Journal of Molecular Graphics and Modelling
27 (6)2009, Pages 735-743

CoMFA and CoMSIA studies on thiazolidin-4-one as anti-HIV-1 agents

Vanangamudi Murugesan, Yenamandra S. Prabhakar and Seturam B. Katti*

Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow-226001, India

E-mail: setu_katti@yahoo.com

Tel: 0522-2620586; Fax: 0522-2620586

Abstract

Comparative molecular field analysis (CoMFA) and Comparative molecular similarity indices analysis (CoMSIA) were performed on thiazolidin-4-one class of compounds as HIV-1 Reverse Transcriptase (HIV-1 RT) inhibitors using global minima and crystal structure conformations. Results obtained from the crystal structure based model yielded superior statistical data (r^2_{cv} values of 0.683 for CoMFA and 0.678 for CoMSIA) when compared to those obtained by the global-minima based model (r^2_{cv} values of 0.625 and 0.654 for CoMFA and CoMSIA, respectively). The models were validated using an external test set of 47 compounds. The predictive r^2 values for the crystal-based CoMFA and CoMSIA models were 0.735 and 0.739, respectively, while the corresponding predictive r^2 values for the global minima-based CoMFA and CoMSIA models were 0.654 and 0.635, respectively. 3D contour maps generated from these models provide the regions in space where interactive fields may influence the activity. The superimposition of contour maps on the active site of HIV-1 Reverse Transcriptase additionally helped in understanding the structural requirements of these inhibitors. The results provide insight for predictive and diagnostic aspects of this class of HIV-1 RT inhibitors for better activity.

Keywords: HIV-1RT, CoMFA, CoMSIA, NNRTIs, Thiazolidin-4-ones

1. Introduction

HIV-1 Reverse Transcriptase is an obligatory enzyme in the replicative cycle of HIV-1 virus [1]. It has been successfully exploited as a drug target and today there are a number of drugs in the market that elicit anti-HIV-1 activity by inhibiting this enzyme [2,3]. Presently two classes of HIV-1 RT inhibitors have been developed: non-nucleoside reverse transcriptase inhibitors (NNRTIs) and nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs). Between these two major classes, NNRTIs interact non-competitively with an allosteric site on the enzyme leading to a conformational change in the enzyme, resulting in decreased affinity for the substrate [4]. They are highly specific, and show lower cellular toxicity compared to the NRTIs. Drug-resistant viruses of NNRTIs point out the need for continuing the search for novel anti-HIV-1 drugs [5]. In drug design and

discovery research quantitative structure–activity relationship (QSAR) studies provide rational inputs for molecular modifications [6,7]. In this endeavor, various computational strategies, such as quantum mechanical applications, 2D QSAR/3D QSAR and structure based drug design approach have been applied to several classes of NNRTIS [8-13]. In previous explorations, we examined quantitative structure-activity relationships (QSAR) of NNRTIs to find the common structural features among them. These studies provided information on the hydrophobic property with less extended or compact structural templates for better HIV-1 RT inhibitory activity [14-16]. Recently we used FlexX program to successfully dock a set of thiazolidin-4-ones into HIV-1 RT binding pocket. The calculated binding energies, based on the docked structures, agree well with the experimentally observed inhibitory activities (EC_{50}) [17]. From our QSAR and docking studies, we have designed and synthesized different diarylsubstituted thiazolidin-4-ones as HIV-1 RT inhibitor [18-23]. Some of them have (EC_{50} value 0.02μM) equipotent to clinically used drugs namely, nevirapine and delaviridine. In order to get better potency and gain higher level of understanding of HIV-1 RT inhibitory activity of these analogues, we add up the compounds of Pietro Monforte's [24-28] and our group [18-23] for the 3D-QSAR study (Figure 1a). We have performed three dimensional quantitative structure activity relationship (3D-QSAR) using the comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) techniques [29,30]. The above two techniques have been successfully applied in many instances to guide the design of new bioactive molecules. In the present investigation, a CoMFA / CoMSIA based 3D QSAR study has been carried out on thiazolidin-4-ones to determine the influence of steric, electrostatic, and hydrophobic fields of these compounds on their HIV-1 inhibitory activity. Furthermore, these fields have been mapped onto the NNRTIs binding pocket of HIV-1 RT for the better understanding of these interactions.

2. Methodology

2.1. Data set

The chemical structures considered in this study have originated from our group [18-23] as well as several others [24-28]. For these analogues the EC_{50}, effective concentration required to achieve 50% (EC_{50}) protection of MT-4 cells against the cytopathic effect of HIV-1 is reported from the same laboratory under one experimental protocol [31]. These were transformed into the corresponding log ($1/EC_{50}$) values as shown in Table 1. For the purpose of CoMFA/CoMSIA analysis all the analogues have been randomly divided into training set (136 compounds) and test set (47 compounds).

2.2. Molecular Modelling and Alignment

The 3D- structure building and all modelling studies were performed using the SYBYL program package, version 7.3 [32] on a silicon graphics fuel workstation with IRIX 6.5 operating system. In this study two different conformations were generated for the most active molecule (compound **64**) of the data set. Of this, one conformation of compound 64 corresponds to global minimum from simulated annealing. The other conformation is generated using

the X-ray crystal structure of 7-chloro-1-(2,6-difluorophenyl)-1H,3H-thiazolo[3,4-a]benzimidazole (Figure 1c; 7-Cl-TBZ) reported by Nicolo et al [33]. The global minimum energy conformation of the most active compound (compound **64**) was obtained by using the cleanup procedure within SYBYL and energy minimized using MAXIMIN2 followed by simulated annealing [34]. The most active compound **64** was subjected to simulated annealing by heating at 1000 K for 1000 fs to allow for exploration of conformational space, and then followed by exponential quenching at 200 K for 1500 fs. Fifty such cycles generated 50 low-temperature conformations, which were then energy ranked. The 50 lowest energy conformers were further minimized, at which point the lowest energy conformation was selected as global minima conformation.

