J Comput Aided Mol Des (2008) 22:69–80
DOI 10.1007/s10822-007-9161-8

Molecular docking studies on tetrahydroimidazo-[4,5,1-jk] [1,4]-benzodiazepinone (TIBO) derivatives as HIV-1 NNRT inhibitors

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

Received: 30 June 2007 / Accepted: 3 December 2007 / Published online: 28 December 2007

Abstract At present, chemotherapy seems to be the main weapon in the arsenal of remedies for the ongoing crusade against AIDS. The mode of binding of the TIBO family of inhibitors has been of interest because these compounds do not fit the two-hinged-ring model as generally observed in the NNRTIs. Flexible docking simulations were performed with a series of 53 TIBO derivatives as NNRTIs. Binding preferences as well as the structural and energetic factors associated with them were studied. A good correlation ($r^2 = 0.849$, $q^2 = 0.843$) was observed between the biological activity and binding affinity of the compounds which suggest that the identified binding conformations of these inhibitors are reliable. Further screening of PubChem database yielded novel scaffolds. Our studies suggest that modifications to the TIBO group of inhibitors might enhance their binding efficacy and hence, potentially, their therapeutic utility.

N. S. Sapre (✉) · N. Pancholi
Department of Applied Chemistry, Shri GS Institute of Technology and Sciences, 3/6 Manoramaganj, Indore 452001, MP, India
e-mails: sukusap@yahoo.com; nsapre@sgsits.ac.in

S. Gupta
Department of Chemistry, Govt. P.G. College, MHOW, MP, India
e-mail: swagataagupta@yahoo.co.in

N. Pancholi
e-mail: nilanjanakpancholi@yahoo.com

N. Sapre
Department of Computer Science, SD Bansal College of Technology, Indore, MP, India
e-mail: neelimasapre@yahoo.com

Keywords TIBO · Structure based drug design (SBDD) · HIV-1 non-nucleoside reverse transcriptase inhibitors (HIV-1 NNRTIs) · pIC_{50} · Multiple linear regression (MLR) · Artificial neural network (ANN)

Introduction

Though the modern research and techniques have lead to tremendous advancement in our understanding of the HIV-1 life cycle and the molecular mechanisms underlying its pathogenesis, yet it has not been possible to fully control the HIV-1 infection for long period of time, not to mention eradicate it [1, 2]. The introduction of Highly Active Antiretroviral Therapy (HAART) has a dramatically positive effect on the natural history of HIV-1 disease in the developed world [3, 4]. However, incomplete suppression leading to drug resistance has often impaired the HAART efficacy. Thus, an approach to the issue of rapid emergence of NNRTI resistance could be attempted by providing potent inhibitors in order to inhibit not only the wild type HIV-1, but also pre-existing resistant viral variants found at low levels, as a result of de novo mutations during ongoing virus replication, or at high levels, as a consequence of NNRTI treatment failure.

Non-nucleoside HIV-1 reverse transcriptase inhibitors (NNRTIs) are a structurally diverse group of 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 qualify as an NNRTI, a compound should interact specifically with an allosteric binding site, which is physically separated from the catalytic domain or the substrate binding site, of the HIV-1 RT, but not HIV-2 (or

any other retrovirus) at a concentration that is significantly lower than the concentration required to affect normal cell viability [5–8]. Pioneering structural analysis, followed by many other experimental evidences, have demonstrated that all NNRTIs, independently of their structure, bind to the hydrophobic pocket, located in the p66 subunit, approximately 10 Å from the polymerase binding site [9–11]. Resistance to NNRTIs develop rapidly in the clinical setting, primarily by altering the binding of NNRTIs to RT [12]. Structural and computational approaches suggest that most mutations are such that larger side chain amino acids are replaced with a smaller side chain [13]. Also, the torsional flexibility ("wiggling") of the inhibitors can generate numerous conformational variants and the compactness of the inhibitors permit significant repositioning and reorientation (translation and rotation) within the pocket ("jiggling") [14].

Along with nucleoside reverse transcriptase inhibitors (NRTIs) and protease inhibitors (PIs), the nonnucleoside reverse transcriptase inhibitors (NNRTIs) have gained a definitive and important place in the treatment of HIV-1 infections due to their antiviral potency, high specificity, low toxicity, and are in rapid developmental phase [15]. The rapid emergence of drug-resistant HIV-1 strains induced by the randomly screened first generation drugs has lead to a broad spectrum of second generation NNRTIs based on rational-based drug design and pharmacokinetic evaluations, which may further lead to the future development of a third generation of NNRTIs [16].

Various computational strategies, such as quantum mechanical applications, QSPR, 2D QSAR/3D QSAR and structure based drug design (SBDD) approach have been used to study the ligand properties and their interaction with the receptor. Hannoongbua et al. [17, 18] have used electronic and molecular properties of the inhibitors as structural descriptors obtained from quantum chemical calculations for studying TIBO derivatives. A detailed QSAR study of TIBO compounds along with other anti-HIV compounds is presented in a review by Hansch et al. [19] 3D-QSAR models based on CoMFA and CoMSIA have been used to predict the steric and electrostatic properties of the binding structure of the TIBO complexes [20]. Molecular docking methods based on binding affinity predictions provide a deeper insight into the binding site structure elucidation. Molecular docking, as an optimization problem, has played an important role in the understanding of drug/receptor interactions. Recently, Barreca et al. [21] have performed successful docking studies with a three-dimensional common feature pharmacophore model involving benzimidazol-2-one system as a scaffold using the X-ray structure of RT/non-nucleoside inhibitor (NNRTI) complexes. Rationally designed NNRTIs deduced from 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 robust anti-HIV activity against the wild-type and drug-escaped mutants without cytotoxicity [22].

