Supplementary Information

Integrated Machine Learning, Molecular Docking, 3D-QSAR Based Approach for Identification of Potential Inhibitors of Trypanosomal N-Myristoyltransferase

Nidhi Singh[a,b], Priyanka Shah[a,c], Hemlata Dwivedi[d], Shikha Mishra[a], Renu Tripathi[d] , Amogh A Sahasrabuddhe[a] Mohammad Imran Siddiqi[a,b]*

[a]*Molecular and Structural Biology Division, CSIR-Central Drug Research Institute, Lucknow 226031, India*

[b]*Academy of Scientific and Innovative Research (AcSIR), CSIR-Central Drug Research Institute, Campus, Lucknow 226031, India;*

[c]*Current affiliation, Medgenome pvt Ltd. Bangalore*

[d]*Division of Parasitology, CSIR-Central Drug Research Institute, Lucknow, Uttar Pradesh, 226031, India.*

*Corresponding author: mi_siddiqi@cdri.res.in

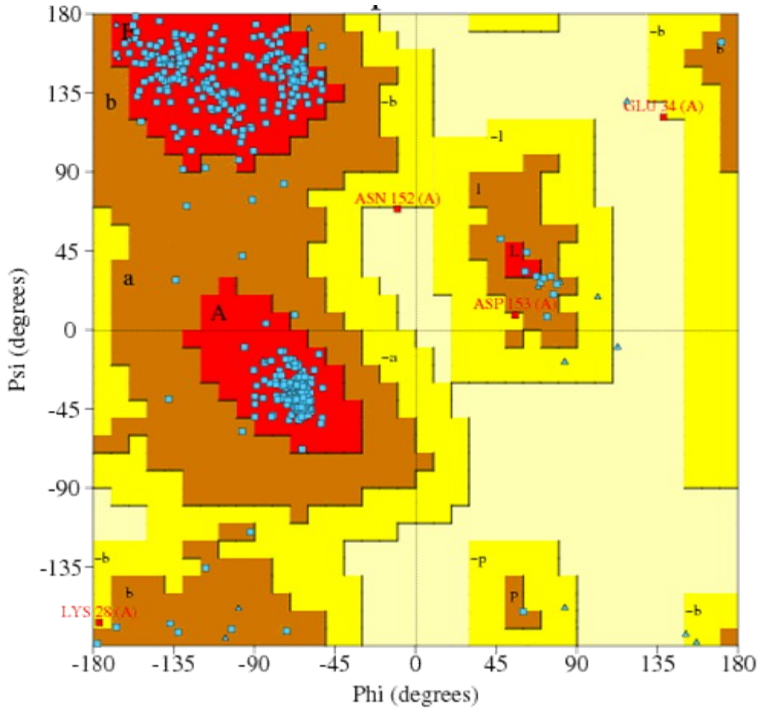


Figure S1: Ramachandran plot diagram shown for modeled N-myristoyl transferase of *T.brucei*

