

## A LINEAR QSAR AND DOCKING APPROACH TO MODEL INHIBITORY ACTIVITY OF 2-ARYLBENZOXAZOLE DERIVATIVES AS CHOLESTERYL ESTER TRANSFER PROTEIN (CETP) INHIBITORS

MUKESH YADAV<sup>1\*</sup>, ANURAJ NAYARISSERI<sup>1</sup>, SHOBHA JOSHI<sup>2</sup>, ANKITA GUPTA<sup>3</sup>, PURVA HOLKAR<sup>4</sup>, ASHISH RAJPUT<sup>5</sup>, RISHABH JAIN<sup>5</sup>

<sup>1</sup>Bioinformatics Research Laboratory, Eminent Biosciences, Indore Madhya Pradesh, India

<sup>2</sup>Govt. Degree College, Sanwer, Indore, Madhya Pradesh, India

<sup>3</sup>Department of Bioinformatics, Banasthali University, Rajasthan, India

<sup>4</sup>Dept. Biotechnology and Bioinformatics Sub-Centre, Devi Ahilya Vishwavidyalaya, Indore, India

<sup>5</sup>Pravara Institute of Medical Science, Pravara Rural University, Maharashtra, India

\*Corresponding author. Email: [mukesh@eminentbio.com](mailto:mukesh@eminentbio.com), Contact no. +91 731 4256149

Received: January 23, 2011; Accepted: March 31, 2011

**Abstract-** QSAR studies of thirty 2-arylbenzoxazole derivatives are carried out to probe their inhibitory activity against Cholesteryl Ester Transfer Protein (CETP). QSAR models have been obtained using multiple linear regression (MLR) analysis after manifestation of forward selection algorithm to cull significant descriptors out of descriptor's pool. QSAR models are elected with adherent set of statistical parameters with  $R^2=0.9431$  and  $R^2=0.9069$ . Validation of modeling includes method of Y-Scrambling and in addition to this, some other methods of validation. Moreover, QSAR approach of 2-arylbenzoxazoles are also attempted, supported and validated by flexible docking studies as well. The search strategies include evolutionary algorithm and edited form of Gehlhaar PLP scoring function. The same set of thirty candidates from 2-arylbenzoxazole derivatives is processed in molecular docking and their docking scores are found in agreement with QSAR studies reported. Remarkable CETP inhibitory activity is exhibited by a 2-arylbenzoxazole derivatives molecule 29C with most comprehensive set of interactions with rerank scores -102.167 and RMSD values 1.104. Hydrogen bond interactions along with hydrophobic and electrostatic interactions are mapped to confirm their potencies.

**Key words** - QSAR, 2-arylbenzoxazole, Cholesteryl Ester Transfer Protein (CETP), Docking

### Introduction

Chemoinformatics has added a new advancement in drug discovery protocols by availing predictions from lead identification to lead optimization. Computational and statistical studies have promoted approximations towards biological activity profiles of candidates to save time and economical issues of efforts in finding biologically active candidates against CETP inhibitors. Computational forecasting uses the basic laws of molecular mechanics and molecular dynamics ruled by thermodynamical parameters to solve structural aspects and secondary valency forces to map drug receptor interactions.

Recently, coronary heart diseases have appeared as a serious health problem to human kind. They are reported mainly due to the plaque (made up of triglycerides) formation in the coronary arteries, obesity, hypertension, diabetes, smoking, high stress level, insufficient physical activity, high cholesterol level and in some cases of women menopause [1-8]. Apart from these, other common factors are also known to increase possibility of

coronary heart diseases. Cholesteryl ester transfer protein (CETP) is mainly responsible for reverse cholesterol transportation (transportation of triglycerides from peripheral tissues including coronary arteries to the liver) [9-10]. CETP has inverse affinity with high-density lipoprotein (HDL) cholesterol. Many researchers have reported HDL and its role in beneficial properties as anti-oxidant and in anti-atherogenicity. Clinical trials have revealed that CETP is considerably inhibited by the drugs mainly dalcetrapib, torcetrapib, and anacetrapib [11-15]. Inhibition of CETP increases the level of HDL cholesterol which further helps to set reverse transport of cholesterol. Unfortunately torcetrapib was failed in clinical trial at phase III and others were not so potent to increase the HDL level. Cameron (2009) explained that the arylbenzoxazole derivatives can be potent drug targets to inhibit CETP.

Quantitative structure activity relationship (QSAR) has proved its usefulness in chemoinformatics. QSAR studies establish relationship of structural

features in terms of molecular descriptors with biological activity[16], which further assist drug design community to synthesize new molecules with optimized structures of desired biological activity. Molecular docking is another field of cheminformatics which helps in computational virtual screening to find the lead compounds. Molecular docking started with Fischer's lock and key theory, where, every receptor has its unique ligand to catalyse the reaction [17].

## Methodology and experimental

### QSAR Studies

Smith *et al* [18] in 2010 reported a series of derivatives of 2-arylbenzoxazole inhibitors of the cholesterol ester transfer protein (CETP). They studied structure–activity relationship by substituting various functional groups at the 5- and 7-positions of the benzoxazole moiety, which were further found to be effective for CETP inhibition. Thirty structures (H explicit 2D and 3D) of 2-arylbenzoxazole group have been drawn in accordance to schemes A, B and C of the basic structures, using Marvin Sketch (a program developed by ChemAxon Company) [Table-1]. Biological activity has been adopted in its original measures (IC50 in  $\mu\text{M}$ ) from the literature [18]. Calculations of molecular descriptors namely SRW09 (self-returning walk count of order 09 from Walk and Path Indices), G3S (3st component symmetry directional WHIM index / weighted by atomic electrotopological states), Yindex (Balaban Y index), H-047 (H attached count to C (SP<sup>3</sup>) or atom centred fragments) and HATSM (Leverage-weighted autocorrelations) have been performed using E-DRAGON, an electronic version of DRAGON (developed by VCC Labs) table-1. To find the optimal subset of the descriptors which play an important role in determining activity, method of variable selection is required. Chances of redundancy are checked in correlation matrix [Table-2].

