Chemometrics modeling, inverse docking and molecular simulations-driven design and multi-layered prioritization lead identification of novel analogs based on 2-aminobenzimidazole scaffolds for addressing Leishmaniasis

Presented

by

Arpita Biswas

PhD Research Scholar

Dept of Pharm Tech

Jadavpur University, Kolkata-700032

DDD LAB
Drug Discovery and Development Laboratory

Under the guidance of

Dr. Probir Kumar Ojha

Associate Professor

Dept. Of Pharm. Tech

Jadavpur University, Kolkata-700032

It’s a Threat to the mankind

Theme of 2023 NTD is-”Act Now. Act Together. Invest in Neglected Tropical Disease”

- Kala-azar (visceral leishmaniasis)-fatal if untreated, **Fig: 1**.
- **Causes**: Protozoan parasites which are transmitted by the bite of infected female phlebotomine sand-flies.
- Very limited no. of drugs available in the market.
- Having good extent of toxicity.
- Huge population gets affected.

Flowchart of the lifecycle of parasite.(Amastigotes and Promastigotes), **Fig: 2**

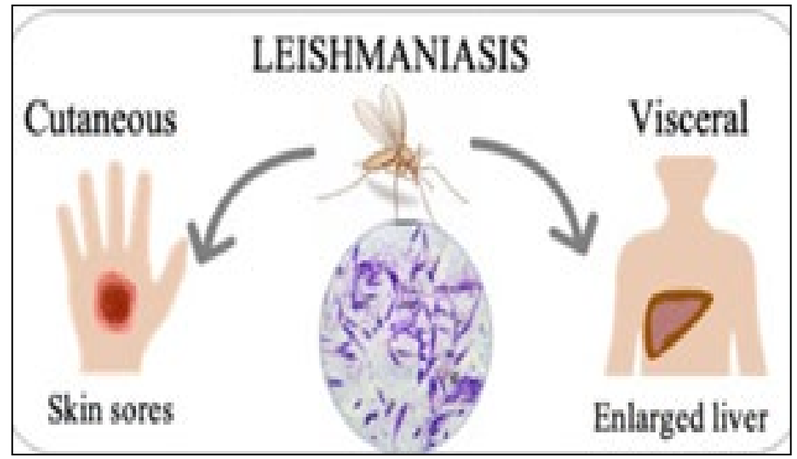


Fig:1

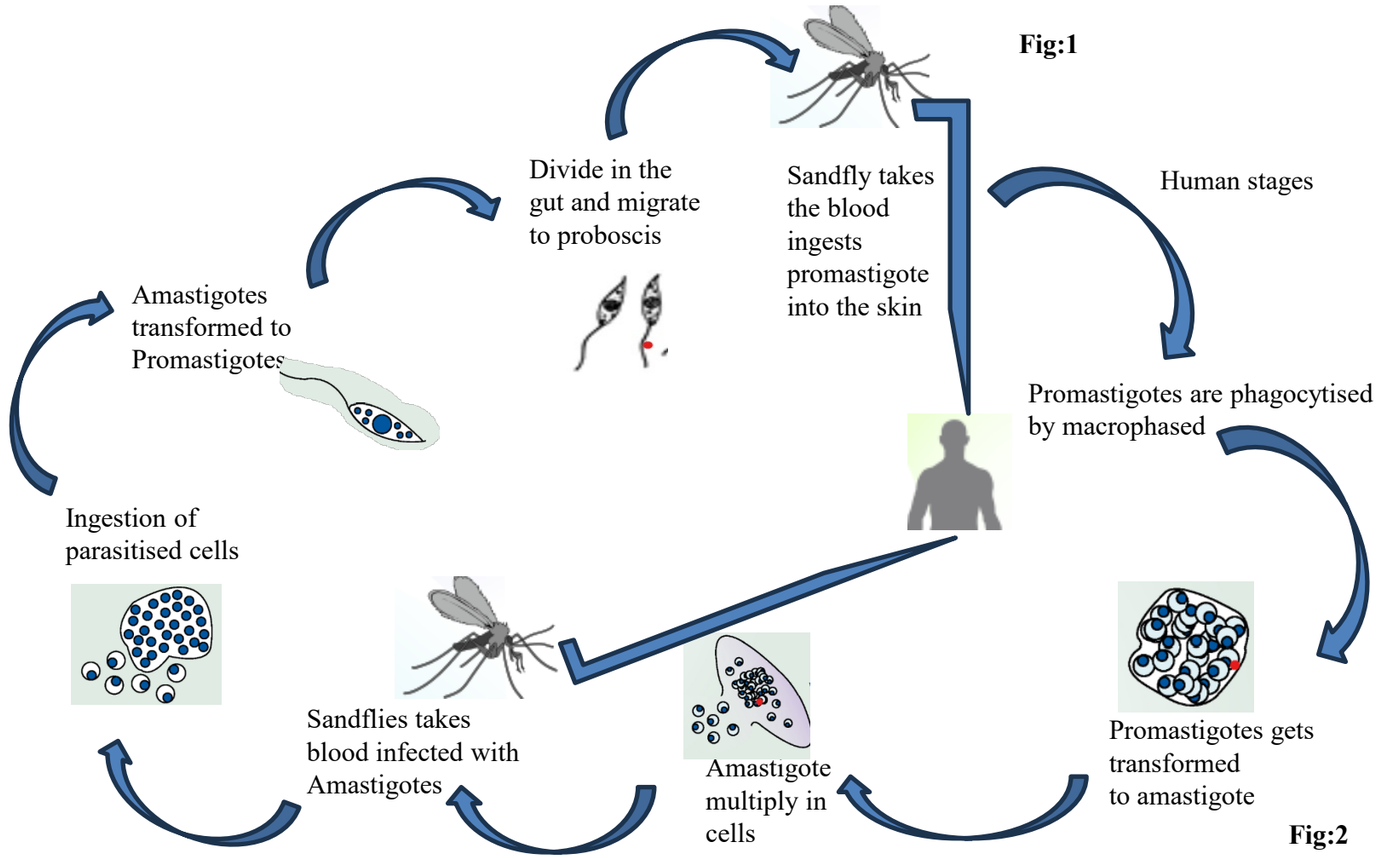


Fig:2

Objective

- Design of new compounds with 2-aminobenzimidazole scaffold to counteract NTD endemic diseases.
- Development of QSAR-regression based predictive models which correlates biological activity with chemicals of molecular structure represented as descriptors is aimed to develop through various statistical approaches.
- Validation of developed models using globally accepted internal and external validation parameters.
- Screening of drugbank compounds using the validated model to develop new chemical compound with least toxicity against leishmaniasis

WORKFLOW

Data set collection from literature with experimental (IC_{50}) value

OECD Principle 1

- constitutional
- ring descriptors
- connectivity index
- functional group counts
- atom centred fragments
- atom type E-states
- 2D atom pairs,
- molecular properties
- ETA indices

Individual compound structure were drawn using Marvin sketch(Version 5.5.0.1)

Descriptor calculation using Alvadesc software

OECD Principle 2

Dataset division using Modified k medoid method

Training set(35)

Test set(12)

Data set division

Model development

Best subset selection

PLS

OECD Principle 3

Applicability domain

OECD Principle 4

Model validation

OECD Principle 5

Explanation of encoded features

Screening of DB Compounds

Compounds within AD subjected to MD

5 Leads selected based on Binding score

17 Analogs were designed incorporating best features

17 derivatives undergoes MD

Based on Binding energy

2 derivatives selected

ADMET

MD Simulation

1 Lead Compound against Leishmaniasis

Results and discussion

PLS Model

Statistical parameters of the model is mentioned in the tabular form as Table:1

$$^{pIC_{50}} = 1.41855 + 2.53073 * \text{Eta\_F\_A} - 0.61002 * \text{B10[C-N]} + 0.08752 * \text{F03[C-N]} + 0.29798 * \text{F06[N-F]}$$

Table:1

Model		Training set(n=35)				Test set(n=12)					
	LV	R^2	$Q^2_{(LOO)}$	$r_m{}^2{}_{(LOO)}$	$MAE_{(LOO)}$	Q^2_{F1}	Q^2_{F2}	$r_m{}^2{}_{(test)}$	$\Delta r_m{}^2{}_{(test)}$	CCC	MAE
PLS	2	0.688	0.611	0.497	0.240	0.735	0.698	0.525	0.235	0.800	0.211

