

Research Journal of Pharmaceutical, Biological and Chemical Sciences

3D-QSAR studies on substituted Purines as HIV-1 TAR inhibitors, A Non-trivial target of HIV-1.

Saikiran Reddy Peddi, Janaiah Chevula, Sree Kanth Sivan, and Vijjulatha Manga*.

Molecular Modeling and Medicinal Chemistry Group, Department of Chemistry, University College of Science, Osmania University, Hyderabad –500007, Telangana, India.

ABSTRACT

HIV-1 TAR RNA is one of the most prominent RNA targets which play an important role in the life cycle of HIV. Use of a selective inhibitor that blocks the interaction of TAR and the virus-encoded protein Tat, which regulates RNA transcriptase processivity, is one possible way in keeping the virus from proliferating. In order to identify the key structural requirements for enhancing HIV-1 TAR inhibitory activity, a series of substituted purine derivatives were selected to establish a 3D-QSAR model using Comparative Molecular Similarity Indices Analysis (CoMSIA) method. Firstly, the lowest energy conformer of most potent compound was used as template for molecular superimposition of all ligands by atom based RMS fit method. Secondly, CoMSIA fields were calculated using SYBYL-X 2.1. Thirdly, all compounds were divided into training set and test set, a PLS analysis was performed and finally a reliable QSAR model was generated. The model yielded satisfactory results with good statistical reliability evident from leave-One-Out cross validated correlation coefficient (q^2_{loo}) of 0.524, non-cross-validated correlation coefficient, (r^2_{ncv}) of 0.968 and predicted correlation coefficient (r^2_{pred}) of 0.502. These results ensure that the obtained CoMSIA model can be used as tool to guide the design of novel potent HIV-1 TAR antagonists.

Keywords: HIV-1 TAR RNA, 3D-QSAR, CoMSIA, Purine derivatives.

**Corresponding author*

INTRODUCTION

According to World Health Organization estimates, 42 million people are living with HIV/AIDS worldwide. With growing resistance of the retrovirus HIV-1 (human immunodeficiency virus, type 1) to current drugs, there is need for research on other HIV-1 targets. The HIV-1 genomic RNA itself presents possibilities. One of the notable RNA targets is the HIV-1 TAR RNA motif, which plays an important role in the life cycle of HIV [1, 2]. The transcriptional transactivation of the HIV-1 genome requires a specific interaction between the highly conserved TAR RNA hairpin fragment with the viral Tat protein and cellular factors (PTEFb-cyclin T1-CDK9 kinase complex). Six-nucleotide loop and the three-nucleotide bulge of TAR RNA are involved in the formation of this complex [3, 4]. Therefore, compounds that can bind to the bulge or the loop of TAR are of great therapeutic interest since disruption of the ternary complex formation leads to abortive mRNA synthesis and consequently inhibits viral replication. A number of recent studies have identified small-molecule ligands that bind to TAR RNA, thus inhibiting Tat binding or Tat transactivation [5-8]. Recently, a series of purine derivatives were reported as potent inhibitors [9, 10]. Several computational approaches are employed in development and optimization of inhibitors. In present article we report atom based 3D QSAR study using CoMSIA [11] methodology on substituted purine derivatives. Partial least square (PLS) analysis [12] was carried out and contour maps were generated subsequently. The PLS analysis in combination with contour maps enabled us to identify the correlation and explain the observed variation in activity respectively.

METHODOLOGY

A dataset comprising 25 substituted purine derivatives were taken from the literature [9, 10]. The dataset was divided into training and test set of 19 and 6 compounds respectively as shown in Table 1. All further studies were performed using the same training and test sets. The EC_{50} values were converted into pEC_{50} ($-\log EC_{50}$) for use in 3D-QSAR analysis. Structural sketches and refinement of the entire set of HIV-1 TAR antagonists were accomplished using SYBYL-X 2.1 molecular modeling software [13] and their 3D structures were generated using the same. All compounds were minimized under the Tripos standard (TS) force field [14] with Gasteiger-Hückel atomic partial charges [15]. Minimizations were done using conjugate gradient method, in which calculations were set to terminate at an energy gradient value of 0.001 Kcal/mol. The best active compound among substituted purines was taken as template (compound 1) and subsequently all other compounds were aligned with this template using SYBYL-X 2.1 (Figure 1). The other parameters for calculation were set to default values.