Two descriptors have been found in close correlation with each other, in step-wise regression analysis on passing from one variable to penta-variable regression we obtained two models 4(a) and 4(b) for tetra-variant set of descriptors. We have used forward selection algorithm [19] with targeting descriptor size five [Table-3] to pull out significant descriptor combinations so as to achieve maximum R<sup>2</sup> in multiple regression analysis. We have selected tetra-variable and penta-variable models with excellent statistics R<sup>2</sup> values are 0.9223 and 0.9529 respectively (Eq. 1 and 2). We have, therefore, examined these two models in detail.

#### MODEL- 4 (b)

$$\text{IC50} = 89.975 - 0.068 (\pm 0.013) \text{SRW09} - 36.373 (\pm 8.054) \text{G3s} - 56.124 (\pm 14.806) \text{Yindex} + 0.319 (\pm 0.078) \text{H-047} \dots \text{Eq. (1)}$$

$$N=30, R^2=0.9223, R^2A= 0.9098, S.E.=0.3196, F=74.1870$$

Here and thereafter N is the number of compounds used, S.E. is standard error, R<sup>2</sup> is multiple regression coefficient, R<sup>2</sup>A is adjusted R<sup>2</sup> and F is Fisher's statistics.

#### MODEL- 5

$$\text{IC50} = 101.291 - 0.083 (\pm 0.011) \text{SRW09} - 44.445 (\pm 6.594) \text{G3s} + 3.398 (\pm 0.826) \text{HATS4m} - 46.622 (\pm 11.801) \text{Yindex} + 0.382 (\pm 0.063) \text{H-047} \dots \text{Eq. (2)}$$

$$N=30, R^2=0.9529, R^2A=0.9431, S.E.= 0.31964, F=97.1503$$

### Validation of QSAR Models

A statistically significant model with parameters in required agreement does not guarantee to predictive powers as well. Once QSAR models are qualified on the basis of acceptable values of R<sup>2</sup>, R<sup>2</sup>A, S.E. and F-statistic, it needs to be evaluated for its optimum predictive powers using other validation parameters to fall in line. Validation of present QSAR study is attributed to various validation parameters, like to develop a correlation of estimated (predicted) and experimental (observed) biological activity, method of Y-scrambling, residual sum of squares (RSS), model sum of squares (MSS), and Pogliani's factor "Q". QSAR analysis and the suitability of the proposed models are predicted by calculating and comparing estimated (predicted) and experimental (observed) biological activity, using models 4b and 5 such a comparison is shown in [Table-4]. Perusal of residue of this comparison confers to good predictive powers of the proposed models [Fig-1 and 2].

Validation by Y-scrambling method is a strong validation parameter and its successful use has been reported by many researchers. Statistical methods are susceptible towards producing good results as by chance [20]. Therefore, in order to validate any chance of modeling by chance we processed our models for Y-scrambling validation. Using the models 4(b) and 5 (eq. 1 and 2) respectively, validation leads to result that in 10 iterations after shuffling endpoint values. Thus statistical significant models (4b and 5) have also been successfully validated by this measure of Y-scrambling method [Fig-3 and 4].

In addition, all statistical parameters yielded satisfactory numerical measures within limits and favored the predictive power of the QSAR modeling using the selected equations 1 and 2. All the

validation parameters as mentioned above RSS, MSS and Q, have been calculated and the values reported are found in support to the predictive powers of the proposed models [Table-5].

### Docking Studies and Validation

Now a days docking is frequently used to predict the binding orientations of small molecules of drug candidates to their protein targets in order to predict the affinity of the small molecule [21]. Molecular docking is performed using Molegro Virtual Docker (MVD). The 3D structure of CETP was retrieved from Protein Data Bank (PDB ID 2OBD) in pdb file format [22]. The source of this protein structure is *Homo sapiens* and its structure determination was achieved by Qiu, X. *et al* using X-ray crystallography. We have selected thirty molecules of 2-arylbenzoxazole derivative as inhibitors for molecular docking.

Previous studies revealed that CETP has unpaired sulfhydryl residue near the active site region. Cysteine residue is present at position 333 and is unpaired. We detected 5 cavities using probe with parameters of minimum cavity volume= 10 and probe size= 1.20. The cysteine residue was found near the active site which has volume 1437 (Å<sup>3</sup>) [23].

Validation of accuracy of docking studies is reported to associate with redocking protocols of the candidates in the active site of CETP through redocking [24]. Redocking is considered as an important step in refinements of Root Mean Square Deviation (RMSD) values which hereby signifies proper orientation (pose) of a ligand when compared with crystallographic structure of a binary complex involving protein and ligand. As a conclusive part of redocking we expect, generated results should yield RMSD values below 1.5 Å. Successful redocking has been performed for the selected set of thirty 2-arylbenzoxazole inhibitors and their corresponding rerank score with their respective RMSD (after redocking) have been produced in the table-6.

