Chapter 6

Chapter 6

AI-Enabled Ultra-large Virtual Screening Identifies Potential Inhibitors of Choline Acetyltransferase for Theranostic Purposes

6.1. Introduction

Finding potentially selective ChAT tracer compounds holds a significant amount of enthusiasm among researchers, as it can play a pivotal role in exemplifying the functional role of the enzyme in the brain and peripheral systems. Despite these efforts, very few ChAT inhibitors have been identified and reported in the literature to date having some drawbacks as discussed earlier. Computer-aided virtual screening is one of the conventional approaches to overcome the above limitations, given that it basically works as a prefilter to screen a selection of molecules based on a predefined criterion (1). Structure-based virtual screening is one such approach that utilizes the three-dimensional (3D) structure of the protein to dock a library of compounds to its active site and select the top-scoring compounds based on the predicted binding scores, as seen from the great success of the **chapter 5 (Study 3)**. However, the traditional methodology has limitations due to the vast expansion of chemical databases (2). AI has demonstrated tremendous applicability in various domains with the pharmaceutical industry being one of the most widely integrated in the process of drug discovery and development. AI typically implies a methodology that can empower computers to mimic human intelligence, like learning and thinking abilities, which it uses to perform complex tasks (3). The benefits of incorporating AI in drug discovery are immense, as they can drastically reduce the time and money required for the discovery of new drugs (4). Deep docking (DD) is one such tool that harnesses the power of the deep neural network (DNN) algorithm with the objective of overcoming the limitations faced by traditional structure-based virtual screening, thereby significantly accelerating the process with up to 100-fold data reduction and 6000-fold data

enrichment of top-scoring molecules (5, 6). We were motivated to design our **Chapter 6 (Study 4)** to tap into the domain of an ultra-large library to screen billions of compounds in order to search for some novel scaffolds that might have potential ChAT inhibitory activity, opening up newer possibilities.

6.2. Objective

The objectives of this study are as follows:

- **AI-assisted Structure based virtual screening:** To virtually screen a an ultra large library of ~1.3B compounds using AI-assisted high-throughput structure based virtual screening against ChAT.
- **In-silico ADME property prediction:** To get top hits with optimum pharmacokinetic properties and with good BBB permeability, we performed in-silico ADME prediction on the obtained hits from AI-assisted virtual screening.
- **Molecular dynamics study:** To perform molecular docking against ChAT followed by extensive molecular dynamics simulation study to evaluate the binding mode of the compound with the target. Followed by per-residue hydrogen-bond interaction network analysis to elucidate its interaction with the catalytic residue HIS324 which indicates the compound may inhibit ChAT if tested in-vitro.
- **Tanimoto similarity search:** To check the novelty of the chemical scaffold of the identified hits, we compared with the previously reported known ChAT inhibitors.

6.3. Results and Discussion

6.3.1. AI-Enabled Virtual Screening Using Deep Docking

Despite the discovery of ChAT over 80 years ago, very few legitimate compounds have been reported as inhibitors of ChAT, having severe drawbacks like covalent binding to the receptor, poor BBB permeability, which hinders its bioavailability in the brain (7). Due to these

limitations, further investigation of ChAT ligands as potential therapeutics or diagnostics is warranted. We applied AI-assisted deep docking (AI-DD) and virtually screened 1.3 billion compounds from the ZINC20 database into the ChAT active cavity using the standard MPI-Vina docking protocols. The overall workflow of the current study is depicted in **Figure 6.1**.

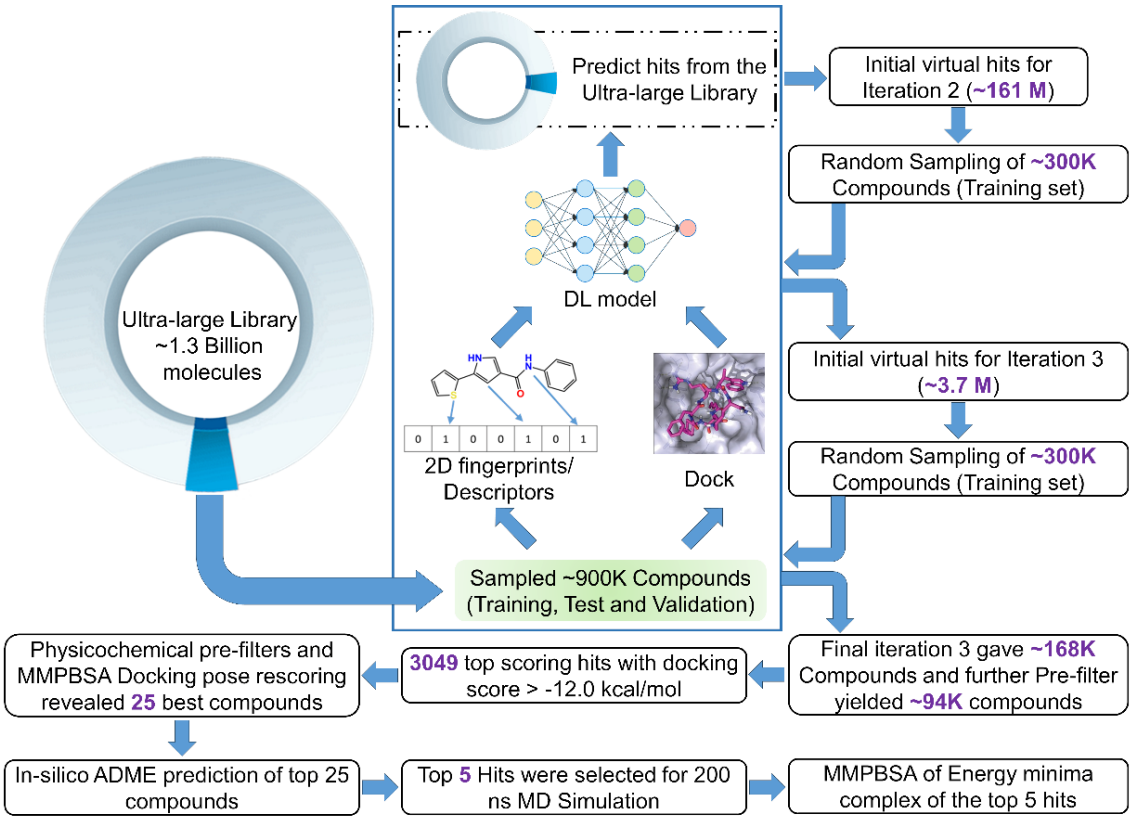

Figure 6.1. Overall workflow diagram of the Deep Docking based virtual screening protocol.

The statistical parameters for the best model of the iterations are listed in **Table 6.1** for the test set and in **Table 6.2** for the training and validation sets. In Iteration 1, with a cutoff of −10.10 kcal/mol, the model showed low precision (0.0683) but high recall (0.9029) and area under the curve (AUC) (0.9593) on the test set, indicating a sensitive but imprecise model, with poor generalization on the validation set (precision of 0.1290, recall of 0.7216). Thus, ~116 million hits were obtained from 1.3 billion compounds library. Followed by, iteration 2, using a cutoff of −10.20 kcal/mol, achieved a better balance with improved precision (0.7903), recall (0.9129), and AUC (0.9994) on the test set, while the validation set also improved (precision

0.3559, recall 1.0). A total of 3,739,639 hit compounds (~3.7 million) were obtained from 116 million compounds. In the final model, which is Iteration 3 with a cutoff of −11.40 kcal/mol, the model exhibited a precision of 0.4909, recall of 0.9643, and an AUC of 0.9999 on the test set, demonstrating high sensitivity and nearly perfect discrimination. While the precision dropped compared to earlier iterations, the recall remained strong. An increase in the recall value is a good indication that ensures that the proportion of actual positive cases is correctly identified by the model and is retained during the deep docking screening, minimizing the loss. Thus, a tradeoff between the precision and recall value can be observed in the third iteration (8). The training set performance (precision of 0.995, recall of 0.9962) and validation set performance (precision of 0.0865, recall of 1.0) suggest that the model captures positives effectively, albeit at the cost of precision. As each iteration is successive to the previous one, the final result reflects incremental improvements, with the third iteration offering the most comprehensive capture of relevant features across the data.

Table 6.1. Model test parameters for the Deep Docking iterations.

Iterations test set	Model Cutoff (kcal/mol)	Model Precision	Model Recall	Model AUC
1	-10.10	0.0683	0.9029	0.9593
2	-10.20	0.7903	0.9129	0.9994
3	-11.40	0.4909	0.9643	0.9999

Table 6.2. Model train and validation parameters for the Deep Docking iterations.

Iteration	Training Accuracy	Training Recall	Training Precision	Validation Accuracy	Validation Recall	Validation Precision
1	0.9967	0.9956	0.9978	0.9593	0.7216	0.1290
2	0.9928	0.9899	0.9956	0.9892	1.0	0.3559
3	0.9960	0.9962	0.995	0.9992	1.0	0.0865

The model statistics, denoted by the receiver operating characteristic (ROC) curve, are shown

in **Figure 6.2**. Overall, the AI-DD screening protocol, after the third iteration, identified 168,447 compounds from the 1.3 billion chemicals database. These were further reduced to 94,188 compounds after applying a molecular weight filter (150−450 Dalton), which were further screened by docking on ChAT using MPI-Vina. The docking results identified 3049 compounds as having a high docking score of above −12 kcal/mol, with docking score values ranging up to −13.5 kcal/mol. The compounds were further screened based on their ClogP <5, hydrogen bond donor <5, hydrogen-bond acceptor (HBA) <10, and rotatable bonds <10 which resulted in the top 215 compounds.

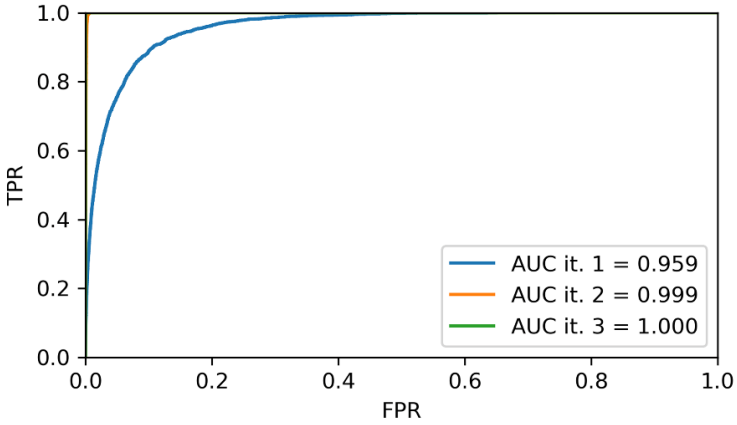

Figure 6.2. Receiver Operating Characteristic (ROC) curve for the three iterations, colour coded as iteration 1 (blue), iteration 2 (orange) and iteration 3 (green); TPR – True Positive Rate, FPR – False Positive Rate, AUC – Area Under the Curve.

6.3.2. Molecular Mechanics Poisson−Boltzmann Surface Area (MMPBSA) Rescoring of the Docking Poses

The utilization of AI-DD and the structure-based virtual screening protocol dramatically reduced the time and expense of identifying putative ChAT ligands. Structure-based virtual screening is primarily dependent on the obtained docking score, which utilizes low-accuracy force-field-based scoring functions that represent the binding affinity as a total sum of various classical mechanics terms like electrostatic interactions and van der Waals, and is excellent for

rapid screening of large data set, but has its limitations as these parameters are not practically transferable from computer models to actual molecules. However, as the data library is reduced to a manageable size, we employed a more accurate and rigorous method for rescoring the final binding pose based on the implicit solvent models, the Poisson−Boltzmann (PB) equation, which is called the MMPBSA method. Rescoring the docking pose using the MMPBSA method drastically improves the outcome, as it is more relevant to the system within the biological medium and helps to discriminate between correct and incorrect docking poses (9). MMPBSA for the top 215 AI-DD hit compounds (**shown in Table 6.3**) revealed the top 25 probable hits for the target protein ChAT with good binding free energy based on the MMPBSA rescoring.