The structures of all inhibitors were constructed from the template molecule by using the SKETCH option function in SYBYL. Energy minimizations were performed using the Tripos force field [35] with a distance-dependent dielectric function and the Powell conjugate gradient algorithm with a convergence criterion of 0.01 kcal/mol Å using 1000 iterations. Partial atomic charges were calculated using the Gastieger-Huckel method. In both conformations the most potent compound **64** was chosen as template molecule to fit the remaining training and test compounds by using the "database align" function in SYBYL. The reference atoms in compound **64** used for alignment were as follows (i) C1' of the benzene ring; (ii) S1, C2, N3 and C4 of the thiazolidin-4-one ring; and (iii) C2" of the pyrimidine ring (Figure 1b). The two different conformations of alignment, namely global minima-based alignment and crystal structure-based alignment are shown in Figures 1d and 1e respectively.

2.3. CoMFA and CoMSIA Analysis

In deriving the CoMFA and CoMSIA descriptor fields, a 3D cubic lattice with grid spacing of 1 Å and extending to 4 Å units beyond the aligned molecules in all directions was created to encompass the aligned molecules. CoMFA descriptors were calculated using an sp^3 carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0 to generate steric (Lennard-Jones 6-12 potential) field energies and electrostatic (Coulombic potential) fields with a distance-dependent dielectric at each lattice point. Steric and electrostatic fields generated were scaled by the CoMFA-Standard method in SYBYL with default cutoff energy of 30 kcal/mol. The minimum column filtering was set to 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points whose energy variation was below this threshold.

CoMSIA similarity indices were derived according to Klebe et al [30]. With the same lattice box as was used for the CoMFA calculations, with a grid spacing of 1 Å employing a C^{1+} probe atom with a radius of 1.0 Å as implemented in SYBYL. CoMSIA similarity indices (A_F) for a molecule j with atoms i at a grid point q were calculated using equation 1 as follows:

$$A_{F,K}^{q}(j) = -\sum_{i=1}^{n} \omega_{probe,k}\, \omega_{ik}\, e^{-\alpha r i q^2} \qquad (1)$$

Being an extension to the CoMFA approach which has two fields, the CoMSIA method incorporates five different physicochemical properties (steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor) denoted as k in equation 1, which were evaluated using the probe atom. A Gaussian type distance-dependence was used between the grid point q and each atom i in the molecule. A default value of 0.3 was used as the attenuation factor (α). In CoMSIA, the steric indices are related to the third power of the atomic radii, the electrostatic descriptors are derived from partial atomic charges, the hydrophobic fields are derived from atom based parameters [36], and the hydrogen-bond donor acceptor atoms within a putative protein environment are derived from experimental values [37].

2.4. PLS Analysis

The 3D QSAR models of CoMFA and CoMSIA descriptors were derived using PLS regression as implemented in the SYBYL package [38]. The cross-validated coefficient, q^2, was calculated using equation 2

$$r^2_{cv} = 1 - \frac{\sum(Y_{predicted} - Y_{observed})^2}{\sum(Y_{observed} - Y_{mean})^2} \qquad (2)$$

Where $Y_{predicted}$, $Y_{observed}$, and Y_{mean} are predicted, actual, and mean values of the target property (pEC_{50}), respectively. $\sum (Y_{predicted} - Y_{observed})^2$ is the predictive sum of squares (PRESS). The optimum number of components (ONC) which corresponds to the lowest standard error of prediction (SEP) were used to generate the final PLS regression models. CoMFA and CoMSIA coefficient maps were generated by interpolation of the pair wise products between the PLS coefficients and the standard deviations of the corresponding CoMFA or CoMSIA descriptor values. The boot strapping analysis for 100 runs and the number of cross-validations (e.g., leave-half-out and leave 20% out) were carried out and confirmed by the average value for 50 runs from each cross validation [39].

2.5. Predictive r^2 value

The activity of the test set was predicted by the CoMFA and CoMSIA models using the predict property command. The predictive r^2 was based only on molecules not included in the training set and is computed using the following equation

$$r^2_{pred} = (SD - PRESS)/SD \qquad (3)$$

Where SD is the sum of the squared deviations between the biological activity of molecules in the test set and the mean biological activity of the training set molecules and PRESS is the sum of the squared deviations between predicted and actual activity values for every molecule in the test set.

3. Results and Discussion

3.1. CoMFA and CoMSIA Statistical Results

The CoMFA and CoMSIA models were developed from 136 compounds (thiazolidin-4-one derivatives) using global minima and crystal structure based alignment structures at 1 Å grid spacing. Compounds **63** and **70** were found as common outliers in all CoMFA and CoMSIA models. While compound **63** has C4"-Cl, compound **70** has C4"-OH group at the C4" position of the 2-pyrimidinyl substituent at the N3 of the thiazolidinone ring (Figure 1b). These two compounds were excluded from both the models. The results of the CoMFA and CoMSIA statistical analysis are summarized in Table 2**.** For 134 compounds, with the lattice step size at 1.0 Å, the global minima based model yielded (r^2_{cv} = 0.625 and r^2_{ncv} = 0.852) and (r^2_{cv} = 0.654 and r^2_{ncv} = 0.793) for the CoMFA model 1 and CoMSIA model 2, respectively. For the same compounds crystal structure based model gave r^2_{cv} = 0.683 and r^2_{ncv} = 0.876 for CoMFA (model 3) and r^2_{cv} = 0.678 and r^2_{ncv} = 0.817 for CoMSIA (model 4). The experimental and predicted values of the compounds from CoMFA (model 3) and CoMSIA (model 4) are given in Table 1**.** The graphs of the observed versus predicted activities from crystal structure based CoMFA and CoMSIA models are shown in Figure 2. Thus the crystal structure based model performed better in terms of prediction than the global minima-based model. For the CoMSIA models, statistical parameters showed significance for steric, electrostatic and hydrophobic features, to influence the activity of the compounds. To further evaluate the robustness and statistical confidence of the derived models, bootstrapping analysis for 100 runs was performed (Table 2). Cross-validation analysis was applied to the compounds in the training set to investigate the stability of the CoMFA and CoMSIA models. The training set model was cross validated using two (leave-half-out) and five (leave 20% out) cross-validation groups 50 times each. The average and standard deviation values of r^2_{cv} are shown in Table 2. When two cross-validation groups were used, the average r^2_{cv} values for CoMFA and CoMSIA were respectively 0.682 and 0.675 (standard deviations 0.005 and 0.011). By using five cross-validation groups the average r^2_{cv} and SD values were 0.622 and 0.036 for CoMFA, and those for the CoMSIA were 0.641 and 0.033, respectively. Thus, two and five cross-validation analyses of the training set composition for each run were consistent for the CoMFA as well as CoMSIA models. The bootstrapped r^2 were 0.897 for CoMFA and 0.841 for CoMSIA suggesting that a good internal consistency exists within the underlying data set.