### Results and Discussions

Statistically significant models are proposed by step-wise-multiple regression analysis following the method of maximum R<sup>2</sup>. Regression if performed on SRW09, G3S, Yindex, H-047 and HATS4m molecular descriptors in accordance with rule of thumb.

#### Model-4B

A close look to this model (eq. 1) indicates a very significant value of R<sup>2</sup> (R<sup>2</sup>=0.9223) as well as R<sup>2</sup>A (R<sup>2</sup>A=0.9098). The model reveals that the coefficient of parameters SRW09, G3S and Yindex are negative, it means that the activity IC<sub>50</sub> is inversely related to these descriptor. On the other hand coefficient of H-047 is positive indicating that it is directly related to the activity IC<sub>50</sub>, it means that

increase in values of this will increase the value of IC<sub>50</sub>.

#### Model-5

Addition of descriptor HATS4m to model-4b has further improved the statistics R<sup>2</sup> and R<sup>2</sup>A to 0.9529 and 0.9223 respectively. The quality of model with addition to descriptor HATS4m is also favored by F-ratio 97.150 when compared to model 4b with F-ratio 74.1870. The nature of coefficients for all descriptors remains same as in model 4b and HATS4m shows positive coefficient correlation.

To Study the discussion on the basis of results obtained as above presence or advancement of self returning walks of different length (SRW09), electrotopological state in terms of whim descriptor (G3S) and Balaban informations related to neighbour degree and edge multiplicity (Yindex) favour biological activity of 2-arylbenzoxazole inhibitors towards inhibition of CETP. Structural modifications in new candidates should have given a special concern with 3D geometry and symmetry elements in terms of above descriptors. Furthermore H attached count to C (SP<sup>3</sup>) or atom centred fragments (H-047) and Leverage-weighted autocorrelations (HATSM) should be minimized in terms of structures to increase biological activity of 2-arylbenzoxazole inhibitors.

Docking studies 2-arylbenzoxazole inhibitors on the active site pocket of CETP produce promising results both in terms of rerank score and their RMSD values to figure out spatial arrangements of atoms favoring maximum interactions on behalf of secondary valency forces. Result of docking studies has proved that the molecules numbered 29C shows rerank score and RMSD value as -102.167 (1.104) respectively (Table-6). The molecule 29C has been reported with appreciable IC<sub>50</sub> values of 0.16 (µM). All the poses of molecule 29C (chosen as best in docking studies) and its interactions in the active pocket of CETP have been illustrated in figures 5, 6 and 7. New efforts in the line of synthesis of 2-arylbenzoxazole inhibitors as CETP inhibitors should be associated with symmetry and 3D geometry concerns along with presence of atoms responsible for Hydrogen bond interactions and other secondary valency forces to map receptor ligand interaction scales.

### CONCLUSION

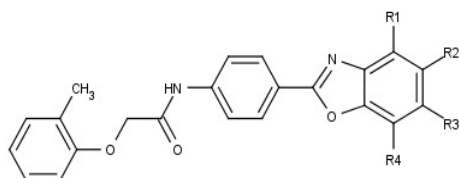
The QSAR studies and the docking approach of 2-arylbenzoxazoles as CETP inhibitors can be successfully modeled using four parametric as well as five parametric equations. Though the four parametric equation is enough to model the biological activity IC<sub>50</sub> statistically, as well as it shows good predictive powers also. However, pentavariation model (eq-2) is furnishing better predictive power besides the expected statistical values of R<sup>2</sup> and R<sup>2</sup>A. Therefore, for modeling the

QSAR in the present study eq. 2 is preferred over the eq. 1. Molecular docking of this class on the active site of CETP confirmed two amino acids namely Arg201 and Asp470 to be active participants in hydrogen bonding. Electrostatic surface further reveals stabilization of complex of molecule 29C with CETP showing attractive charges of oxygen atoms in active site pocket.