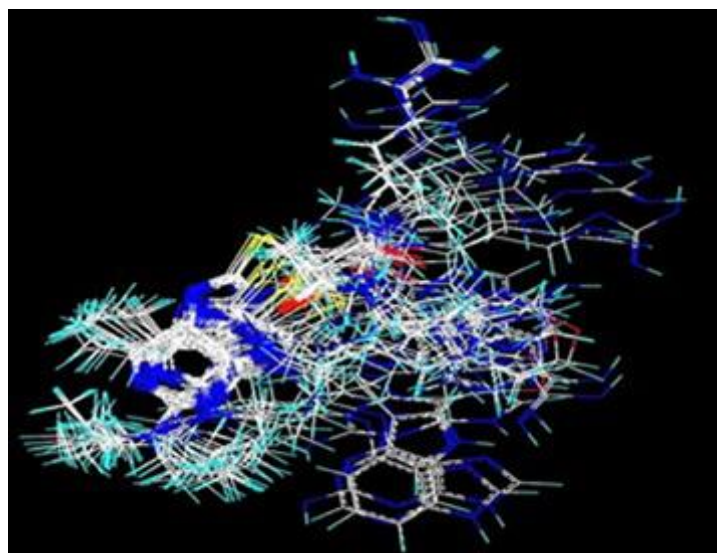
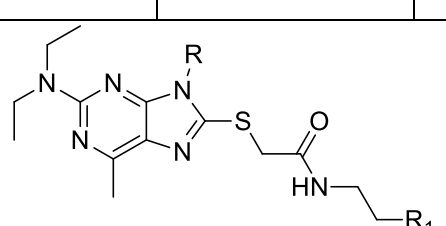
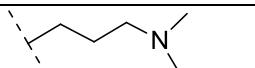
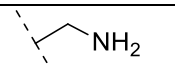
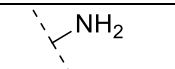
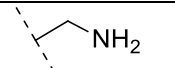
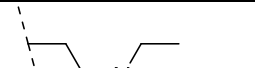
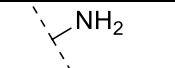
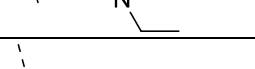
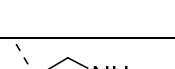
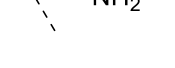
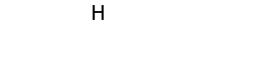
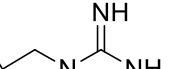
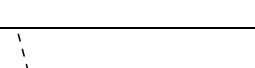
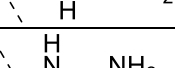
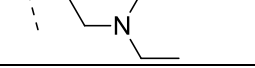
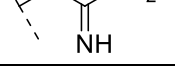
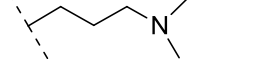
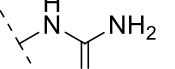
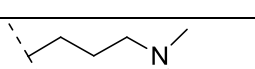
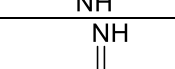


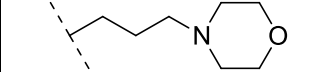
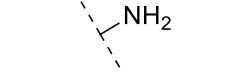
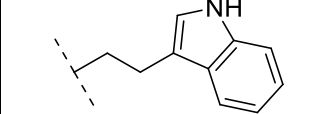
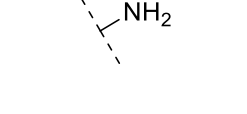
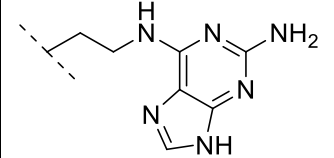
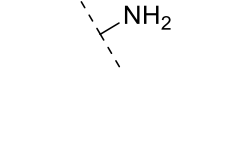
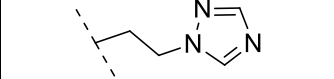
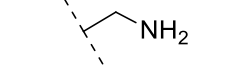
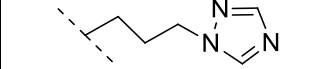
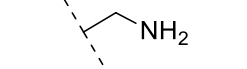
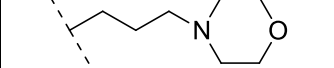
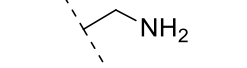
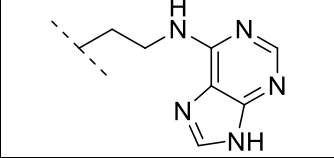
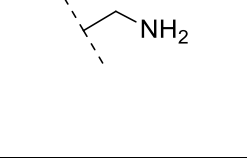
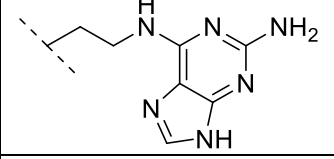
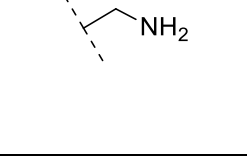
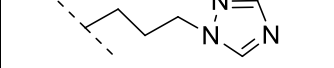
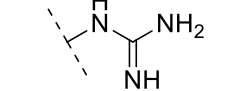
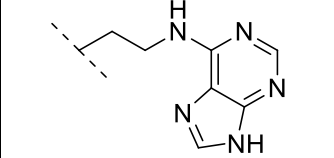
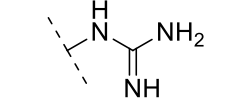
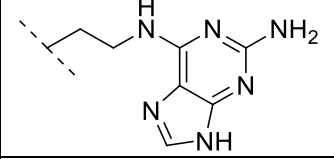
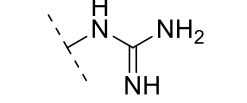
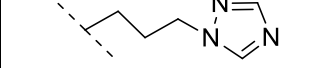
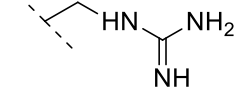
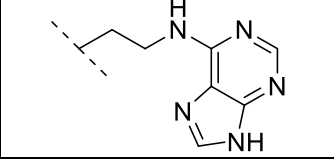
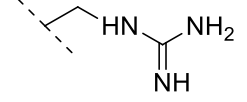
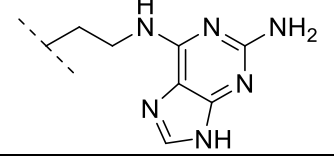
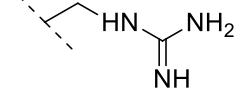
Figure 1: Atom based alignment of substituted purine derivatives.