Descriptors Contribution

4 Descriptors are found to have contribution towards developed model:

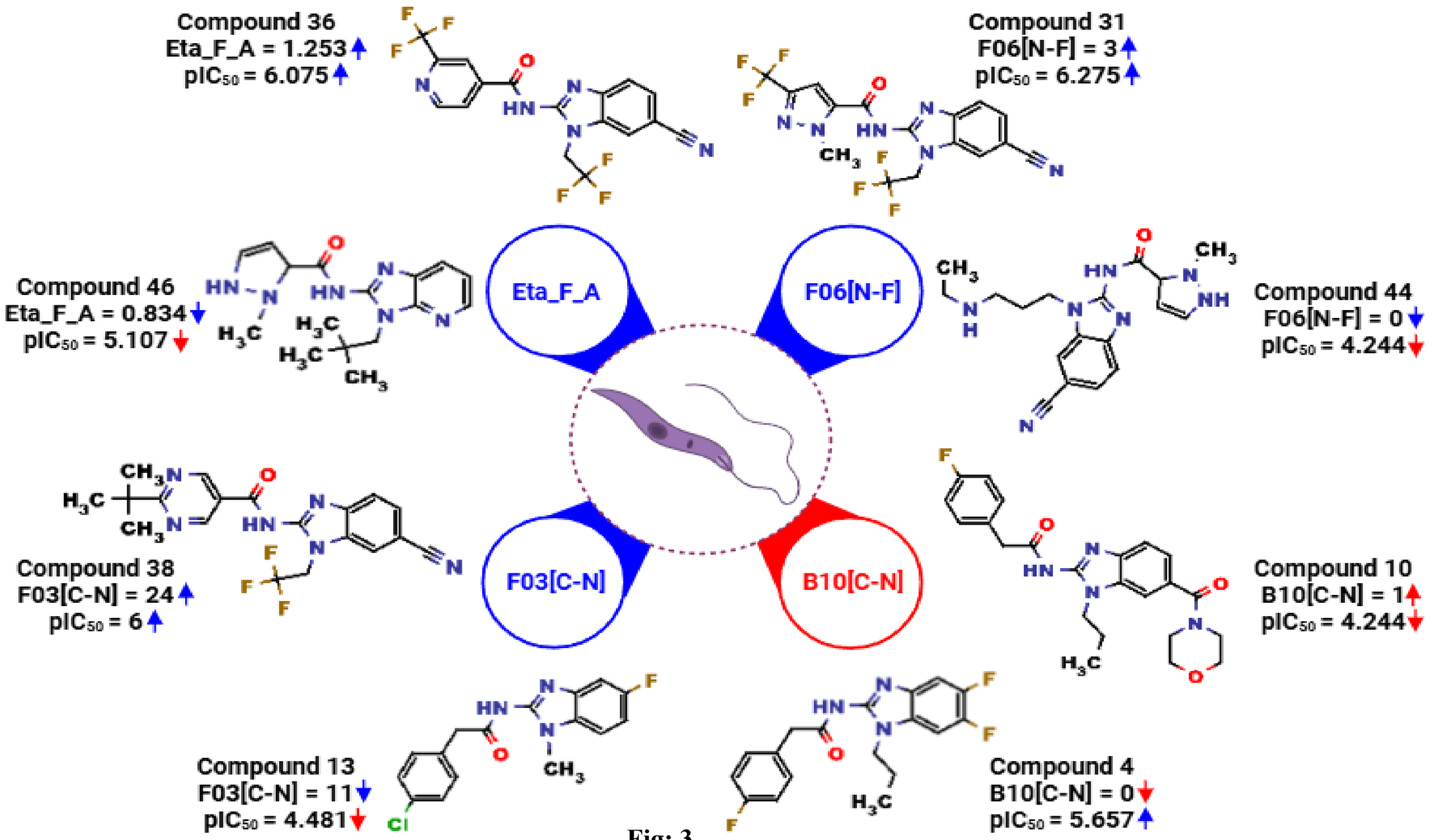


Fig: 3

PLS PLOT

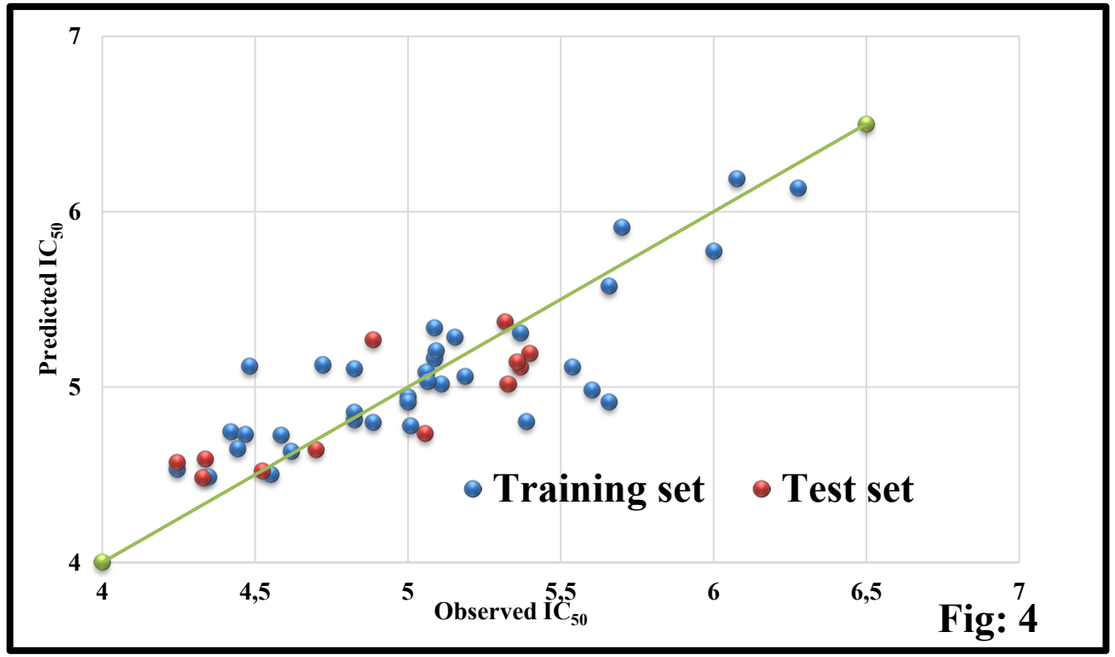


Fig: 4

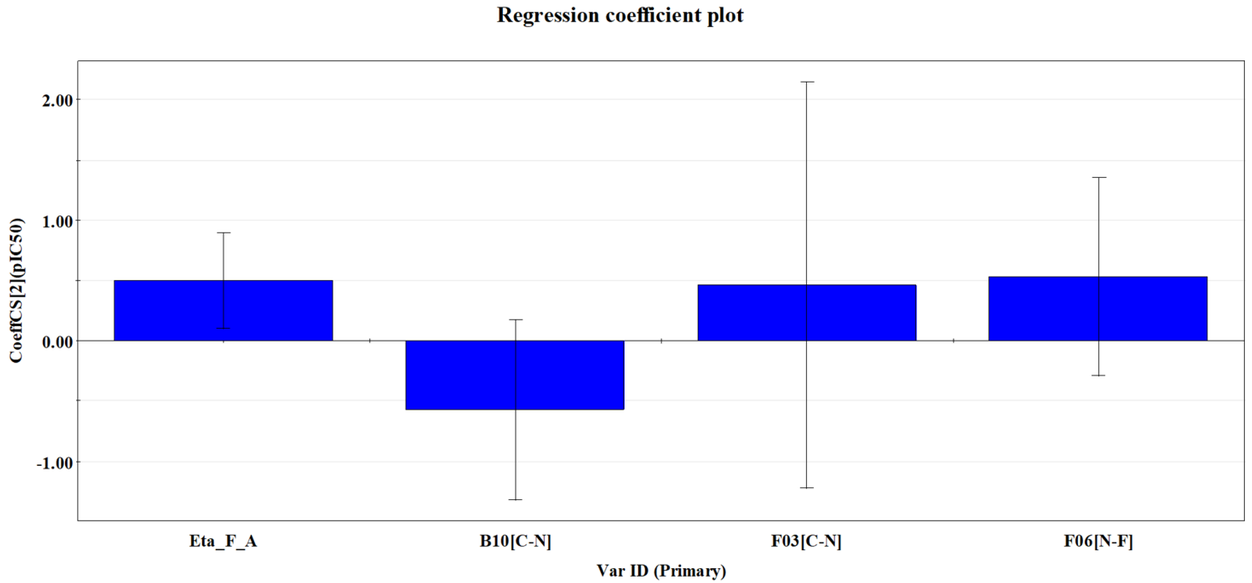


Fig: 5

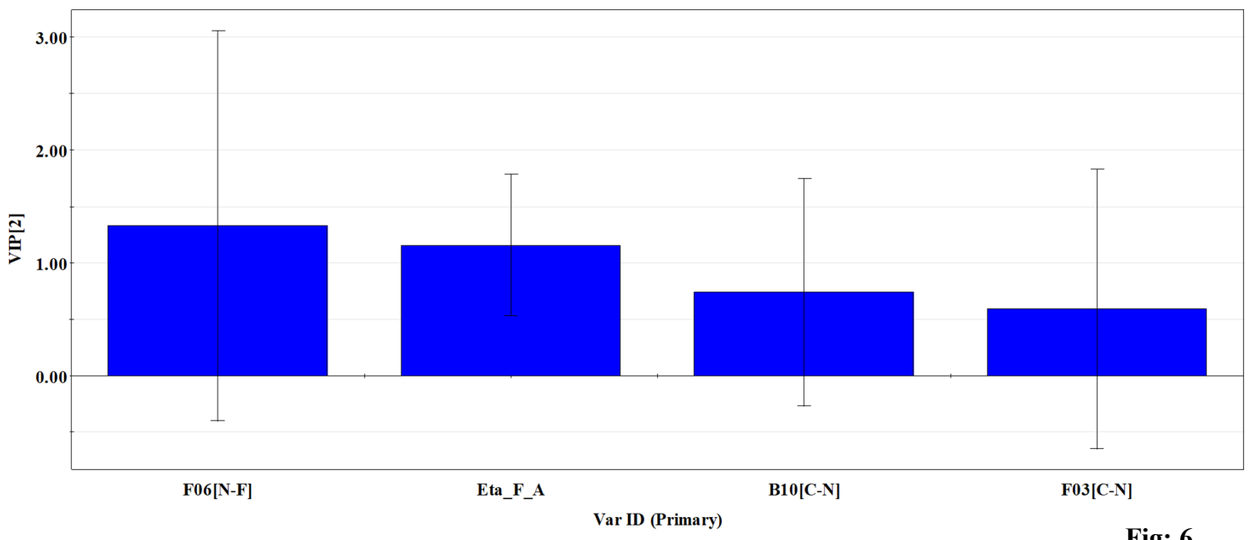


Fig: 6

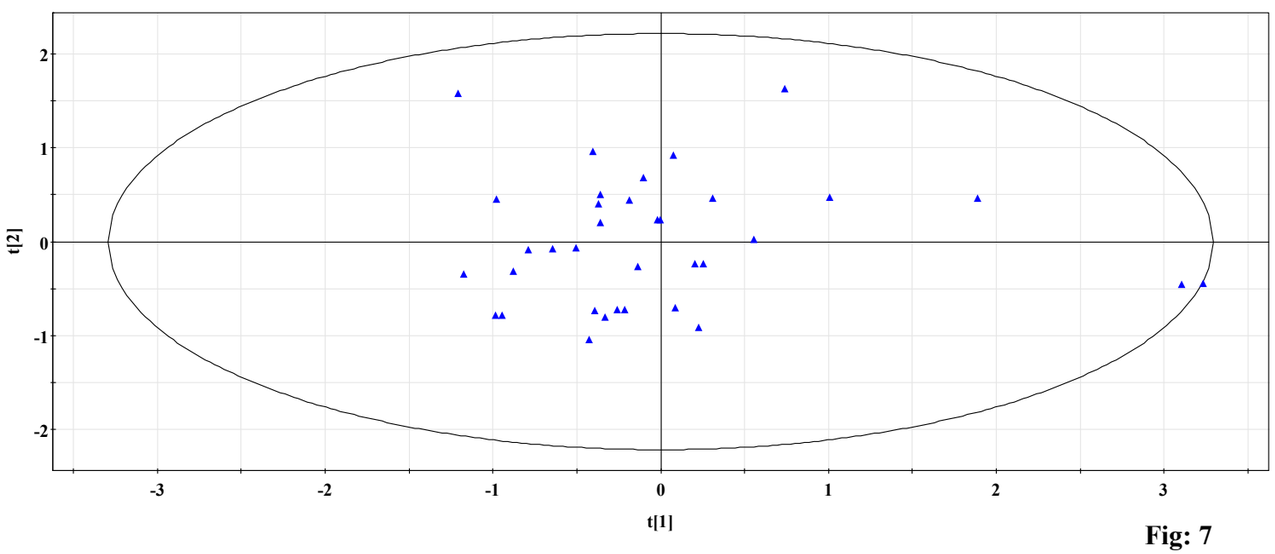


Fig: 7

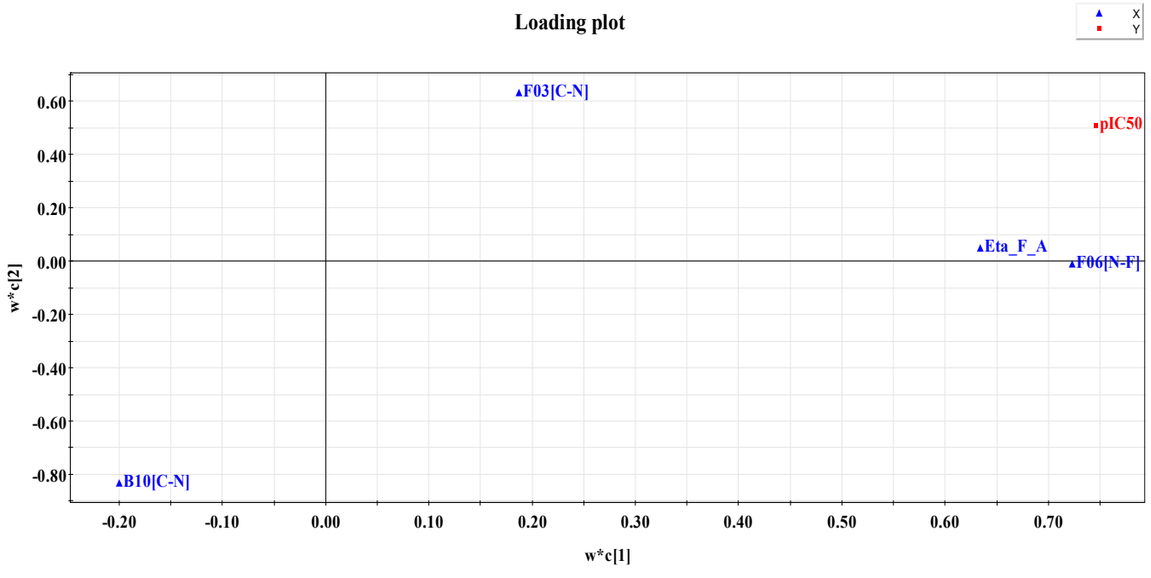


Fig: 8

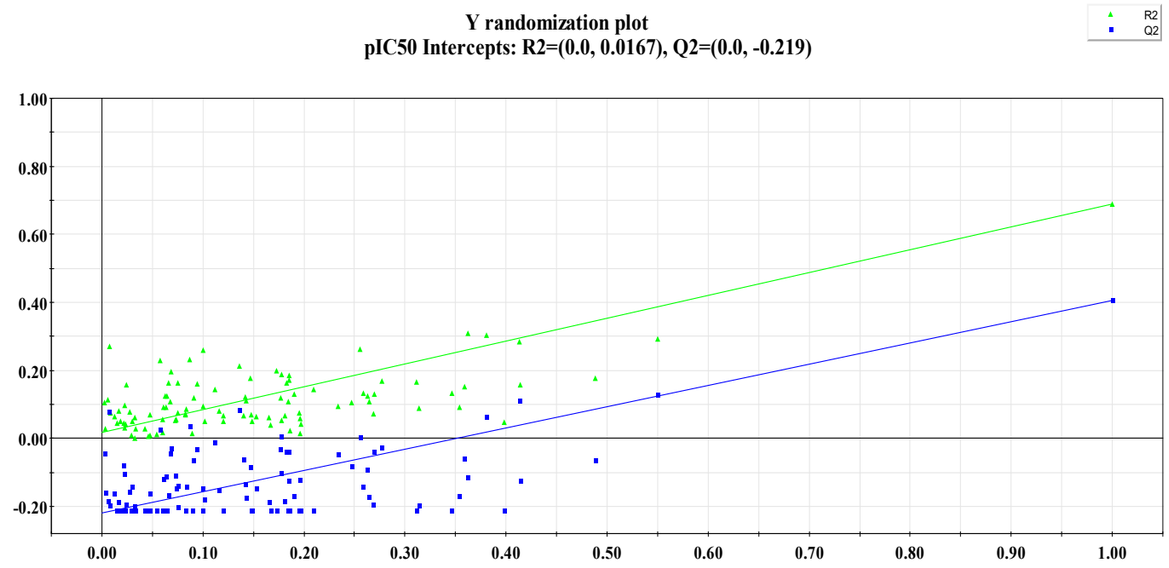


Fig: 9

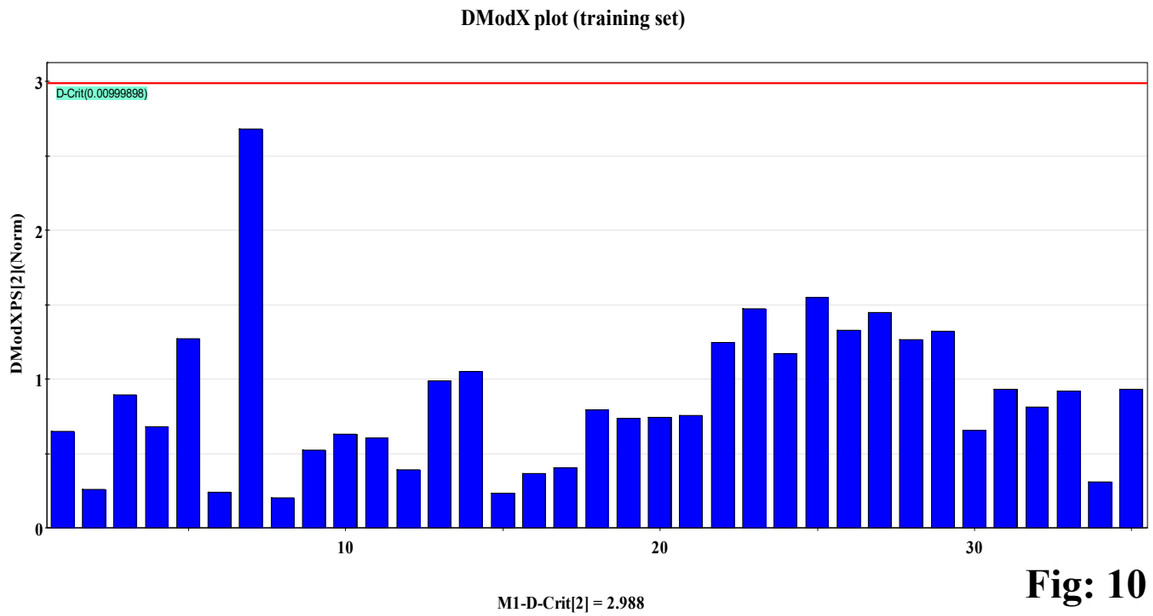


Fig: 10

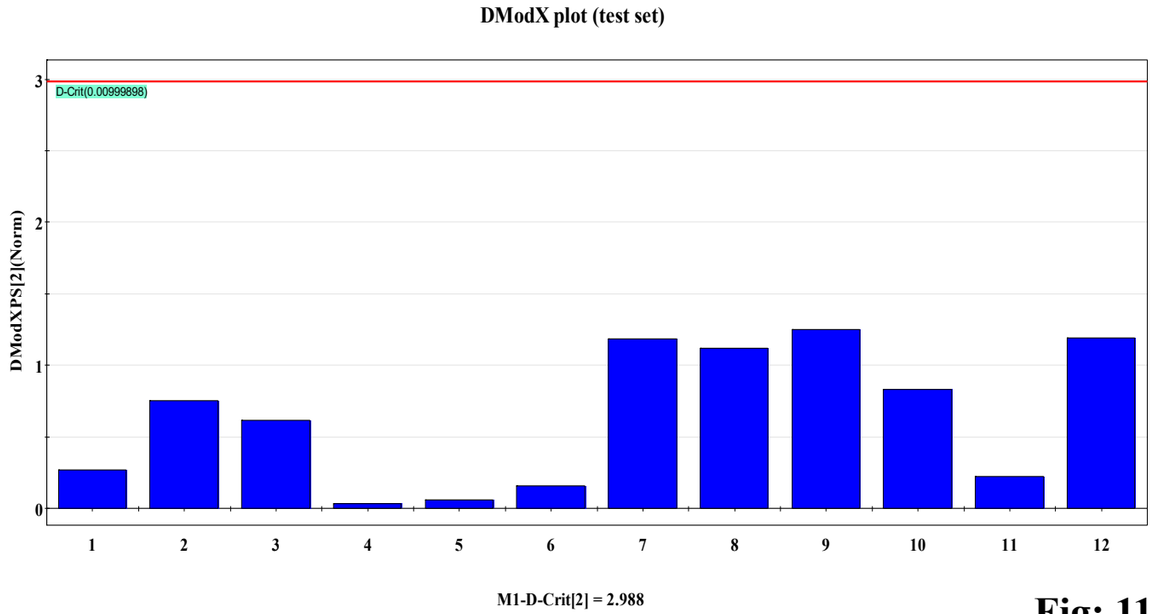


Fig: 11

Step-wise generation of Lead Analog against leishmaniasis

- The PLS Model generated was used to screen the DrugBank compounds based on AD.
- The compounds obtained again screened using R05 violation.
- The screened compounds now subjected to molecular docking using 15 targeted proteins.
- The docking results shows that the best binding score was obtained against TR.
- The screened compounds were assessed for their localization within the binding site of 15 putative protein targets to prioritize leads compounds based on their binding affinity.
- Inverse Docking, a tabular form of presented with target proteins against leishmaniasis, represented as,

Table:2

Target Proteins	Highest Binding score
TR	-6.58
AdoMetDC	-6.45
ARG	-5.95
FPS	-6.37
GspS	-6.40
NDkb	-6.49
ODC	-6.40
OPB	-6.42
S14D	-6.31
S24CM	-6.09
SpdS	-6.45
SS	-6.47
TS	-6.45
TXN	-6.47
TXNPx	-5.12

- 124 Compounds were found with highest dock score exceeding experimental TR inhibitor with TRL190
- Considering binding energy, top 5 leads against TR was identified, as mentioned on the table below., Table:3
- Positively correlated features escalating TR inhibitory activity obtained from QSAR study incorporated rationally to the generation of analogs.
- Thus resulting in 17 Analogs generation.