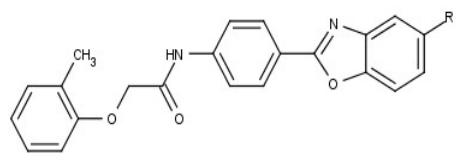
## References

- [1] Montaye M., Bacquer D.De., Backer G.De and Amouyel P. (2000) *Eur. Heart J.*, 21(10), 808–813.
- [2] Verdecchia P., Reboldi G., Angeli F., Gattobigio R., Bentivoglio M., Thijs L., Staessen J.A., and Porcellati C. (2005) *Hypertension*, 46, 386-392.
- [3] Haffner S.M., Lehto S., Ronnema T., Pyorala K., and Laakso M. (1998) *N. Engl. J. Med.*, 339(4), 229- 234.
- [4] Cullen P., Schulte H. and Assmann G. (1998) *Eur. Heart J.*, 19(11), 1632–1641.
- [5] Ahmad M., Khan M.A., Khan A.S. (2009) *J. Ayub Med. Coll. Abbottabad*, 21(2), 56-59.
- [6] Powell K.E., Thompson P.D., Caspersen C.J. and Kendrick J.S. (1987) *Annu. Rev. Public Health.*, 8, 253-287.
- [7] Shestov D.B., Deev A.D., Klimov A.N., Davis C.E. and Tyroler H.A. (1993) *Research Clinics Prevalence Follow-up Study. Circulation.*, 846-853.
- [8] Säfström K., Nielsen N.E., Björkholm A., Wiklund G., Swahn E. and the IRIS study group. (1998) *Eur. Heart J.*, 19(6), 899–907.
- [9] Rothblat G.H., Mahlberg F.H., Johnson W.J. and Phillips M.C. (1992) *J. Lipid Res.*, 33, 1091–1097.
- [10] Reichl D. and Miller N.E. (1989) *Arteriosclerosis*, 9, 785–797.
- [11] Tollefson J.H., Liu A., Albers J.J. (1988) *Am. J. Physiol.*, 225, 894-902.
- [12] Parthasarathy S., Barnett J. and Fong L.G.(1990) *Biochim. Biophys. Acta Journal.*, 1044, 275–283.
- [13] Stein E.A., Roth E.M. et al. (2010) *Eur. Heart J.*, 31, 480-488.
- [14] Millar J.S., Brousseau M.E., Diffender M.R. et al. (2006) *Arterioscler. Thromb. Vasc. Biol.* 26, 1350-1356.
- [15] Krishna R., Anderson M. S., Bergman A. J., Jin B., Fallon M., Cote J., Rosko K., Chavez- Eng C., Lutz R., Bloomfield D. M., Gutierrez M., Doherty J., Bieberdorf J., Chodakewitz F., Gottesdiener K. M., Wagner J. A. (2007) *The Lancet*, 370, 1907-1914.
- [16] Saigo H., Kadowaki T. and Tsuda K. (MLG 2006) *Proceedings of the International Workshop on Mining and Learning with Graphs* 2006, 85-96.
- [17] Selassie C.D. (2003) *Medicinal Chemistry and Drug Discovery*, 1, 1- 48.
- [18] Smith C. J., Ali A., Chen L., Hammonda M. L., Anderson M. S., Chen Y., Eveland S. S., Guo Q., Hyland S. A., Milot D. P. , Sparrow C. P., Wright S. D. and Sinclair P. (2010) *J. Bioorg. Med. Chem. Lett.*, 20, 346–349.
- [19] Kohavi R. and John G.H. (97) *Artificial Intelligence.* (1-2), 273-324.
- [20] Alexander T., Paola G. and Gombar V.K. (2003) *QSAR Comb. Sci.*, 22 (1), 69-77.
- [21] Schulz-Gasch T. and Stahl M. (2004) *Drug Discov. Today.* 1 (3), 231-239.
- [22] Qiu X., Mistry A., Ammirati M.J., Chrnyk B.A., Clark R.W., Cong Y., Culp Js., Danley D.E., Freeman T.B., Geoghegan K.F., Griffor M.C., Hawrylik S.J., Hayward C.M., Hensley P., Hoth L.R., Karam G.A., Lira M.E., Lloyd D.B., McGrath K.M., Stutzman-Engwall K.J., Subashi A.K., Subashi T.A., Thompson J.F., Wang I.K., Zhao H., Seddon A.P. (2007) *Na. Struct. Mol. Biol.* 14, 106-113.
- [23] Epps D.E. and Vosters A.F. (2002) *Chem. Phys. Lipids.* 114(2) 113–122.
- [24] De Azevedo Jr W.F. (2010) *Curr. Drug. Targets.*, 11(3), 327-334.

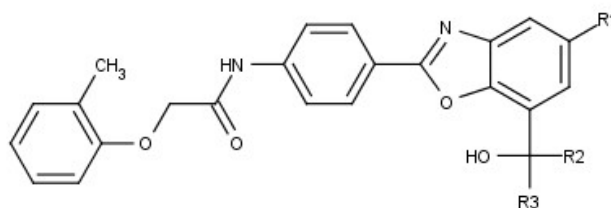
Table 1- Derivatives of 2-Arylbenzoxazoles used their molecular structures and molecular descriptors.



**SCHEME- A**



**SCHEME-B**



**SCHEME-C**

| Compound | R 1      | R 2 | R 3             | R 4             | IC50( $\mu$ M) |
|----------|----------|-----|-----------------|-----------------|----------------|
| 1A       | H        | Cl  | H               | H               | 1.1            |
| 2A       | H        | H   | NO <sub>2</sub> | H               | 3.2            |
| 3A       | H        | H   | F               | H               | 1.9            |
| 4B       | Br       |     |                 |                 | 1.3            |
| 5B       | SMe      |     |                 |                 | 2.9            |
| 6B       | CoMe     |     |                 |                 | 1.3            |
| 7B       | CH(OH)Me |     |                 |                 | 3.4            |
| 8B       | Vinyl    |     |                 |                 | 2.8            |
| 9B       | Ethynyl  |     |                 |                 | 2              |
| 10A      | H        | CN  | Me              | H               | 1.9            |
| 11A      | H        | Br  | H               | Me              | 0.51           |
| 12A      | H        | CN  | H               | Me              | 0.06           |
| 13A      | H        | CN  | H               | CN              | 0.27           |
| 14A      | H        | Cl  | H               | NO <sub>2</sub> | 0.57           |
| 15A      | H        | Br  | H               | F               | 0.91           |
| 16A      | H        | CN  | H               | F               | 0.062          |
| 17A      | H        | Br  | H               | COMe            | 0.38           |
| 18C      | CN       | H   | Me              |                 | 0.046          |
| 19C      | Br       | Me  | Me              |                 | 0.044          |
| 20C      | CN       | Me  | Me              |                 | 0.028          |
| 21C      | Br       | Me  | Et              |                 | 0.11           |

|     |    |    |            |  |       |
|-----|----|----|------------|--|-------|
| 22C | CN | Me | Et         |  | 0.031 |
| 23C | Br | Me | n-Pr       |  | 0.2   |
| 24C | CN | Me | n-Pr       |  | 0.058 |
| 25C | Br | Me | i-Pr       |  | 0.21  |
| 26C | CN | Me | i-Pr       |  | 0.08  |
| 27C | Br | Me | Ethynyl    |  | 0.094 |
| 28C | Br | Me | 1-Propynyl |  | 0.21  |
| 29C | CN | Me | 1-Propynyl |  | 0.16  |
| 30C | H  | Me | Me         |  | 0.44  |