Table 6.3. Top scoring hits obtained from the Deep Docking.

Sr. No.	**SMILES**	**Docking Score**	**MMPBSA Rescoring**
1	FC1=CC=C2C(=O)C\C(OC2=C1)=C3/O/C(N=N3)=C4/CCC5=C(CCC5)C4	-12	-35.2
2	FC1=CC=C2O\C(CC(=O)C2=C1)=C3\O\C(N=N3)=C4\CCC5=C(CCC5)C4	-12.1	-32.24
3	O=C(NCC1=CC=C(C=C1)C2=NC3=CC=CC=C3[NH]2)C4=CC(=O)C5=C(O4)C=CC=C5	-12.1	-32.05
4	O=C1C\C(OC2=C1C=CC=C2)=C3/O/C(N=N3)=C4/CCC5=C(CCC5)C4	-12	-31.77
5	CC(OC1=CC=CC2=C1C=CC=C2)C3=NC(=NO3)C4=CC=C(C=C4)C(=O)N5CCN(C)CC5	-12.2	-30.55
6	CC1=N[N](CC2=CC=C(C=C2)C3=NN=C(O3)C4=CC(=O)C5=CC=C(F)C=C5O4)C(=C1)C	-12.4	-29.56
7	O=C(CO[C]1=[C]=[C]2CCCC2=CC1)NCC3=CC=CC(=C3)N4C(=O)NC5(CCCC5)C4=O	-12	-28
8	CC(OC1=CN=C2C=CC=CC2=C1)C(=O)N3CCN(CC3)CC4=NC(=NO4)C5=CC=CC=C5	-12	-26.66
9	CC(OC(=O)C1=CC(=CC=C1F)N2CCNC2=O)C(=O)NC3=CC=C4C=CC=CC4=C3	-12	-26.34
10	O=C1CCC2=C(N1)C=C(C=C2)C3=NN=C(O3)C4=CC5=C(C=CC=C5)C=C4	-12.1	-25.95

11	ClC1=CC2=C(ON=C2N3CCC(CC3)C(=O)NCCNC4=NC5=C(C=CC=C5)C=C4)C=C1	-12.2	-25.1
12	O=C(COC1=CC=C2CCCCC2=C1)NCC3=CC=CC(=C3)N4C(=O)NC5(CCCC5)C4=O	-12.2	-25.06
13	CC1CN(CC(C)O1)C(=O)C2=CC=C(CNC(=O)C3=CC(=O)C4=CC=C(F)C=C4O3)C=C2	-12.1	-24.9
14	CC1=C(F)C=CC(=C1)N2CC(CC2=O)C3=NC(=NO3)C4=CC5=CC=CN=C5C=C4	-12.1	-24.47
15	CC(C)C1=CC=CC(=C1)C2=CC=C(O2)C(=O)N3CCN(CC3)C4=NC=NC5=C4C=N[N]5C	-12.2	-23.09
16	FC1=C(C=CC=C1)N2CCN(CC2)C(=O)CC3=NOC(=N3)C4CC5=CC=CC=C5C(=O)O4	-12.1	-23.05
17	O=C(CCNC(=O)C1=CC2=C(C=CC=C2)C=C1)N3CCCC(C3)C4=NN=C5CCCCC[N]45	-12.2	-22.75
18	FC1=CC(=C(C=C1)C2=CC(=CC=C2)C3=NC(=NO3)C4=NC5=C(CNCC5)C=C4)F	-12.1	-22.62
19	O=C(NCC1=CC=NC(=C1)N2CCCCC2)C(=O)NC3=CC4=C(C=CC=C4)C=C3	-12	-22.59
20	CC(OC1=CC=C2C=CC=CC2=C1)C(=O)N3CCC(CC3)C4C(=O)NC5=C4C=CC=C5	-12.7	-22.53
21	FC1=CC(=CC=C1)C2CC(=NO2)NC(=O)N3CCN(CC3)C4=NC5=CC=CC=C5O4	-13	-22.36
22	CC(C)(C)C1=CC=C(CCC(=O)NC2=N[NH]C(=N2)C3=NC4=C(C=CC=C4)C=C3)C=C1	-12.1	-22.31
23	FC1=CC(=C(C=C1)N2CC(CC2=O)C3=NC(=NO3)C4=CC5=C(C=CC=C5)C=C4)F	-12.2	-22.31
24	CC1=CC(=CC=C1)NC(=O)N2CCN(CC2)C(=O)COC3=CC=C4C=CC=CC4=C3	-12	-21.98
25	[O-]C1=[NH+]C=CC2=CC(=CC=C12)NC(=O)C3CCN(CC3)C(=O)C4=C(F)C=C(F)C=C4	-12.4	-21.91
26	CC(OC1=CC=C2C=CC=CC2=C1)C3=NN=C(O3)C4=CC=C5N(C)C(=O)CCC5=C4	-12	-21.55
27	FC1=CC=C2C(=O)C=C(OC2=C1)C(=O)NC3=CC=CC(=C3)C(=O)N4CCC5=C4C=CC=C5	-12.2	-21.42
28	CC1=CC=C(C=C1NC(=O)CC2=CC3=C(O2)C=CC4=C3C=C	-12	-21.38

	C=C4)N5CCNC5=O		
29	FC1=CC(=C2OCC(CC2=C1)NC(=O)N3CCC(=CC3)C4=C[NH]C5=NC=CC=C45)F	-12.5	-21.22
30	CC(OC1=CC=CC2=C1C=CC=C2)C(=O)NC3=CC(=CC=C3)N4CCCCC4=O	-12	-20.99
31	O=C(NCC1=CC(=O)OC2=CC3=C(OCO3)C=C12)NC4=CC5=C(C=CC=C5)C=C4	-12.3	-20.91
32	CC1=N[N](C(=C1)C)C2=NN=C(C=C2)N3CCN(CC3)C(=O)NC4=CC=C5C=CC=CC5=C4	-12.1	-20.87
33	CC(NC(=O)C1=N[NH]C2=CC=CC=C12)C(=O)[N]3C=CC(=C3)C4=CC=CC5=C4C=CC=C5	-12.4	-20.75
34	[O-]C1=[N+](C=C(C=C1)C2=NN=C(O2)C3=NC4=C(C=CC=C4)C=C3)C5=CC=CC(=C5)F	-12	-20.69
35	OC1=CC=C(O)[N]1C2=CC=CC(=C2)C3=NN=C(O3)C4=NC5=C(C=CC=C5)C=C4	-12.2	-20.39
36	CN1C(=O)NC2=C1CC(=C=N2)C3=CC(=CC=C3)C(=O)NC4=CC=C5CCCCC5=C4	-12	-20.16
37	CC(OC1=CC=CC2=C1C=CC=C2)C(=O)NC3CCCN(C3)C4=NC(=CC(=N4)C)O	-12	-20.15
38	O=C1CCCN1C2=CC=CC(=C2)C3=NN=C(O3)C4=NC5=C(C=CC=C5)C=C4	-12	-20.04
39	CC(OC1=CC=CC2=C1C=CC=C2)C(=O)NC3=CC(=CC=C3)N4C(=O)NC(C)(C)C4=O	-12	-19.73
40	CC(N1CCN(CC1)CC2=NC(=C3C=CC=CC3=N2)N)C(=O)NC4=CC5=C(C=CC=C5)C=C4	-12	-19.58
41	O=C(NC1=CC=C2C=CC=CC2=C1)C(=O)N3CCC(CC3)C4=NC(=N[NH]4)C5=CC=CC=C5	-12.2	-19.44
42	CC(C)(C)C1=CC=C(CCC(=O)NC2=NN=C([NH]2)C3=NC4=C(C=CC=C4)C=C3)C=C1	-12.1	-19.39
43	FC1=CC=CC(=C1)C2=CC(=NO2)NC(=O)N3CCC(=CC3)C4=C[NH]C5=C4C=CC=N5	-12.3	-19.13
44	CC1=C(Br)C=CC(=C1)CCC(=O)NC2=NN=C([NH]2)C3=NC4=C(C=CC=C4)C=C3	-12.1	-19
45	C[N]1N=C(C=C1NC(=O)C2=CC3=C(OC4=CC=CC=C4C3=O)C=C2)C5=CC=CC=C5C	-12	-18.98

46	CC1=CC=C(C=C1)C(=O)N2CCC(CC2)NC(=O)C3=NC4=C(C=CC=C4)C=C3	-12	-18.97
47	O=C1C=C(OC2=CC=CC=C12)C3=NN=C(O3)C4=CC=C5C=CC=CC5=C4	-12.2	-18.92
48	CC1=C(F)C=C(CNC(=O)C2CCN(CC2)C(=O)C3CC3C4=CC=C5C=CC=CC5=C4)C=C1	-12.2	-18.87
49	CN1C(=O)OC2=C1C=CC=C2C3=NN=C(O3)C4=CC=C5C=CC=CC5=C4	-12	-18.82
50	CC1(CCC2=CC=CC=C2C1)C3=NC(=NO3)CC(=O)N4CCN(CC4)C5=C(F)C=CC=C5	-12	-18.81
51	[O-]C1=C2C=CC(=CC2=NC3=[N+]1CCCCC3)C(=O)NCCNC4=NC5=C(C=CC=C5)C=C4	-12.5	-18.71
52	FC1=C(F)C(=CC=C1)C2CC2C(=O)NC3=NN=C([NH]3)C4=NC5=C(C=CC=C5)C=C4	-12.4	-18.68
53	FC1=CC=CC(=C1F)C2CC2C(=O)NC3=NN=C([NH]3)C4=NC5=C(C=CC=C5)C=C4	-12.1	-18.58
54	CC1=C(OC2=C(F)C=CC=C12)C3=NC(=NO3)CC(=O)N4CCN(CC4)C5=C(F)C=CC=C5	-12	-18.46
55	FC1=CC=C(NC(=O)CN2CCN(CC2)C(=O)NC3=CC4=C(C=CC=C4)C=C3)C=C1	-12	-18.23
56	FC1=CC(=C(C=C1)C(=O)N2CCC(CC2)C(=O)NC3=CC4=C(C=CC=C4)C=C3)F	-12.5	-18.12
57	FC1=CC=CC=C1C2=CC(=NO2)NC(=O)N3CC[N]4C(=NN=C4C5CCCC5)C3	-12.2	-18.08
58	FC(F)(F)C1=NC(=NO1)C2=CC(=CC=C2)C3=NN=C(O3)[C]4=[C]=[C]5CCCC5=CC4	-12	-17.95
59	C1CC(CC(N1)C2=NC(=NO2)C3=NC4=C(C=CC=C4)C=C3)C5=CC=CC=C5	-12.2	-17.87
60	O=C(NC1=NC2=C(C=CC=C2)C=C1)C3CCN(CC3)C(=O)C4=CC=CC5=C4C=CC=C5	-12.5	-17.79
61	[O-]C1=CC=[NH+]C2=C1C=CC(=C2)NC(=O)C3CCN(CC3)C(=O)C4=C(F)C=C(F)C=C4	-12.2	-17.73
62	FC1=CC=C2N(CCC2=C1)C(=O)NCC3=CC=CC(=C3)N4C(=O)NC5(CCCC5)C4=O	-12.2	-17.66

