure adopted here. Thus template-based molecular docking studies using different templates of the ligand associated with crystal structure followed by ‘Tabu clustering’ (though computationally more expensive) has provided a reasonably satisfactory model and can be instrumental in providing better docking results.

## 5. Conclusion

In this work, template-based molecular docking studies were carried out to explore the different ligand efficiencies based on scoring functions and heavy atoms count. The alignment of the ligands in the putative binding site of HIV-1 RT enzyme using above protocol could be used to facilitate design of newer and

**Table 3.** Experimental, predicted and residual values of pIC<sub>50</sub> for TIBO derivatives.

Sl. no.	Observed activity	Predicted activity	Residual
1	7.47	6.72	-0.75
2	8.37	8.33	-0.04
3	8.24	8.05	-0.19
4	8.30	7.51	-0.79
5	7.47	7.80	0.33
6	7.02	7.34	0.32
7	5.94	5.02	-0.92
8	6.73	6.21	-0.52
9	7.53	7.18	-0.35
10	6.00	5.30	-0.70
11	5.18	5.72	0.54
12	4.22	4.30	0.08
13	3.80	4.57	0.77
14	7.60	6.89	-0.71
15	6.50	6.73	0.23
16	3.07	3.48	0.41
17	3.24	4.21	0.97
18	4.30	4.11	-0.19
19	4.05	3.28	-0.77
20	4.46	4.47	0.01
21	4.90	4.75	-0.15
22	4.54	5.13	0.59
23	5.40	5.04	-0.36
24	4.15	4.05	-0.10
25	7.34	7.52	0.18
26	6.80	6.84	0.04
27	4.64	4.46	-0.18
28	4.50	4.38	-0.12
29	4.13	4.37	0.24
30	4.92	4.92	0.00
31	6.84	7.17	0.33
32	6.80	6.95	0.15
33	7.92	7.86	-0.06
34	7.64	8.23	0.59
35	4.25	3.98	-0.27
36	5.65	5.97	0.32
37	4.84	5.72	0.88
38	4.59	4.93	0.34

more effective NNRTIs. The docking simulations carried could satisfactorily reproduce a bound complex from the crystal structure of RT/TIBO (PDB code: 1 REV). The binding affinities and the ligand efficiencies of the TIBO derivatives were well predicted by the docking algorithm used. It is worth mentioning here that the molecular docking may give rise to very different lowest energy orientations in the active site, even for very similar compounds (which though are expected to have the same binding mode). The above disparity might be due to the fact that the protein is not considered flexible in our present docking studies. Also, binding interactions upon binding

require a deeper understanding. On the basis of correlations obtained, 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 ligand efficiencies and a more favourable binding mode of TIBO derivative's top ranking compounds.

### Acknowledgements

The authors wish to thank Dr R C Saraswat and P K Sen, Shri G S Institute of Technology and Sciences, (SGSITS), Indore.

### References

- O'Meara J A, Yoakim C, Bonneau P R, Bös M, Cordingley M G, Déziel R, Doyon L, Duan J, Garneau M, Guse I, Landry S, Malenfant E, Naud J, Ogilvie W W, Thavonekham B and Simoneau B 2005 *J. Med. Chem.* **48** 5580
- Zheng Y H, Lovsin N and Peterlin B M 2005 *Immunol. Lett.* **97** 225
- Tomasselli A G and Henrikson R L 2000 *Biochem. Biophys. Acta.* **1477** 189
- Silvestri R and Maga G 2006 *Expert Opinion on Therapeutic Patents* **16** 939
- Geitmann M, Unge T and Danielson U H 2006 *J. Med. Chem.* **49** 2375
- Palella F J, Delaney K M, Moorman A C, Loveless M O, Fuhrer J, Satten G A, Ashman D J and Holmberg S D 1998 *N. Engl. J. Med.* **338** 853
- Palella F J, Chmiel J S, Moorman A C and Holmberg S D 2002 *AIDS* **16** 1617
- Tavel J A, Miller K D and Masur H 1999 *Clin. Infect. Dis.* **28** 643
- Das K, Levi P J, Hughes S H and Arnold E 2005 *Prog. Biophys. Mol. Biol.* **88** 209
- Zhou Z, Lin X and Madura J D 2006 *Infectious Disorders – Drug Targets* **6** 391
- De Clercq E 1998 *Antiviral Res.* **38** 153
- De Clercq E 2004 *Chem. Biodiv.* **1** 44
- De Clercq E 2005 *J. Med. Chem.* **48** 1297
- Kohlstaedt L A, Wang J, Friedman J M, Rice P A and Steitz T A 1992 *Science* **256** 1783
- Ren J, Esnouf R, Garman E, Somers D, Ross C, Kirby I, Keeling J, Darby G, Jones Y, Stuart D and Stammers D 1995 *Nature Struct. Biol.* **2** 293
- Ding J, Das K, Tantillo C, Zhang W, Clark A D J, Pauwels R, Moereels H, Koymans L, Janssen P A J, Smith R H J, Kroeger Koepke R, Michejda C J, Hughes S H and Arnold E 1995 *Structure* **3** 365
- Deeks S G 2001 *J. AIDS* **26** S25
- DeClercq E 2001 *Current Medicinal Chem.* **8** 1543
- Campiani G, Ramunno A, Maga G, Nacci V, Fattorusso C, Catalanotti B, Morelli E and Novellino E 2002 *Curr. Pharmaceutical Des.* **8** 615