3.2. Predictivity of 3D QSAR Models

The predictive power of the CoMFA and CoMSIA models were checked from a set of 47 test compounds not included in the model development. The crystal structure based CoMFA (model 3) and CoMSIA (model 4), have shown test set r^2 values of 0.735 and 0.739 respectively (Table 2). The external Predictivity of crystal structure based models was better than the global minima based models, which showed 0.654 for CoMFA model 1 and 0.635 for CoMSIA

model 2. The observed and predicted activity values for crystal structure based CoMFA model 3 and CoMSIA model 4 in test set compounds are listed in Table 1. The corresponding graph of observed versus predicted activities for test set compounds by crystal structure based CoMFA model 3 and CoMSIA model 4 is depicted in Figure 2.

3.3. CoMFA contours

The earlier investigations on 2,3-diaryl-1,3-thiazolidin-4-ones have suggested that these compounds express the anti-HIV-1 activity by inhibiting HIV-1 RT [28,40]. The information of HIV-1 reverse transcriptase binding pocket is available in the form of co-crystal structure of enzyme-nevirapine complex (PDB code 1VRT) [41]. The nevirapine binding pocket of the enzyme has very well accommodated the conformation of compound **64** originating from the crystal structure of 7-Cl-TBZ. In this, 2-(2',6'-dichlorophenyl), 1,3-thiazolidin-4-one and 3-(4",6"-dimethylpyrimidin-2-yl) moieties of compound 64, respectively occupy the methylpyridine, cyclopropyl and pyridine regions of nevirapine. In view of this, to understand the relationship between CoMFA fields and the surroundings of ligand binding site of the enzyme, the CoMFA contours were mapped onto the active site of the enzyme (PDB code 1VRT). It has provided an opportunity to analyze the CoMFA fields with respect to the amino acid residues surrounding the active site. The mapping was carried out by superimposing the structure of the most active compound **64** together with its contours developed from the CoMFA and CoMSIA model onto the structure of Nevirapine in the binding site of HIV-1 reverse transcriptase enzyme. In order to replicate the overlap with the active site, the 2-(2',6'-dichlorophenyl), 1,3-thiazolidin-4-one and 3-(4",6"-dimethylpyrimidin-2-yl) structural units of compound 64 were mapped on to nevirapine's pharmacophore units, namely methylpyridine, cyclopropyl and pyridine. The steric and electrostatic CoMFA contour maps obtained by model 3 are shown in Figure 3a, b. To aid the visualization, the most potent compound **64** is overlaid on the map. The steric contour maps wherein green and yellow contours indicate sterically favorable and unfavorable areas respectively. Blue and red contours in the electrostatic contour map indicate areas where positive and negative charge substitutions favor activity. The activity of a compound is cumulative effect of its functional groups and their interactions with the potential receptor. To highlight the substructural regions of thiazolidinone scaffold in eliciting the activity, selected compounds are chosen to exemplify their interactions with the receptor.

The two green contour maps located at the C2' and C6' positions of phenyl moiety at the C2 position of thiazolidinone scaffold indicate that steric bulk at these positions is important for HIV-1 RT inhibitory activity. One large round green contour map located at C2' position of phenyl ring, which is situated near to TYR188 amino acid residue suggesting that more bulky group at this position is well tolerated due to hydrophobic interaction between substitutions of C2' phenyl with aromatic portion of the amino acids residue Tyr188. Secondly a small U-Shaped green contour map at C6' position of phenyl ring, co-localised with the small red contour map indicating a favorable effect of steric bulk with high electron density groups such as $–OCH_3$. This steric area is located near to

hydrophobic amino acid Tyr181 and aliphatic side chain of Leu100 indicating the steric interaction with these amino acid side chains. This is corroborated by the biological activity of compounds having steric substituents at C2' and C6' positions of the phenyl ring (eg **41, 44, 47, 61, 62, 64, 67, 68, 71, 75, 76, 102, 105, 107, 139, 150, 154, 155, 157**) are more active than molecules lacking substituent at these same position (eg compounds **111, 113, 115, 117, 162,171**). Furthermore it may be inferred from the model that bulky substitutions at positions C3' and C4' of this phenyl ring (compounds **113,116,117, & 121**) lead to decrease in the activity as is evident from the contour map wherein corresponding position is shown in yellow colour. Moreover, in these furan based analogues (e.g. compounds **113,116,117, 121** etc) the presence of methylene moiety between the thiazolidinone ring and the heteroaryl moiety of R_2 group has led to a distortion in their butterfly-like conformation. This may also have lowered the activity of these compounds. The large green colour contour map encompassing on to the N3 substituted phenyl ring of C3" and C4" position, C5" and C6" position of the pyridin-2-yl ring, and C4" and C5" position of the pyrimidin-2-yl ring. This contour area close to phenyl portion of Tyr318 residue, which prefer hydrophobic interaction with the methyl substitution of five and six membered of the aromatic and heteroaromatic rings. The most active molecule **64** and other molecules with higher activity (eg compounds **10, 12, 28, 54, 67, 75,** etc) have methyl group on heteroaromatic ring close to large green steric favorable area. While molecules, which lack such methyl group in this region is observed as less active molecules (eg compounds **1, 2, 60, 92, 101,** and **175**). However, a few compounds (**33, 38, 48, 53,** etc) have deviated from this observation. This may be attributed to the nature of substituents on C2-aryl moiety of thiazolidinone ring. The furan-2-ylmethyl, pyrrol-2-ylmethyl, thiophen-2-ylmethyl and pyridin-2-ylmethyl group containing compounds (**109, 111, 112, 113, 148, 163, 179, 181,** and **183**) have less activity, which fall in to the sterically disfavorable three yellow contour maps due to distraction of butterfly-like conformation by insertion of $–CH_2-$ bridge between N3 atom of thiazolidinone and five or six membered rings. As a result steric portion of heteroaromatic-2-yl-methyl group is placed far away from the hydrophobic portion of Tyr318 and close to the side chain of Lys101 and Lys103 residue. A small three tiny yellow contour maps and two small yellow contour maps placed close to C5 position of the thiazolidinone ring indicate that steric bulky groups are restriction at this position. These results suggest that methyl group substitution (compounds **84, 85, 86** and **87**) and extension of carbon atom (compound **159** and **162**) at C5 position of the thiazolidinone ring decrease the HIV-1 RT inhibitory activity. A large yellow contour map placed behind the plane of C6" position of the pyrimidine ring, suggest that large bulky group decreased the HIV-1 RT inhibitory activity (compounds **136** and **137**). There are five yellow colour contour maps representating unfavorable steric environment. Out of these three are behind the plane and two are placed above the plane of the heteroaromatic ring. The side chain of Tyr318, Pro236 and His235 are near to these five yellow maps.