63	FC1=CC=CC(=C1N2CC(CC2=O)C3=NC(=NO3)C4=CC5=C(C=CC=C5)C=C4)F	-12	-17.65
64	[O-]C1=C2C=CC(=CC2=NC3=[N+]1CCC3)C(=O)NCCNC4=NC5=C(C=CC=C5)C=C4	-12.1	-17.39
65	[O-]C1=[N+](CC(=O)N2CCC(=CC2)C3=C[NH]C4=CC(=CC=C34)F)N=NC5=CC=CC=C15	-12	-16.83
66	CC1=CC=C(CCC(=O)NC2=NN=C([NH]2)C3=NC4=C(C=CC=C4)C=C3)C=C1Cl	-12.1	-16.59
67	FC1=CC=C2C(=C1)[NH]C=C2C3=CCN(CC3)CC4=NC(=O)C5=CC=C(C=C5N4)C(F)(F)F	-12.5	-16.55
68	FC1=CC=C2OC(=NC2=C1)N3CCN(CC3)CC4=NC(=NO4)C5=CC(=CC=C5)C(F)(F)F	-12	-16.51
69	CC(OC1=CN=C2C=CC=CC2=C1)C(=O)N3CCN(CC3)C4=N[N]5C=NN=C5C=C4	-12.3	-16.41
70	CC1CC2=CC=CC=C2CN1C(=O)CSC3=NN=C(O3)C4=CC=C5C=CC=CC5=C4	-12	-16.37
71	CC(=O)N1CCC2=C1C=CC(=C2)NC(=O)C(=O)N3CCC(=CC3)C4=C[NH]C5=CC(=CC=C45)F	-12	-16.3
72	FC(F)(F)C1=CC(=CC=C1)C2=NOC(=N2)CN3CCN(CC3)C4=CN=C5C=CC=CC5=N4	-12.3	-16.27
73	FC1=CC(=CC=C1)C(=O)NC2CCN(CC2)C(=O)CC3=CC=C4[NH]C5=CC=CC=C5C4=C3	-12	-16.26
74	CC1CCCC2=CC=CC=C2C1NC(=O)N3CCC(CC3)C(=O)NC4=CC=C5OCOC5=C4	-12.2	-16.25
75	CC1CCC2=C(C1)C3=CC(=CC=C3[NH]2)C(=O)OC(C)C(=O)C4=CC=C5N(CCC5=C4)C(C)=O	-12.1	-16.13
76	CN1C(=O)CCC2=C1C=CC(=C2)NC(=O)N3CCC(=CC3)C4=C[NH]C5=C4C=CC(=C5)C	-12	-16.12
77	O=C(NC1CCC2=C(C1)C=CC=C2)C(=O)NC3=CC=C(C=C3)C4=NN=C5CCCCCC[N]45	-12.1	-16.12
78	BrC1=C[N](CC2=CC=CC(=C2)C3=NN=C(O3)C4=CC(=O)C5=C(O4)C=CC=C5)N=C1	-12.2	-16.07
79	CC(C)C1=NC(=NC=C1)N2CCN(CC2)C(=O)NC3=NC4=C(C=CC=C4)C=C3	-12.1	-16.05

80	CC(NC(=O)C1=C2C=CC=CC2=N[NH]1)C(=O)[N]3C=CC(=C3)C4=C(F)C=CC=C4	-12	-15.91
81	CC(OC1=CN=C2C=CC=CC2=C1)C(=O)N3CCC(CC3)C4C(=O)NC5=C4C=CC(=C5)Cl	-12	-15.83
82	O=C1CC(CN1C2=CC=CC=C2)C3=NC(=NO3)C4=NC5=C(C=CC=C5)C=C4	-12	-15.77
83	O=C(NC1CCN(CC1)C(=O)NC2=CC=CC=C2)C3CC4=C(C=CC=C4)C(=O)O3	-12.5	-15.55
84	FC(F)(F)C1=CC=CC(=C1)N2CC(CC2=O)C3=NC(=NO3)C4=NC5=C(C=CC=C5)C=C4	-12.5	-15.45
85	CC(OC1=CN=C2C=CC=CC2=C1)C(=O)N3CCN(CC3)C4=NC5=CC=CC=C5O4	-12.4	-15.34
86	C[N]1N=C(C=C1NC(=O)N2CC[N]3C(=NN=C3C4CCCC4)C2)C5=CC=CC=C5C	-12	-15.23
87	[O-]C1=[NH+]C=C(C(=O)N2CCC(CC2)C(=O)NC3=CC=C(F)C(=C3)Cl)C4=CC=CC=C14	-12	-15.21
88	CC1CN(CCN1C2=NC=C[N]3C=NN=C23)C(=O)C(C)OC4=CC=C5C=CC=CC5=C4	-12	-14.93
89	CC1=CC=CC(=C1)C[N]2N=CC=C2NC(=O)C3=CC4=C([O-])[N]5C=CC=CC5=NC4=[N+]3C	-12	-14.84
90	O=C(NCCNC1=NC2=C(C=CC=C2)C=C1)C3=CN=C(C=C3)[N]4C=NC5=CC=CC=C45	-12.2	-14.71
91	CC(O[C]1=[C]=[C]2CCCC2=CC1)C(=O)N3CCN(CC3)C(C)C4=NC(=NO4)C5=CC=CC=C5	-12.1	-14.58
92	FC1=CC=C2C(=O)C=C(OC2=C1)C3=NN=C(O3)C4=CC=C5C=CC=CC5=C4	-12.2	-14.46
93	CC1=C(N=C(O1)C2=C(F)C=CC=C2)C(=O)N3CCC(CC3)C4=NC(=N[NH]4)C5=CC=CC=C5	-12.3	-14.42
94	O=C(NC1=NC(=C[NH]1)C2=CC=CC=C2)C3CCN(CC3)C(=O)C4=CC=CC5=C4C=CC=C5	-12	-14.38
95	CC1=C(OC2=CC=CC=C12)C3=NC(=NO3)CC(=O)N4CCN(CC4)C5=C(F)C=CC=C5	-12.2	-14.31
96	FC1=CC=CC2=C1N=C(C=C2)C(=O)NC3CCN(CC3)C(=O)NC4=CC=CC=C4	-12.4	-14.22
97	O=C(NC1=CC=C(CC[N]2N=NC(=N2)C3=CC=CC=C3)C=C1	-12.2	-13.91

	)N4CCC5=C(CC4)C=CC=C5		
98	CC1=C(C)N=C2C=C(C=CC2=N1)C3=NN=C(O3)C4=NC5=C(C=CC=C5)C=C4	-12.2	-13.76
99	FC1=CC=C(C2CC2C(=O)NC3=NN=C([NH]3)C4=NC5=C(C=CC=C5)C=C4)C(=C1)F	-12.5	-13.63
100	NC(=O)N1CCC2=C1C=C(NC(=O)CC3=CC4=C(O3)C=CC5=C4C=CC=C5)C=C2	-12	-13.5
101	FC(F)(F)C1=CC(=CC=C1)C(=O)NC2CCN(CC2)C(=O)NC3=CC=C4OCOC4=C3	-12.2	-13.47
102	FC1=CC=C2C(=O)C=C(OC2=C1)C3=NC(=NO3)C4CN(C(=O)C4)C5=CC=CC=C5	-12.3	-13.41
103	FC(F)(F)C1=CC(=CC=C1)N2CC(CC2=O)C3=NC(=NO3)C4=CC=C5C=NC=CC5=C4	-12	-13.22
104	O=C(NCC1=CN=C(C=C1)N2CCC3=C(C2)C=CC=C3)NC4=N[N]5C=CC=CC5=C4	-12	-13.22
105	CC(OC1=CC=C2C=CC=CC2=C1)C(=O)N3CCN(CC3)CC4=NN=N[N]4C5=CC=CC=C5	-12.3	-13.13
106	CC(OC1=CN=C2C=CC=CC2=C1)C(=O)N3CCC(CC3)C4=NN=C5C=CC=C[N]45	-12.1	-12.9
107	CC1=CC=C(NC(=O)C(=O)N2CCC(CC2)C3=NOC4=C3C=CC(=C4)F)C=C1C(F)(F)F	-12.1	-12.86
108	FC1=CC(=CC=C1)C(=O)NC2CCN(CC2)C(=O)C3=CC4=C(C=C3)C(=O)CC45CCNCC5	-12.2	-12.79
109	CC(OC1=CN=C2C=CC=CC2=C1)C(=O)N3CCC(CC3)C(=O)NC4=CC=C5OCOC5=C4	-12.2	-12.79
110	FC1=CC=C2OC(=CC(=O)C2=C1)C3=NN=C(O3)C4=CC=C5C=CC=CC5=C4	-12.3	-12.7
111	[O-]C1=[N+]2CCCC2=NC3=C1C=C(NCC4=NC(=NO4)C5=CC(=CC=C5)C(F)(F)F)C=C3	-12.1	-12.68
112	CC(NC(=O)C1=N[NH]C2=CC=CC=C12)C(=O)[N]3C=CC(=C3)C4=C(F)C=CC=C4	-12.1	-12.52
113	O=C(NC1=CC2=C(OCCCO2)C=C1)N3CCC(=CC3)C4=C[NH]C5=CC=CC=C45	-12.1	-12.48
114	C[N]1C(=NC2=C1C=CC(=C2)F)NC(=O)C3=CC=C(C=C3)C4=NN=C(O4)C5=CC(=CC=C5)C	-12	-12.39

115	FC(F)(F)C1=CC=CC(=C1)C2=NOC(=N2)CN3CCCC(C3)C4=NN=C5CCCCC[N]45	-12.7	-12.37
116	FC1=CC=C2C(=O)C=C(OC2=C1)C3=NC(=NO3)CC4=CC5=C(OCC5)C=C4	-12	-12.18
117	CC1=N[N](C=C1)C2=CC=CC(=C2)C3=NN=C(O3)C4=NC5=C(C=CC=C5)C=C4	-12	-12.12
118	O=C(NC1=NN=C(O1)C2=CC=C3CCCCC3=C2)C4COC5=CC=CC=C5O4	-12.5	-12.01
119	CC(=O)N1CCCC2=C1C=CC(=C2)C3=NOC(=N3)COC4=CC=C5C=CC=CC5=C4	-12.1	-11.88
120	CC(OC(=O)C1CCN(CC1)C2=NC3=CC=CC=C3O2)C(=O)NC4=CC=C5C=CC=CC5=C4	-12	-11.56
121	FC1=CC=C(C=C1)N2CCN(CC2)C3=CC(=CC=N3)CNC(=O)NC4=N[N]5C=CC=CC5=C4	-12.3	-11.43
122	O=C1C=C(OC2=CC=CC=C12)C3=NN=C(O3)C4=CC=C5C=CC=CC5=N4	-12.2	-11.31
123	O=C(COC1=CC2=C(C=CC=C2)C=C1)NC3=N[NH]C(=N3)C4=NC5=C(C=CC=C5)C=C4	-12.3	-11.27
124	[O-]C1=C2C=CC=CC2=CC(=[NH+]1)C3=NN=C(O3)C4=NC5=C(C=CC=C5)C=C4	-12	-11.26
125	O=C(NCC1=CN=C(C=C1)N2CCC3=C(C2)C=CC=C3)NC4=N[N]5C=CC=CC5=C4	-12	-11.24
126	O=C1OC(CC2=NC(=NO2)C3=CC=C(C=C3)C(=O)C4=CC=CC=C4)C5=C1C=CC=C5	-12	-11.02
127	CC1=NC2=C(C=CC(=C2)F)C=C1C(=O)N3CCN(CC3)CC4=NC(=NO4)C5=CC=CC=C5	-12	-10.91
128	FC1=CC=CC=C1C2=NC(=NO2)C3=CC(=CC=C3)C(=O)NC4=N[N]5C=CC=CC5=C4	-12.2	-10.91
129	FC1=CC=C(C=C1)C2CC(CCO2)NC(=O)C(=O)NC3=CC4=C(C=CC=C4)C=C3	-12.1	-10.67
130	FC1=CC(=CC=C1)C(=O)NC2CCN(CC2)C(=O)C3=CC4=C(OC5=CC=CC=C5C4=O)C=C3	-12.4	-10.62
131	O=C(NC1CCC2=C(C1)C=CC=C2)C(=O)NC3=CC(=CC=C3)C4=CC5=C(OCC5)C=C4	-12.2	-10.49
132	O=C(NCC1CCCN(C1)C2=NC=C[N]3C=NN=C23)C4=NC5=	-12.2	-10.39