There is a set of compounds commonly referred to as TIBO (Tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepinone) derivatives developed by Pauwels et al. [23] which act as NNRTIs, have gained a prominent place in the treatment of HIV-1 infections. The mode of binding of the TIBO family of inhibitors is of interest because these compounds do not fit the two-hinged-ring model, like other NNRTIs. Although Cl-TIBO is chemically very different from other NNIs, it has achieved remarkable spatial equivalence and shape complementarity with other NNIs on binding to reverse transcriptase. Comparison of the different RT-NNI complexes suggest modifications to the TIBO group of inhibitors which might enhance their binding and hence, potentially, their therapeutic efficacy [24, 25]. A TIBO compound, Tivirapine has already entered into the clinical trials [26]. Till now several crystal structures of TIBO/RT complexes have been solved [10, 24, 27, 28]. These structures have provided valuable insight into the binding orientations and specific interactions of TIBOs in the binding pocket, but the binding complexes of many other TIBOs are not yet known. Although it is assumed that these structurally similar TIBOs bind in a similar orientation to RT, further exploration of the binding structure of other TIBOs is necessary to understand the mechanism of inhibition and to aid in the design of more potent inhibitors that are effective against RT mutants which emerge when HIV is exposed to a NNRTI. Some new TIBO-like derivatives have been designed by contraction of the β-ring (diazepine ring) of TIBO derivatives from 7- to 6-membered ring by Bahram et al. [29]. The total free energy of docking of these compounds indicated that some of these derivatives were bonded to the receptor stronger than the 7-membered derivatives. Recently, binding energy analysis for wild-type and Y181C mutant HIV-1 RT/8-Cl TIBO complex structures using quantum chemical calculations based on the ONIOM method has been reported [30]. In the present docking studies performed using MolDock, we have been able to achieve better alignment with the crystal coordinates in terms of RMSD (0.269 Å) as compared to 0.808 Å by Autodock3 [31], 0.590 Å by FlexX [32] and 0.590 Å by MATADOR [33]. Based on this, we were able to derive robust predictive models using MLR as well as ANN subsequently used as a virtual screen for determining potential hit candidates from PubChem compound database [34].

Computational methods

Molecular structures

TIBO or tetrahydroimidazo-[4, 5,1-jk][1,4]-benzodiazepinone and its derivatives developed by Pauwels et al. [23] along with their biological activities were taken for docking studies. The molecular structures were drawn and geometrically optimized with MM2 force field using chemdraw and then exported to MolDock where they were further prepared along with the proteins by the docking engine. The molecules were prepared by assigning bonds, bond orders, hybridization, charges, creating explicit hydrogens and flexible torsions in the ligand. The bond orders, the number of hydrogens attached to the atoms, and their hybridization were recognized and also the aromatic rings were detected. The charges were set according to the following scheme. The charge of the ligand atoms were set as follows: 0.5 charge for N atoms in $–C(NH_2)_2$ HIS (ND1/NE2), 1.0 charge for N atoms in $–N(CH_3)_2$, $–(NH_3)$, −0.5 charge for O atoms in –COO, $–SO_4$, $–PO_2$, $–PO_2–$, −0.66 charge for O atoms in $–PO_3$, −0.33 charge for O atoms in $–SO_3$, −1.0 charge for N atoms in $–SO_2NH$. The charge templates of the protein atoms were set as follows: 0.5 charge for N atoms in ARG (NH1/NH2), 1.0 charge for N atoms in LYS (N), −0.5 charge for O atoms in ASP (OD1/OD2) and GLU (OE1/OE2). The structure of HIV-1 RT protein (PDB code 1REV) was obtained from Protein Data Bank [Research Collaboratory for Structural Bioinformatics(RCSB) [35]. PubChem compound database (http://www.pubchem.ncbi.nlm.nih.gov) was used to find putative novel hit compounds.

Docking simulations

MolDock software is based on a new heuristic search algorithm that combines differential evolution with a cavity prediction algorithm [36]. The guided differential evolution algorithm incorporates differential evolution optimization technique along with a cavity prediction algorithm. Previously, differential evolution (DE) has been successfully applied to molecular docking [37, 38]. Fast and accurate identification of potential binding modes (poses) can be achieved using the predicted cavities during the search process. The MolDock docking scoring function is based on a piecewise linear potential (PLP) introduced by Gehlhaar et al. [39, 40] and further extended in GEMDOCK by Yang and Chen [41]. Here, the docking scoring function takes hydrogen bond directionality into account. E_{PLP} is a piecewise linear potential using two different sets of parameters: one for approximating the steric (van der Waals) term between atoms and the other stronger potential for hydrogen bonds. Moreover, a re-ranking procedure has been applied to the highest ranked poses to further increase the docking accuracy. On an average, 10 docking runs were made to obtain a high docking accuracy. Only the ligand properties are represented in the individuals, because the protein remained rigid during the docking process. MolDock automatically identifies potential binding sites (cavities) using a flexible cavity detection algorithm as there is no dependence on the orientation of the target molecule and an arbitrary number of directions can be used. The fitness of a candidate solution is derived from the docking scoring function, E_{score} and is defined by the following energy terms

$$E_{score} = E_{inter} + E_{intra} \tag{1}$$

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

$$E_{inter} = \sum_{i \in ligand} \sum_{j \in protein} \left[E_{PLP}(r_{ij}) + 332.0 \frac{q_i q_j}{4 r_{ij}^2} \right] \tag{2}$$

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

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

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

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, parametrized according to the hybridization types of the bonded atoms. θ is the torsional angle of the bond. The last term, E_{clash}, assigns a penalty of 1,000 if the distance between two atoms (more than two bonds apart) is less than 2.0 Å. Thus, the E_{clash} term punishes the infeasible ligand conformations.

Hardware and software

Molegrow Virtual Docker 2007.2.2 [42] was run on a Windows XP based pentium IV 2.66 GHz PC (with 512 MB of memory).

Results and discussion

Our specific aims for the present study were threefold: Firstly, to estimate the relative binding affinities in the context of structure-based drug design (SBDD) using

available data, secondly, to understand the variations in binding affinities through interpretation of results related to structural and energetic correlations, as derived from the simulations. Thirdly, using the best model obtained as a virtual screen to navigate a large chemical database and retrieve probable potential hit candidates.

The chemical structures of 53 TIBO derivatives are given in Table 1 along with their biological activities, 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. It also records binding affinity ($E_{binding}$) in kJ/mol, MolDock score and a more stringent rerank score (both in arbitrary energy units).

The enzyme is a heterodimer (p66/p51), with domains labeled "fingers", "thumb", "palm" and "connection" in both subunits, and a ribonuclease H domain in the larger subunit only. The binding of NNRTIs occur at a catalytic and allosteric site, which has a inner hydrophobic core stacked mainly with TYR181, TYR188 amino acid residues. The other amino acids involved in the hydrophobic interactions with NNRTIs are VAL106, VAL108, PHE227, TRP229 and GLU138 (P51 subunit). Amino acid residues VAL179 and LYS101 are associated with hydrogen-bonding interactions [43, 44]. Most of the NNRTI inhibitors assume a similar butterfly-like shape and bind to HIV-1 RT in a very similar manner. Important differences occur in the conformation of amino acid residues that form the binding pocket [28]. To date, a number of crystal structures of unaligned wild type and mutant HIV-RT varieties complexed with different inhibitors are available. These crystal structures provide valuable clues for designing new inhibitors by providing deep knowledge about the pattern of interaction in the binding pocket [21]. Thus, it is specifically possible to delineate the amino acid residues involved in different interactions. Since we are dealing with TIBO derivatives, as a natural choice, we have selected 9-Cl-TIBO/HIV-RT complex (PDB CODE-1REV) for our present studies.