The electrostatic contour map constructed from CoMFA model 3 is shown in Figure 3b. A large red colour contour surrounding the electronegative nitrogen

atom of the N1" pyridine and pyrimidine ring indicate that the presence of electron rich nitrogen would increase the inhibitory activity (compounds **44, 47, 49, 61, 62, 67** etc), when compared to the phenyl ring compounds **11, 19, 23, 26,** etc. This observation indicates the interaction between the electron rich nitrogen atom of pyridine and pyrimidine ring with side chain of Leu100. Two small red contour maps located at the C6' position of phenyl moiety at the C2 of the thiazolidinone scaffold placed near Lys101 indicated that the electronegative groups would increase the inhibitory activity. In most of the highly active compounds (**41, 64, 150, 47,** and **45),** this position is occupied by the electron rich substituents such as –F, –Cl, -Br, -OMe and CF_3. Small blue contour map located at nitrogen atom of N3-pyridin-3-yl moiety indicate that electronegative nitrogen atom is unfavorable at this position (compound **92**). The small and large blue contour maps around the heteroaromatic ring indicate that an electropositive group needs to be present at this position for HIV-1 RT inhibitory activity. Most of the highly active compounds have $-CH_3$ group at this position, and the least and moderate active compounds lack electropositive substitution at this position as exemplified by compounds **134, 135, 143, 144, 151, 166** and **168** having an electronegative ($-CF_3$) substitution at this position. Two tiny blue contour maps located on the C4' position of phenyl ring and N3 substituted C3"-phenyl, C4"-Pyridin-2-yl and C4"-pyrimidin-2-yl rings, implies that the presence of electropositive group would increase the HIV-1 RT inhibitory activity.

3.4. CoMSIA contours

The crystal structure based CoMSIA model 4 was used to generate contour plots. The steric and electrostatic contour maps of CoMSIA are very similar to those of the CoMFA. Figure 4 displays the hydrophobic plot represented by orange and white contours. Orange contours indicate regions where hydrophobic groups on ligands are favored and white contours represent areas where hydrophobic groups are unfavored (or favorable for hydrophilic groups on ligands). Two orange contour maps, one is located on the C2' position phenyl ring and another is located on the N3 substituted C3"- phenyl, C6"-pyridine, C4"-pyrimidine and C4"-thiazole ring. This indicate that most of the highly active compounds (**28, 61, 64, 71, 75, 154, 155, 157** etc) carry the small hydrophobic halo groups (Cl, F, and Br) at C2' position of phenyl ring and non-polar group ($-CH_3$) at N3 substituted heteroaromatic ring position of C3"-phenyl, C6"-pyridin-2-yl, C4"-pyrimidin-2-yl and C4"-thiazole which could provide a hydrophobic environments around this region due to hydrophobic interaction of halo groups with the Tyr188 residue and methyl group with Tyr 318. A small hydrophilic contour map located at the C3' position of phenyl ring near to the side chain of Trp229, which suggest that hydrophilic group like –OMe more favorable for HIV-1 RT inhibitory activity. But unexpectedly the compound **117** has shown the least activity in the whole data set due to the absence of hydrophobic halo groups at C2', C6' diortho position of phenyl ring. The large white contour map is encircling at N3 substituted phenyl ring C4" of position phenyl ring, C5" position of pyridine ring and C5" of pyrimidine ring. This observation shows that moderate activity of compounds (**93, 94, 97, 104, 125, 126, 127, 132, 152, 174**) having

hydrophobic group such as Cl, Br and Me orient towards in to the encircling white contour map. Amino acid hydroxy group of Tyr318 are very close to this contour. The round white contour surround at the carbonyl group of the thiazolidinone ring which could provide hydrophilic environment with the polar group of Lys101.

4. Conclusions

CoMFA and CoMSIA based 3D QSAR study of 183 thiazolidin-4-ones as HIV-1 inhibitors was carried out using the global minimum energy and crystal structure conformations of the compounds. A high bootstrapped r^2 values for CoMFA and CoMSIA models with small standard deviation indicated the existence of strong relevance for the conformations of the compounds used to build the model. Also, the models are validated with the external test set compounds. In addition to steric and electrostatic fields, hydrophobic field is also important for the HIV-1 inhibitory activity of the compounds. CoMFA results suggested that halo substitution at C2' and C6' position of C2 substituted phenyl ring and small methyl substitution at N3 substituted aromatic and heteroaromatic ring are required to increase the inhibitory activity. Moreover the electronegative nitrogen atom at N3 substituted N1"-pyridin-2-yl and N1"-pyrimidin-2-yl derivatives are more active than the N3 substituted phenyl derivatives. The results obtained from the CoMSIA model revealed the importance of hydrophobic group present on C2'-phenyl and N3 substituted C3"-phenyl, C6"-pyridin-2-yl, C4"-pyrimidin-2-yl and C4"-thiazole ring, respectively, for the inhibitory activity. Interpretation of the CoMFA and CoMSIA contours mapped into the active site of HIV-1 Reverse Transcriptase provided a better understanding of the activity of the compounds vis-à-vis probable interactions with HIV-1 Reverse Transcriptase. Hence the study provides predictive and diagnostic value for the modification of thiazolidin-4-one analogues.

Acknowledgement

One of the authors (V.Murugesan) thanks CSIR, New Delhi for Junior Research fellowship. The authors thank the Director, CDRI for support. CDRI communication number: 7569.

Supplementary data

Structural information and biological activity of compounds listed in Tables 1 and structure files of all compounds in mol2 (Tripos) format.

References:

1. Turner, B.G., and Summers, M.F. Structural Biology of HIV. J. Mol. Biol. 1999, 285, 1-32.
2. Pauwels, R. New non-nucleoside reverse transcriptase inhibitors (NNRTIs) in development for the treatment of HIV infections. Curr. Opin. Pharmacol. 2004, 4, 437-446.