	C(C=CC=C5)C=C4		
133	FC1=CC=C(OC2=NC=C(C=C2)C3NOC(=N3)C4=CC(=NO4)C5=C(F)C=CC(=C5)F)C=C1	-12	-10.23
134	CC(OC1=CC=C2C=CC=CC2=C1)C3=NN=C(O3)C4CC5=C(C=CC=C5)C(=O)O4	-12	-9.97
135	CC(OC1=CC=C2C=CC=CC2=C1)C(=O)N3CCN(CC3)C4=NC=NC5=C4C=N[N]5C	-12	-9.83
136	FC(F)(F)C1=CC=CC(=C1)[N]2C=CC(=N2)C(=O)N3CCC(CC3)C4=NN=C5CCCC[N]45	-12.5	-9.75
137	O=C(COC1=CC2=C(C=CC=C2)C=C1)NCC3=CC(=CC=C3)N4C(=O)NC5(CCCC5)C4=O	-12.5	-9.72
138	OC1=C2C=CC=CC2=C(O)[N]1CC3=CC=CC(=C3)C(=O)NC4=NC5=C(C=CC=C5)C=C4	-12	-9.64
139	[O-]C1=[N+](N=C(C=C1)C2=NN=C(O2)C3=NC4=C(C=CC=C4)C=C3)C5=CC=C(F)C=C5	-12	-9.61
140	O=C1CCCC2=C1C=C(NC(=O)N3CCC(=CC3)C4=C[NH]C5=NC=CC=C45)C=C2	-12.1	-9.5
141	[O-]C1=C(C=CC=[N+]1C2=CC=C(F)C=C2)C3=NC(=NO3)C4=NC5=C(C=CC=C5)C=C4	-12.1	-9.5
142	FC1=CC(=C(C=C1)C2CC2C(=O)NC3=NN=C([NH]3)C4=NC5=C(C=CC=C5)C=C4)F	-12.4	-9.37
143	FC(F)(F)C1=CC=C2C(=O)N=C(CN3CCC(=CC3)C4=C[NH]C5=CC=CC=C45)NC2=C1	-12.2	-9.29
144	ClC1=CC(=CC(=N1)N2CCCC2)C3=NN=C(O3)C4=NC5=C(C=CC=C5)C=C4	-12.3	-9.01
145	ClC1=C(C=CC=C1)C2=CC(=NO2)C(=O)N3CCC(CC3)C4=NN=C([NH]4)C5=CC=CC=C5	-12.2	-8.86
146	CC1=CC=CC(=C1)C2=NC(=C(C)[NH]2)C(=O)N3CCC(CC3)C4=NOC5=C4C=CC(=C5)F	-12.1	-8.83
147	CC1=CC(=N[N]1C2=CC=CC(=C2)C(F)(F)F)C(=O)NC3=NN=C([NH]3)C4=C(F)C=CC=C4F	-12.1	-8.81
148	C1CC(CN(C1)C2CCNCC2)C3=NC(=NO3)C4=NC5=C(C=CC=C5)C=C4	-12.1	-8.75
149	FC(F)(F)C1=CC(=CC=C1)C2=NOC(=N2)CN3CCCC(C3)C4=	-12	-8.63

	NN=C5C=CC=C[N]45		
150	CC(OC1=CN=C2C=CC=CC2=C1)C(=O)N3CCC(=CC3)C4=CC=C5OCCOC5=C4	-12.2	-8.55
151	FC1=CC=C(NC(=O)C2CCN(CC2)C(=O)NC3C4CCC5=C(C=CC=C5)C34)C=C1Cl	-12.1	-8.49
152	OC1=CC2=C([NH]1)C=CC(=C2)NC(=O)C3CCN(CC3)C(=O)C4=C(F)C=C(F)C=C4	-12.1	-8.29
153	CC1=CC=CC(=C1)C2=CC=CC(=C2)C(=O)NC3=NN=C([NH]3)C4=C(F)C=CC=C4	-12	-8.27
154	FC1=CC=C2C(=C1)[NH]C=C2C3=CCN(CC3)C(=O)CCNC(=O)C4=CC5=C(C=CC=C5)C=C4	-12	-8.25
155	O=C1CCCCC2=CC(=CC=C12)C(=O)N3CCN(CC3)CC4=NC(=NO4)C5=CC=CC=C5	-12	-7.66
156	FC1=C(C(=CC=C1)F)C2=NN=C(NC(=O)C3=N[N](C=C3)C4=CC(=CC=C4)C(F)(F)F)[NH]2	-12	-7.57
157	FC1=CC=C(OC2=NC=C(C=C2)C3NOC(=N3)C4=CC(=NO4)C5=C(F)C=CC(=C5)F)C=C1	-12	-7.29
158	FC1=CC2=C(C=C1)C(=NO2)C3CCN(CC3)C(=O)COC4=CC5=C(C=CC=C5)C=C4	-12	-7.24
159	CC1=CC(=CC=C1C2=C(F)C=C(C=C2)C3=NN=C(C[N]4N=NC(=N4)C5=CC=CC=C5)O3)F	-12.1	-7.1
160	FC1=CC=C(C=C1)C2CC(CCO2)NC(=O)N3CCC(=CC3)C4=CC=C5OCCOC5=C4	-12.1	-7.01
161	O=C1CCCCC2=CC=C(NC(=O)N3CCC(CC3)NC(=O)C4=CC=CC=C4)C=C12	-12.2	-6.87
162	CC1=NC(=CC=C1)NC2=NN=C(C=C2)N3CCN(CC3)C(=O)NC4=CC=C5C=CC=CC5=C4	-12.8	-6.86
163	O=C(NC1=CC=CC(=C1)C(=O)NC2=NC3=C(C=CC=C3)C=C2)C4CC5=C(O4)C=CC=C5	-12.3	-6.84
164	FC1=CC(=C2CCN(CC2=C1)C(=O)CSC3=NN=C(O3)C4=CC5=C(C=CC=C5)C=C4)F	-12	-6.74
165	CC1=NN=C2[N]1C=CN=C2N3CCN(CC3)CC4=NC(=NO4)C5=CC=CC(=C5)C(F)(F)F	-12.5	-6.73
166	CC(OC1=CC=C2C=CC=CC2=C1)C(=O)N3CCC(=CC3)C4=C[NH]C5=C4C=CC=N5	-12.2	-6.72
167	FC1=CC(=C(F)C=C1)C2CC2C(=O)NC3=NN=C([NH]3)C4=N	-12.3	-6.53

	C5=C(C=CC=C5)C=C4		
168	CC1=CC=CC(=C1)C(=O)NC2CCN(CC2)C(=O)NC3=CC=C4CCCCC(=O)C4=C3	-12.5	-6.4
169	OC1=C2C=CC=C(F)C2=C[N]1C3CCN(C3)C(=O)NC4=CC5=C(C=CC=C5)C=C4	-12	-6.28
170	CC1OC2=C(NC1=O)C=CC=C2C(=O)N3CCC(CC3)NC(=O)C4=CC(=CC=C4)C	-12	-6.19
171	FC1=CC2=C(C=C1)C(=NO2)C3CCN(CC3)C(=O)C4=CC=C([NH]4)C5=CC=CC=C5	-12.1	-6.11
172	CC(OC1=CC=CC2=C1C=CC=C2)C(=O)NC3=CC4=C(C=C3)N=C5CCC[N+]5=C4[O-]	-12.4	-5.98
173	FC(F)(F)C1=CC(=CC=C1)N2CC(CC2=O)C3=NC(=NO3)[C]4=[C]=[C]5C(=O)CCC5=CC4	-12	-5.95
174	CC(OC1=CC=C2C=CC=CC2=C1)C(=O)N3CCC(CC3)C(=O)NC4=CC(=CC=C4)F	-12.3	-5.88
175	CC1=CC2=C(C=C1)C(=C[NH]2)C3=CCN(CC3)C(=O)NC4=CC=C(C[N]5C(=CC=C5O)O)C=C4	-12	-5.41
176	CC1=N[N](CC2=CC=CC(=C2)C3=NN=C(O3)C4=CC(=O)C5=CC=C(F)C=C5O4)C(=C1)C	-12.1	-5.29
177	CC1=CC=CC=C1C2=NC(=NO2)CNC(=O)C3=CC4=C(C=C3)C5=CC=CC=C5O4	-12	-5.26
178	CC1=CC=CC(=C1)C(=O)NC2CCN(CC2)C(=O)C3=[NH+]C(=C4C=CC=CC4=C3)[O-]	-12.2	-5.23
179	CC(N1CCCC(C1)C2=NN=C3CCCCC[N]23)C(=O)NC4=CC=C5C=CC=CC5=C4	-12.1	-5.21
180	CC1=CN=C(NC(=O)C2CCN(CC2)C(=O)C3=CC4=C([NH]C5=C4CCCC5)C=C3)C=C1	-12.1	-4.85
181	CC1=CC(=CC=C1[N]2C=CC(=N2)C(=O)N3CCC(CC3)N4CCC5=CC=CC=C5C4)F	-12.1	-4.64
182	CC1=N[N](CC2=CC=CC(=C2)C3=NN=C(O3)C4=CC(=O)C5=CC(=CC=C5O4)F)C(=C1)C	-12.2	-4.53
183	CC1=NC(=NO1)C2=CC=CC(=C2)C3=NN=C(O3)C4=CC=C5C=CC=CC5=C4	-12.1	-4.44
184	FC(F)(F)C1=CC(=NC=C1)C2=NOC(=N2)CC3=CC4=C(OCC5=CC=CC=C5C4=O)C=C3	-12.2	-4.38
185	CC1CN(CC(C)O1)C2=CC=NC(=C2)C3=NN=C(O3)C4=CC=	-12.1	-4.26

	C5C=CC=CC5=C4		
186	CC1=CN=C(NC2=NN=C(C=C2)N3CCN(CC3)C(=O)NC4=CC=C5C=CC=CC5=C4)C=C1	-12.1	-4.24
187	FC1=CC=C(C=C1)C(=O)NC2CCN(CC2)C(=O)NC3=CC=C4OC5(CCCC5)OC4=C3	-12.1	-4.06
188	O=C(NC1CCC2=C(C1)C=CC=C2)C(=O)N3CCC(CC3)C(=O)N4CCC5=CC=CC=C45	-12.8	-3.84
189	FC(F)(F)C1=CC=CC(=C1)[N]2C=C(N=N2)C3=NN=C(O3)C4=NC5=C(C=CC=C5)C=C4	-12.1	-3.83
190	[O-]C1=C(C=CC=[N+]1C2=CC=C(F)C=C2)C3=NN=C(O3)C4=CC5=C(C=CC=C5)C=C4	-12.1	-3.43
191	O=C(NC1=N[N](CCC2=CC=NC=C2)C=C1)N3CCC(=CC3)C4=C[NH]C5=CC=CC=C45	-12	-3.36
192	CC1=CC(=CC=C1[N]2C=C(N=N2)C(=O)N3CCC(CC3)C4=NC(=NO4)C5=CC=CC=C5)F	-12.1	-3.34
193	O=C1CC(CN1C2=CC3=C(CCC3)CC2)C(=O)NC4=NC=C(C=N4)C5=CC=CC=C5	-12	-3.29
194	[O-]C1=C(C=CC=[N+]1CC2=CC=C(F)C(=C2)F)C3=NC(=NO3)C4=NC5=C(C=CC=C5)C=C4	-12	-2.94
195	OC1=CC=C(O)[N]1CC2=CC=C(NC(=O)N3CCC(=CC3)C4=C[NH]C5=CC=CC=C45)C=C2	-12.2	-2.94
196	CN1CCC2=C(C=C(NC(=O)N3CCC(=CC3)C4=C[NH]C5=C4C=CC(=C5)C)C=C2)C1=O	-12	-2.76
197	FC(F)(F)C1=CC(=CC=C1)C(=O)NC2CCN(CC2)C(=O)N[C]3=[C]=[C]4CCCC4=CC3	-12.3	-2.48
198	FC1=CC=C2C(=O)C=C(OC2=C1)C3=NC(=NO3)C4=NC5=C(C=CC=C5)C=C4	-12	-2.41
199	FC1=C(C=CC=C1)[N]2C=CC(=N2)C(=O)N3CCC(CC3)C4=NC(=N[NH]4)C5=CC=CC=C5	-12	-2.39
200	CC1=CC(=NC2=C1C=CC=C2)C3=NN=C(O3)C4=NC5=C(C=CC=C5)C=C4	-12.3	-2.27
201	CC1=N[N](CC2=CC=CC(=C2)C3=NN=C(O3)C4=CC(=O)C5=CC=CC=C5O4)C(=C1)C	-12.1	-2.15
202	FC1=CC=C2C(=O)C=C(OC2=C1)C(=O)OCC(=O)C3=CC4=C	-12.1	-2.04