Table: 3

DrugBank ID	Binding energy
DB12269	-6.58185
DB01705	-6.38635
DB12457	-6.32147
DB03231	-6.0676
DB04260	-6.01953
MWW	-4.70903

Structures of 17 Analogs with 5 Lead Compounds

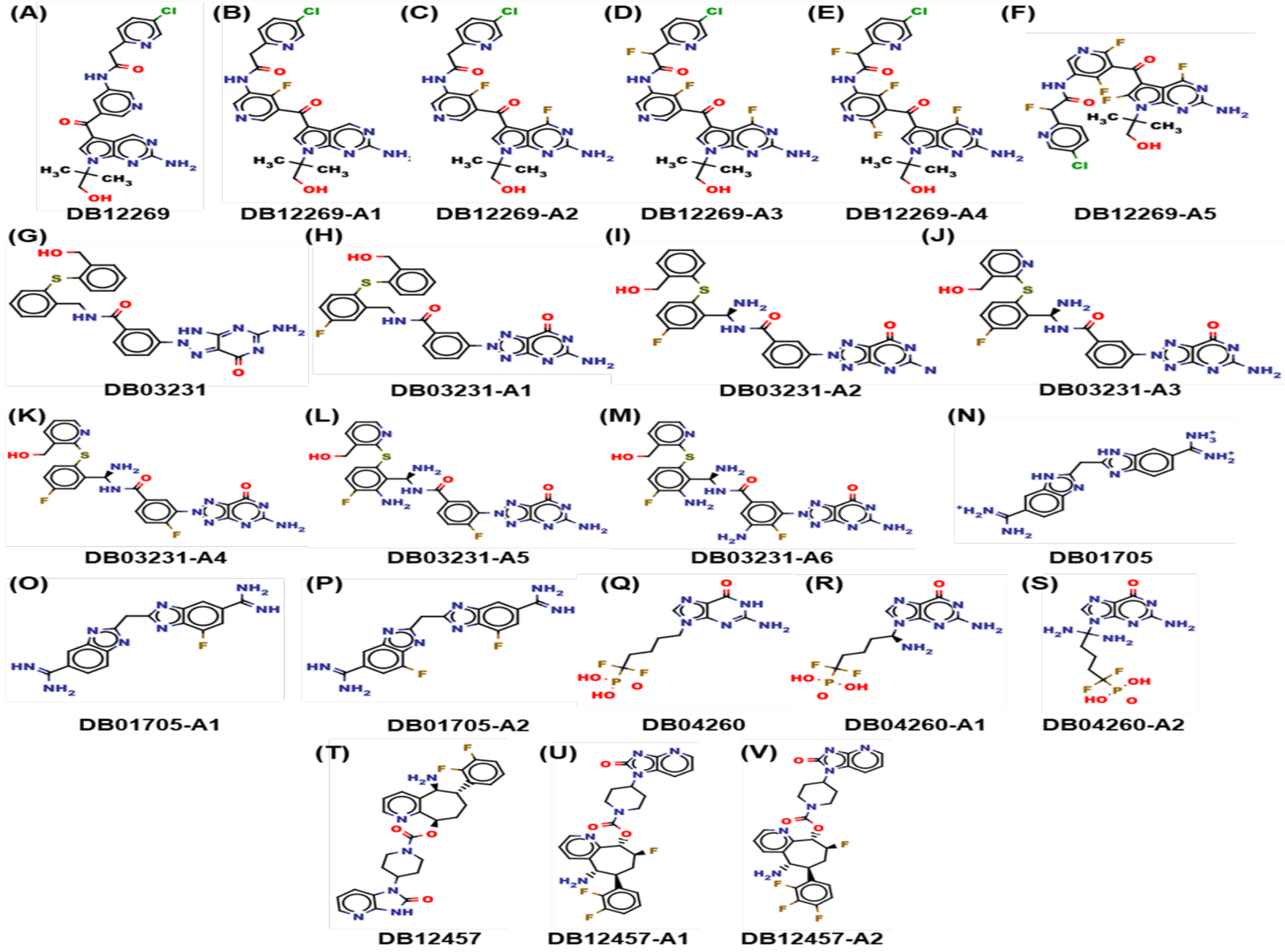
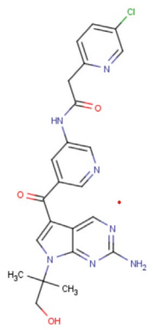
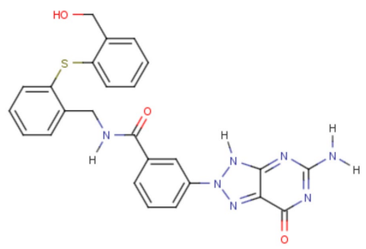


Fig: 12

Flowchart to Lead Analog compound generation



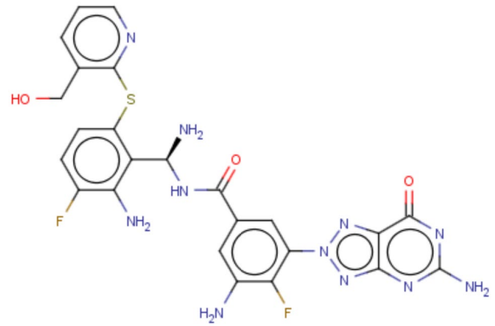
DB12269



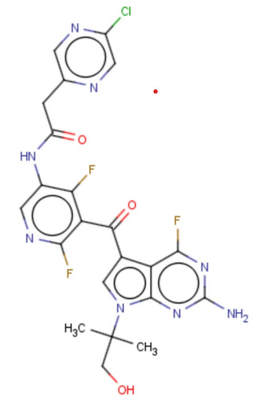
DB03231

17 Analogues were generated

Based on MD

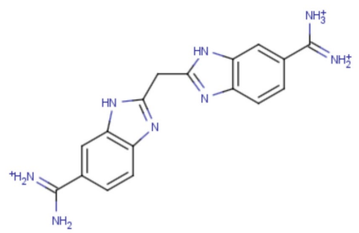


DB03231-A6 (M)

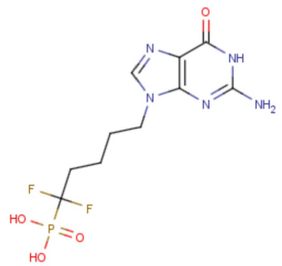


DB12269-A4 (E)

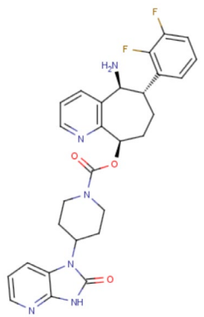
Based on ADMET study and MDS



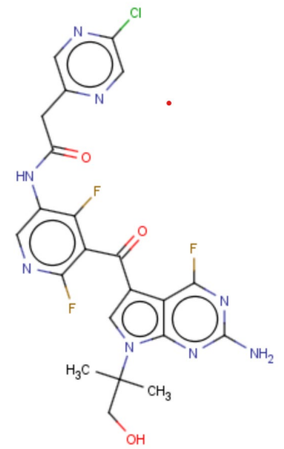
DB01705



DB04260



DB12457



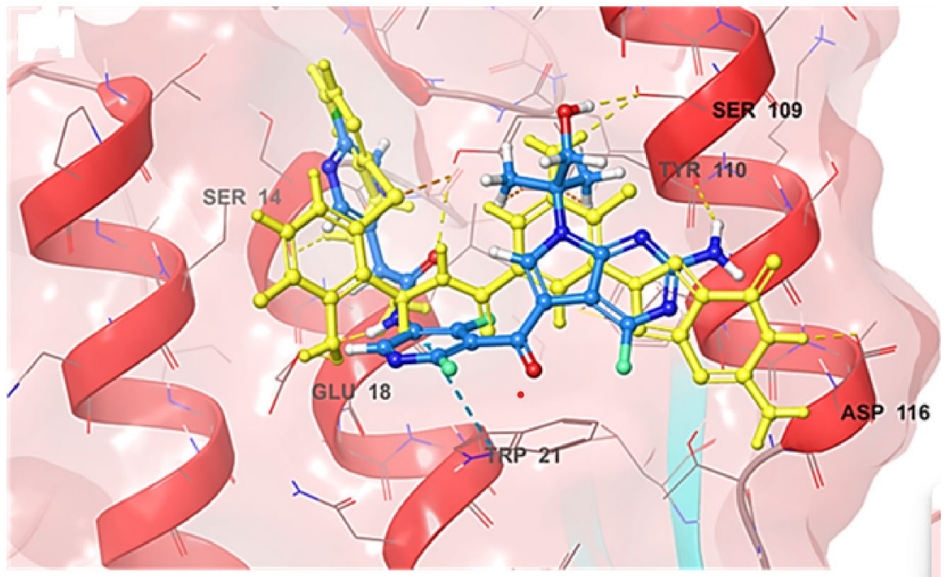
DB12269-A4

Fig: 13

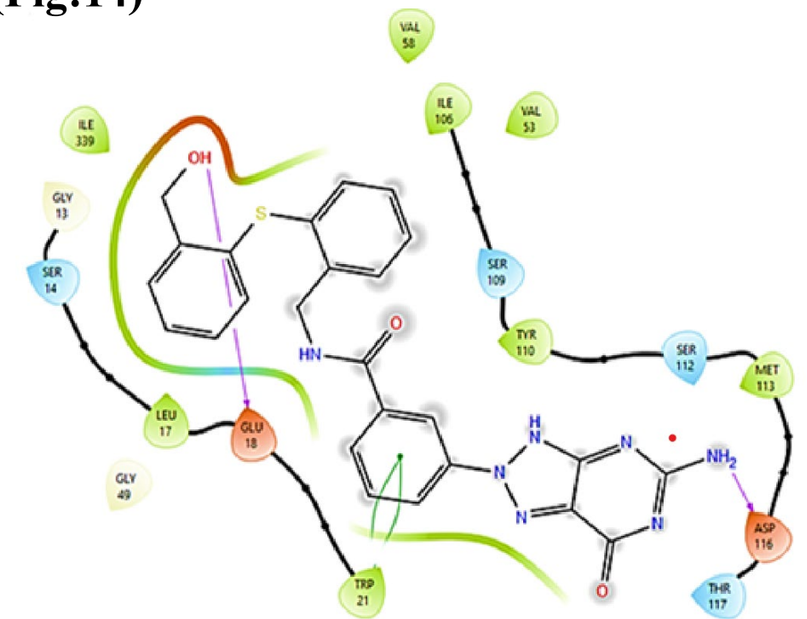
Depiction of Binding score of parent lead and their derivatives along with the binding pose and 2D interaction **(Fig:14)**