3. Clercq, E.D. The design of drugs for HIV and HCV. Nature Rev. Drug Discov. 2007, 6**,** 1001–1018.
4. Esnouf, R., Ren, J., Ross, C., Jones, Y., Stammers, D., and Stuart, D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. Structural Biology. 1995, 2, 303-308.
5. Safadi, Y.E., Boudou, V.V., and Marquet, R. HIV-1 reverse transcriptase inhibitors. Appl. Microbiol. Biotechnol. 2007, 75, 723-737.
6. Hansch, C. The physicochemical approach to drug design and discovery (QSAR). Drug Develop. Res. 1981, 1, 267-309.
7. Kubinyi, H. (Ed). 3D QSAR in drug design volume 1: theory, methods and applications. ESCOM, Leiden, 1994.
8. Garg, R., Gupta, S.P., Gao, H., Babu, M.S., Debnath, A.K., and Hansch, C. Comparative quantitative structure-activity relationship studies on anti-HIV drugs. Chem. Rev. 1999, 99, 3525-3601.
9. Hannongbua, S. Structural information and drug–enzyme interaction of the non-nucleoside reverse transcriptase inhibitors based on computational chemistry approaches. Top. Heterocycl. Chem. 2006, 4, 55-84.
10. Pungpo, P., Saparpakorn, P., Wolschann, P., and Hannongbua, S. Computer-aided molecular design of highly potent HIV-1 RT inhibitors: 3D QSAR and molecular docking studies of efavirenz derivatives. SAR and QSAR in Environ. Res. 2006, 17, 353-370.
11. Jorgensen, W.L., Caro, J.R., Rives, J.T., Basavapathruni, A., Anderson, K.S., and Hamilton, A.D. Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. Bioorg. Med. Chem. Lett. 2006, 16, 663-667.
12. Carlsson, J., Boukharta, L., and Aqvist, J. Combining docking, molecular dynamics and the linear interaction energy method to predict binding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. J. Med. Chem. 2008, 51, 2648-2656.
13. Ravichandran, V., Kumar, .B.R.P., Sankar, S., and Agrawal, R.K. Predicting anti-HIV activity of 1,3,4-thiazolidinone derivatives: 3D-QSAR approach, Eur. J. Med. Chem. article in press.
14. Prabhakar, Y.S, Solomon, V.R., Rawal, R.K, Gupta, M.K., and Katti, S.B. CP-MLR/PLS directed structure-activity modeling of the HIV-1RT inhibitory activity of 2,3-diaryl-1,3-thiazolidin-4-ones. QSAR. Comb. Sci. 2004, 23, 234-244.
15. Prabhakar, Y.S, Rawal, R.K., Gupta, M.K., Solomon, V.R., and Katti, S.B. Topological descriptors in modeling the HIV inhibitory activity of 2-aryl-3-pyridyl-thiazolidin-4-ones, Comb. Chem. & High Through. Screen. 2005, 8, 431-437.
16. Rawal, R.K., Prabhakar, Y.S., and Katti, S.B. Molecular surface features in modeling the HIV-1 RT inhibitory activity of 2-(2,6-disubstituted phenyl)-3-(substituted pyrimidin-2-yl)-thiazolidin-4-ones. QSAR. Comb. Sci. 2007, 26, 398-406.

17. Rawal, R.K., Kumar, A., Siddiqi, M.I., and Katti, S.B. Molecular docking studies on 4-thiazolidinones as HIV-1RT inhibitors. J. Mol. Model. 2007, 13, 155-161.
18. Rawal, R.K., Prabhakar, Y.S., Katti, S.B., and Clercq, E.D. 2-(aryl)-3-furan-2-ylmethyl-thiazolidin-4-ones as selective HIV-RT Inhibitors. Bioorg. Med. Chem. 2005, 13, 6771-6776.
19. Rawal, R.K., Tripathi, R., Katti, S.B., Pannecouque, C., and Clercq, E.D. Design, synthesis and evaluation of 2-aryl-3-heteroaryl-1,3-thiazolidin-4-ones as anti-HIV agents. Bioorg. Med. Chem. 2007, 15, 1725-1731.
20. Rawal, R.K., Tripathi, R., Katti, S.B., Pannecouque, C., and Clercq, E.D. Synthesis and evaluation of 2-(2,6-dihalophenyl)-3-pyrimidinyl-1,3-thiazolidin-4-one analogues as anti-HIV-1 agents. Bioorg. Med. Chem. 2007, 15, 3134-3142.
21. Rawal, R.K., Tripathi, R., Katti, S.B., Pannecouque, C., and Clercq, E.D. Synthesis and biological evaluation of 2,3-diaryl substituted-1,3-thiazolidin-4-ones as Anti-HIV agents. Medicinal Chemistry. 2007, *3*, 355-363.
22. Rawal, R.K., Tripathi, R., Katti, S.B., Pannecouque, C., and Clercq, E.D. Design and synthesis of 2-(2,6-dibromo-phenyl)-3-heteroaryl-1,3-thiazolidin-4-ones as Anti-HIV agents. Eur. J. Med. Chem. 2008, article in press
23. Rawal, R.K., Tripathi, R., Kulkarni, S., Paranjape, R., Katti, S.B., Pannecouque, C., and Clercq, E.D. 2-(2,6-dihalo-phenyl)-3-heteroaryl-2-ylmethyl-1,3-thiazolidin-4-ones: Anti-HIV agents, Chem. Biol. & Drug Des. 2008, 72, 147-154.
24. Barreca, M.L., Balzarini, J., Chimirri, A., Clercq, E.D., Luca, L.D., Holtje, H.D., Holtje, M., Monforte, A.M., Pietro, M., Pannecouque, C., Rao, A., and Zappala, M. Design, synthesis, structure-activity relationships and molecular modeling studies of 2,3-diaryl-1,3-thiazolidin-4-ones as potent Anti-HIV Agents. J. Med. Chem. 2002, 45, 5410-5413.
25. Rao, A., Carbone, A., Chimirri, A., Clercq, E.D., Monforte, A.M., Pietro, M., Pannecouque, C., and Zappala, M. Synthesis and anti-HIV activity of 2,3-diaryl-1,3-thiazolidin-4-(thi)one derivatives. IL Farmaco. 2002, 57, 747-751.
26. Rao, A., Carbone, A., Chimirri, A., Clercq, E.D., Monforte, A.M., Monforte, P., Pannecouque, C., and Zappala, M. Synthesis and anti-HIV activity of 2,3-diaryl-1,3-thiazolidin-4-ones. IL Farmaco. 2003, 58, 115-120.
27. Rao, A., Balzarini, J., Carbone, A., Chimirri, A., Clercq, E.D., Monforte, A.M., Monforte, P., Pannecouque, C., and Zappala, M. Synthesis of new 2,3-diaryl-1,3-thiazolidin-4-ones as anti-HIV agents. IL Farmaco. 2004, 59, 33-39.
28. Rao, A., Balzarini, J., Carbone, A., Chimirri, A., Clercq, E.D., Monforte, A.M., Monforte, P., Pannecouque, C., and Zappala, M. 2-(2,6-dihalophenyl)-3-(pyrimidin-2-yl)-1,3-thiazolidin-4-ones as non-nucleoside HIV-1 reverse transcriptase inhibitors. Antiviral Res. 2004, 63, 79-84.