	(CC3)C5=C(CCCC5)C4		
203	CC1=CC=C2OC(=CC(=O)C2=C1)C3=NC(=NO3)C4=NC5=C(C=CC=C5)C=C4	-12	-2.02
204	[O-]C1=[NH+]C(=NC2=CC(=CC=C12)C(F)(F)F)CN3CCC(=CC3)C4=C[NH]C5=CC(=CC=C45)F	-12.1	-1.92
205	FC1=CC2=C(C=C1)C(=NO2)C3CCN(CC3)C(=O)C4=N[N](C=C4)C5=CC=CC=C5F	-12.1	-1.71
206	NC1CCN(CC1)C(=O)C2=CC=CC(=C2)C3=NC(=NO3)C4=CC5=C(C=CC=C5)C=C4	-12	-1.36
207	FC1=CC(=C(C=C1)[N]2C=CC(=N2)C(=O)N3CCC(CC3)C4=NC(=N[NH]4)C5=CC=CC=C5)F	-12.1	-1.14
208	CC1=CC=C(C=C1)[N]2N=C(C=C2C)C3=NN=C(O3)C4=NC5=C(C=CC=C5)C=C4	-12	-1.13
209	CC(OC1=CC=C2C=CC=CC2=C1)C3=NN=C(O3)C4=CC5=C(C=C4)N(CC5)C(C)=O	-12	-0.98
210	FC1=CC=C(C=C1)[N]2C=CC(=N2)C(=O)N3CCN(CC3)C4=NC5=C(C=CC=C5)C=C4	-12.1	-0.88
211	FC(F)(F)C1=CC=C(C=C1)N2CCC(NC(=O)C(=O)NC3=CC4=C(C=CC=C4)C=C3)C2=O	-12	-0.55
212	O=C1CN(C(=O)CCCC2=NC(=NO2)C3=CC4=C(C=CC=C4)C=C3)C5=CC=CC=C5N1	-12.3	-0.42
213	CC1CCC2=C(C1)C3=C([NH]2)C=CC(=C3)C(=O)N4CCC(CC4)C(=O)NC5=NC=C(C)C=C5	-12.4	-0.42
214	CC1=CC=CC(=C1)C(=O)NC2CCN(CC2)C(=O)C3=CC4=C(C=CC=C4)C(=O)O3	-12	-0.17
215	O=C(CC1=CC=C2C=CC=CC2=C1)NC3=N[NH]C(=N3)C4=NC5=C(C=CC=C5)C=C4	-12.2	-0.08

6.3.3. In Silico Pharmacokinetic Study

To understand the bioavailability and blood−brain barrier permeability of the hits, we performed in silico pharmacokinetic analysis of the top 25 hits obtained from the MMPBSA rescoring of the docked poses, utilizing DruMAP (Drug Metabolism and Pharmacokinetics Analysis Platform) web server facility. From the ADME parameters, the top 5 hit molecules

were CPD1 (ZINC001329075107), CPD2 (ZINC000685404542), CPD3 (ZINC001258168283), CPD4 (ZINC000102819103), and CPD5 (ZINC001255982897), as shown in **Figure 6.3**. These were selected primarily based on their brain-to-plasma partition coefficient ($K_{p,brain}$) and unbound brain-to-plasma partition coefficient ($k_{p\_uu\_brain}$) values, which are indicative of BBB permeability, and other parameters such as solubility at pH 7.4 (d_sol74) and intrinsic clearance (CL_{int}). The apparent permeability coefficient (Papp_human_caco2) for human intestinal absorption with Caco-2 and the fraction of the dose absorbed (F_a) were also calculated.

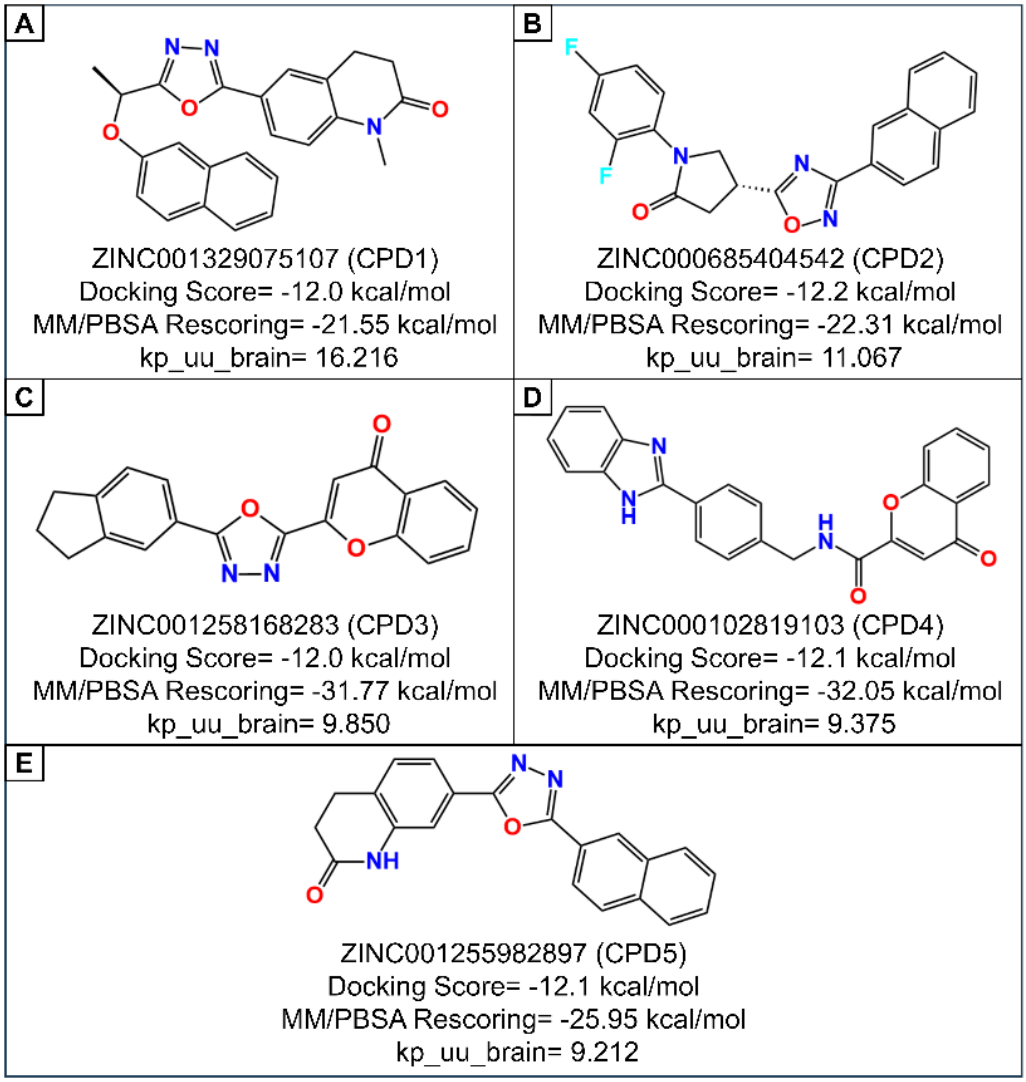

Figure 6.3. Top five novel hit compounds identified as potential ChAT inhibitors.

6.3.4. Molecular Dynamics Simulations

MD simulations were carried out to understand the mechanism of interaction for the enzyme−ligand complex formation, the structural changes, and the stability of the system with respect to time. The first docking pose was used as the starting point for the complex with ChAT to be simulated and analyzed for various parameters. α-NETA, a well-documented ChAT inhibitor, was taken as a reference standard to be compared for the MD simulation.

The protein RMSD plots of the simulated systems are shown in **Figure 6.4A**. The protein (ChAT) RMSD values range from 0.15 to 0.32 nm. All of the 5 simulated and reference systems converged within 20 ns and were quite stable with no sudden variation or increase in RMSD values, thereby depicting a stable complex formation process during the simulation time period. The RMSD plot for the ligand is shown in **Figure 6.4B**, with RMSD values ranging between 0.1 and 0.3 nm for the top-hit compounds 2, 3, 4, and 5, whereas for the top-hit compound 1, it was slightly in the upper range (0.4 nm). All of the hit compounds were relatively stable and showed a good amount of interaction during the complex formation with the ChAT. In comparison to the reference compound α-NETA, the ligand RMSD was higher side in the range of 0.15−0.52 nm, which is due to the compact size of the reference ligand.

Next, the trajectory of the protein was analyzed by RMSF, which measures the average deviation or fluctuation of the position per amino acid residue from their average positions over the course of the simulation, providing insights into the flexible loops and rigid regions of the protein. The protein RMSF plots for the ChAT protein complexed with the reference and hit compounds are shown in **Figure 6.4C**. The RMSF values ranged between 0.1 and 0.5 nm, indicating that the complexed units had an ideal balance between flexibility and rigidity. It is noteworthy that the C- and N-termini of the ChAT protein have a higher RMSF from 0.1 nm and reach up to 1.1 nm, which most likely reflects its exposure to the solvent front and, therefore, becoming more structurally flexible. Overall, the ChAT protein displayed ideal behavior for complex formation with a good amount of flexibility for the ligand to

accommodate in the binding tunnel during the complex formation process and have a good geometrical conformation, retained during the simulation. Likewise, the RMSF values for the ligand were determined and are shown in **Figure 6.4D**. The ligand RMSF values for α-NETA were found to be between 0.02 and 0.21, and for the top-hit compounds 2, 3, 4, and 5, they were between 0.01 and 0.12 nm, which are lower with respect to the reference compound α-NETA. For the hit compound 1, the RMSF value was higher than the others and the reference and ranged from 0.02 to 0.22 nm, indicative of higher flexibility and constant adjustment in the binding tunnel of the ChAT. Overall, all of the hit compounds had a good range of flexibility, indicating constant changes in the geometry in order to find its energy minima complex conformation during the simulation.

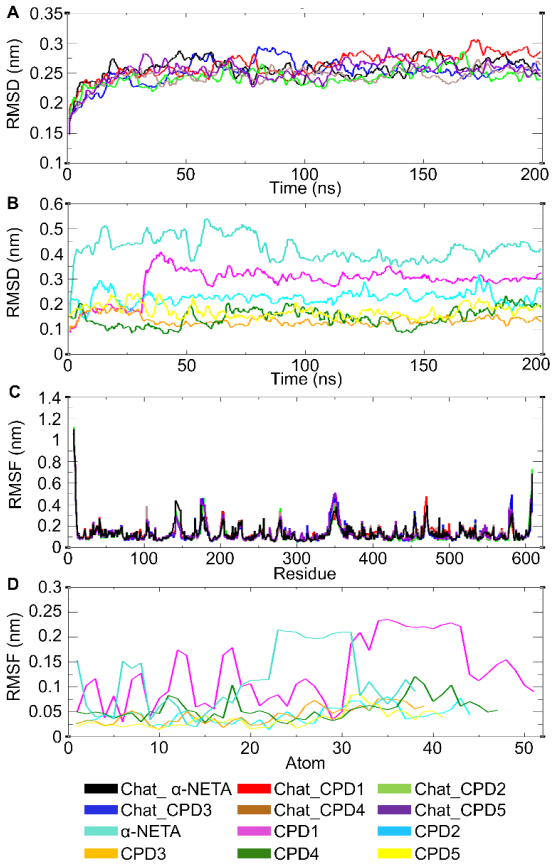

Figure 6.4. Molecular dynamics simulation analyses for the 200 ns simulation trajectory

A) represents the RMSD plots for the protein ChAT (PDB ID: 2FY3) during the complex formation with α-NETA (black) CPD1 (red), CPD2 (green), CPD3 (blue), CPD4 (brown) and

CPD5 (Violet); **B)** shows the RMSD plot for the reference α-NETA (turquoise) and the top hits CPD1 (magenta), CPD2 (cyan), CPD3 (orange), CPD4 (deep green) and CPD5 (yellow) during the complex formation with ChAT; **C)** shows root mean squared fluctuation (RMSF) of the protein (ChAT) throughout the simulation during the complex formation with α-NETA (black), CPD1 (red), CPD2 (green), CPD3 (blue), CPD4 (brown) and CPD5 (Violet); **D)** represents the RMSF plot for the top hits α-NETA (turquoise), CPD1 (magenta), CPD2 (cyan), CPD3 (orange), CPD4 (deep green) and CPD5 (yellow) during its complex formation with ChAT protein.