The three dimensional Quantitative Structure Activity Relationships (3D-QSAR) play a crucial role in drug discovery and design. Like any other popular QSAR methods, CoMSIA studies also searches for sites on molecules capable of being modified into specific ligands with better activity. A Gaussian-type function based on distance is employed in CoMSIA so as to avoid enormous changes in the potential energy of the grid points

near the molecular surface. Hence unlike CoMFA, CoMSIA generates more stable models [16-18]. In present study, the influence of steric, electrostatic, hydrophobic and hydrogen-bonding interactions on inhibitory activity was explored by employing CoMSIA methodology using partial least square analysis (PLS). A 3D cubic lattice of dimension 4 Å was created and at each lattice intersection of a regularly spaced grid of 2.0 Å steric and electrostatic potential fields were calculated. A sp^3 hybridized carbon atom with +1 charge was used as probe atom to calculate steric (Lennard-jones potential) and electrostatic (coloumbic potential) field energies. These fields were truncated at +30.0 Kcal/mol. Similarly the rest of the descriptors in CoMSIA i.e., Hydrophobic, Hydrogen bond donor and Hydrogen bond acceptor fields were calculated with attenuation factor of 0.3.

Table 1: Structures of substituted purines along with their Experimental and Predicted pEC_{50} values.

Compound	R	R ₁	EC ₅₀ (μ M)	pEC ₅₀	Pred pEC ₅₀ (CoMSIA)
					
1			0.2	6.6989	6.7308
2	H		2.0	5.6989	5.7452
3	H		0.5	6.3010	6.2382
4			1.3	5.8860	6.0324
5*			1.4	5.8538	6.1122
6	H		0.7	6.1549	6.1426
7			0.7	6.1549	5.9400
8			7.2	5.1426	5.3199
9			3.2	5.4948	5.3772
10*			13.8	4.8601	5.4434
11			6.9	5.1611	5.0292

12			12.4	4.9065	4.9083
13			1.6	5.7958	5.7752
14			20.0	4.6989	4.6401
15*			11.5	4.9393	5.3044
16			8.9	5.0506	5.2113
17			11.8	4.9281	5.0054
18*			5.5	5.2596	5.3035
19			6.5	5.1870	5.1737
20*			5.4	5.2676	5.1801
21			7.8	5.1079	5.0829
22			6.1	5.2146	5.1923
23			5.8	5.2365	5.2857
24*			13.0	4.8860	5.1316
25			8.3	5.0809	5.0686

*test set compounds

A linear correlation between the CoMSIA fields and the biological activity values was established using partial least squares method [19] which was performed in two stages. In first stage cross-validation analysis using the leave-one-out (LOO) method was performed so as to obtain optimum number of components and the corresponding cross-validation coefficient, q^2 [20]. The value of q^2 with lowest standard error of estimate and minimal number of components was considered and accepted. In this process column filtering value was set to 2.0 Kcal/mol so as to reduce noise and speed up the process. In second stage the final PLS model was derived using optimum number of components with no validation method [21, 22]. Finally field contribution maps were used to graphically interpret the CoMSIA results. Furthermore to validate the derived CoMSIA model predictive correlation coefficient, r^2_{pred} was calculated using the test set according to the formula,

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

Where SD is the sum of the squared deviations of each experimental value from the mean, and PRESS is the sum of the squared differences between the predicted and actual affinity values for every compound.

RESULTS AND DISCUSSION

A statistically significant 3D-QSAR model was obtained using properly selected training set of 19 ligands. Results of the statistical analysis are presented in Table 2. In the CoMSIA model, initial PLS analysis of the aligned training set was done using a default (σ min) data filter of 2.0 Kcal/mol and the Tripos standard field. The CoMSIA study revealed a cross validated q^2 of 0.524 with optimum number of component 3, a conventional r^2 of 0.968 with a standard error of 0.109 for training set. Predicted activities and experimental values are listed in Table 1 and the correlation between the predicted activities and the experimental values is depicted in Figure 2. From Figure 2 it is evident that the activities predicted by CoMSIA model are in good agreement with the experimental data. This implies that the CoMSIA model is a reliable model with a valid predictive ability. The predictive power of CoMSIA model was further examined using a test set of 6 compounds not included in the training set. The predicted pEC_{50} values are in good agreement with the experimental data within a statistically tolerable error range, with a predicted correlation coefficient (r^2_{pred}) of 0.502. Thus the results yielded imply that the generated QSAR model can be conveniently used to design novel and more potent HIV-1 TAR antagonists.