Table 2 - Correlation matrix of descriptors used for QSAR modeling

|                       | IC50( $\mu\text{m}$ ) | SRW09           | H-047    | G3s      | Yindex   | HATS4m |
|-----------------------|-----------------------|-----------------|----------|----------|----------|--------|
| IC50( $\mu\text{m}$ ) | 1                     |                 |          |          |          |        |
| SRW09                 | -0.881015             | 1               |          |          |          |        |
| H-047                 | 0.7604502             | -0.67217        | 1        |          |          |        |
| G3s                   | 0.3652698             | -0.57655        | 0.484673 | 1        |          |        |
| Yindex                | -0.825654             | <b>0.807389</b> | -0.6062  | -0.58035 | 1        |        |
| HATS4m                | -0.223236             | 0.357624        | -0.34932 | 0.024863 | 0.164688 | 1      |

Table 3- Regression parameters and quality of correlations for different models

| Model  | R <sup>2</sup> | R <sup>2</sup> A | SE      | F-statistics |
|--|----------------|------------------|---------|--------------|
| Univariable model<br>(1) SRW09                               | 0.77619        | 0.76819          | 0.5126  | 97.10424     |
| Bivariable model<br>(2) SRW09, H-047                         | 0.82783        | 0.81508          | 0.45784 | 64.91222     |
| Trivariable model<br>(3) SRW09, H-047, G3s                   | 0.87358        | 0.85899          | 0.3998  | 59.88807     |
| Tetravariable model<br>(4a) SRW09, H-047, G3s, Yindex        | 0.91972        | 0.90688          | 0.3249  | 71.60299     |
| (4b) SRW09, H-047, G3s, HATS4m                               | 0.9223         | 0.90987          | 0.31964 | 74.18706     |
| Pentavariable model<br>(5) SRW09, H-047, G3s, Yindex, HATS4m | 0.95292        | 0.94311          | 0.25394 | 97.15037     |

Table 4- Comparison of observed and estimated biological activity (IC50) using models 4(b) and 5 (eq. 1 and 2) with their residue.

| Compound | Observed IC50( $\mu\text{m}$ ) | Model-4(b) |         | Model-5   |         |
|----------|--------------------------------|------------|---------|-----------|---------|
|          |                                | Estimated  | Residue | Estimated | Residue |
| 1A       | 1.1                            | 1.5069     | -0.4069 | 1.2441    | -0.1441 |
| 2A       | 3.2                            | 2.4893     | 0.7107  | 2.7372    | 0.4628  |
| 3A       | 1.9                            | 2.2868     | -0.3868 | 2.1163    | -0.2163 |
| 4B       | 1.3                            | 1.6967     | -0.3967 | 1.4038    | -0.1038 |
| 5B       | 2.9                            | 3.1563     | -0.2563 | 3.1207    | -0.2207 |
| 6B       | 1.3                            | 1.2436     | 0.0564  | 1.401     | -0.101  |
| 7B       | 3.4                            | 2.7677     | 0.6323  | 2.9999    | 0.4001  |
| 8B       | 2.8                            | 2.7546     | 0.0454  | 2.6901    | 0.1099  |
| 9B       | 2                              | 1.9012     | 0.0988  | 1.979     | 0.021   |
| 10A      | 1.9                            | 1.9969     | -0.0969 | 2.1078    | -0.2078 |
| 11A      | 0.51                           | 0.3502     | 0.1598  | 0.5054    | 0.0046  |
| 12A      | 0.06                           | -0.3944    | 0.4544  | -0.0171   | 0.0771  |
| 13A      | 0.27                           | -0.1154    | 0.3854  | 0.145     | 0.125   |
| 14A      | 0.57                           | 0.5285     | 0.0415  | 0.593     | -0.023  |
| 15A      | 0.91                           | 0.8364     | 0.0736  | 0.8119    | 0.0981  |
| 16A      | 0.062                          | 0.2336     | -0.1716 | 0.6142    | -0.5522 |
| 17A      | 0.38                           | 0.0259     | 0.3541  | 0.0775    | 0.3025  |
| 18C      | 0.046                          | 0.1471     | -0.1011 | 0.2182    | -0.1722 |
| 19C      | 0.044                          | 0.3672     | -0.3232 | 0.2811    | -0.2371 |
| 20C      | 0.028                          | 0.1618     | -0.1338 | 0.0821    | -0.0541 |
| 21C      | 0.11                           | 0.3273     | -0.2173 | 0.2214    | -0.1114 |
| 22C      | 0.031                          | 0.1862     | -0.1552 | 0.0197    | 0.0113  |
| 23C      | 0.2                            | -0.0055    | 0.2055  | -0.0477   | 0.2477  |
| 24C      | 0.058                          | 0.239      | -0.181  | 0.1209    | -0.0629 |
| 25C      | 0.21                           | 0.3719     | -0.1619 | 0.176     | 0.034   |
| 26C      | 0.08                           | -0.0939    | 0.1739  | -0.4279   | 0.5079  |
| 27C      | 0.094                          | 0.509      | -0.415  | 0.3743    | -0.2803 |
| 28C      | 0.21                           | 0.0712     | 0.1388  | 0.0168    | 0.1932  |
| 29C      | 0.16                           | 0.2914     | -0.1314 | 0.1651    | -0.0051 |
| 30C      | 0.44                           | 0.4355     | 0.0045  | 0.5431    | -0.1031 |