Furthermore, the RoG, another crucial parameter in the MD analysis, was evaluated in order to monitor the simulation trajectory. This determines the geometric distribution of atoms within the protein during simulation. The RoG values of the simulated complexes are shown in **Figure 6.5A**. The values ranged between 2.51 and 2.56 nm. Taken together, the RMSD and RoG values correlated well, indicating that the simulated systems had a steady folding mechanism during the 200 ns simulation and were similar to the reference with similar dynamics throughout the simulation.

Determining the stability of the hydrophobic core of a simulated system is another important criterion. This provides insight into how nonpolar amino acids might play a role in stabilizing the protein under physiological conditions by shielding the nonpolar amino acids within the hydrophobic core. This was evaluated here as the SASA of the complex, which is depicted in **Figure 6.5B**. The SASA values were in the range of 250−270 nm^2, which together with the graph, indicated that all simulated complexes had similar SASA profiles during the simulation time.

Hydrogen bond formation is also a crucial aspect for understanding protein−ligand interactions during the complex formation and for reaching equilibrium. The average distance of H-bond formation is ≤0.35 nm (10, 11). Assessing the number of H-bonds formed within the cutoff

range during the simulation time provides crucial details about the complex formation. The HBN for the simulated complexes are shown in **Figure 6.5C**. The results indicate that α-NETA maintained 1 H-bond on average during the simulation, whereas the ChAT_1 and ChAT_2 complexes formed on average 1−2 H-bonds throughout the simulation time, and the ChAT_3 and ChAT_5 complexes formed 2−3 H-bonds during the simulation time. However, the ChAT_4 complex formed 2−5, and hence, exhibited the highest H-bonding throughout the simulation. The HBD was also calculated (**Figure 6.5D**) and was around 0.3 nm, which is within the range of the optimal H-bond formation.

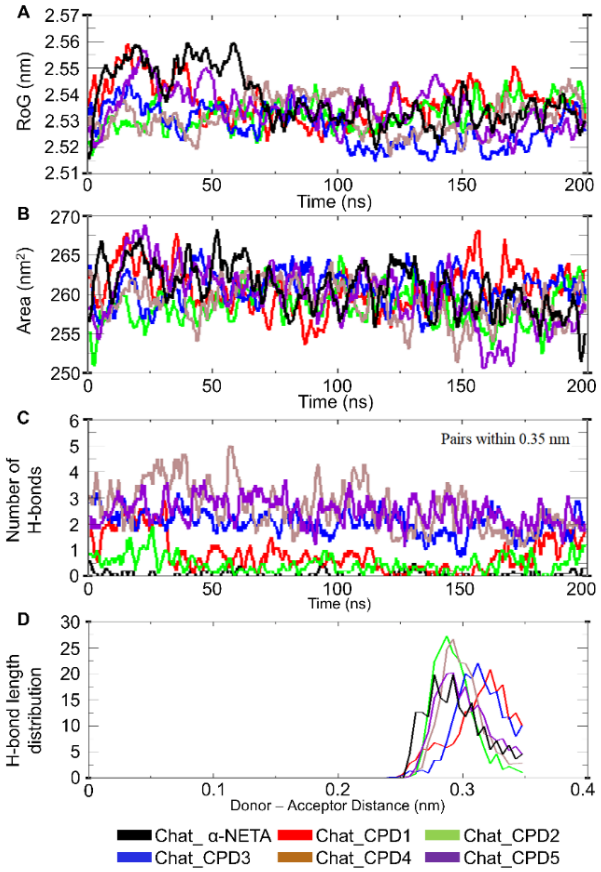

Figure 6.5. Molecular dynamics simulation analyses for the 200 ns simulation trajectory A) RoG for the ChAT_α-NETA (black), ChAT_CPD1 complex (red), ChAT_CPD2 complex (green), ChAT_CPD3 complex (blue), ChAT_CPD4 complex (brown) and ChAT_CPD5 complex (violet); **B)** denotes the changes in total solvent accessible surface area (SASA) of the ChAT_α-NETA (black), ChAT_CPD1 complex (red), ChAT_CPD2 complex (green),

ChAT_CPD3 complex (blue), ChAT_CPD4 complex (brown) and ChAT_CPD5 complex (violet); **C)** shows the average HBN formed between the ChAT_α-NETA (black), ChAT_CPD1 complex (red), ChAT_CPD2 complex (green), ChAT_CPD3 complex (blue), ChAT_CPD4 complex (brown) and ChAT_CPD5 complex (violet); **D)** Displays the average HBD maintained which was found to be close to 0.3 nm and is within the cut-off.

Furthermore, analyzing the per-residue hydrogen-bond interaction network throughout the simulation can help us identify the crucial residues that significantly contribute to the strong binding of the ligands under investigation. The hydrogen-bond donor and acceptors were identified based on geometric considerations using the VMD plugin; the cutoff criterion was set to <3.5 Å, and the cutoff angle was set to < 180°. The percentage of hydrogen-bond occupancy provides an insight into how many times the hydrogen bond was formed during the MD simulation and a higher occurrence is directly related to its stability. The top 10 most frequently occurring H-bonds between the ligands and ChAT during the MD simulation are represented in **Figure 6.6**. Here, it can be observed from **Figure 6.6A** for the ChAT_α-NETA complex that the occupancy of the key catalytic amino acid residue HIS324 was about 66.27%, acting as an H-bond acceptor, and another H-bond formation was observed where the HIS324 acts as an H-bond donor and α-NETA acts as an H-bond acceptor with an occupancy of 51.40%. Other important amino acids like TYR85, TYR436, SER438, SER540, and TYR552 contributed significantly to the H-bond formation and the stability of the ChAT_α-NETA complex formation during the simulation. Similarly, for the ChAT_CPD1 (**Figure 6.6B**), it can be seen that the catalytic amino acid residue HIS324 has an occupancy of 63.67%, where it acts as an H-bond donor and at 58.48% occupancy, it acts as an H-bond acceptor. Also, amino acid residues SER540, GLN541, SER438, and TYR552 contributed significantly to ChAT_CPD1 complex formation. For the complex ChAT_CPD2, shown in **Figure 6.6C**, it can be observed that HIS324 has occupancies of 86.63 and 79.94%, where it acts as an H-bond donor and

acceptor, respectively, interacting with CPD2. Other amino acid residues like SER540, TYR552, SER438, GLN541, and TYR85 interacted significantly through H-bond formation. Likewise, for the ChAT_CPD3 complex shown in **Figure 6.6D**. HIS324 had an occupancy of 67.17% and acted as an H-bond donor, along with other residues GLY329, GLN541, SER438, TYR552, SER540, ASN95, GLY553, and SER540, which play major roles in H-bond formation and stabilization of the CPD3 in the binding cavity of ChAT during complex formation. For the ChAT_CPD4 complex shown in **Figure 6.6E**, it can be observed that HIS324 acted as an H-bond donor with an occupancy of 51.20%, and SER440, SER438, TYR552, GLY553, SER540, and GLN541 played significant roles in H-bond formation and stabilizing the complex formation process. For the ChAT_CPD5 complex, the H-bond occupancy is shown in **Figure 6.6F**, where HIS324 has H-bond occupancies of 88.82 and 87.82% for the donor and acceptor, respectively. Overall, the five compounds maintained equal or higher H-bond occupancies throughout the simulation.

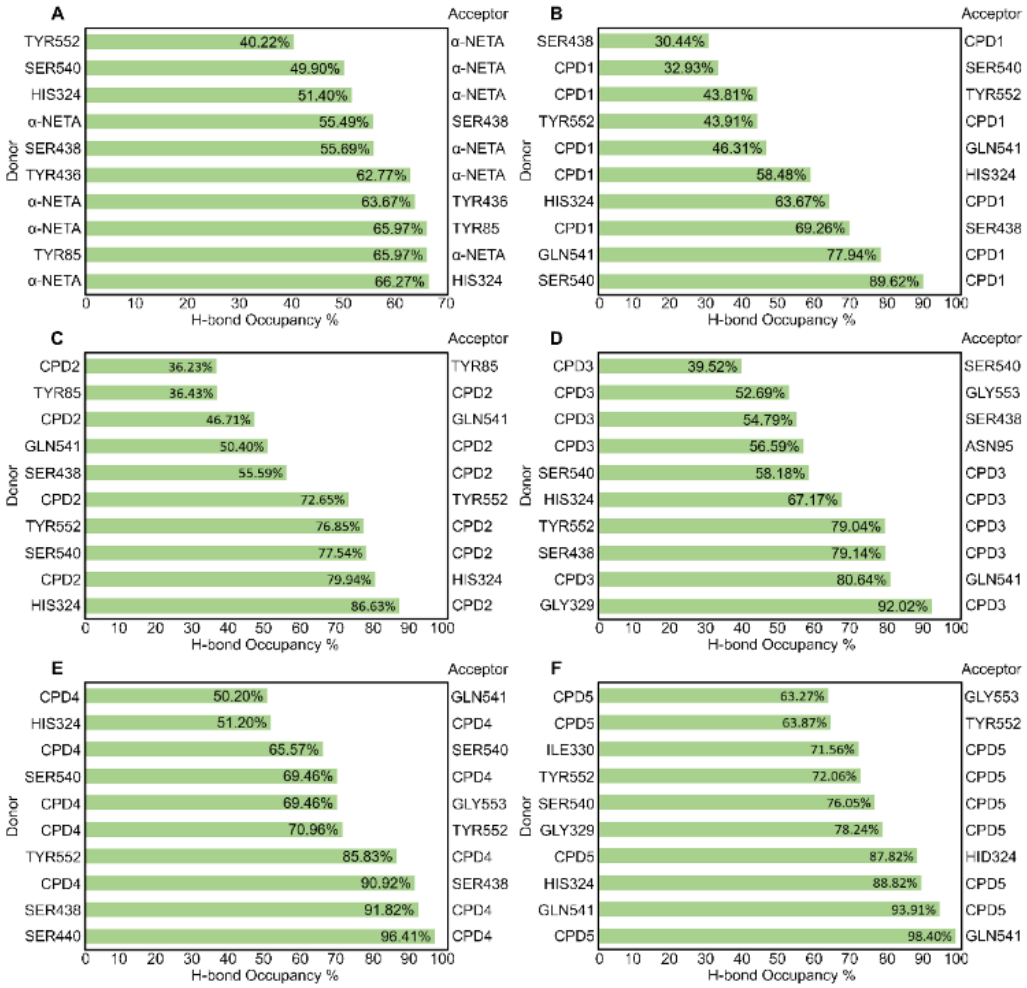

Figure 6.6. Per-residue hydrogen-bond occupancy analysis of the A) ChAT_ α-NETA complex; B) ChAT_ CPD1 complex; C) ChAT_ CPD2 complex; D) ChAT_ CPD3 complex; E) ChAT_ CPD4 complex; F) ChAT_ CPD5 complex.

To reduce the dimensionality of the data and obtain only the most essential information with respect to the overall dynamic motions of the system, we performed PCA. The eigenvalues of the first 10 principal components (PC) (eigenvectors) of the trajectories of the system were plotted for the simulated ChAT_ α-NETA and ChAT_1 to ChAT_5 complexes (**Figure 6.7A−F**). The graphical illustrations indicate that the first two principal components retained most of the details of the motions of the system.