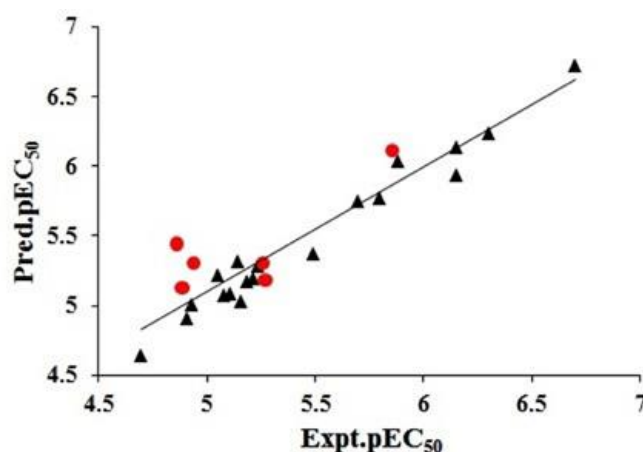


Figure 2: Scatter plot of experimental versus predicted pEC_{50} values for the CoMSIA model.

Graphical Interpretation of Fields

In order to graphically visualize the contributions of various fields, CoMSIA contour maps were generated. These contour maps help in reasoning out the differences in the biological activities of various compounds. The generation of contour maps involves various field types of $StDev \cdot coefficients$ to show the favourable and unfavourable interactions between ligands and receptors in the active site. All contours which are generated represent default 80 and 20% level contribution for favoured and disfavoured regions. For better understanding of results compound 1 is numbered and divided into Anchor, Linker and Activator regions which are depicted in Figure 3. The numbering pattern and division of other compounds is same as

that of compound 1. In the CoMSIA model, the fractions of the steric, electrostatic, hydrophobic and hydrogen-bond acceptor and donor fields were 11.4%, 14.2%, 19.4%, 20.8% and 34.2% respectively. The CoMSIA contour maps of the steric and electrostatic fields are shown in Figures 4(a), (b), (c) and (d) in which both best active compound (compound 1) and least active compound (compound 14) are displayed. In the CoMSIA steric field, the green contours indicate bulky groups that can increase the activity, while yellow contours represent a bulky group that results in a decrease in the activity. Similarly, the blue contours represent areas where the addition of less electronegative groups will enhance the biological activity, while more electronegative substituents near the red regions will enhance the biological activity. From Figure 4(a) we can see a large region of green contour near N, N-dimethyl propan-1-amine at N9 position of purine and propan-1-amine group of activator region and a large yellow contour close to green contour. However none of the groups in the compound 1 are close to this yellow contour. Therefore it is concluded that presence of bulky groups at anchor and activator regions increases the biological activity. This is due to the fact that in compound 1 the propan-1-amine group which is located in activator region is flipped and oriented towards the green contour close to N, N-dimethyl propan-1-amine group at N9 position of purine. The remaining compounds in the dataset showed decreased activities than compound 1. In Figure 4 (b) we can observe a large bulky favoured green contour close to position 9 and anchor region, a large yellow contour close to this green contour and a small green contour near activator region. Even though the contours in compound 14 are very similar to compound 1 the activity is greatly decreased. This is due to the fact that unlike in compound 1 the propan-1-amine group located in activator region in compound 14 is not flipped, orienting the bulky groups onto steric favoured green contours. In the contour map of the electrostatic field in CoMSIA model (shown in Figure 4(c)), a red contour close to position 1 of the purine moiety and a huge blue contour near anchor region (i.e., at position 2 and 9 of purine moiety) and activator region (methylene groups of propan-1-amine) indicate that an electronegative substituent at position 1 and less electronegative groups near position 2 and 9 of purine moiety as well near methylene groups of propan-1-amine can increase the activity. The electrostatic red contours of compound 14 are similar to that of compound 1 but same is not true in case of blue contours. The groups possessing electronegative atoms such as sulphur, oxygen and nitrogen atoms are protruding onto the blue contours. So there is a fall in activity in case of compound 14. Figure 5(a) displays a CoMSIA contour map overlay of hydrophobic and hydrophilic fields in combination with compound 1. The presence of grey contours near linker region (oxygen of carbonyl group and dimethyl sulfane groups) and nitrogen of activator region in compound 1 suggests that hydrophilic groups at these positions are beneficial to the activity. The hydrophobic yellow contours though present are not much effective in predicting biological activity as they are oriented far away from the linker and activator groups. Further when we observe Figure 5(a) we can notice that the first methylene group of N, N-dimethyl propan-1-amine at position 9 is penetrating into the yellow contour suggesting that a hydrophobic group at this position can increase the activity. In case of compound 14 there is a hydrophobic contour close to position 9 where there is a hydrophilic group and moreover the hydrophilic groups at linker position are onto hydrophobic contours (Figure 5(b)). Therefore the activity decreased in case of compound 14. Figures 5(c) and (d) depict the contribution of donor fields while 5(e) and (f) depict the contribution of acceptor fields in enhancing the biological activity. When we observe the Figure 5(c) we can notice the activator group along with its chain has flipped away from the donor fields and oriented towards substituent at position 9 of purine moiety in compound 1. Hence much information cannot be explored though contribution of donor fields is significant but presence of cyan contours close to substituent at position 9 of purine moiety indicates that substituting a donor group in this region will increase the activity. In case of compound 14 this type of flipping is not observed. In compound 14 the hydrogen bond donor NH group in activator region is close to hydrogen bond donor disfavoured violet contour and the carbonyl group which is a hydrogen bond acceptor in the same region is close to hydrogen bond donor favoured cyan contour. Hence the activity is decreased in case of compound 14. Large hydrogen bond acceptor contours are noticed in both compound 1 and 14 which are depicted in Figure 5(e) and (f). From Figure 5(e) it is observed that there is a large magenta contour in compound 1 near purine moiety surrounding N9 position indicating that hydrogen bond acceptor groups at this position are beneficial to the activity where as a moderate red contour which is near the activator region in compound 1 suggests that hydrogen bond donor groups at this region can increase the activity. In case of compound 14 the activity is decreased though most of the hydrogen bond acceptor groups in anchor region are in close proximity to hydrogen bond acceptor favoured violet contours. This is due to the fact that hydrogen bond donors in linker and activator regions are in close proximity to hydrogen bond acceptor favoured violet contours. Compounds 3, 6 and 7 exhibited good activity values in range of 6.155-6.301 which are close to best active compound. The reason behind such good activity values is that all the contours in these compounds are very similar to compound 1 even though there are no substituents at position 9 in compounds 3 and 6. The only exception was a major part of activator