Table: 4

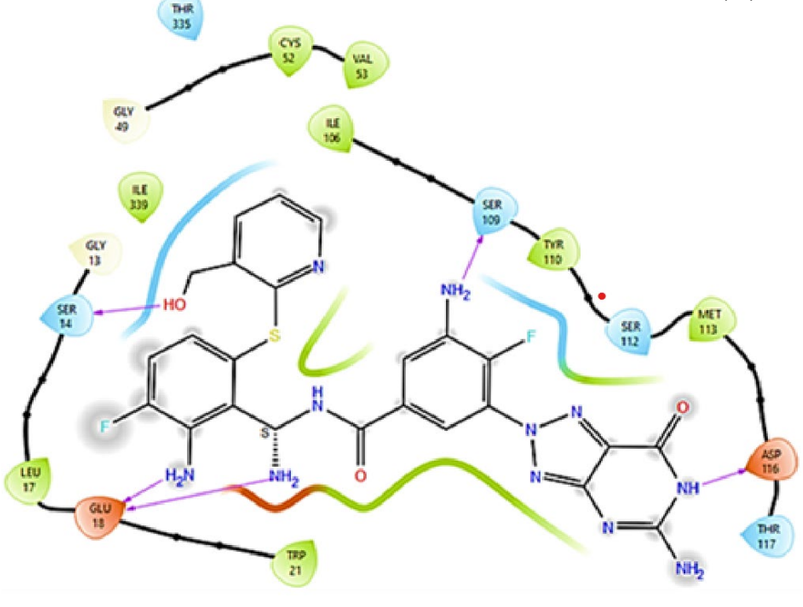
DrugBank ID	G(kcal/mol)
DB03231	-6.06
DB03231-A6	**-6.80**
DB12269	-6.58
DB12269-A4	**-6.67**
TRL190	-4.70



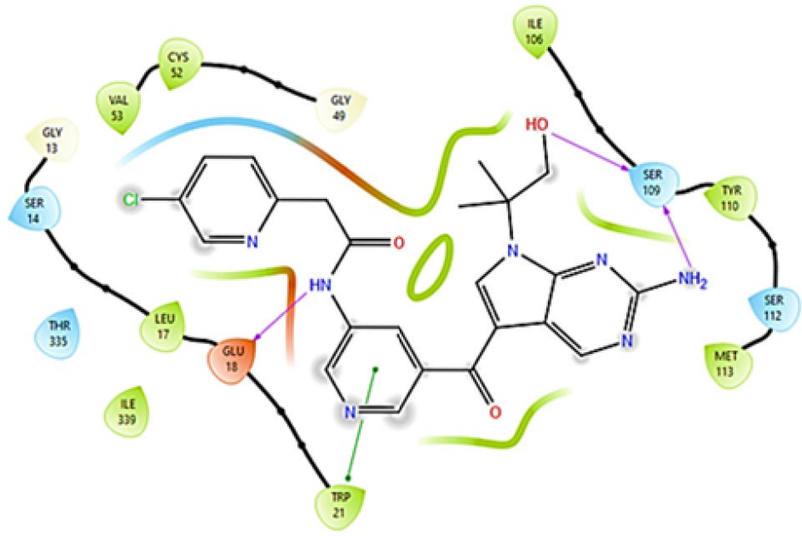
(A) Zoomed view depicting the best docked pose of DB03231-A6 and DB12269-A4 at the TR binding site



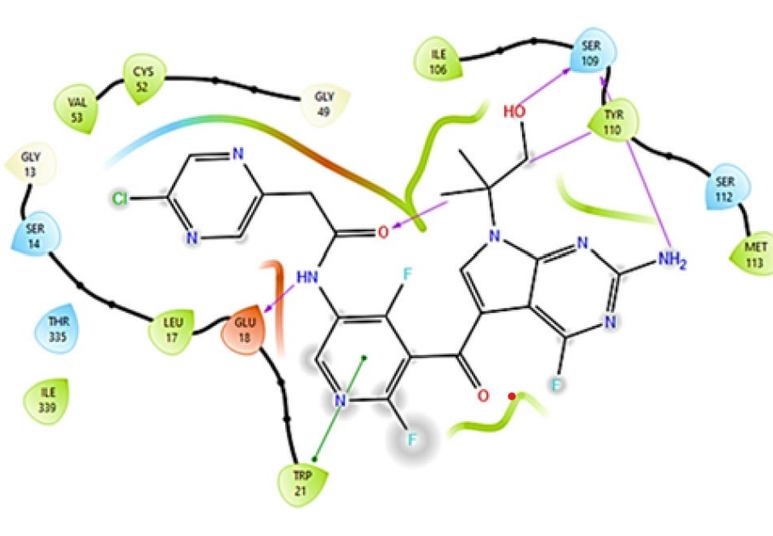
(B) 2D protein-ligand interactions as traced in DB03231



(B1) 2D protein-ligand interactions as traced in DB03231-A6



(C) 2D protein-ligand interactions as traced in DB12269



(C1) 2D protein-ligand interactions as traced in DB12269-A4

ADMET STUDY

Table: 5

Property / Parameter / Endpoint	DB12269-A4	DB03231-A6
Water solubility (log mol/L)	-2.989	-2.907
Caco2 permeability (log Papp in 10^{-6} cm/s)	0.936	0.962
Intestinal absorption (% Absorbed)(human)	75.855	46.7
Skin Permeability (log Kp)	-2.735	-2.735
Hepatotoxicity	0.52	0.52
Carcinogenicity	**0.59**	**0.63**
Mutagenicity	0.64	0.55
Cytotoxicity	0.72	0.72
Predicted LD50 (mg/kg)	**2500**	**2000**
Toxicity Class	**5**	**4**

Analysis of Cα RMSD

The top-docked poses of DB03231-A6 and DB12269-A4 in complex with Trypanothione reductase underwent MDS and were analysed to found the best Analog compound through RMSD and RMSF analysis

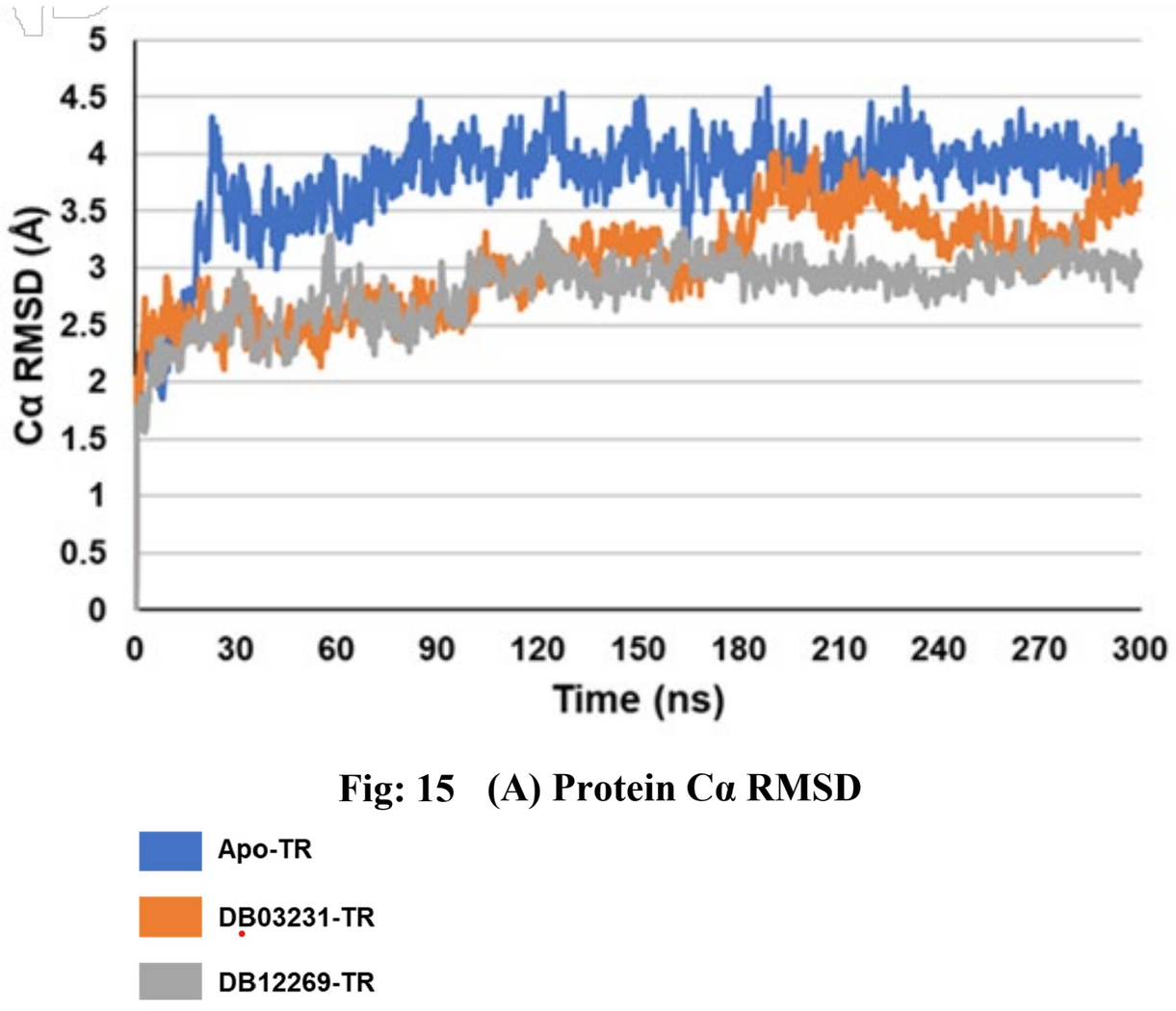


Fig: 15 (A) Protein Cα RMSD

- The RMSDmean of the Cα atoms in the DB03231-A6-TR and DB12269-A4-TR complexes were significantly lower (3.04Å and 2.81Å, respectively) compared to the apo-TR form (3.77 Å).
- RMSD trajectories were consistent up to 177ns, after which DB03231-A6-TR exhibited more fluctuations than DB12269-A4-TR.
- Beyond 100ns, the apo-TR showed reduced fluctuations with a maximum RMSD difference of 1.40Å.
- DB12269-A4-TR demonstrated even lower fluctuations (0.78Å), while DB03231-A6-TR exhibited fluctuations (1.45Å) comparable to those observed in the apo-TR form.