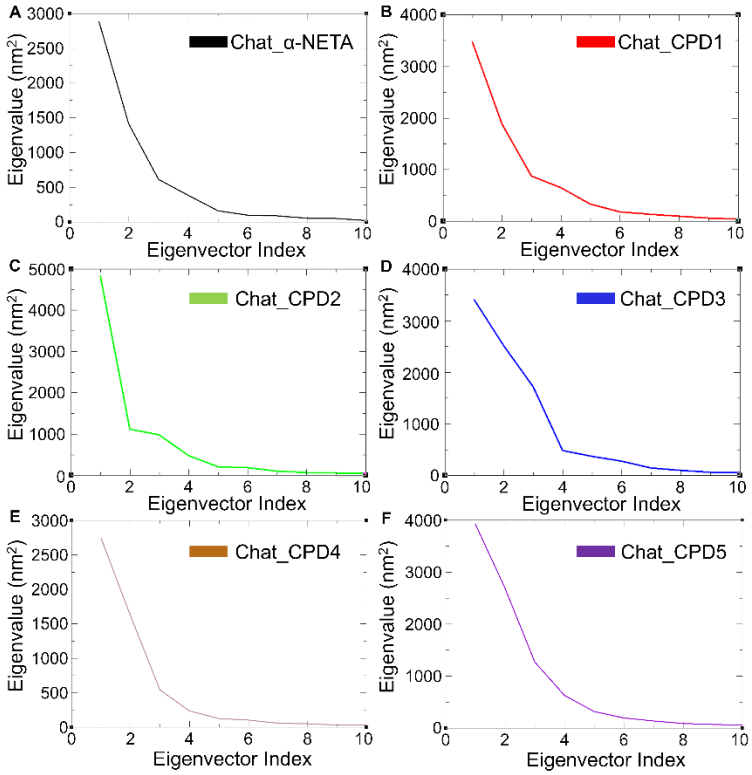

Figure 6.7. Eigenvalue calculation for the first 10 principal components, which indicates that the first two principal components eigenvector accounts for most of the motions of the system. **A)** shows the eigenvalue for the ChAT_α-NETA complex **B)** shows the eigenvalue for the ChAT_CPD1 complex **C)** shows the eigenvalue for the ChAT_CPD2 complex **D)** shows the eigenvalue for the ChAT_CPD3 complex **E)** denotes the eigenvalue for the ChAT_CPD4 complex **F)** indicates the eigenvalue for the ChAT_CPD5 complex.

Therefore, we used the first two PC and plotted the two-dimensional (2D) projection (**Figure 6.8**) to gain crucial insights into the conformational changes occurring in the protein during the simulation. The simulated complexes ChAT_α-NETA (black), ChAT_1 complex (red), ChAT_2 complex (green), ChAT_3 complex (blue), ChAT_4 complex (brown), and ChAT_5 complex (violet) seem to have a widespread distribution in the conformational space as observed from the graph, where each dot represents a particular geometrical conformation taken up by the system for the simulated complexes denoting more structural changes taking place as the dots are observed to be more widespread, thus undergoing various geometrical changes

during complex formation in order to reach its energy minima conformation, which is ideal for protein−ligand interactions and can act as probable inhibitors for ChAT proteins leading to the wide conformational changes that it undergoes during complex formation.

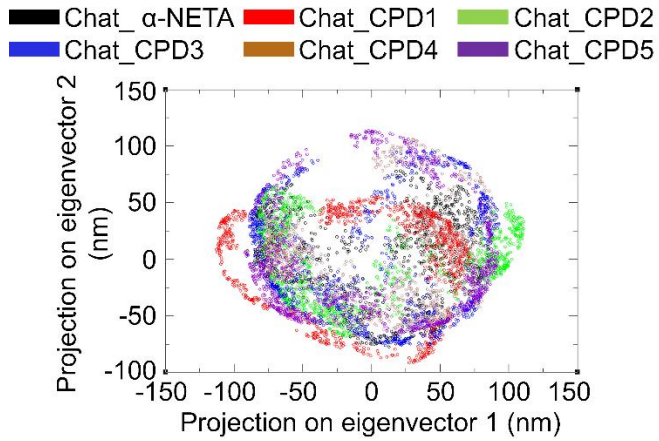

Figure 6.8. The first two PCs were used to make the 2D projections of the simulated systems ChAT_α-NETA complex (**black**), ChAT_CPD1 complex (**red**), ChAT_CPD2 complex (**green**), ChAT_CPD3 complex (**blue**), ChAT_CPD4 complex (**brown**) and ChAT_CPD5 complex (**violet**). Each dot denotes a specific conformation of the system that it underwent during the simulation time period.

Moreover, we also performed a FEL analysis of the obtained trajectories. To obtain the energy minima landscape, here, we utilized RMSD and RoG as two different reaction coordinates. The changes in the Gibbs free energy (ΔG) values of the system were calculated between 0 and 9.70 kJ/mol for all simulated complex ChAT_α-NETA and ChAT_1 to ChAT_5 (**Figure 6.9A−F**). The shape and size of the funnel in the 3D plot are indicators of the overall stability of the system. For example, cluster formation with bluish points in the center of the 2D plot indicates that the complex is stable, which is also visible in the form of a single funnel in the 3D diagrams, indicating the stability of the system during the folding process.

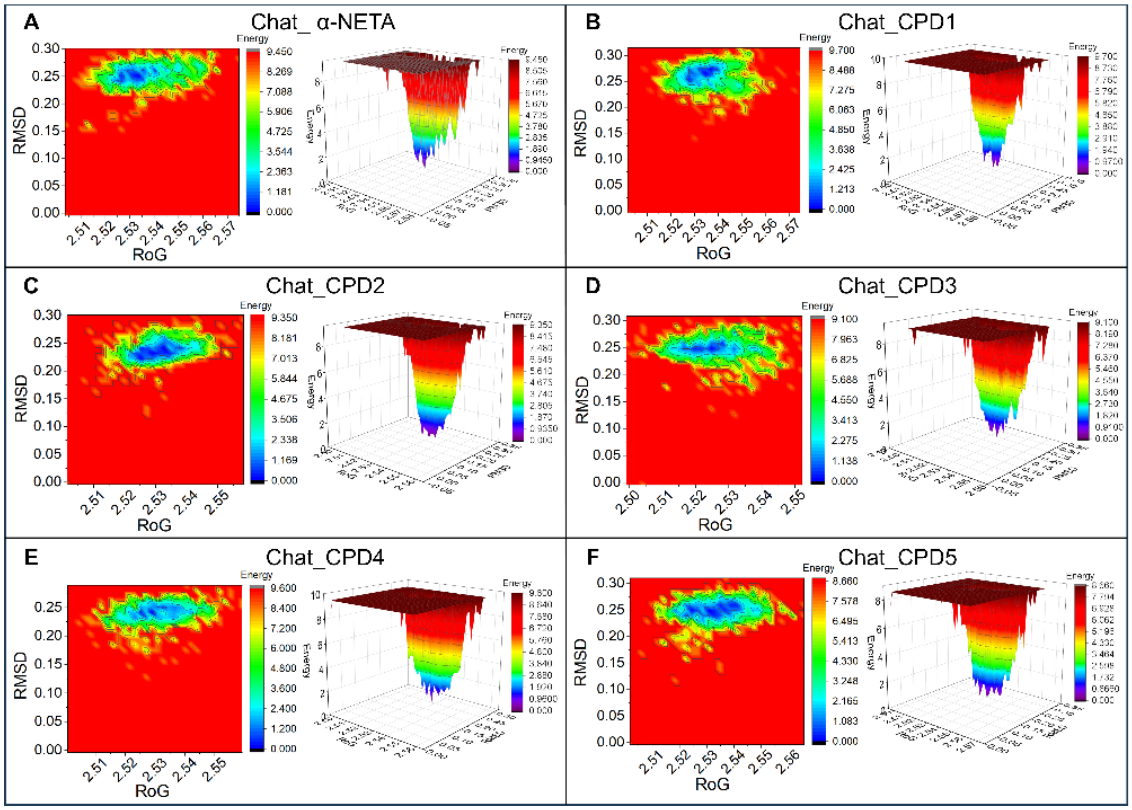

Figure 6.9. FEL plot 2D and 3D representations with RMSD and RoG taken as the two different variables that signifies the folding process of the simulated complex. The energy terms are color-coded, where the purple color denotes the lowest energy and the red color the highest energy states in kJ/mol. **A)** ChAT_α-NETA; **B)** ChAT_CPD1; **C)** ChAT_CPD2; **D)** ChAT_CPD3; **E)** ChAT_CPD4; **F)** ChAT_CPD5.

6.3.5. Molecular Mechanics Poisson−Boltzmann Surface Area Analysis

To evaluate the binding free energy of the top-hit molecules under investigation with ChAT, we used MMPBSA calculations to determine the Gibbs free energy (ΔG). The complex structures were selected based on FEL analysis, where the energy minima conformation and its two adjacent frames were taken. The obtained ΔG values suggested that the van der Waals, electrostatic, and nonpolar contributions to the solvation-free energy (ENSURF) positively contributed to the overall binding interactions, whereas the electrostatic contribution to the solvation-free energy seemed to have a negative impact (**Figure 6.10**). For the reference compound in **Figure 6.10A**, it can be seen that the EEL and EGB contributed negatively.

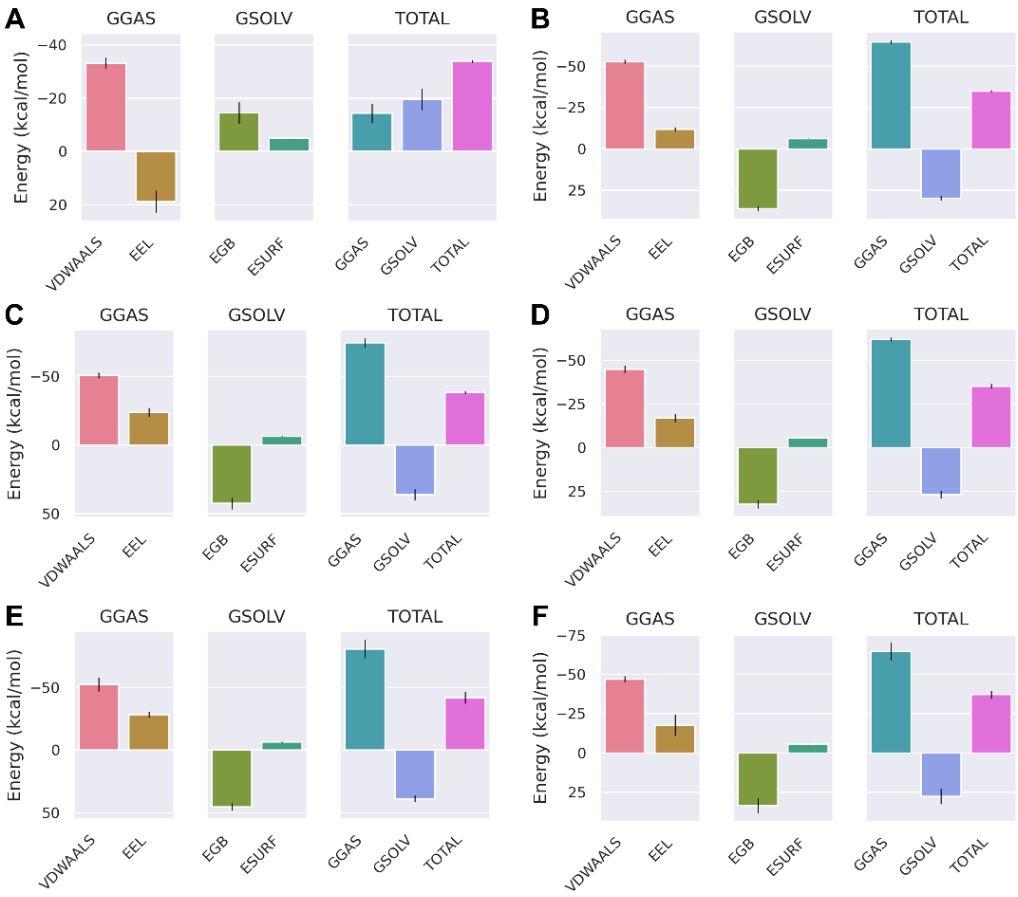

Figure 6.10. In silico Binding Free Energy (Kcal/mol) calculation (ΔG) by MM-PBSA. A) Represents the ΔG for ChAT_α-NETA complex; B) Represents the ΔG for ChAT-CPD1 complex; C) Shows the ΔG for ChAT- CPD2 complex; D) Shows the ΔG for ChAT- CPD3 complex; E) Shows the ΔG for ChAT- CPD4 complex; F) Shows the ΔG for ChAT-5. complex. **MM-PBSA**= Molecular mechanics Poisson–Boltzmann surface area; EGB= the electrostatic contribution to the solvation free energy calculated by PB or GB; ESURF= nonpolar contribution to the solvation free energy calculated by an empirical model; GGAS= Gibbs free energy into a gas-phase term; GSOLV= Gibbs free energy into a solvation term.