Table 5- Different validation parameters calculated for QSAR models

| Model | RSS    | MSS     | R      | S.E.   | Q      |
|-------|--------|---------|--------|--------|--------|
| 1     | 7.3574 | 25.5154 | 0.8810 | 0.5126 | 1.7187 |
| 2     | 5.6596 | 27.2131 | 0.9099 | 0.4578 | 1.9873 |
| 3     | 4.1558 | 28.7170 | 0.9347 | 0.3998 | 2.3378 |
| 4-a   | 2.6390 | 30.2337 | 0.9590 | 0.3249 | 2.9517 |
| 4-b   | 2.5542 | 30.3185 | 0.9604 | 0.3196 | 3.0045 |
| 5     | 1.5477 | 31.3250 | 0.9762 | 0.2539 | 3.8441 |

Table 6- Docking results of 2-arylbenzoxazole inhibitors with their docking scores rerank scores and RMSD values.

| COMPOUND | MOLDOCK SCORE  | RERANK SCORE   | RMSD         |
|----------|----------------|----------------|--------------|
| 1A       | -112.04        | -93.168        | 1.33         |
| 2A       | -108.29        | -60.754        | 1.183        |
| 3A       | -112.8         | -89.203        | 0.653        |
| 4B       | -106.45        | -77.199        | 0.838        |
| 5B       | -115.76        | 83.625         | 1.628        |
| 6B       | -120.53        | -91.72         | 1.128        |
| 7B       | -117.43        | -90.497        | 1.724        |
| 8B       | -117.2         | -82.18         | 1.21         |
| 9B       | -116.09        | -80.75         | 0.77         |
| 10A      | -28.867        | -78.423        | 0.914        |
| 11A      | -113.81        | -91.432        | 2.514        |
| 12A      | -118.67        | -96.298        | 0.127        |
| 13A      | -125           | -91.839        | 0.896        |
| 14A      | -111.42        | -96.786        | 1.494        |
| 15A      | -108.34        | -89.562        | 2.226        |
| 16A      | -118.42        | -97.284        | 0.263        |
| 17A      | -118.83        | -87.428        | 1.333        |
| 18C      | -121.87        | -94.676        | 0.903        |
| 19C      | -122.9         | -98.025        | 3.147        |
| 20C      | -130.82        | -97.657        | 1.436        |
| 21C      | -124.78        | -75.031        | 1.955        |
| 22C      | -123.16        | -99.3          | 1.489        |
| 23C      | -130.79        | -98.338        | 1.543        |
| 24C      | -126.48        | -88.245        | 0.712        |
| 25C      | -123.89        | -97.075        | 0.432        |
| 26C      | -130.51        | -88.782        | 1.464        |
| 27C      | -123.16        | -86.98         | 4.784        |
| 28C      | -132.52        | -97.49         | 7.643        |
| 29C      | <b>-131.18</b> | <b>-102.17</b> | <b>1.104</b> |
| 30C      | -120.57        | -87.057        | 0.617        |



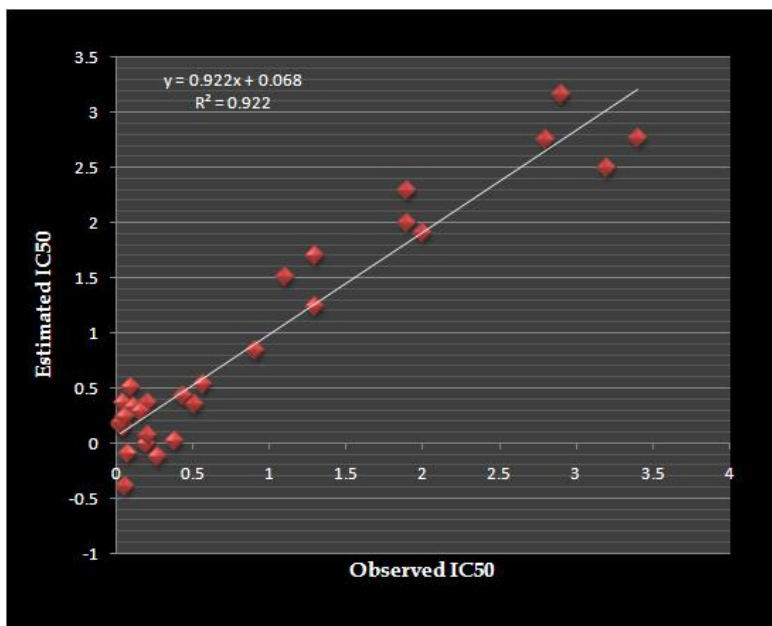


Fig. 1- Correlation of observed and estimated biological activity (IC50) using model 4b