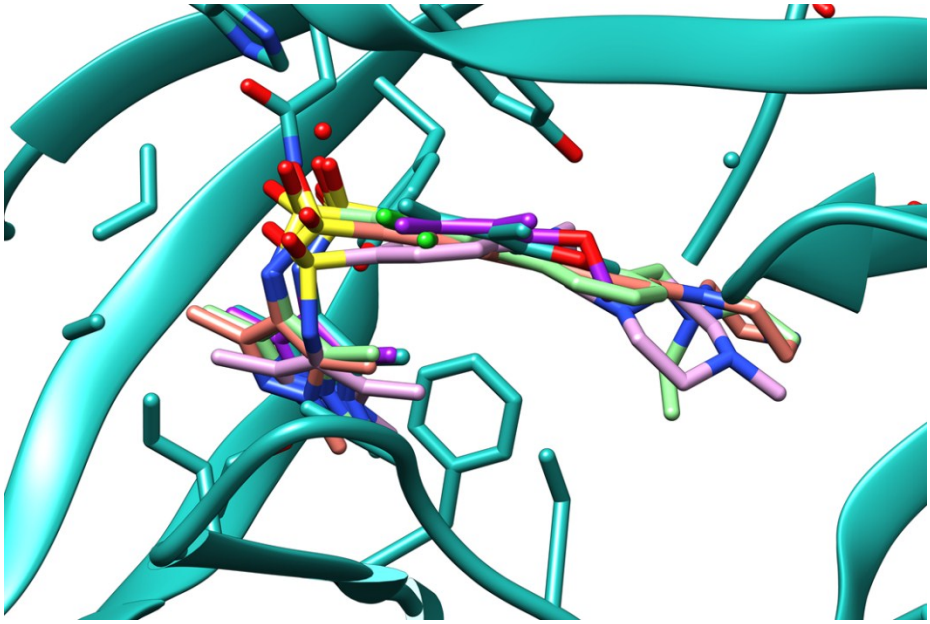
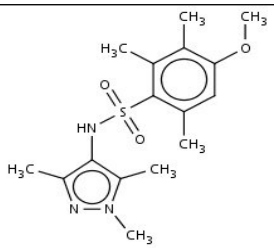
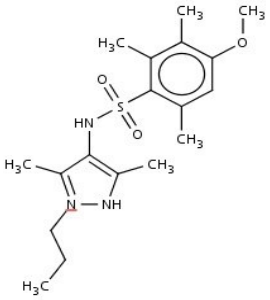
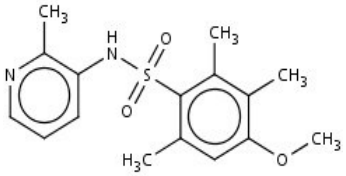
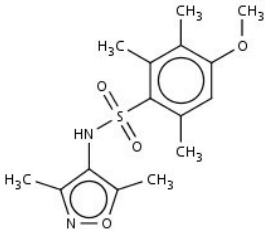
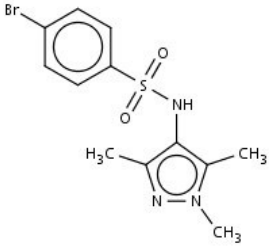
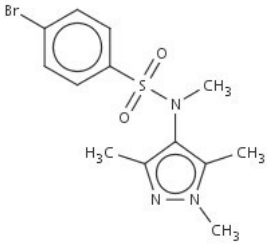
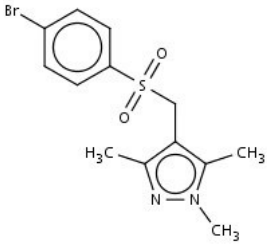
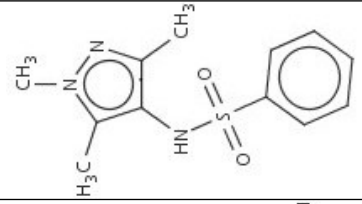
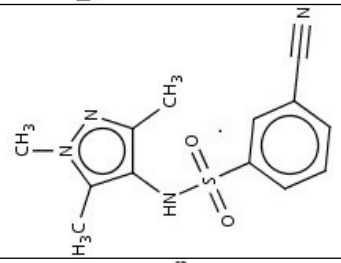
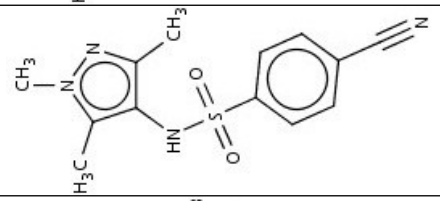
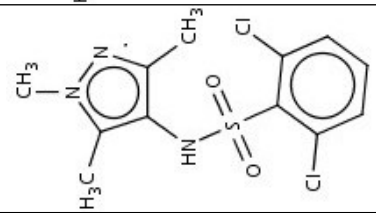
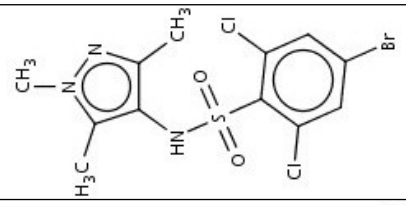
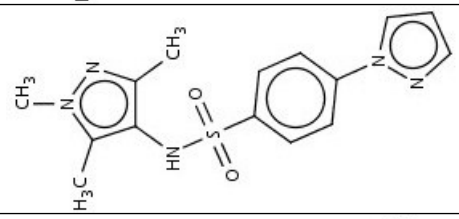
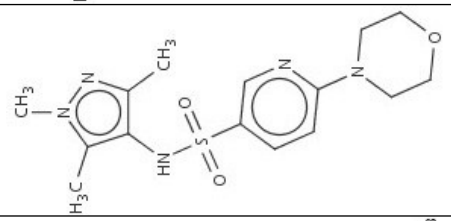
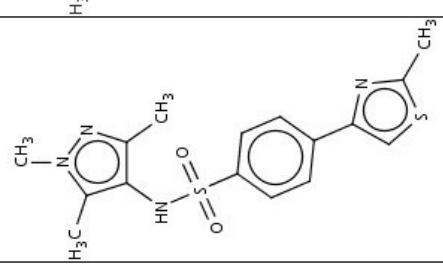
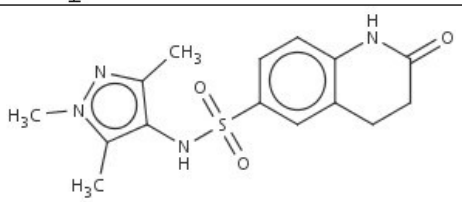
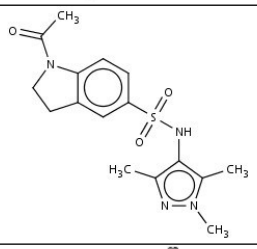
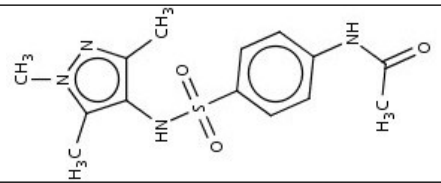
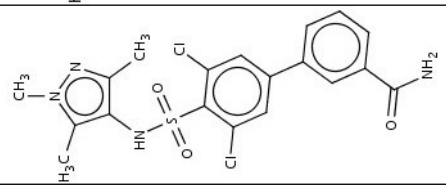