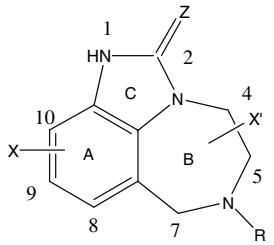
Validation of the docking method

To validate the docking protocol, 9Cl-TIBO from the complex (1REV) was 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). They were prepared as per the algorithm of the docking engine by assigning bond orders, charges and flexible torsions. Bonds of the ligand were checked to ascertain the flexibility. The protein was protonated and sidechains were minimized using Nedler–Mead simplex algorithm. Potential binding sites were identified using the built-in cavity detection algorithm (MolDock optimizer). Search space was allocated and docking was performed. The cavities found by the cavity detection algorithm were actively used by the search algorithm (guided differential evolution) to focus the search during the docking simulation. The following parameters were fixed: No. of runs = 10, population size = 50, crossover rate = 0.9, scaling factor = 0.5, and max iterations = 2,000. It was followed by pose clustering so that the best-scoring pose was not missed out. Five top scoring poses were returned. Fitness evaluation was done by MolDock [Grid] scoring functions (grid resolution = 0.30) based on PLP and GEMDOCK and were again reranked to increase the docking accuracy. The encouraging root mean square deviation (RMSD) of 0.269 Å between the docked and crystal ligand coordinates indicate a very good alignment of the experimental and calculated positions (Out of the five best poses, four were well below 1.3 Å RMSD). Figure 1 shows the best fit redocked coordinates with respect to the crystal structure. Therefore, this docking simulation protocol is used to dock all the other TIBOs into RT to find their binding structures and predict their relative affinities. Several favorable interactions between the ligand and the receptor were clearly observed; one hydrogen bond formed between the five member ring of ligand and the backbone of LYS101 residue of the receptor binding site. The favorable electrostatic interaction between LYS101, LYS103, and the nitrogen atoms in the imidazolone ring; and the favorable hydrophobic interaction between TYR181, PRO95, PHE227, LEU234, VAL106, TYR188, and the DMA fragment of the ligand at N6 position of the diazepine ring.

Docking of the molecule set

After the validation of the docking method using 9Cl-TIBO, a dataset of 53 molecules belonging to TIBO derivatives with varied activity range (4.0–8.52) were docked into the same coordinates of the crystal structure. The same docking protocol was used for all the docking calculations. The docked 3D-structures of TIBO derivatives were scored, reranked and then compared with the X-ray crystallographic structure of 9-Cl-TIBO. The beauty of this series was that all the compounds share a basic common pattern. The result demonstrates that the docking simulation can dock these TIBO inhibitors back into the NNRTI allosteric binding site very well as per expectation that similar compounds bind in a similar way. Although, most of the compounds prefer one single binding position

Table 1 Anti-HIV-1 activity (inhibitory concentration pIC_{50}), rerank scores and binding affinity ($E_{binding}$ in kJ/mol) of tetrahydroimidazo-[4, 5, 1-jk][1,4]-benzodiazepinone (TIBO derivatives)



No.	X	Z	R	X′	pIC_{50}^{a}	Binding affinity[b]	Moldock score[c]	Reranking score[c]
1	8-F	S	DMA[d]	5-Me(S)	8.24	−24.9082	−125.000	−100.159
2	8-SMe	S	DMA	5-Me(S)	8.30	−24.0819	−126.229	−84.775
3	8-OMe	S	DMA	5-Me(S)	7.47	−23.7163	−126.510	−100.800
4	8-OC_2H_5	S	DMA	5-Me(S)	7.02	−23.3485	−124.217	−89.577
5	8-$CONH_2$	O	DMA	5-Me(S)	5.20	−22.0340	−133.014	−96.857
6	8-Br	O	DMA	5-Me(S)	7.33	−24.0970	−117.451	−90.643
7	8-Br	S	DMA	5-Me(S)	8.52	−25.2689	−122.608	−96.456
8	8-I	O	DMA	5-Me(S)	7.06	−24.2886	−122.586	−86.421
9	8-I	S	DMA	5-Me(S)	7.32	−24.8787	−120.200	−69.297
10	8C=–CH	O	DMA	5-Me(S)	6.36	−22.5076	−134.635	−107.130
11	8C=–CH	S	DMA	5-Me(S)	7.53	−22.0992	−130.437	−102.936
12	8-CH_3	O	DMA	5-Me(S)	6.00	−21.8323	−124.364	−100.164
13	8-$N(CH_3)_2$	O	CPM[e]	5-Me(S)	5.18	−20.3497	−126.740	−90.264
14	9-NH_2	O	CPM	5-Me(S)	4.22	−20.6174	−115.112	−76.010
15	9-F	S	DMA	5-Me(S)	7.60	−24.8918	−130.513	−102.772
16	9-Me	O	DEA[f]	5-Me(S)	6.50	−23.0641	−110.884	41.017
17	H	O	$CH_2CH_2CH{=}CH_2$	5-Me(S)	4.30	−20.5919	−109.298	−88.001
18	H	O	$CH_2CH_2CH_3$	5-Me(S)	4.05	−20.3046	−109.965	−77.689
19	H	O	2-MA[S(+)]	5-Me(S)	4.72	−21.0433	−108.793	−82.047
20	H	O	CPM	5-Me(S)	4.36	−20.1291	−101.478	−71.925
21	H	O	$CH_2CH{=}CHMe(Z)$	5-Me(S)	4.46	−20.7922	−106.435	−84.223
22	H	O	$CH_2CH_2CH_2CH_3$	5-Me(S)	4.00	−19.8649	−115.139	−81.201
23	H	O	DMA	5-Me(S)	4.90	−20.8096	−117.129	−88.638
24	H	O	$CH_2C(Me){=}CHMe(E)$	5-Me(S)	4.54	−21.1422	−117.786	−87.795
25	H	O	DMA[R(+)]	5-Me(S)	4.66	−20.7882	−119.291	−82.609
26	H	O	DMA[S(+)]	5-Me(S)	5.40	−21.5057	−116.421	−58.463
27	H	O	$CH_2C(C_2H_5){=}CH_2$	5-Me(S)	4.43	−21.5001	−124.061	−87.929
28	8-Cl	S	DMA	H	7.34	−24.6669	−118.465	−93.413
29	H	O	2-MA[g]	5,5-di-Me	4.64	−20.5888	−109.719	−69.346
30	H	O	2-MA	4-Me	4.50	−20.5378	−115.374	−94.688
31	H	O	C_3H_7	4-$CHMe_2$	4.13	−19.4336	−116.363	−91.484
32	H	O	2-MA	4-$CHMe_2$	4.90	−20.3312	−128.392	−106.378
33	H	O	2-MA	4-C_3H_7	4.32	−20.1040	−129.831	−95.152
34	H	O	DMA	7-Me	4.92	−20.7655	−121.453	−90.577
35	8-Cl	O	DMA	7-Me	6.84	−23.4125	−125.338	−86.005
36	9-Cl	O	DMA	7-Me	6.80	−23.9026	−120.708	−87.949
37	H	S	C_3H_7	7-Me	5.61	−20.6850	−110.045	−35.973
38	8-Cl	S	DMA	7-Me	7.92	−24.3636	−121.443	−93.671