29. Cramer, R.D.III., Patterson, D. E., and Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc. 1988, 110, 5959-5967.
30. Klebe, G., Abraham, U., and Mietzner, T. Molecular similarity indices in a comparative analysis (COMSIA) of drug molecules to correlate and predict their biological activity. J. Med. Chem. 1994, 37, 4130-4146.
31. Pauwels, R., Balzarini, J., Baba, M., Snoeck, R., Schols, D., Herdewijn, P., Desmyter, J., and Clercq, E.D. Rapid and automated tetrazolium-based colorimetric assay for the detection of anti-HIV compounds. J. Virol. Methods. 1988. 20, 309-321.
32. SYBYL, version 7.3, 2006, Tripos associates, St. Louis (MO), USA.
33. Nicolo, F., Bruno, G., Scopelliti, R., Grasso, S., Rao, A., and Zappala, M. 7-Chloro-1-(2,6-difluorophenyl)-1H,3H-thiazolo[3,4-a]benzimidazole and 1-(2,6-difluorophenyl)-6-methyl-1H,3H-thiazolo[3,4-a]benzimidazole. Acta. Crystallog. Sect. C. 2001, 57, 572-574.
34. Barakat, M.T., and Dean. P.M. Molecular structure matching by simulated annealing. I. A comparison between different cooling schedules. J. Comput. Aided. Mol. Des. 1990, 4, 295- 316.
35. Clark, M., Cramer, R.D.III., and Opdenbosch, N.V. Validation of the general-purpose tripos 5.2 force field. J. Comput. Chem. 1989, 10, 982-1012.
36. Viswanadhan, V.N., Ghose, A.K., Revenkar, G.R., and Robins, R.K. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring antibiotics. J. Chem. Inf. Comput.Sci. 1989, 29, 163-172.
37. Klebe, G. The use of composite crystal-field environments in molecular recognition and the de novo design of protein ligands. J. Mol. Biol. 1994, 237, 212-235.
38. Stahle, L., and Wold, S. 6 multivariate data analysis and experimental design in biomedical research. Prog. Med. Chem. 1988. 25, 291-338.
39. Cramer, R.D.III., Bunce, J.D., and Patterson, D.E. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. Quant. Struct. Act. Relat. 1988, 7, 18-25.
40. Barreca, M. L., Chimirri, A, Luca, L.D., Monforte, A.M., Monforte, P., Rao, A., Zappala, M., Balzarini, J., Clercq, E.D., Pannecouquec, C., and Witvrouw, M. Discovery of 2,3-diaryl-1,3-thiazolidin-4-ones as Potent Anti-HIV-1 Agents. Bioorg. Med. Chem. Lett. 2001. 11, 1793-1796.
41. Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D., and Stammers, D. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. Nat.Struct.Biol. 1995. 2, 293-302.

Table 1 Observed/Experimental and CoMFA/CoMSIA predicted HIV-1 inhibitory activity (EC_{50}) of 2,3-diaryl-1,3-thiazolidin-4-one derivatives (Figure 1a)

cpd[a]	Log(1/EC_{50})			cpd[a]	Log(1/EC_{50})			cpd[a]	Log(1/EC_{50})		
	Obs	Pred			Obs	Pred			Obs	Pred	
		CoMFA model 3	CoMSIA model 4			CoMFA model 3	CoMSIA model 4			CoMFA model 3	CoMSIA model 4
1	5.64	6.37	6.17	33t	5.24	5.36	4.94	65t	7.42	7.67	7.18
2	6.40	5.77	5.82	34	6.06	6.30	6.18	66	6.77	7.08	6.71
3t	6.21	6.35	6.22	35	5.67	5.41	5.23	67	7.62	7.71	7.66
4	6.19	6.23	6.34	36	6.49	6.48	6.41	68	7.25	7.64	7.80
5	5.88	5.76	5.83	37	5.48	5.63	5.78	69t	6.58	6.75	7.25
6	6.51	6.29	6.14	38	5.81	5.43	5.36	70[b]	4.48	-	-
7	6.21	6.18	6.26	39	6.54	6.80	6.60	71	7.09	6.45	6.37
8t	5.84	5.67	5.77	40t	5.52	5.93	5.67	72t	7.36	7.03	6.72
9t	7.10	6.85	6.54	41	7.30	6.73	6.86	73t	6.61	5.93	5.93
10	6.67	6.76	6.66	42	5.87	6.17	6.19	74	6.83	6.61	6.30
11	6.16	6.17	6.19	43	5.20	5.53	5.57	75	7.23	6.41	6.45
12	6.81	6.77	6.32	44	7.47	7.02	6.85	76	7.52	6.62	6.62
13t	6.68	6.65	6.43	45	6.18	6.02	5.94	77	5.19	5.91	5.50
14	6.11	6.14	5.99	46t	6.46	6.93	7.09	78t	4.94	6.46	5.86
15	5.85	5.78	5.28	47	7.35	6.47	6.43	79	5.36	5.34	5.11
16	5.48	5.58	5.41	48	5.36	5.58	5.44	80	6.55	6.69	6.10
17t	4.93	5.49	5.01	49	6.69	6.78	6.69	81	5.19	5.79	5.59
18t	6.14	6.23	5.70	50	5.23	5.76	5.76	82	6.24	6.06	5.78
19	5.88	6.12	5.81	51	6.63	6.69	6.95	83t	5.46	5.36	5.93
20	5.26	5.64	5.40	52t	5.95	6.14	6.27	84	5.42	5.94	6.29
21	6.55	6.84	6.87	53	5.47	5.57	5.56	85	5.51	5.34	6.02
22t	6.50	6.71	6.99	54	6.27	6.75	6.79	86	5.72	5.56	5.97
23	6.01	6.20	6.50	55	6.04	5.84	5.91	87	5.88	6.14	6.30
24	6.89	6.85	6.45	56	6.11	6.75	7.05	88t	5.69	5.70	6.16
25t	6.61	6.73	6.57	57t	6.04	5.95	5.93	89	6.75	6.62	6.20
26	6.27	6.18	6.09	58	6.64	6.51	6.60	90t	6.56	6.54	6.32
27t	6.72	7.26	6.78	59t	6.24	6.36	6.73	91	6.07	5.94	5.84
28	7.07	7.15	6.90	60	5.44	5.79	6.25	92	4.02	3.83	4.00
29	6.62	6.64	6.43	61	7.36	7.23	6.95	93	5.75	5.92	5.57
30t	5.94	6.45	6.43	62	7.13	7.17	7.07	94	5.67	5.82	5.69
31	5.83	6.56	6.65	63[b]	4.29	-	-	95t	5.14	5.23	5.21
32	5.72	5.77	6.03	64	7.77	7.74	7.05	96t	5.82	5.81	5.47