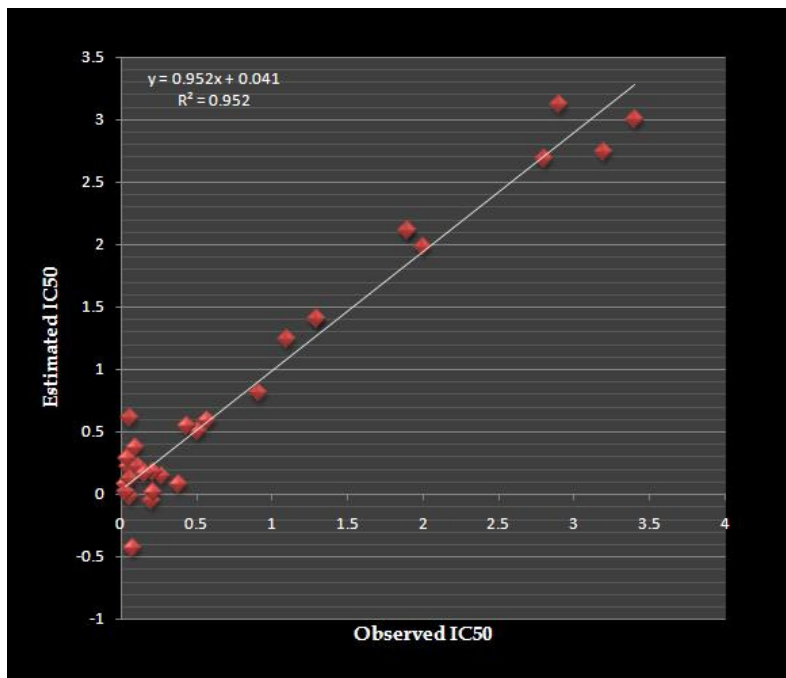


Fig. 2- Correlation of observed and estimated biological activity (IC50) using model 5