Table 1 continued

No.	X	Z	R	X′	pIC_{50}^{a}	Binding affinity[b]	Moldock score[c]	Reranking score[c]
39	9-Cl	S	DMA	7-Me	7.64	−24.4459	−116.505	−39.006
40	H	O	DMA	4,5-di-Me(cis)	4.25	−20.8737	−120.301	−93.203
41	H	S	DMA	4,5-di-Me(cis)	5.65	−22.0467	−121.115	−91.587
42	H	S	CPM	4,5-di-Me(trans)	4.87	−20.4605	−99.553	17.814
43	H	S	DMA	4,5-di-Me(trans)	4.84	−20.5012	−124.931	−28.077
44	9-Cl	O	DMA	5,7-di-Me(R,R-trans)	6.64	−24.0043	−141.084	−112.370
45	9-Cl	S	DMA	5,7-di-Me(R,R-trans)	6.32	−23.5978	−141.328	−112.993
46	H	S	DMA	4,7-di-Me(trans)	4.59	−20.4486	−133.811	−109.217
47	9-Cl	O	DMA	5-Me(S)	6.74	−21.5528	−119.268	−83.007
48	H	O	C_3H_7	5-Me(S)	4.22	−20.0591	−108.315	−78.349
49	H	S	C_3H_7	5-Me(S)	5.78	−20.7547	−106.696	−85.829
50	H	O	2-MA	5-Me(S)	4.46	−20.6721	−118.379	−92.302
51	H	O	DMA	5-Me(S)	5.48	−21.0160	−117.593	−81.798
52	9-Cl	S	CPM	5-Me(S)	7.47[h]	−20.3021	−114.596	−93.299
53	H	S	CPM	5-Me(S)	7.22[h]	−21.0228	−105.660	−87.392

[a] $pIC^{50} = -\log IC_{50}$ (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)

[b] $E_{binding}$ in kJ/mol

[c] Ranking in arbitrary energy units

[d] DMA = 3,3-dimethylallyl

[e] CPM = cyclopropylmethyl

[f] DEA = 3,3-diethylallyl

[g] 2-MA = 2-methylallyl

[h] Data points not included

deeply buried in the allosteric binding pocket multiple binding orientations were also observed. The most interesting observation was that two major clusters of binding conformations were observed occupying two 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. 2a). The other major cluster observed can rather be referred to as 'flip fit', which occupied an overlapping binding orientation representing a flip with respect to the imidazolone ring (Fig. 2b). Thus there is no apparent binding difference between highly active and lesser active ligands. Therefore, the binding position cannot specifically differentiate the activity.

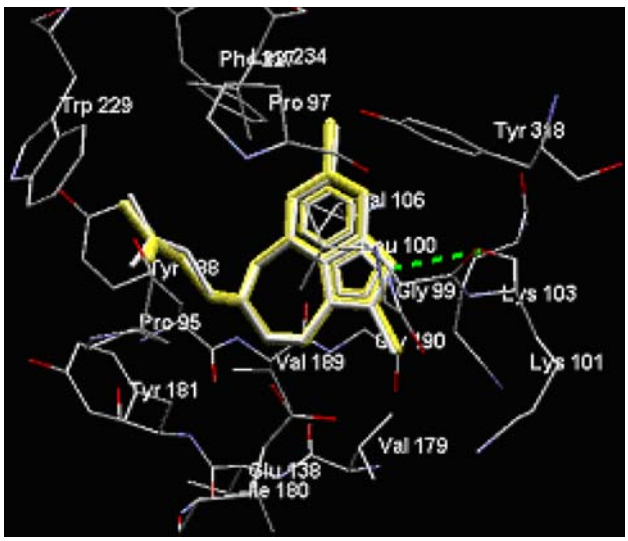


Fig. 1 Conformation of 9-Cl TIBO crystal structure [PDB code: 1REV] (*yellow*) as compared to redocked conformation of 9-Cl TIBO (*white*)

As the TIBO compounds have the same basic structure, as expected, similar interactions are observed. Hydrogen bonding interactions (1 or 2) were observed between the imidazolone ring of ligand and the LYS101 residue (Fig. 3a). The TIBO compounds seemed to be stabilized by extensive hydrophobic interactions (Fig. 3b). The major contributor to the hydrophobicity seemed to be the diazepine ring and the group attached to it at 6C. Among the groups attached to the diazepine ring, DMA seemed to be the most hydrophobic in orientation and perhaps that is the reason of high activity of compounds with DMA group. The amino acid residues involved in the hydrophobic interactions with diazepine ring and 6N were TYR181, TYR185, TYR229, LEU234, VAL106, GLY190, VAL189