(Continued)

Table-1. (Continued)

cpd[a]	Log(1/EC_{50})	cpd[a]	Log(1/EC_{50})	cpd[a]	Log(1/EC_{50})

	Obs	Pred			Obs	Pred			Obs	Pred	
		CoMFA model 3	CoMSIA model 4			CoMFA model 3	CoMSIA model 4			CoMFA model 3	CoMSIA model 4
97	5.91	5.71	5.59	126	5.82	6.00	5.99	155	7.70	7.71	7.46
98	5.31	5.14	5.11	127	5.95	5.75	6.14	156	6.49	6.43	6.45
99	6.57	7.23	6.97	128t	5.49	5.36	5.64	157	7.22	7.24	7.50
100t	7.19	7.13	7.09	129t	5.57	6.42	6.04	158	5.28	5.34	5.56
101	4.27	4.79	5.21	130	6.59	6.18	6.17	159	5.25	4.95	5.74
102	7.00	6.54	6.42	131	6.34	5.65	5.71	160t	4.43	4.67	5.33
103t	5.85	6.11	6.35	132	4.59	4.73	4.56	161	4.35	4.17	3.98
104	5.29	5.46	5.96	133t	6.00	6.12	6.13	162	4.25	4.02	3.68
105	7.28	6.97	6.84	134	5.89	6.08	6.24	163	5.67	5.64	5.16
106	7.05	7.00	6.80	135	5.30	5.37	5.64	164t	5.30	5.62	5.27
107	7.38	6.92	6.92	136	5.21	5.09	5.17	165	4.88	5.04	4.80
108t	6.69	6.36	5.97	137	5.41	5.03	5.29	166	5.17	4.85	5.11
109	5.54	5.63	5.62	138t	5.19	4.90	4.97	167t	5.15	4.71	5.22
110	6.24	6.23	6.10	139	7.52	7.21	7.06	168	4.44	4.21	4.76
111	4.36	4.51	4.36	140t	7.70	7.10	7.18	169t	3.60	4.56	4.26
112	5.19	5.59	5.51	141	6.40	6.59	6.72	170	4.59	4.96	4.87
113	4.38	4.31	4.29	142t	5.37	5.33	5.14	171	4.13	5.03	4.80
114t	4.24	5.15	5.08	143	5.37	5.24	5.26	172t	5.99	6.29	6.00
115	4.07	3.97	4.00	144	4.59	4.63	4.79	173	6.02	6.20	6.11
116	3.58	3.75	4.00	145	6.30	6.03	6.53	174	5.89	5.63	5.65
117	3.45	3.64	3.49	146	6.24	6.17	6.17	175	5.37	5.34	5.60
118t	5.16	4.37	4.61	147	6.68	6.35	6.43	176	6.11	6.07	5.74
119	4.54	5.02	4.93	148	6.08	6.13	6.51	177t	5.70	5.86	5.83
120	5.11	5.16	4.61	149t	6.66	6.55	6.61	178	5.04	5.15	5.31
121	4.33	4.75	4.86	150	7.05	6.96	7.12	179	6.00	5.76	5.68
122	5.26	5.14	4.81	151	5.34	5.39	6.00	180	5.36	5.46	5.76
123	5.82	6.05	6.15	152	5.99	6.20	6.41	181	5.28	4.86	5.28
124t	6.12	5.88	6.27	153t	6.62	6.36	7.11	182t	5.05	4.94	5.24
125	5.87	5.31	5.79	154	7.40	6.99	7.35	183	4.76	4.33	4.76

[a] the Suffix‘t’ after the compound number indicate the test set analysis; complete structural information is given in supplementary data. [b] outlier not included in model development.

Table 2. Summary of CoMFA and CoMSIA Results

PLS statistics	Global minima based model		Crystal structure based model	
	CoMFA model 1	CoMSIA model 2	CoMFA model 3	CoMSIA model 4
r^2_{ncv} [a]	0.852	0.793	0.876	0.817
SEE[b]	0.367	0.434	0.336	0.408

F_{test}[c]	121.61	81.08	149.68	94.36
r^2_{cv}[d]	0.625	0.654	0.683	0.678
SEP[e]	0.584	0.561	0.537	0.542
r^2_{pred}[f]	0.654	0.635	0.735	0.739
PLS components[g]	6	6	6	6
Contribution:				
Steric	0.546	0.142	0.573	0.177
Electrostatic	0.454	0.525	0.427	0.460
Hydrophobic	-	0.332	-	0.363
r^2_{boot}[h]	-	-	0.897	0.841
SEE_{boot}[i]	-	-	0.304	0.382
r^2_{LHO}[j]	-	-	0.682	0.675
SD_{LHO}[k]	-	-	0.005	0.011
r^2_{5cv}[l]	-	-	0.622	0.641
SD_{5cv}[m]	-	-	0.036	0.033

[a] Non-cross-validated correlation coefficient. [b] Standard error of estimate. [c] Ratio of r^2_{ncv} explained to unexplained = $r^2_{ncv}/(1-r^2_{ncv})$. [d] Crossvalidated correlation coefficient after the leave-one-out procedure. [e] Standard error of prediction. [f] Predicted correlation coefficient for the test set of compounds. [g] Optimal number of principal components. [h] Average of correlation coefficient for 100 samplings using the bootstrapped procedure. [i] Average standard error of estimate for 100 samplings using the bootstrapped procedure. [j] Average cross-validated correlation coefficient for 50 runs using the leave-half-out (LHO) group. [k] Standard deviation of average cross-validated correlation coefficient for 50 runs. [l] Average Crossvalidated correlation coefficient for 50 runs using the five cross-validation group. [m] Standard deviation of average cross-validated correlation coefficient for 50 runs.