region in compound 3 and a major part of anchor region in compound 6 were close to green contour. Compounds 2, 4, 5 and 13 showed slight decreased activities (5.69-5.88) to compounds 3, 6 and 7. Absence of bulky aryl group at position 9, mixed features of compounds 14 and 1, hydrogen bond donor contours similar to compound 14 and hydrogen bond acceptor contours similar to compound 14 in compounds 2, 4, 5 and 13 respectively contributed to decreased activity values though most of the features resembled compound 1. Compounds 10, 12, 15, 17 and 24 showed very low activity values in range of 4.86-4.93 close to least active compound. The orientation of activator in compound 15, linker in compound 12 and a part of linker and activator in compound 17, orientation towards disfavoured yellow contours has contributed to their decreased activity values. Moreover the similarity of the contours in these compounds with that of least active compound (compound 14) has drastically affected their biological activities. Furthermore presence of hydrophilic contour across triazole ring in compound 15 and near purine ring at position 9 in compound 24 were one of the reasons for their low activity values. The remaining compounds i.e., compounds 8, 9, 11, 16, 18, 19, 20, 21, 22, 23 and 25 in the series showed moderate biological activity values ranging from 5.05-5.49 as they exhibited the mixed characteristics of both best active and least active compounds. The results of all the contours are summarized in Figure 6. On the whole the detailed contour map analysis helped us in understanding the key structural features responsible for observed inhibitory activity.

Table 2: Statistical results of CoMSIA model.

Statistical Parameters	CoMSIA
q_{loo}^2 ^a	0.524
Number of compounds in training set	19
Number of compounds in test set	6
ONC ^b	3
SEE ^c	0.109
r^2 ^d	0.968
F_{ratio} ^e	151.384
r^2_{pred} ^f	0.502
Fraction of fields contributions	
Steric	11.4 %
Electrostatic	14.2 %
Hydrophobic	19.4 %
Acceptor	20.8 %
Donor	34.2 %

- a – Cross-validation correlation coefficient by leave one out method
- b – Optimum number of principal components
- c – Standard error of estimate
- d – Non cross-validated correlation coefficient
- e – Fisher test value
- f – Predictive correlation coefficient

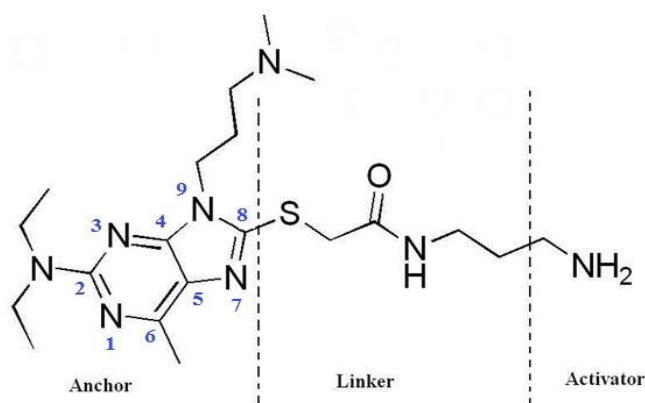


Figure 3: Structure of best active compound showing Anchor, Linker and Activator regions.