Analysis of Protein Ligand RMSD

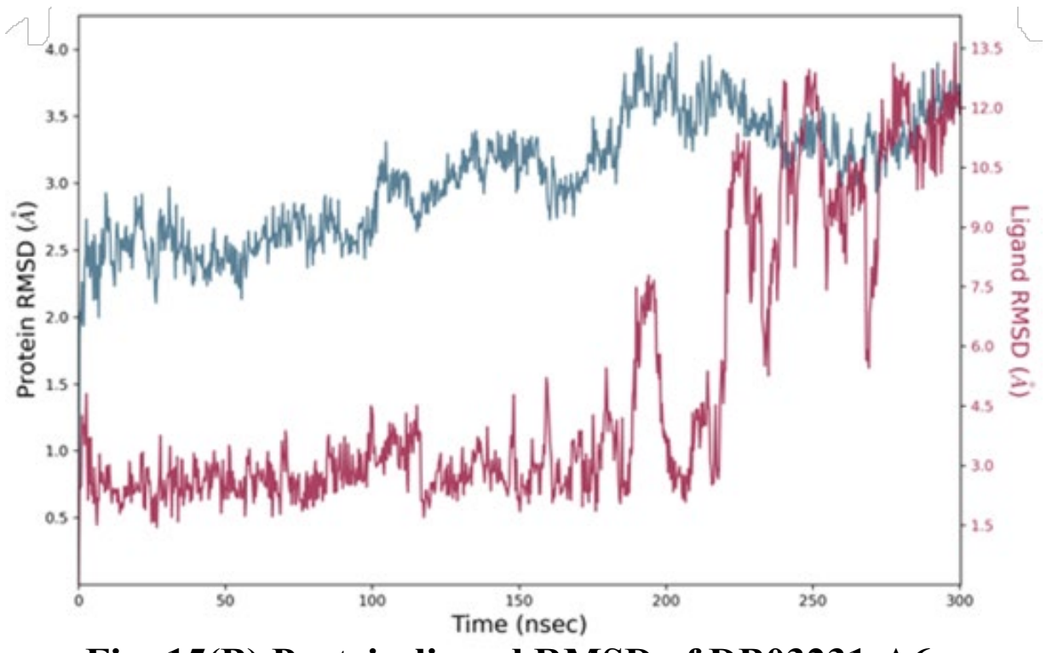


Fig: 15(B) Protein-ligand RMSD of DB03231-A6

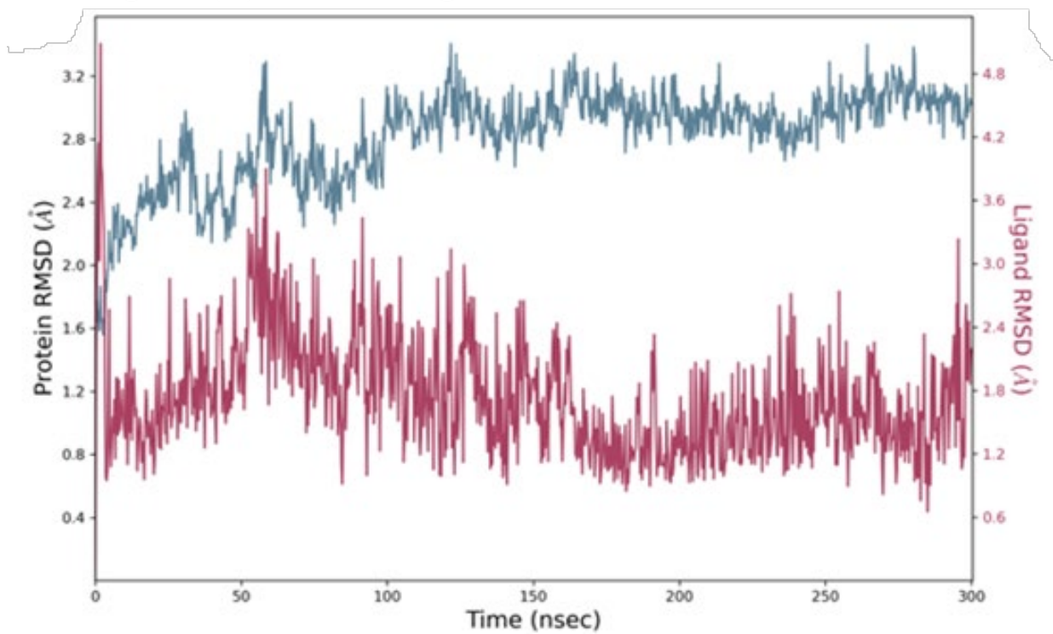


Fig: 15(C) Protein-ligand RMSD of DB12269-A4

- Superior stability of DB12269-A4 at the binding site was observed throughout the simulations.
- The average RMSD difference relative to TR was notably higher for DB03231-A6 (2.91Å) compared to DB12269-A4 (1.15Å),
- After 179 ns, DB03231-A6 exhibited significant fluctuations, although the protein-ligand trajectory converged towards the end of the simulation.
- Conversely, the ligand trajectory of DB12269-A4 did not overlap with the TR trajectory; however, the ligand distance from the binding site remained within the acceptable range (<3.5Å).

Analysis of protein ligand RMSF

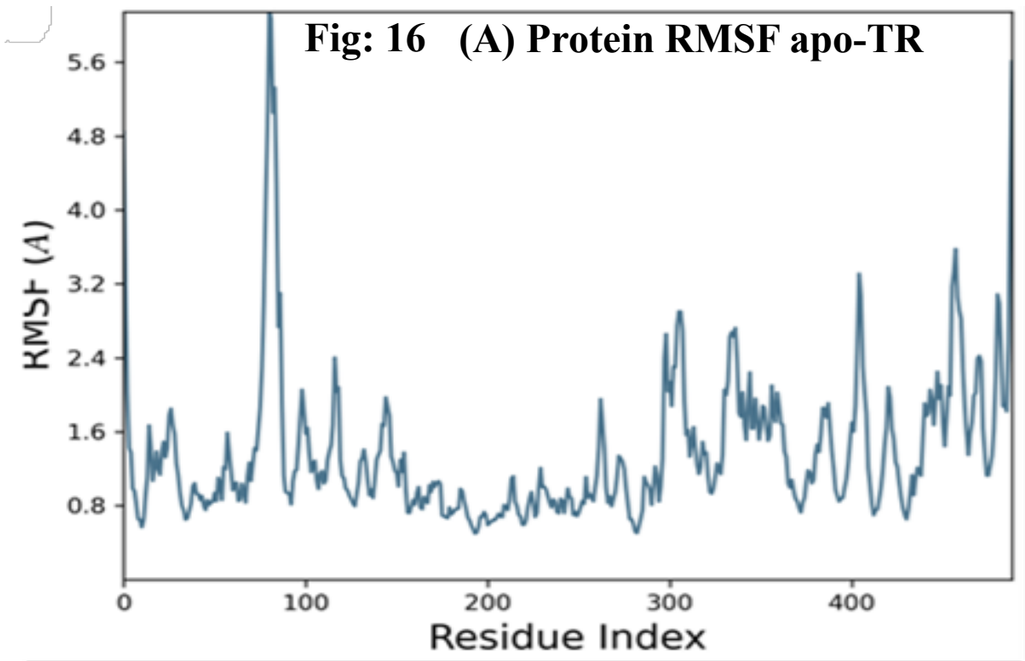


Fig: 16 (A) Protein RMSF apo-TR

- It is found that there are fewer interacting amino acid residues, exist with DB12269-A4.
- Still significant reduction in the flexibility of the target site amino acids is observed more effectively than DB03231-A6.