20. Mager PP 1997 *Med. Res. Rev.* **17** 235
21. Tantillo C, Ding J, Jacobo-Molina A, Nanni R G, Boyer P L, Hughes S H, Pauwels R, Andries K, Janssen P A and Arnold E 1994 *J. Mol. Biol.* **243** 369
22. Sapre N S, Pancholi N, Gupta S and Sapre N 2008 *J. Comp. Chem.* DOI 10.1002/jcc 20931
23. Sapre N S, Pancholi N, Gupta S and Sikarwar A 2007 *Acta Chim. Slov.* **54** 797
24. Sapre N S, Pancholi N, Gupta S, Sikarwar A and Sapre N 2007 *J. Chem. Sci.* **119** 625
25. Barreca M L, Rao A, De Luca L, Zappala M, Monforte A M, Maga G, Pannecouque C, Balzarini J, De Clercq E, Chimirri A and Monforte P 2005 *J. Med. Chem.* **48** 3433
26. D'Cruz O J and Uckun F M 2006 *J. Antimicrob. Chemother.* **57** 411
27. Pauwels R, Andries K, Desmyter J, Schols D, Kukla M J, Breslin H J, Raeymaeckers A, Van Gelder J, Woestenborghs R, Heykants J, Schellekens K, Janssen M A C, De Clercq E and Janssen P A J 1990 *Nature* **343** 47
28. Barre-Sinoussi F, Chermann J C, Rey F, Nugeyre M T, Chamaret S, Gruest J, Dautet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W and Montagnier L 1983 *Science* **220** 868
29. Gallo R C et al 1984 *Science* **224** 500
30. Ren J, Esnouf R, Hopkins A, Ross C, Jones Y, Stammers D and Stuart D 1995 *Structure* **3** 915
31. Zhou Z and Madura J D 2004 *J. Chem. Inf. Comput. Sci.* **44** 2167
32. Bahram H, Mohammad Hossein Tabaei S and Fatehmeh N 2005 *J. Mol. Struct. Theochem.* **732** 39
33. De Clercq E 1995 *Clin. Microbiol. Rev.* **8** 200
34. Das K, Ding J, Hsiou Y, Clark, A J, Moereels H, Koymans L, Andries K, Pauwels R, Janssen P, Boyer P, Clark P, Smith R J, Kroeger S M, Michejda C, Hughes S and Arnold E 1996 *J. Mol. Biol.* **264** 1085
35. Ding J, Das K, Moereels H, Koymans L, Andries K, Janssen P A J, Hughes S and Arnold E 1995 *Nature Struct. Mol. Biol.* **2** 407
36. Saen-oon S, Kuno M and Hannongbua S 2005 *Proteins* **61** 859
37. Wang J, Morin P, Wang W and Kollman P A 2001 *J. Am. Chem. Soc.* **123** 5221
38. Pan Y, Huang N, Cho S and MacKerell A D 2003 *J. Chem. Inf. Comput. Sci.* **43** 267
39. Bender A, Mussa H Y and Glen R C 2004 *J. Chem. Inf. Comput. Sci.* **44** 1708
40. ChemDraw Ultra 7.0.0 (www.cambridgesoft.com)
41. <http://www.molegro.com> (free trial version)
42. Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E 2000 *Nucl. Acids Res.* **28** 235
43. Thomsen R and Christensen M H 2006 *J. Med. Chem.* **49** 3315
44. Storn R and Price K 1995 *Differential evolution – A simple and efficient adaptive scheme for global optimization over continuous spaces*; Technical Report; International Computer Science Institute: Berkley, CA
45. Thomsen R 2003 *Proceedings of the 2003 Congress on evolutionary computation* **4** 2354
46. Gehlhaar D K, Verkhivker G, Rejto P A, Fogel D B, Fogel L J and Freer S T 1995 *Proceedings of the fourth international conference on evolutionary programming* 615
47. Gehlhaar D K, Bouzida D and Rejto P 1998 *Proceedings of the seventh international conference on evolutionary programming* 449
48. Yang J M and Chen C C 2004 *Proteins* **55** 288
49. Michalwicz Z and Fogel D B 2000 *How to solve it: Modern Heuristics* (Berlin: Springer-Verlag)
50. Sapre N S, Gupta S, Pancholi N and Sapre N 2007 *J. Comput. Aided Mol. Des.* **22** 69