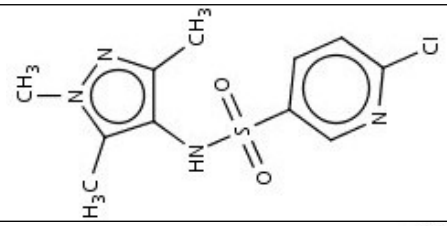
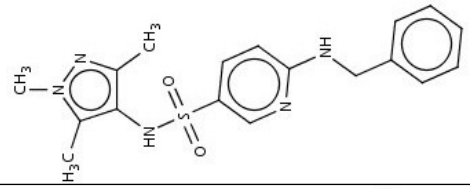
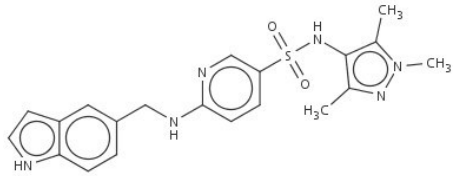
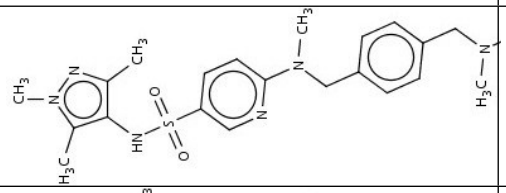
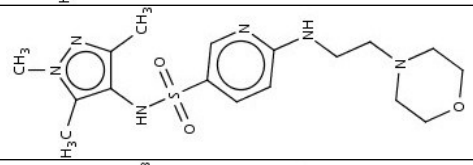
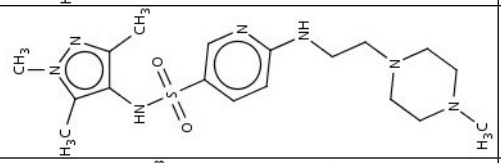
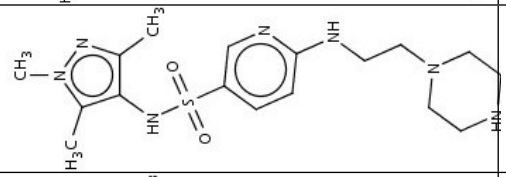
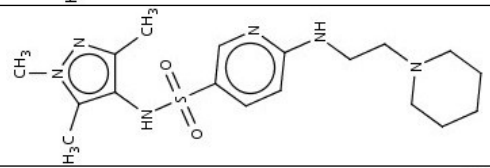
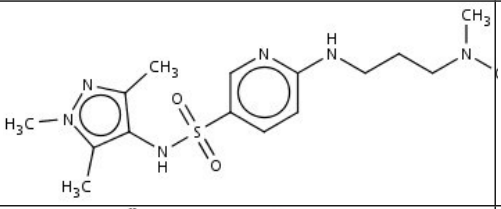
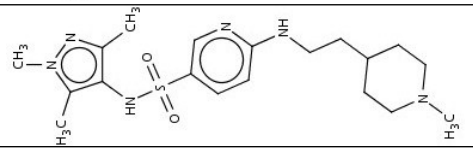
Figure S2: The docked poses of compound 1, 29, 42, 59 and 63 in the *Lm*NMT receptor (PDB accession code: 4A2Z)