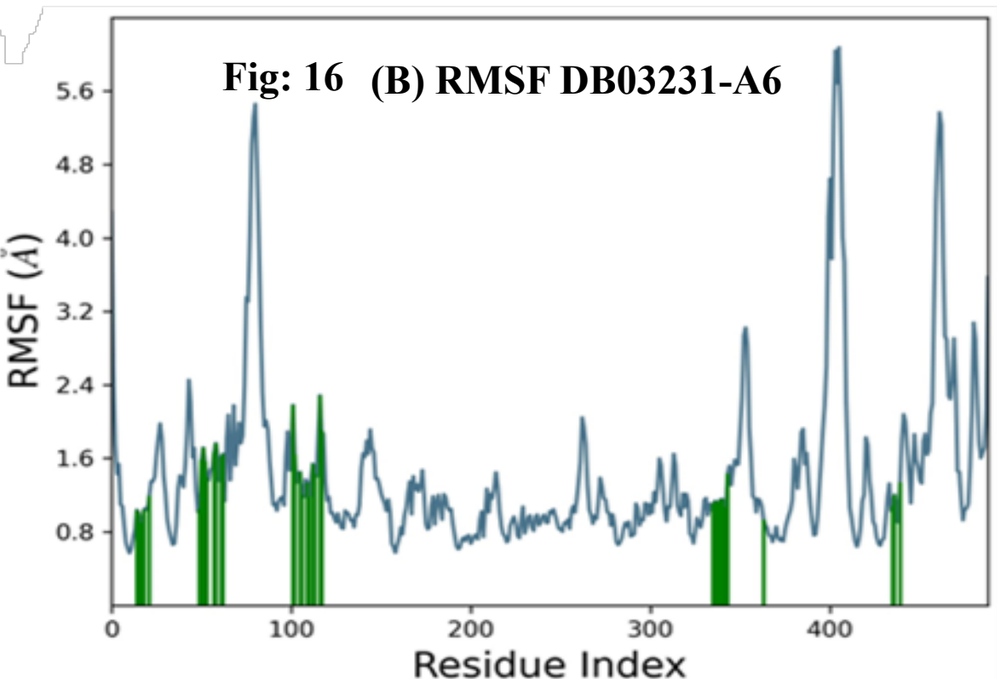


Fig: 16 (B) RMSF DB03231-A6

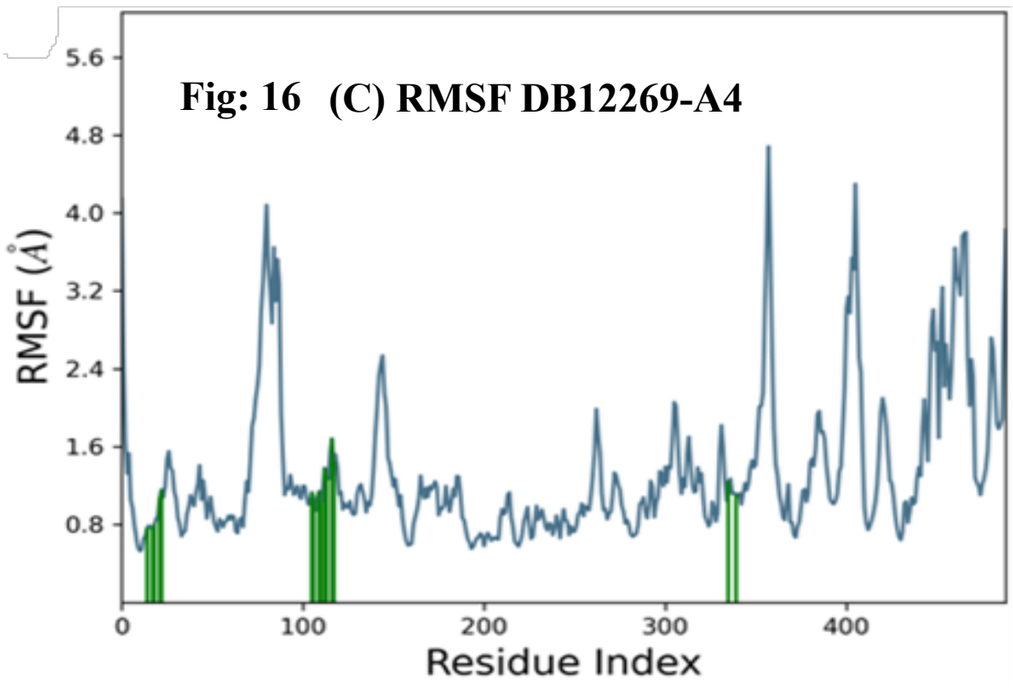


Fig: 16 (C) RMSF DB12269-A4

Analysis of protein-ligand contacts
For DB122269-A4

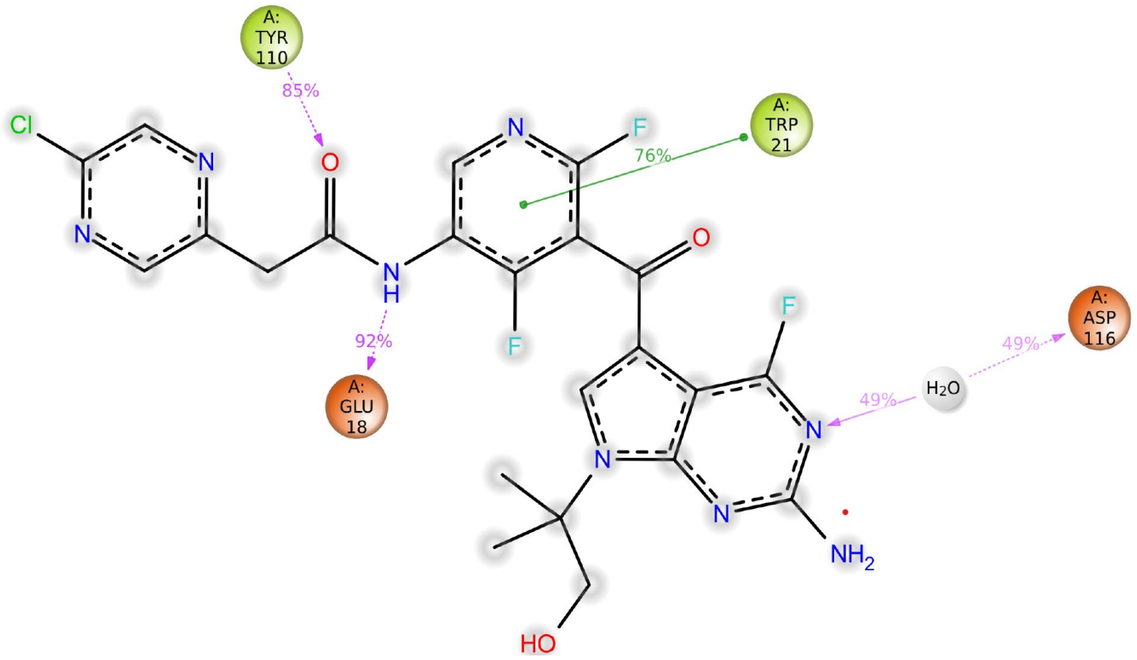


Fig: 17 (A)Ligand atom interactions with the protein residues

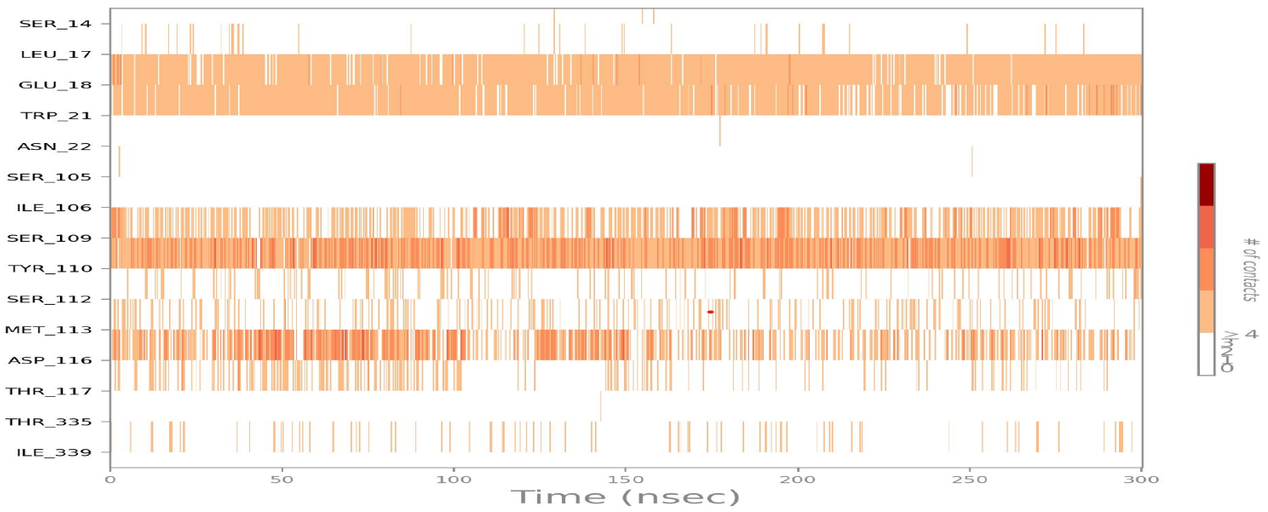


Fig: 17 (B) Interactions and contacts

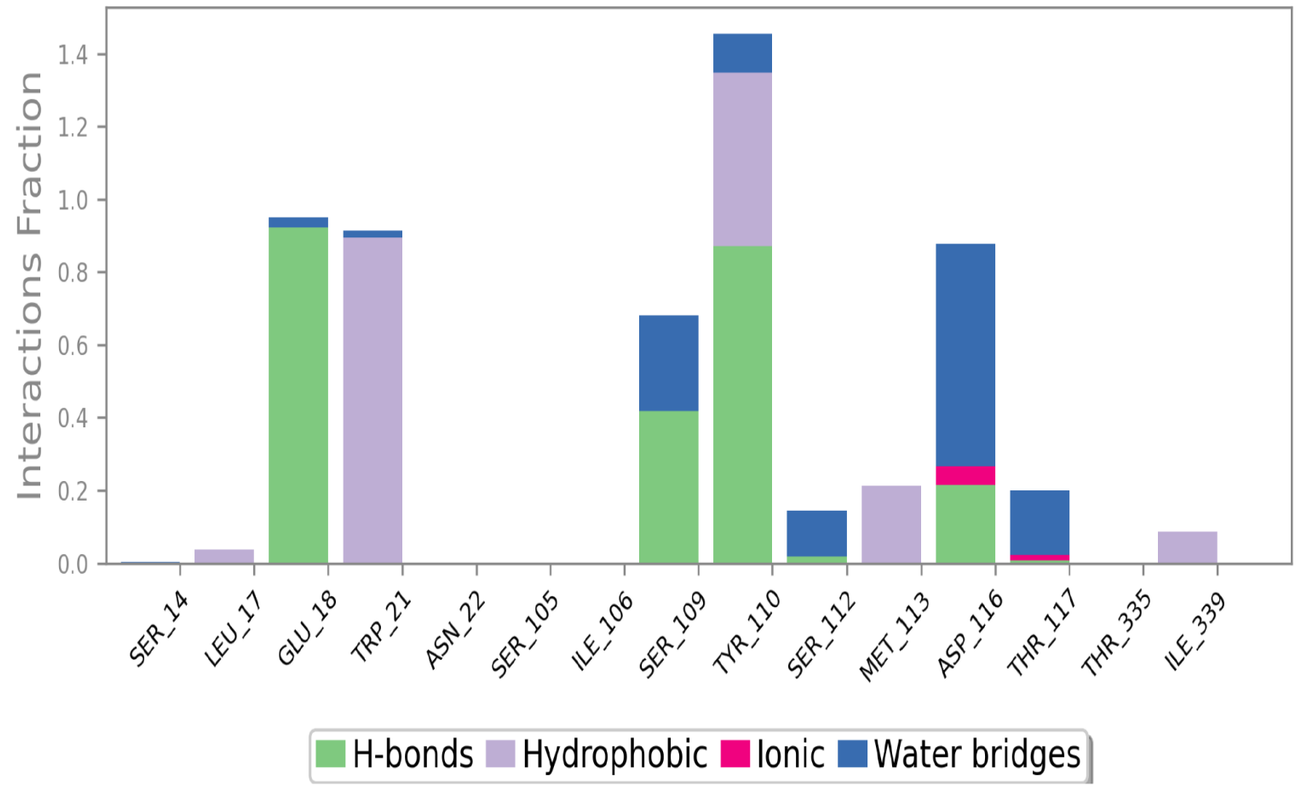


Fig: 17 (C) Type of interactions with ligand

- DB12269-A4 complex exhibited stable interactions with Tyr110 (85%), Trp21 (76%), and Glu18 (92%).
- And water-mediated hydrogen bonding with Asp116 (49%), indicating superior stability.