Figure captions

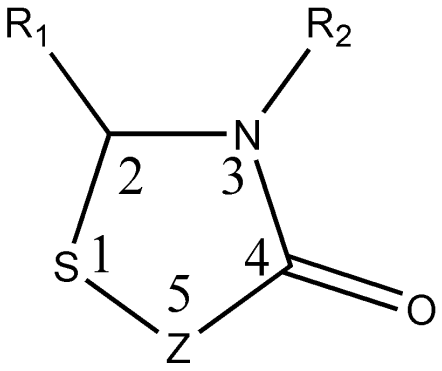
Figure 1. (a) General structure of 2,3-diaryl-1,3-thiazolidin-4-one (b) Compound 64**,** (c) Crystal structure of 7-chloro-1-(2,6-difluorophenyl)-1H,3H-thiazolo[3,4-a]benzimidazole. (d) Alignment pose of thiazolidinones (Table 1) from the global minima conformation, (e) Alignment pose of thiazolidinones (Table 1) from the crystal structure conformation.

Figure 2. Plot of observed versus predicted activities for the training (△) and test set (✹) compounds based on the (a) CoMFA model 3 and (b) CoMSIA model 4.

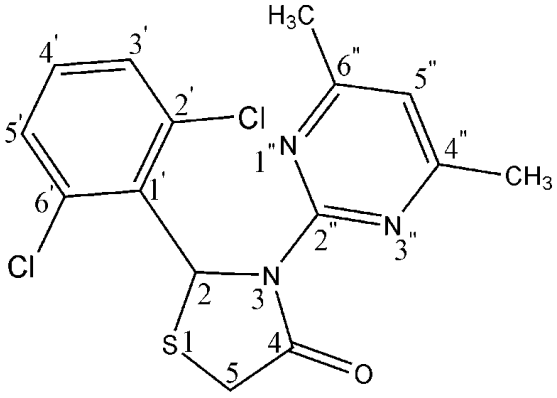
Figure 3. CoMFA contour map of (a) steric and (b) electrostatic fields (StDev X coeff) generated from the crystal structure based model. Color-coding is as follows. Green indicates regions where bulky group increase activity, whereas yellow indicates regions where bulky group decrease activity. Blue indicates that a positive charge favors high affinity, where as red indicates that a negative charge favors high affinity. The contours were mapped on to the active site of HIV-1 RT and compound 64 was displayed.

Figure 4. Mapping of crystal structure based CoMSIA model (StDev X coeff) hydrophobic contour map within the active site of HIV-1 RT/compound 64. Color-coding is as follows. Orange indicates regions where hydrophobic groups increase the activity, whereas white indicates regions where hydrophobic groups decrease the activity.

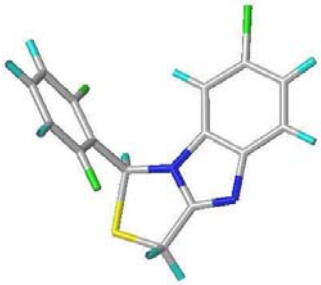
Figure 1



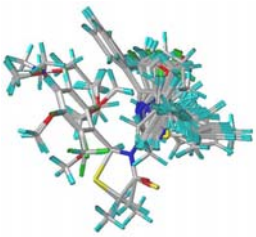
(a)



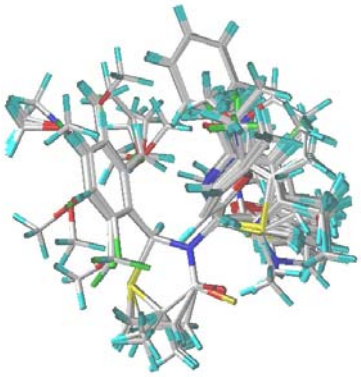
(b)



(c)

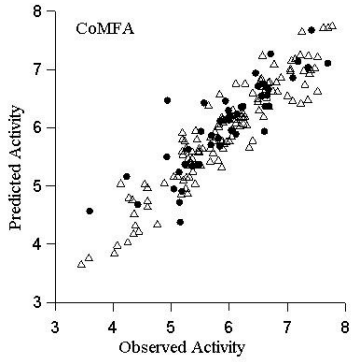


(d)

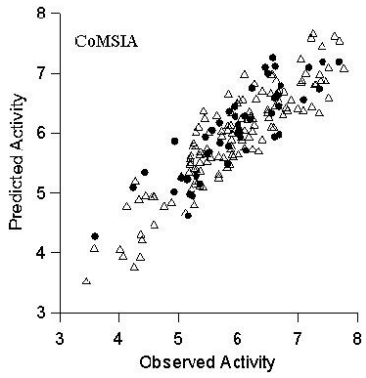


(e)

Figure 2

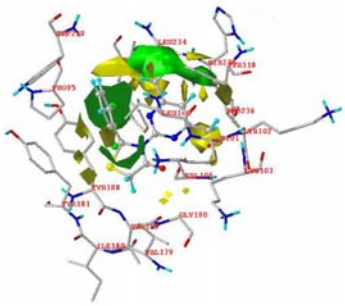


(a)

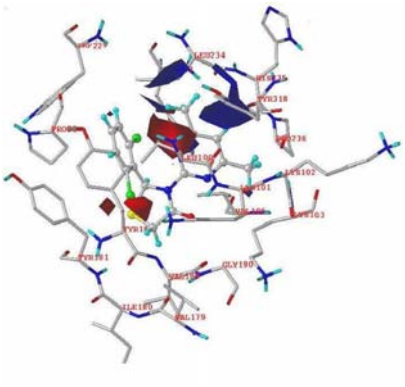


(b)

Figure 3



(a)



(b)

Figure 4