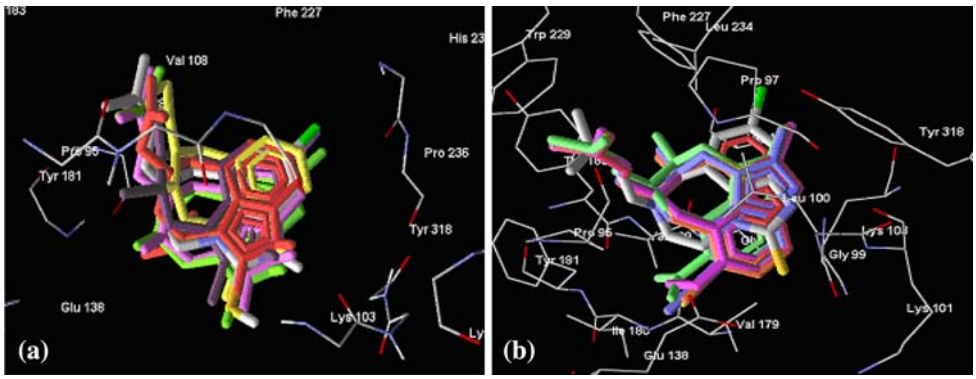


Fig. 2 (**a**) Docked conformation of 9-Cl TIBO crystal structure (*CPK*) with five best fit TIBO derivatives. Compound no. 32 (*red*), 45 (*pink*), 46 (*green*), 52 (*purple*), 53 (*yellow*). (**b**) Docked conformation of 9-Cl TIBO crystal structure (*CPK*) with six flip fit TIBO derivatives. Compound no. 3 (*purple*), 5 (*green*), 7 (*orange*), 8 (*yellow*), 10 (*blue*), 12 (*pink*)

and PRO95. The aromatic moieties of the amino acid residues converge at the inner side of the binding cavity. Comparatively, the 6-member and the 5-member ring are less hydrophobic in nature. The hydrophobic interactions involved with the phenyl ring are PHE227, TYR318 and LEU234. The imidazolone ring seemed to interact hydrophobically with LYS101, LYS103. The surface map of the binding pocket also showed favorable electrostatic interactions with the ligand (Fig. 3c). 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 was played by an aromatic moiety present in the structure. As the contour map exhibiting blue color indicated more positive charge i.e. a nucleophilic attachment would be favorable for electronic interactions. The electrostatic interaction was stabilized by stacking type interaction with another aromatic ring located at the receptor site, such as TYR181 and TYR188.

Correlation between binding affinity and activity

An important application of MolDock is to predict the binding affinity and interaction energies along with the binding conformation of an inhibitor with the receptor. The binding affinity is supposed to be the best choice while identifying the best binder to a given target in a varied ligand dataset. The binding affinity (kJ/mol) of a given 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 −19.43 to −25.27 kJ/mol. The scoring reliability of MolDock was ascertained by correlating binding affinities with the biological activity of the TIBO derivatives. Various models were created using multiple linear regression technique (MLR) and artificial neural networks (ANN) using in-built data analyzer. 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 autogenerated random seed used in the model training was 1265312701. 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, Spearman rank correlation coefficient (rho), MSE is the mean squared error and PRESS is the predictive sum of squares.

Multiple linear regression technique (MLR)

The best model relating biological activity with the binding affinity was derived using MLR (using LOO) and presented below:

$$Activity(\text{pIC}_{50}) = -0.734952(\pm 0.923)E_{\text{binding}} - 10.3891 \quad (4)$$

($n = 51$, $r = 0.922$, $r^2 = 0.849$, $r^2_{\text{adj}} = 0.840$, Spearman (rho) $= 0.870$, $q^2 = 0.843$, MSE $= 0.288$, PRESS $= 14.70$).

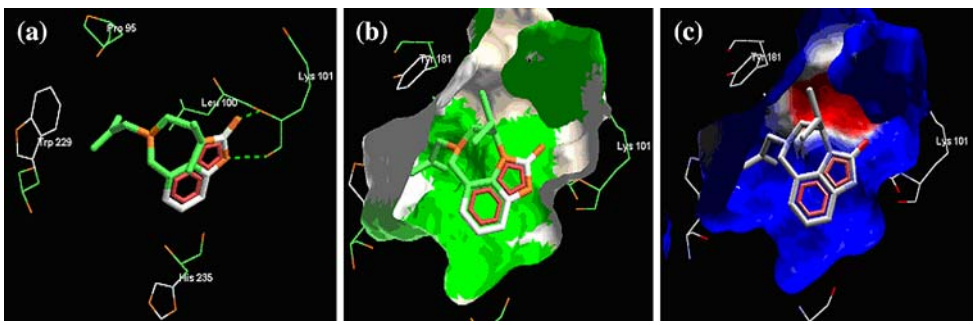


Fig. 3 (**a**) Hydrogen bond interactions (shown by dotted lines), hydrophobic interactions (**b**) and electrostatic interactions (**c**) as exhibited by Compound no. 32 (of average activity)

The normalization of the dataset using the unit variance scaling (UVS) as well as mean centering (MC) did not lead to any significant improvement ($r^2 = 0.852$), but gave a lower value of MSE (0.144). Also results derived from MLR (N-CV) provided similar results ($r^2 = 0.845$ for $N = 10$ and $r^2 = 0.843$ for $N = 5$). A model was also created using three other descriptors (E_{steric}, $E_{torsion}$ and $E_{VdW(LJ12-6)}$) along with $E_{binding}$ ($r^2 = 0.856$). As the coefficient relevance of the aforementioned three descriptors was not significant as compared to $E_{binding}$, the model is not presented here.

Artificial neural networks (ANN)

The algorithm used for training NN model was back-propagation method. Using the same random seed, following parameters were fixed: max training epoch = 2,000, learning rate = 0.50, output learning rate = 0.50, momentum = 0.20, no. of neurons in the first hidden layer = 10, no. of neurons in the second hidden layer = 10, initial weight (±) = 0.90. A slightly better prediction was obtained ($r^2 = 0.855$, Spearman (rho) = 0.882 and MSE = 0.267). Altering the above parameters of neural network did not yield any significant change in the value of r^2.

The above statistical analyses suggest the robustness of the docking procedure adopted herein and the results obtained can be validated by Fig. 4 (correlating biological activity with the binding affinity) and Fig. 5 (a good fit was observed between experimental and predicted activity, MLR as well as ANN).

Table 2 records the observed and the calculated (from MLR as well as ANN) values of pIC_{50} for the training set of TIBO derivatives. The quality of correlation is demonstrated by their respective residual values i.e. the difference between experimental and predicted pIC_{50}.

A very interesting exception was observed in the case of two TIBO compounds (Fig. 2a, no. 52 (purple) and 53 (yellow) with activities 7.22 and 7.47 respectively).