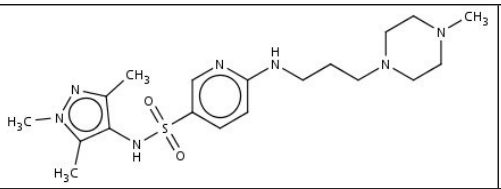
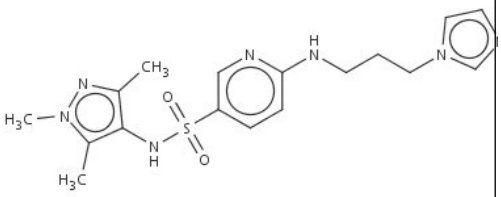
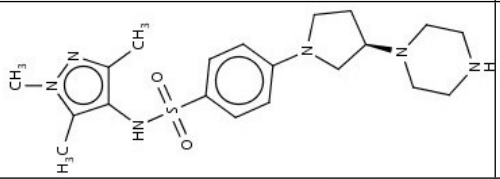
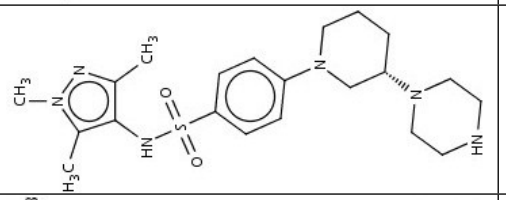
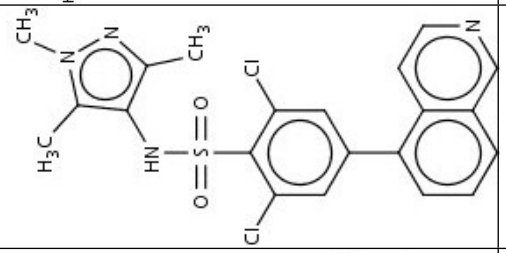
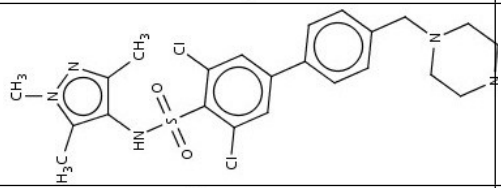
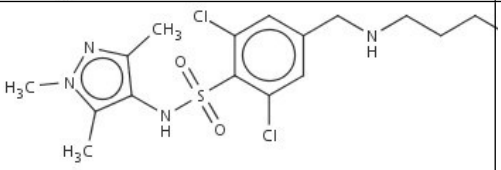
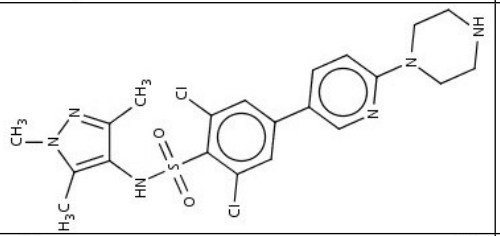
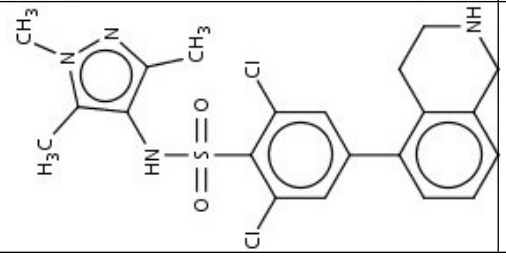
Table S1: The experimental activity and CoMFA predicted activity for compounds used in QSAR model development