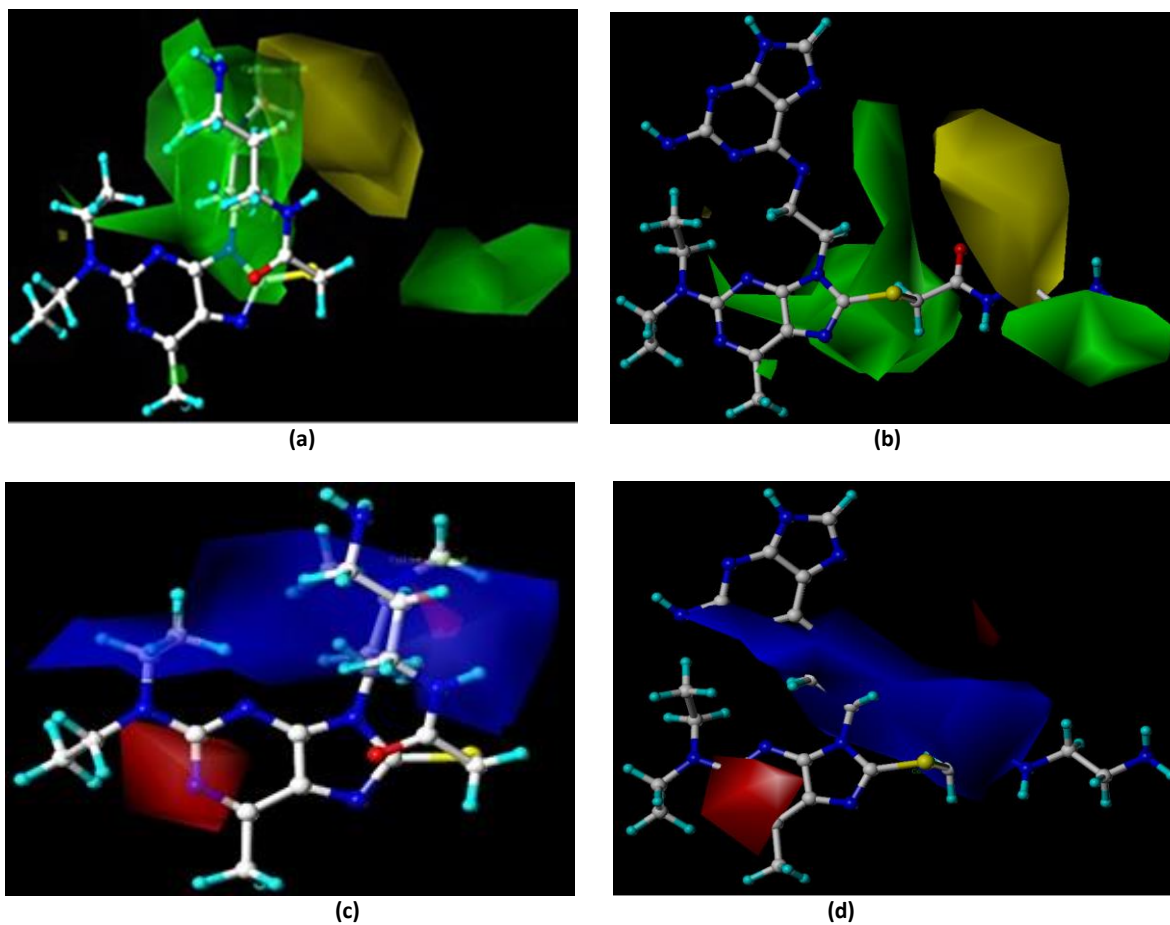
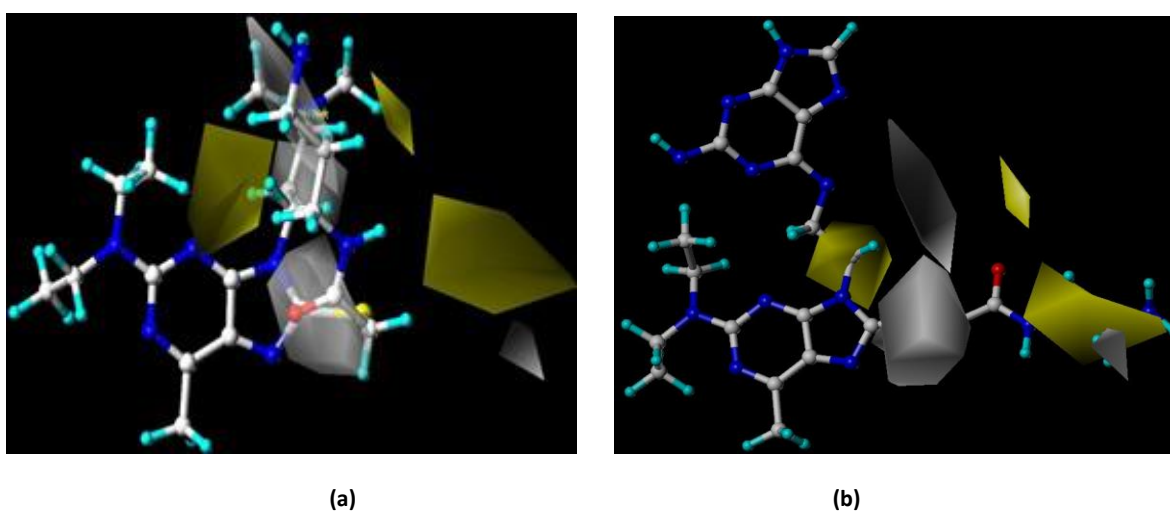


Figure 4: CoMSIA STDEV*COEFF contour maps showing steric and electrostatic features in combination with compound 1 (best active) and 14 (least active) respectively. Figures (a) and (b) represent favourable (green) and unfavourable (yellow) steric fields while Figures (c) and (d) represent favourable (red) and unfavourable (blue) electronegative fields.