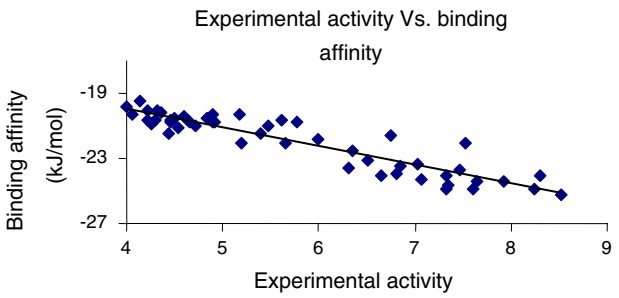


Fig. 4 Graph between experimental activity (pIC_{50}) and binding affinity ($E_{binding}$ in kJ/mol)

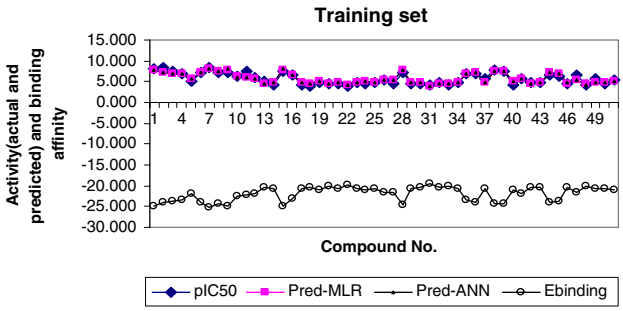


Fig. 5 Graph showing comparison between experimental pIC_{50}, predicted pIC_{50} (MLR and ANN) and the binding affinity ($E_{binding}$ in kJ/mol) of TIBO compounds

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 the binding interaction, which might be related to protein flexibility or some modified interaction upon binding. Inclusion of these two compounds in the regression model gave a comparatively lower r^2 (0.702) value and therefore they were not considered for regression analysis.

Predicting probable novel hit compounds

PubChem compound database (http://www.pubchem.ncbi.nlm.nih.gov), a huge chemical dataset containing more than ten million compounds was searched. A similar structure based search could retrieve 158 compounds. The template used as a query to conduct the search is given in Fig. 7. Further screening following Lipinsky's rule and elimination of compounds with unwanted functional groups (–NO_2, –CHO) yielded 128 compounds. These compounds were finally docked, their respective binding affinities evaluated and their activity predicted. Tivirapine (R-86183), a potent TIBO drug was also docked into the same binding coordinates ($E_{binding} = -25.039$ kJ/mol). On the basis of binding affinities obtained, finally compounds with $E_{binding} < -24.0$ kJ/mol were selected. This yielded 20 compounds ($E_{binding} = -24.023$ to -36.768 kJ/mol and $pIC_{50} > 7.27$), which can serve as promising hit candidates. Table 3 presents the binding affinity and predicted activity of 20 hit compounds from the PubChem compound database. Figure 6 shows the plot of binding affinity and predicted activity (MLR as well as ANN) of the selected hit compounds from the PubChem compound database (Fig. 7).

It is worth mentioning here that 13 out of the 20 compounds had better binding affinities as compared to Tivirapine. The compounds similar to TIBO derivatives specifically had halogen/halogens attached at 8 or 9 position and a bulky chain or a phenyl group either directly

Table 2 Experimental, predicted and residuals values of pIC_{50} and binding affinities (kJ/mol) of TIBO derivatives

No.	Obs. activity	$E_{binding}$ (kJ/mol)	Pred-MLR	Residual (MLR)	Pred-ANN	Residual (ANN)
1	8.240	−24.908	7.917	0.323	7.865	−0.375
2	8.300	−24.082	7.310	0.990	7.235	−1.065
3	7.470	−23.716	7.041	0.429	6.973	−0.497
4	7.020	−23.349	6.771	0.249	6.759	−0.261
5	5.200	−22.034	5.805	−0.605	5.804	0.604
6	7.330	−24.097	7.321	0.009	7.304	−0.026
7	8.520	−25.269	8.182	0.338	8.124	−0.396
8	7.060	−24.289	7.462	−0.402	7.491	0.431
9	7.320	−24.879	7.896	−0.576	7.969	0.649
10	6.360	−22.508	6.153	0.207	6.149	−0.211
11	7.530	−22.099	5.853	1.677	5.874	−1.656
12	6.000	−21.832	5.657	0.343	5.637	−0.363
13	5.180	−20.350	4.567	0.613	4.565	−0.615
14	4.220	−20.617	4.764	−0.544	4.767	0.547
15	7.600	−24.892	7.905	−0.305	7.979	0.379
16	6.500	−23.064	6.562	−0.062	6.507	0.007
17	4.300	−20.592	4.745	−0.445	4.752	0.452
18	4.050	−20.305	4.534	−0.484	4.550	0.500
19	4.720	−21.043	5.077	−0.357	5.080	0.360
20	4.360	−20.129	4.405	−0.045	4.376	0.016
21	4.460	−20.792	4.892	−0.432	4.910	0.450
22	4.000	−19.865	4.211	−0.211	4.216	0.216
23	4.900	−20.810	4.905	−0.005	4.923	0.023
24	4.540	−21.142	5.149	−0.609	5.165	0.625
25	4.660	−20.788	4.889	−0.229	4.891	0.231
26	5.400	−21.506	5.417	−0.017	5.392	−0.008
27	4.430	−21.500	5.412	−0.982	5.428	0.998
28	7.340	−24.667	7.740	−0.400	7.762	0.422
29	4.640	−20.589	4.743	−0.103	4.737	0.097
30	4.500	−20.538	4.705	−0.205	4.713	0.213
31	4.130	−19.434	3.894	0.236	3.897	−0.233
32	4.900	−20.331	4.553	0.347	4.583	−0.317
33	4.320	−20.104	4.386	−0.066	4.395	0.075
34	4.920	−20.766	4.873	0.047	4.885	−0.035
35	6.840	−23.413	6.818	0.022	6.823	−0.017
36	6.800	−23.903	7.178	−0.378	7.189	0.389
37	5.610	−20.685	4.813	0.797	4.817	−0.793
38	7.920	−24.364	7.517	0.403	7.535	−0.385
39	7.640	−24.446	7.578	0.062	7.534	−0.106
40	4.250	−20.874	4.952	−0.702	4.956	0.706
41	5.650	−22.047	5.814	−0.164	5.791	0.141
42	4.870	−20.461	4.648	0.222	4.645	−0.225
43	4.840	−20.501	4.678	0.162	4.686	−0.154
44	6.640	−24.004	7.253	−0.613	7.291	0.651
45	6.320	−23.598	6.954	−0.634	6.988	0.668
46	4.590	−20.449	4.640	−0.050	4.655	0.065
47	6.740	−21.553	5.451	1.289	5.427	−1.313
48	4.220	−20.059	4.353	−0.133	4.349	0.129