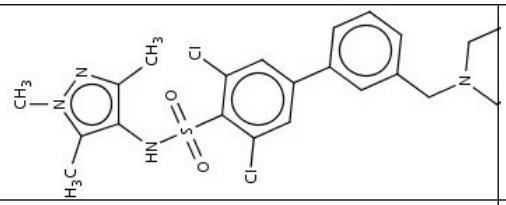
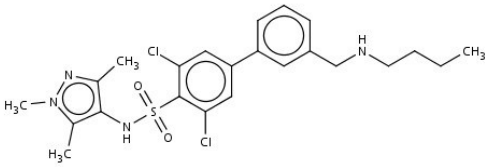
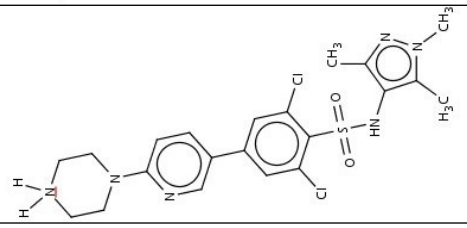
Compound code	Structure	Experimental activity (log IC50)	Predicted activity
1		5.721	5.715
3		4.552	4.556
9		5.000	5.397

10		4.721	4.739
15		4.853	4.916
16		5.000	5.014
20		4.376	4.391
25		4.920	4.983
26		4.552	4.524
27		4.494	4.479
28		6.000	5.988

29		6.468	6.429
30		5.318	5.303
31		5.721	5.725
32		4.920	4.904
33		5.481	5.472
34		5.065	5.117
35		5.552	5.573
36		6.045	6.518

37		4.585	4.525
38		5.886	5.882
39		6.481	6.525
40		6.443	6.618
41		4.420	4.392
42		6.853	6.895
43		6.000	6.002
44		4.853	5.629
46		5.677	5.636
47		7.522	7.535

49		6.958	6.938
50		5. 886	5.909
51		6.958	6.976
52		7.522	7.472
56		4.309	4.337
57		5.920	5.878
58		4.958	4.931
59		6.537	6.543
60		5.366	5.869

61		8.301	8.277
62		7.699	7.736
63		8.699	8.699

S3: Binding Pocket and Contour Map Analysis

CoMFA electrostatic (S3 C) and steric (S3 D) contours are shown in Figure S3. The steric interaction is represented by green (sterically favorable) and yellow (sterically unfavorable) contours, while electrostatic interaction is denoted by red (electronegative charge favorable) and blue (electropositive charge favorable) contours. The most potent compound 63 is embedded in the contour maps. A chunk of yellow contour near pyrazole ring explains the loss of activity by increasing steric bulk around it as seen in case of compound 3 as compared to compound 1. There are two small green contours present at meta position of substituents at rings attached next to sulfonamide group. This can explain the difference in activity of compounds 57 and 61, 58 and 62, 59 and 63. The majority of cavity portion is surrounded by sterically unfavorable yellow contours. The plausible reason for this is presence of amino acid residues with bulky side chains such as Phe 88 and Phe 232, Tyr 217 and Tyr 345 and His 219. The large difference in potency of compounds 41 and 42 despite of only difference in having morpholine instead of methylpiperazine at terminus of ethyl linkage can be explained on the basis of presence of large chunk of favorable electropositive substitution around it. This is also complemented by active site as compound 42 can form salt bridge with carboxyl group of the terminal residue of *Tb*NMT (val446).The compounds 56-63 having electropositive nitrogen in substituent rings falls in blue color region and are more potent as compared to the rest of compounds.