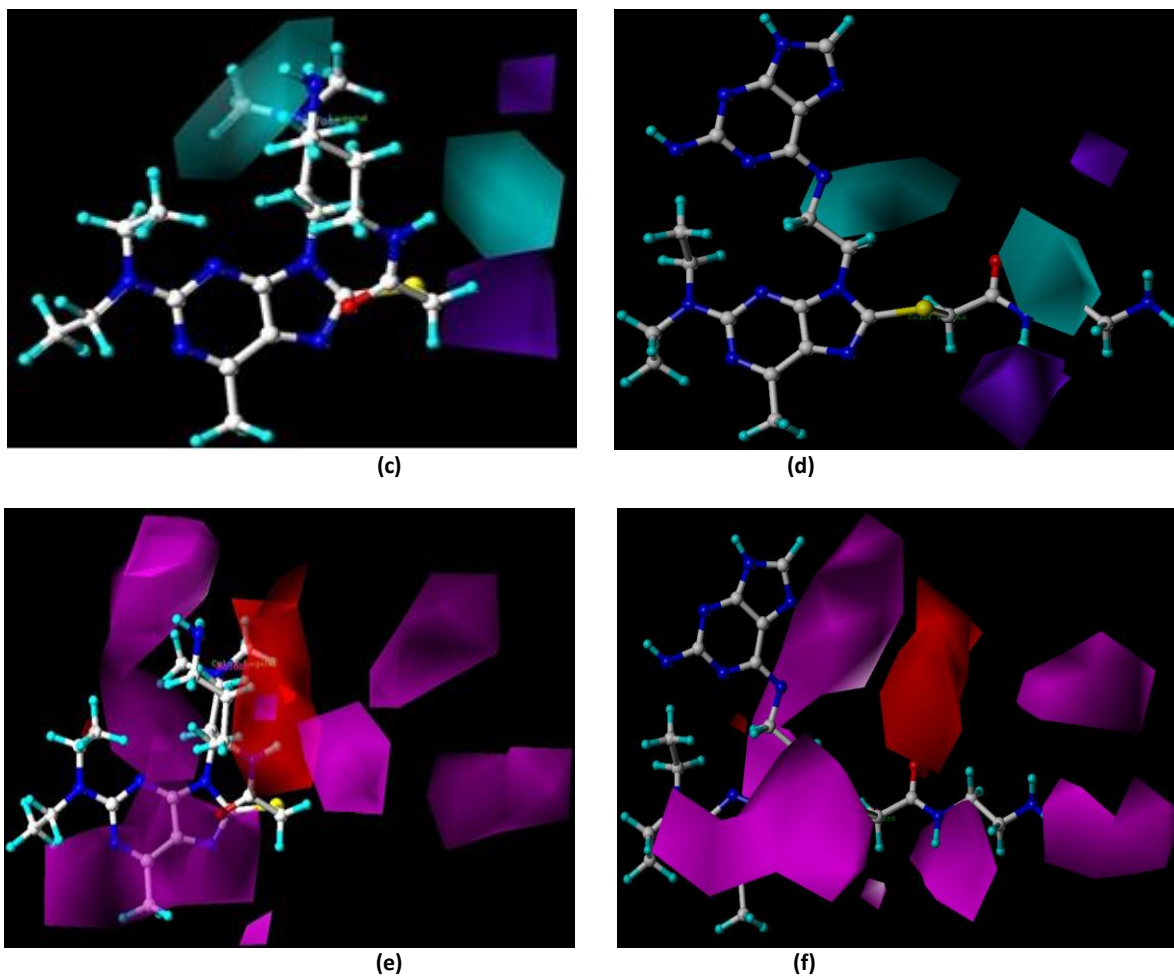


Figure 5: CoMSIA STDEV*COEFF contour maps showing hydrophobic, hydrogen bond donor and hydrogen bond acceptor features in combination with compound 1 (best active) and 14 (least active) respectively. Figures (a) and (b) represent favourable (yellow) and unfavourable (grey) hydrophobic fields, Figures (c) and (d) represent favourable (cyan) and unfavourable (purple) hydrogen bond donor fields and Figures (e) and (f) represent favourable (magenta) and unfavourable (red) hydrogen bond acceptor fields respectively.