Table 2 continued

No.	Obs. activity	$E_{binding}$ (kJ/mol)	Pred-MLR	Residual (MLR)	Pred-ANN	Residual (ANN)
49	5.780	−20.755	4.865	0.915	4.830	−0.950
50	4.460	−20.672	4.804	−0.344	4.813	0.353
51	5.480	−21.016	5.057	0.423	5.063	−0.417

Table 3 Binding affinities and predicted values of pIC_{50} of compounds from PubChem database

No.	Comp ID	$E_{binding}$ (kJ/mol)	Pred-MLR	Pred-ANN
54	10061731	−30.863	12.29	8.99
55	10087945	−24.023	7.27	7.29
56	10365317	−24.602	7.69	7.63
57	10376208	−25.659	8.47	8.16
58	10380506	−25.406	8.28	8.04
59	1038250	−31.970	13.11	9.00
60	10428650	−36.768	16.63	9.03
61	10517380	−25.585	8.41	8.12
62	10589822	−25.629	8.45	8.14
63	3010438	−25.126	8.08	7.90
64	452838	−24.957	7.95	7.82
65	460857	−25.528	8.37	8.10
66	465251	−26.666	9.21	8.53
67	465287	−25.429	8.30	8.05
68	465288	−24.097	7.32	7.33
69	465289	−24.938	7.94	7.81
70	504502	−30.705	12.18	8.98
71	5388404	−26.606	9.16	8.51
72	6451105	−24.214	7.41	7.40
73	6451106	−24.686	7.75	7.67

attached to 6-N or to 4-C or to 5-C or to the terminal end of the chain attached to 6-N. The tri-fluoro group in compound no. 54, 60 and 70 showed better interaction with TYR318. The phenyl group attached to 6-N (cpd no. 60) and the phenyl group attached to 5-C of diazepine ring (cpd no. 66) showed better alignment and interaction with TRP229 and PHE227. A similar interaction has been observed in compound no. 71 with a long chain attached to 6-N. Interestingly, compound no. 59, 61, 62 showed different scaffolds from that of TIBO, where the diazepine ring has either been replaced by a 6-member ring or there is no ring at all. Here also the phenyl group attached to 6-member ring showed better interaction with TRP229 and PHE227 residues. Our findings suggest that NNRTIs interacting more closely with these conserved residues TYR318, TRP229 and PHE227 residues in the NNRTI-specific pocket of HIV-1 RT have better binding affinities and targeting these conserved amino acids for drug-design purposes may yield drugs that are effective in escaping

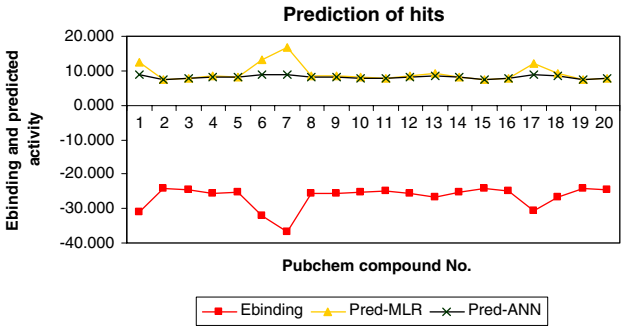


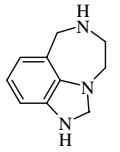
Fig. 6 Graph showing the binding affinity ($E_{binding}$ in kJ/mol) and their respective predicted pIC_{50} (MLR and ANN) of selected compounds from PubChem database

mutations. Hence, these scaffolds can be used as starting point in the design and synthesis of novel and potential hit compounds.

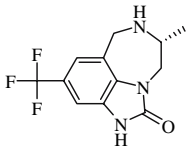
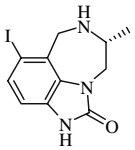
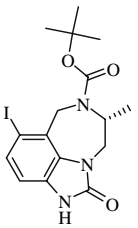
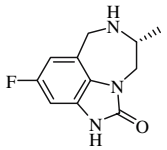
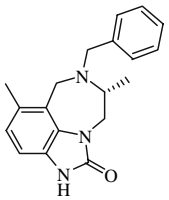
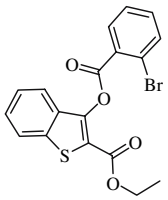
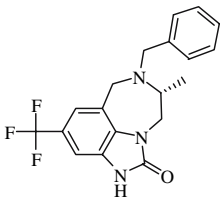
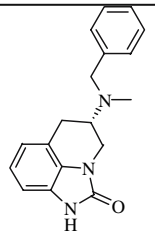
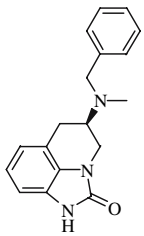
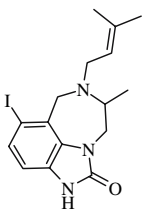
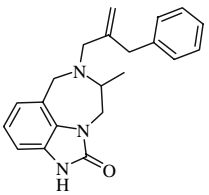
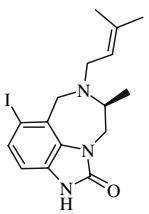
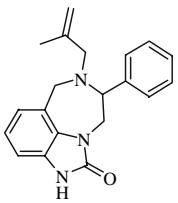
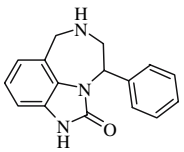
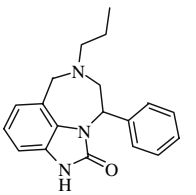
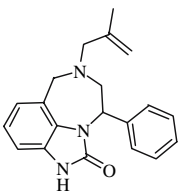
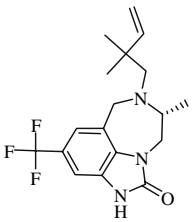
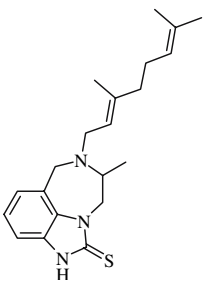
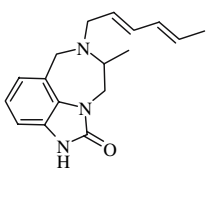
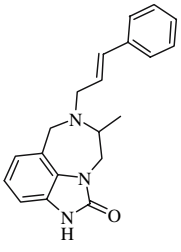
Conclusion