Analysis of protein-ligand contacts
For DB03231-A6

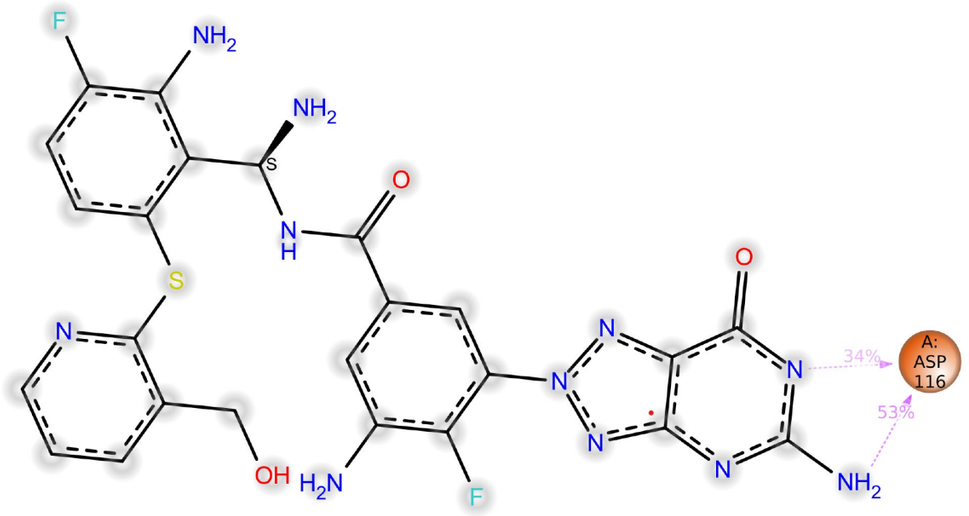


Fig: 18 (A)Ligand atom interactions with the protein residues

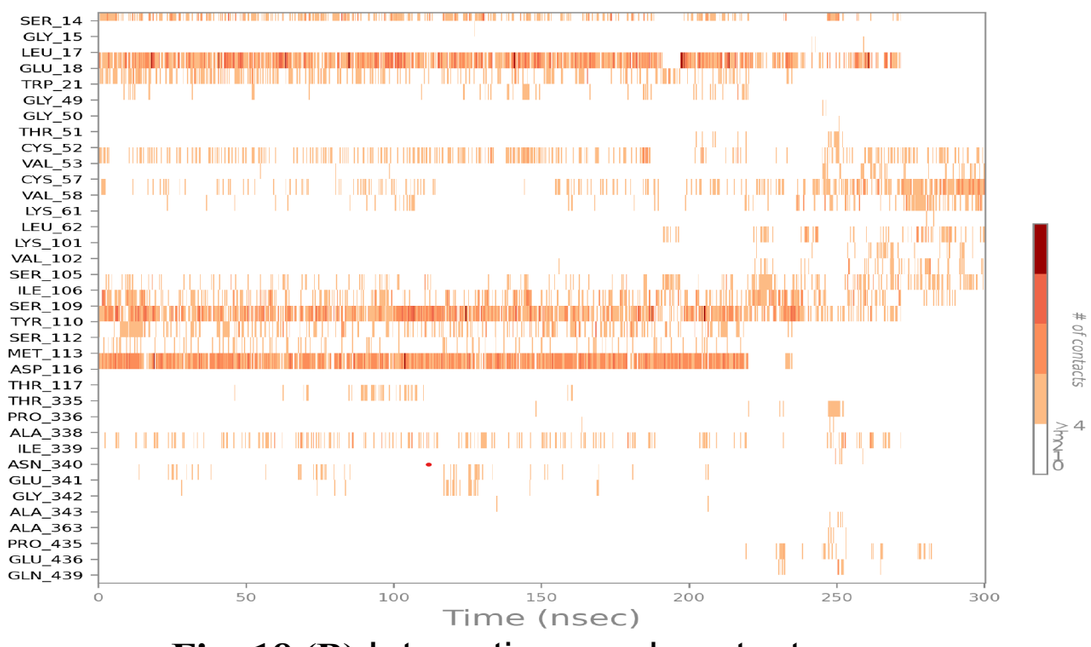


Fig: 18 (B) Interactions and contacts

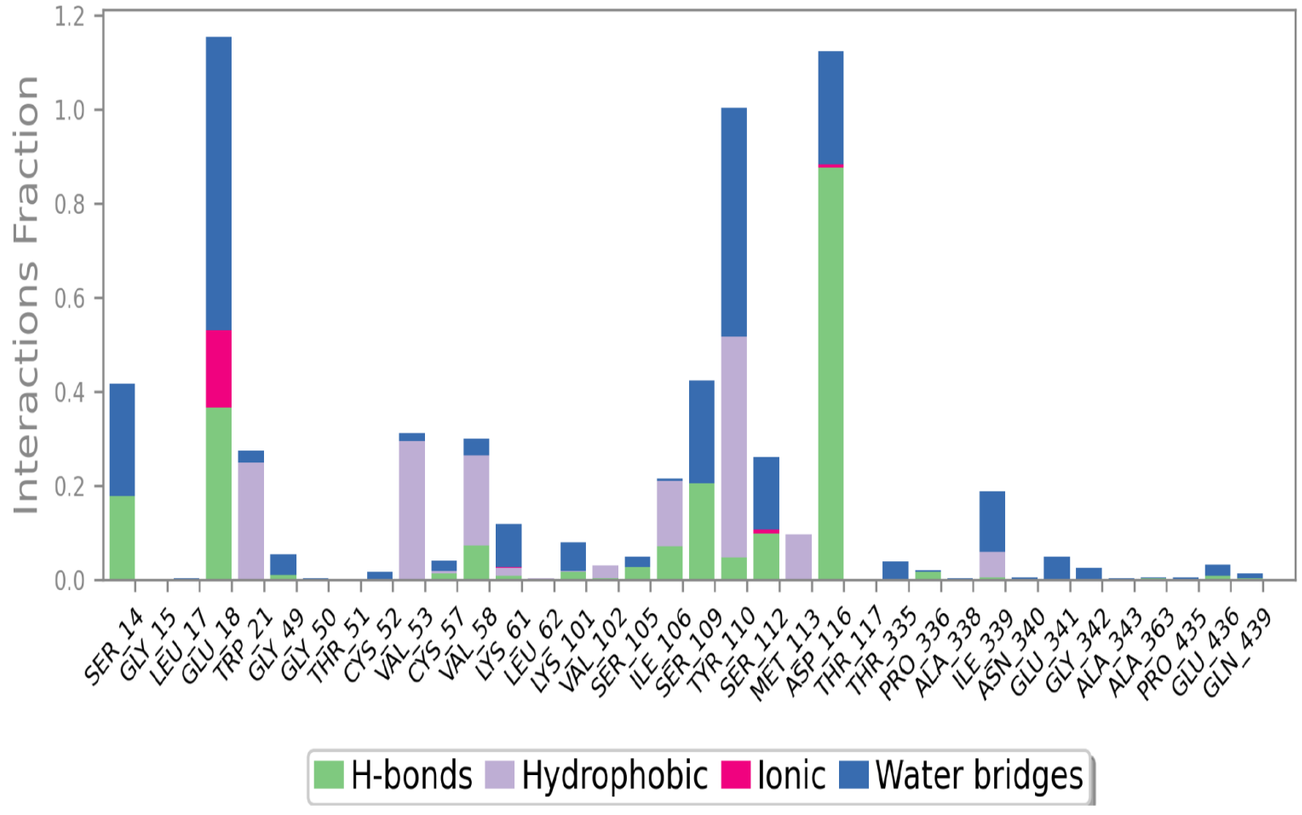


Fig: 18 (C) Type of interactions with ligand

- DB03231-A6 complex, the **nitrogen** atom and the attached **amine group** of the 5-amino-[1,2,3]triazolo[4,5-d]pyrimidin-7-one moiety.

- Formed hydrogen bonds with Asp116 for 34% and 53% of the simulation time, respectively.

Conclusion

- From QSAR model development the structural features revealed that descriptors like F03[C-N], F06[N-F], and Eta_F_A contributes positively in designing the structure and presence of B10[C-N] significantly reduces the efficacy of the compound designed.

- The developed model was deployed to reliably screen 12557 DrugBank compounds which enabled identification of the top five lead compounds i.e. DB12269, DB03231, DB01705, DB04260 and DB12457.

- The current study effectively utilizes simple, interpretable 2D molecular descriptors in QSAR modeling to predict TR inhibitory activity of 2-aminobenzimidazole derivatives with prospective lead modification.

- Positive descriptor features were incorporated and negative features were reduced to further potentiate the screened leads, resulting in the generation of 17 lead analogs.

- Finally through multi-layered screening involving inverse molecular docking and molecular dynamics revealed DB12269-A4 as the most promising TR inhibitor for leishmaniasis.

References

- Ghosh, V., Bhattacharjee, A., Kumar, A. and Ojha, P.K., 2024. q-RASTR modelling for prediction of diverse toxic chemicals towards T. pyriformis. SAR and QSAR in Environmental Research, 35(1), pp.11-30.

- Roy, K., Kar, S. and Das, R.N., 2015. Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. Academic press.

- Lipinski, C.A., 2004. Lead-and drug-like compounds: the rule-of-five revolution. Drug discovery today: Technologies, 1(4), pp.337-341.

- Finally through multi-layered screening involving inverse molecular docking and molecular dynamics revealed DB12269-A4 as the most promising TR inhibitor for leishmaniasis.

- Podder, T., Kumar, A., Bhattacharjee, A. and Ojha, P.K., 2023. Exploring regression-based QSTR and i-QSTR modeling for ecotoxicity prediction of diverse pesticides on multiple avian species. Environmental Science: Advances, 2(10), pp.1399-1422.

THANK YOU.....