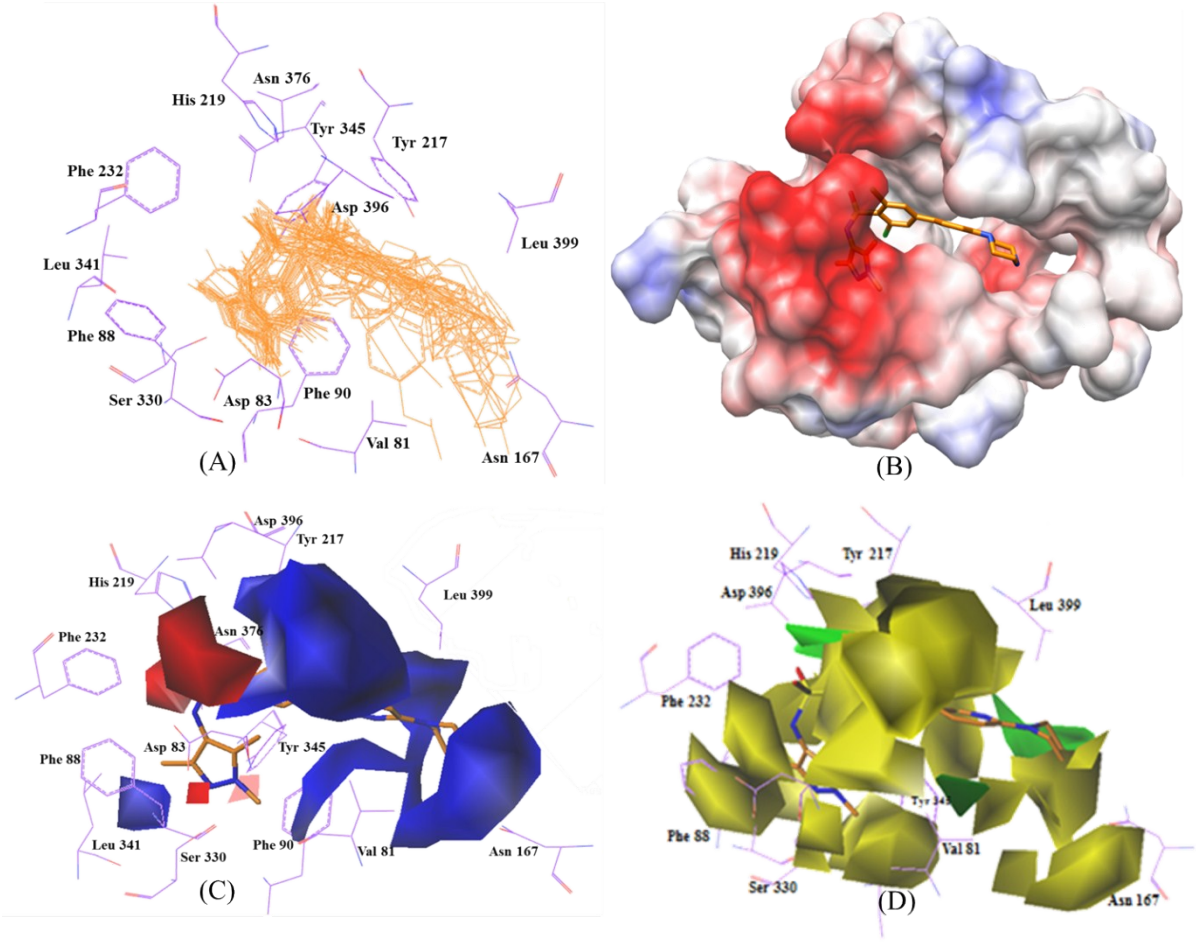


Figure S3: CoMFA steric and electrostatic contour map with most potent compound (63). Green color represents the region where steric bulk is favorable for increase in inhibitory activity and yellow color represents region where steric bulk is unfavourable; blue color represents favourable electropositive substitution and red color represents favourable electronegative substitution.