In this work, molecular docking studies were carried out to explore the binding preferences of TIBO derivatives towards the putative binding site of HIV-1 RT enzyme, to facilitate the design of new and effective NNRTIs. The docking simulations carried could satisfactorily reproduce a bound complex from the crystal structure of RT/TIBO (PDB code: 1REV). The model obtained for the TIBO derivatives using the binding conformations and energies calculated by the docking algorithm yielded a reasonable correlation with the experimental inhibition data. It is necessary to mention here that molecular docking may give very different lowest energy orientations even for very similar compounds (even though they are expected to have the same binding mode). This above disparity might be due to the fact that the protein is not considered flexible in our present docking studies. The hydrophobic interactions seemed to play a pivotal role in binding. Although the relative electrostatic interaction contributes to a lesser extent to the differential binding of these molecules, yet it has a decisive effect on the mutual orientation of the aromatic rings of the inhibitor and the amino acid residues in the binding pocket. The ligand solvation, before and after binding is again an area which requires a deeper insight to interpret the hydrophobicity effectively. On the basis of

Fig. 7 Structure of template and the selected 20 compounds from the PubChem database



Template structure used for sub-structure search with compound similarity ≥ 90%.

			
54. CID 10061731	**55. CID 10087945**	**56. CID 10365317**	**57. CID 10376208**
			
58. CID 10380506	**59. CID 1038250**	**60. CID 10428650**	**61. CID 10517380**
			
62. CID 10589822	**63. CID 3010438**	**64. CID 452838**	**65. CID 460857**
			
66. CID 465251	**67. CID 465287**	**68. CID 465288**	**69. CID 465289**
			
70. CID 504502	**71. CID 5388404**	**72. CID 6451105**	**73. CID 6451106**

correlations obtained and observed interactions, it can be inferred that the present docking protocol was successful in retrieving newer TIBO derivatives and novel scaffolds with favorable binding modes and these compounds will hopefully be useful in the design of novel NNRT inhibitors against wild-type as well as mutant variants of HIV-1 RT.

Acknowledgement The author wishes to acknowledge Dr. R. C. Saraswat, Director, S.G.S. I.T.S, Indore.

References

1. Silvestri R, Maga G (2006) Expert Opin Ther Pat 16:939
2. Matthis G, Torsten U, Danielson U (2006) J Med Chem 49:2375
3. Palella FJ, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA, Ashman DJ, Holmberg SD (1998) N Engl J Med 338:853
4. Palella FJ, Chmiel JS, Moorman AC, Holmberg SD (2002) AIDS 16:1617
5. De Clercq E (1998) Antiviral Res 38:153
6. De Clercq E (2004) Chem Biodivers 1:44
7. De Clercq E (2005) J Med Chem 10:1297
8. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA (1992) Science 256:1783
9. Ren J, Esnouf R, Garman E, Somers D, Ross C, Kirby I, Keeling J, Darby G, Jones Y, Stuart D, Stammers D (1995) Nat Struct Biol 2:293
10. 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
11. Deeks SG (2001) J AIDS 26:S25
12. Tavel JA, Miller KD, Masur H (1999) Clin Infect Dis 28:643
13. Das K, Levi PJ, Hughes SH, Arnold E (2005) Prog Biophys Mol Biol 88:209
14. Zhou Z, Lin X, Madura JD (2006) Infect Disord Drug Targets 6:391
15. De Clercq E (2001) Curr Med Chem 8:1543
16. Campiani G, Ramunno A, Maga G, Nacci V, Fattorusso C, Catalanotti B, Morelli E, Novellino E (2002) Curr Pharm Des 8:615
17. Hannongbua S, Lawtrakul L, Limtrakul J (1996) J Comput Aided Mol Des 10:145
18. Hannongbua S, Pungpo P, Limtrakul J, Wolschann P (1999) J Comput Aided Mol Des 13:563
19. Garg R, Gupta SP, Gao H, Babu MS, Debnath AK, Hansch C (1991) Chem Rev 99:3525
20. Zhou Z, Madura JD (2004) J Chem Inf Comput Sci 44:2167
21. Barreca ML, Rao A, De Luca L, Zappala M, Monforte AM, Maga G, Pannecouque C, Balzarini J, De Clercq E, Chimirri A, Monforte P (2005) J Med Chem 48:3433
22. D'Cruz OJ, Uckun FM (2006) J Antimicrob Chemother 57:411
23. 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
24. Ren J, Esnouf R, Hopkins A, Ross C, Jones Y, Stammers D, Stuart D (1995) Structure 3:915
25. Zhou Z, Madura JD (2004) J Chem Inf Comput Sci 44:2167
26. De Clercq E (1995) Clin Microbiol Rev 8:200
27. 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
28. Ding J, Das K, Moereels H, Koymans L, Andries K, Janssen PA, Hughes SH, Arnold E (1995) Nat Struct Biol 2:407
29. Bahram H, Mohammad Hossein Tabaei S, Fatemeh N (2005) J Mol Struct Theochem 732:39
30. Saen-oon S, Kuno M, Hannongbua S (2005) Proteins 61:859
31. Zhou Z, Madura JD (2004) Proteins 57:493
32. Rawal RK, Kumar A, Siddiqi IS, Katti SB (2007) J Mol Model 13:155
33. Rizzo RC, Wang DP, Tirado-Rives J, Jorgensen WL (2000) J Am Chem Soc 122:12898
34. http://www.pubchem.ncbi.nlm.nih.gov
35. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235
36. Thomsen R, Christensen MH (2006) J Med Chem 49:3315
37. 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, Berkley, CA
38. Thomsen R (2003) Flexible ligand docking using differential evolution. In: Proceedings of the 2003 congress on evolutionary computation, vol 4, pp 2354–2361
39. 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. In: Proceedings of the fourth international conference on evolutionary programming, pp 615–627
40. Gehlhaar DK, Bouzida D, Rejto PA (1998) Fully automated and rapid flexible docking of inhibitors covalently bound to serine proteases. In: Proceedings of the seventh international conference on evolutionary programming, pp 449–461
41. Yang JM, Chen CC (2004) Proteins 55:288
42. http://www.molegro.com (free trial version)
43. Mager PP (1997) Med Res Rev 17:235
44. 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