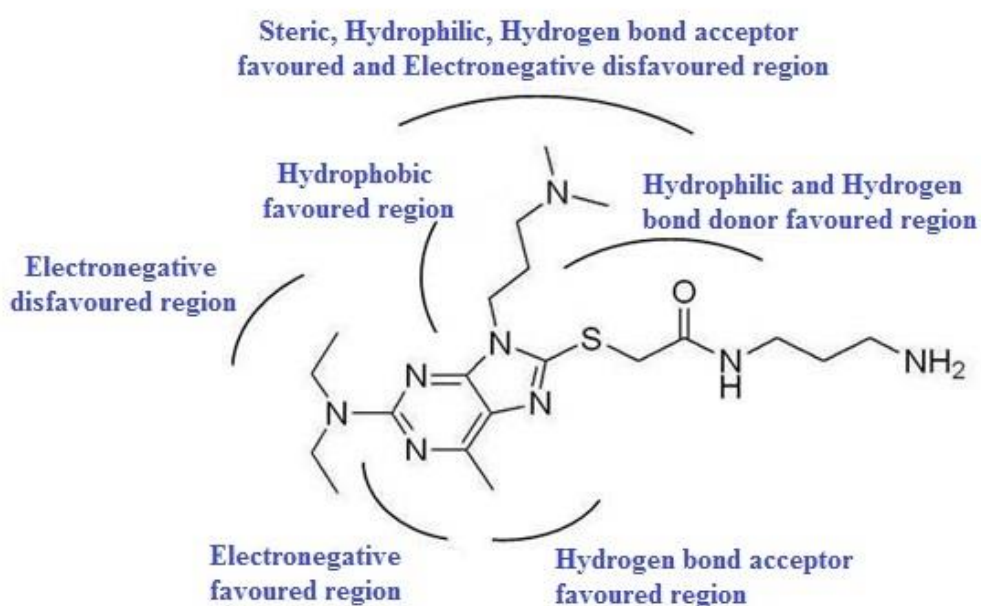


Figure 6: Structural requirements for binding and inhibitory activity of substituted purine derivatives.

CONCLUSIONS

In the present study a robust CoMSIA model has been built using SYBYL-X 2.1 and its predictive ability was validated using a set of 6 compounds. The model yielded good q^2 , r^2 and r^2_{pred} values with small standard error of estimation. Contour maps revealed that inhibitory activity can be improved by substituting bulky hydrophilic hydrogen bond acceptor groups at N9 position of purine moiety and hydrogen bond donor groups at linker and activator regions. The activity can be further improved by substituting more electronegative groups at position 1 and less electronegative groups at position 2 of purine moiety. Furthermore hydrogen bond acceptor groups at position 6 of purine can be beneficial to the activity. The aryl groups at N9 position could affect the biological activity but due to steric hindrance and orientation of these aryl groups towards sterically disfavoured yellow contours in most of the compounds did not significantly increased the activity values. On the whole it is concluded that N9 position of purine moiety plays a crucial role in enhancing the activity. The outcome of the results can be further exploited to design and synthesize more active compounds with new scaffolds.

ACKNOWLEDGEMENTS

We gratefully acknowledge Tripos Inc., USA for providing the software. This research was made possible through grants from DST-SERB (SB/EMEQ-004/2013). The author Saikiran Reddy Peddi (IF150172) would like to thank DST for providing Inspire Fellowship.

REFERENCES

- [1] Aboul-ela F, Karn J, Varani G. *J Mol Biol* 1995; 253(2): 313–332.
- [2] Cheng AC, Calabro V, Frankel AD. *Curr Opin Struct Biol* 2001; 11(4): 478–484.
- [3] Zhou Q, Chen D, Pierstorff E, Luo K, The EMBO 1998; 17(13): 3681–3691.
- [4] Richter S, Cao H, Rana TM. *Biochemistry* 2002; 41(20): 6391–6397.
- [5] Lind KE, Du Z, Fujinaga K, Peterlin BM, James TL. *Chem Biol* 2002; 9(2): 185–193.
- [6] Renner S, Ludwig V, Boden O, Scheffer U, Gobel M, Schneider G. *Chembiochem* 2005; 6(6): 1119–1125.
- [7] Hwang S, Tamilarasu N, Kibler K, Cao H, Ali A, Ping YH, Jeang KT, Rana TM. *J Biol Chem* 2003; 278: 39092–39103.
- [8] Davis B, Afshar M, Varani G, Murchie AI, Karn J, Lentzen G, Drysdale MJ, Bower J, Potter AJ, Starkey ID, Swarbrick T, Aboul-ela F. *J Mol Biol* 2004; 336(2): 343–356.
- [9] Yuan D, He M, Pang R, Lin S, Li Z, Yang M. *Bioorg Med Chem* 2007; 15: 265–272.
- [10] Pang R, Zhang C, Yuan D, Yang M. *Bioorg Med Chem* 2008; 16: 8178–8186.
- [11] Klebe G, Abraham U, Mietzner T. *J Med Chem* 1994; 37(24): 4130–4146.
- [12] Wold S, Johansson A, Cochi M. *PLS-Partial least squares projection to latent structures*. Lieden, ESCOM, 1993, pp. 523-550.
- [13] SYBYL-X 2.1, Tripos Associates, St. Louis (MO), 2013.
- [14] Clark M, Cramer RD III, Opdenbosch NV. *J Comput Chem* 1989; 10(8): 982–1012.
- [15] Gasteiger J, Marsili M. *Tetrahedron* 1980; 36(22): 3219-3228.
- [16] Cramer RD, Patterson DE, Bunce JD. *J Am Chem Soc* 1988; 110(18): 5959–5967.
- [17] Klebe G, Abraham U, Mietzner T. *J Med Chem* 1994; 37(24): 4130–4146.
- [18] Klebe G, Abraham U. *J Comput Aided Mol Des* 1999; 13(1): 1–10.
- [19] Wold S. *Quant Struct-Act Relat* 1991; 10: 191–193.
- [20] Shao J. *J Am Stat Assoc* 1993; 88: 486–494.
- [21] Cruciani G, Baroni M, Clementi S, Costantino G, Riganelli D, Skagerberg B. *J Chemom* 1992; 6(6): 335–346.
- [22] Baroni M, Clementi S, Cruciani G, Costantino G, Riganelli D, Oberrauch E. *J Chemom* 1992; 6(6): 347–356.