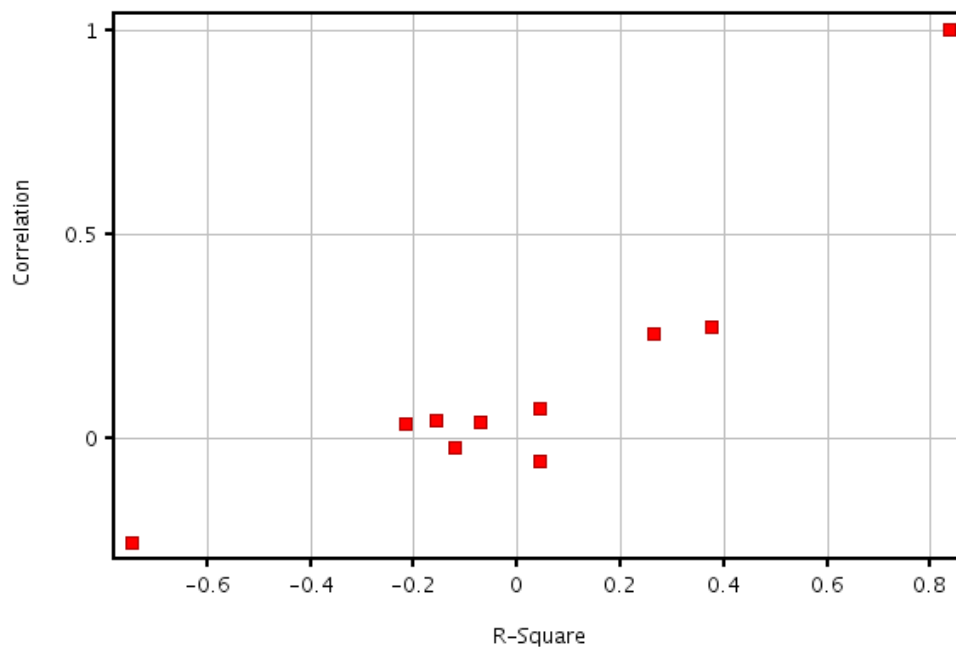


Fig. 3- Graphical Y-scrambling validation for models-4(b) for 10 iterations

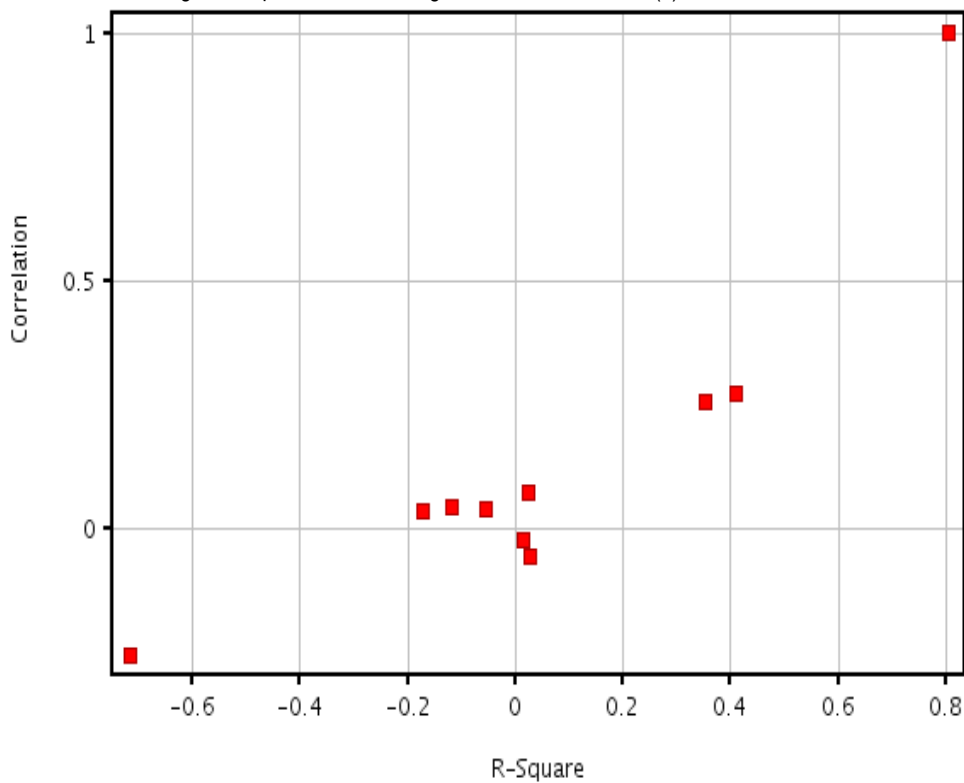


Fig. 4- Graphical Y-scrambling validation for models-5 for 10 iterations

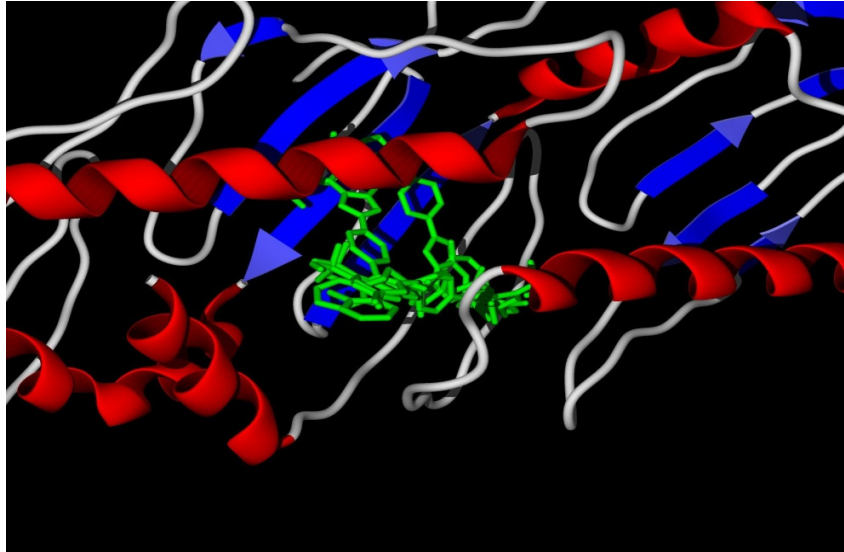


Fig. 5- All poses of molecule 29C and secondary structure of CETP

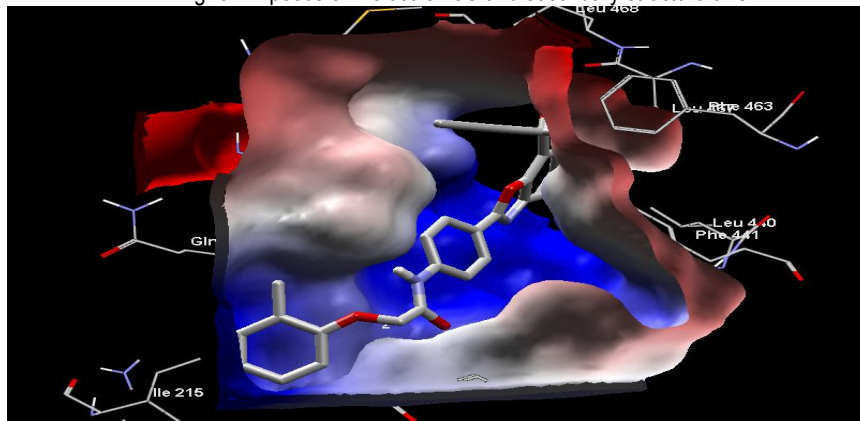


Fig. 6- Best pose of molecule 29C and electrostatic surface of CETP

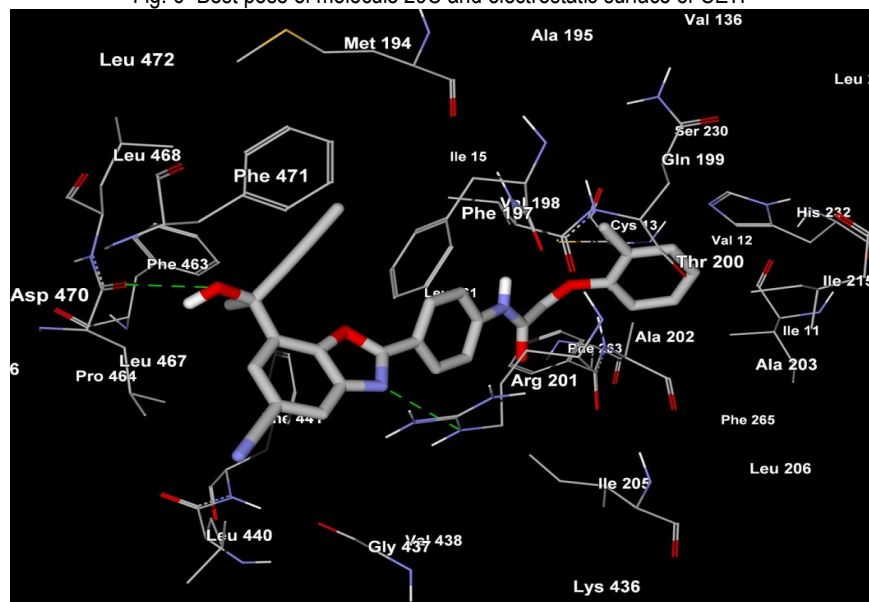


Fig. 7- Most active pose of molecule 29C and hydrogen bond with Arg201 and Asp470 residues of CETP