It is worth noting that all of the top 5 hit compounds displayed good binding (**Table 6.4**). The ΔG scores were averaged based on the values of the energy minima conformation and its two adjacent conformations. The results indicated an ΔG value of −33.84 kcal/mol for α-NETA and all of the hit compounds had higher binding free energy than the reference compound α-NETA, where CPD4 exhibited the best binding free energy with a ΔG value of −41.84 kcal/mol,

followed by CPD2 with a ΔG value of −38.31 kcal/mol. CPD5 also displayed good binding free energy with a ΔG value of −37.14 kcal/mol, while CPD1 had the least ΔG value of −35.02 kcal/mol among the compounds under investigation (**Table 6.4**). A violin plot for the MMPBSA binding free energy was plotted (**Figure 6.11**) as it helps to visualize the distribution of the obtained numeric data in a shape by using a probability density function (PDF); a wider width of the data in the plot indicates the frequently occurring value in the data set, and narrow regions indicate less frequent occurrence values in the data set.

Table 6.4. MMPBSA Binding Free Energy (ΔG) Calculations for the top five hits with ChAT protein

Compound	**Van-der Waals**	**Electrostatic energy**	**EGB**	**ESURF**	**GGAS**	**GSOLV**	**Averaged ΔG (kcal/mol)**
α-NETA	-33.1	18.8	-14.52	-5.02	-14.3	-19.54	-33.84
CPD1	-52.8	-11.95	36.02	-6.28	-64.76	29.74	-35.02
CPD2	-50.66	-23.74	42.4	-6.32	-74.4	36.09	-38.31
CPD3	-44.84	-17.03	32.33	-5.59	-61.88	26.74	-35.14
CPD4	-52.39	-28.34	45.27	-6.38	-80.73	38.89	-41.84
CPD5	-47.05	-17.64	33.32	-5.77	-64.69	27.55	-37.14

MM-PBSA= Molecular mechanics Poisson–Boltzmann surface area; EGB= the electrostatic contribution to the solvation free energy calculated by PB or GB; ESURF= nonpolar contribution to the solvation free energy calculated by an empirical model; GGAS= Gibbs free energy into a gas-phase term; GSOLV= Gibbs free energy into a solvation term.

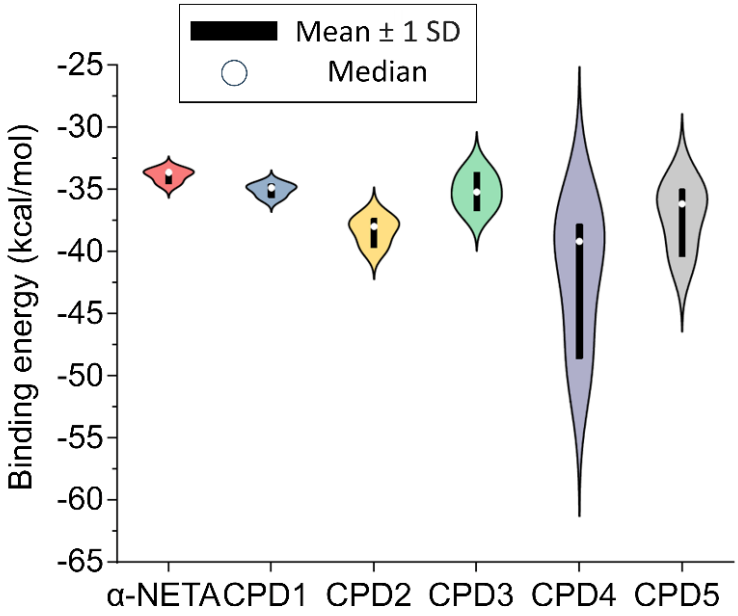

Figure 6.11. Violin plot for the MM/PBSA binding free energy (ΔG) (kcal/mol). A violin plot is referred to a hybrid plot containing the box plot and a kernel density plot. The mean is denoted as an average of the total data with a standard deviation of ±1 depicted by the thick black vertical bar at the center of the plot, and the white circle is the exact middle value (median) in the data set. Mode is denoted by the portion of the violin plot where it is the thickest and is the representation of the data that occurs the most often. The violin shape gives idea of the binding free energy range and longer violin formation like for CPD4 mean the ΔG value can vary more as compared to the other with high probability to be within the mean value. CPD1 shows small violin formation showing less variation in the ΔG value.

6.3.6. Similarity Index Analysis

Quantifying the similarity of the identified novel small-molecule inhibitors with respect to previously reported potent inhibitor compounds is a crucial task in chemoinformatic studies. This is particularly of utmost importance when screening an ultra-large chemical library for the discovery of novel inhibitors. Here, we have compared the chemical diversity of the identified top five hits with the previously reported potent ChAT inhibitors from the literature, as shown

in **Figure 6.12A**. The graphic results indicate that there was very low similarity among the compounds (**Figure 6.12B**), suggesting that in silico virtual screening of an ultra-large library can result in the identification of novel chemical classes of compounds that can potentially bind and inhibit ChAT.

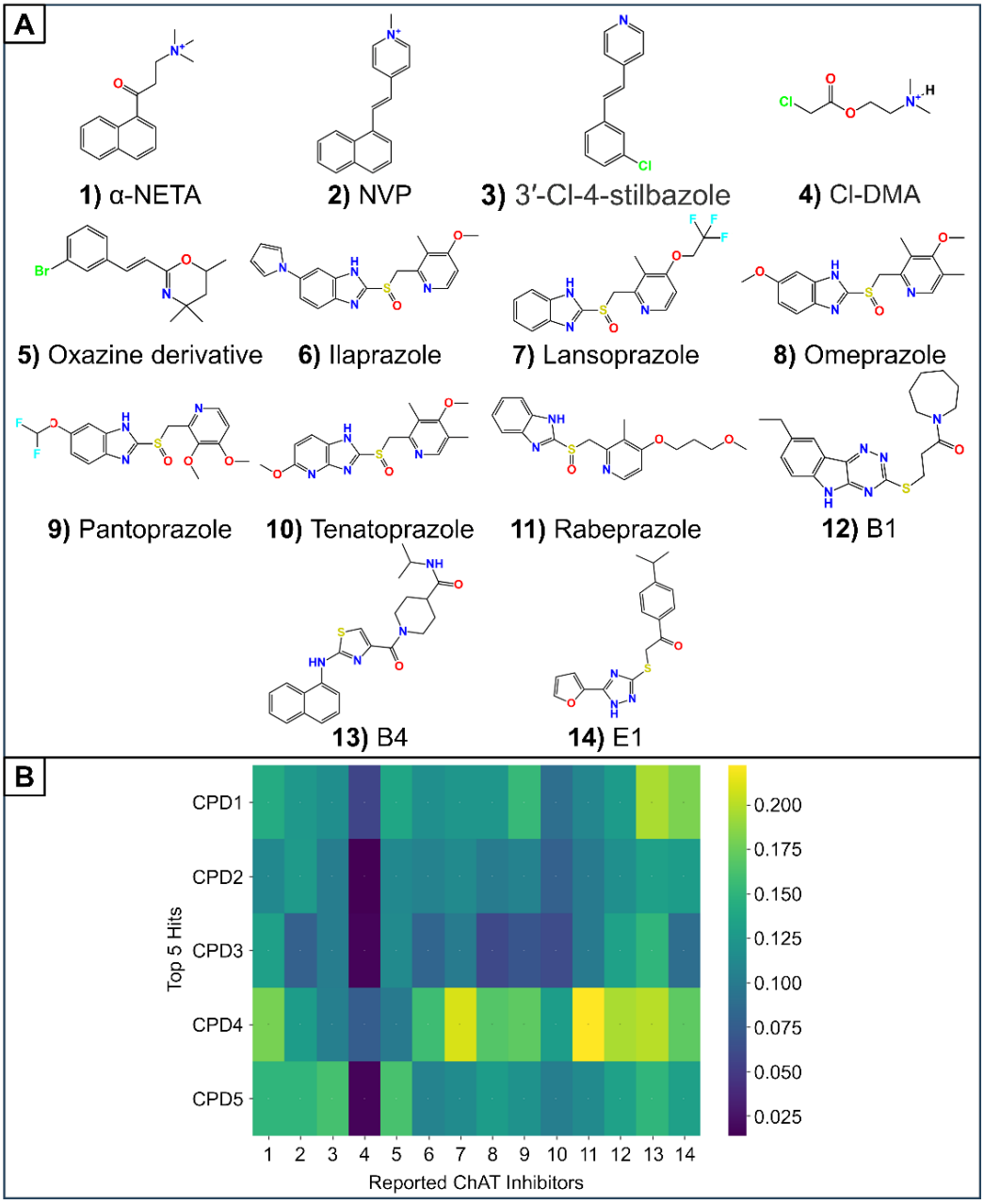

Figure 6.12. A) Previously reported ChAT inhibitors; B) Tanimoto Similarity Heatmap

6.4. Materials and Methods

6.4.1. Deep Docking Protocol

6.4.1.1. Preparation of the Ligand Database

The ZINC20 database is one of the most used ultra-large chemical libraries of 1.3 billion compounds containing aggregates of commercially accessible and annotated compounds. (12) The prepared data set, along with the fingerprints, was acquired from [https://files. docking.org/zinc20-ML/] and was further used for the deep docking accelerated virtual screening (6). The prepared data set consists of 100 prepared SMI files. Each SMI file consists of over 10 million ligand structures. The prepared Morgan fingerprints were calculated using the link. The prepared fingerprints were divided into 100 separate text files with each file containing 10 million ligand structures. Morgan fingerprints are based on circular substructures or "environments" within a molecule, defined by a radius that determines the number of bonds included in the substructure.

6.4.1.2. Molecular Docking

Molecular docking is a quick and inexpensive tool for predicting the binding affinity of the ligands to the receptor proteins. The 3D X-ray crystallographic structure of the human choline acetyltransferase (hChAT) complexed with the choline molecule was downloaded from the Protein Data Bank (PDB) (PDB ID: 2FY3, UniProt ID: P28329) (13, 14) with a resolution of 2.27 Å. The binding site comprising two binding pockets is in the shape of a tunnel with openings on both ends, one for CoA entry and the other for the entry of the choline moiety. HIS324 acts as a catalytic amino acid residue that lies between the two pockets of the binding site. Prior to defining the docking site, the ChAT protein was prepared for docking by adding any missing hydrogen, assigning bond orders, repairing side chains, treating termini, fixing protonation states and side chain amides, and removing water molecules and other ligands using AutoDock Tools. The preprocessed data set of small molecules was downloaded in the PDBQT

format. The grid box was prepared with the cocrystallized ligand as the center coordinates with a box size of 26 Å × 26 Å × 26 Å, a grid point spacing of 1 Å, and the exhaustiveness was set to 8 as the default for all three systems. The compounds were docked into the binding site, and the binding energy (kcal/mol) of the ligands was recorded.

6.4.1.3. Deep Docking Iterations

Deep docking was developed using the Keras Python library to construct and train a feed-forward deep neural network (DNN) classification model (15). It follows a progressive docking approach, where it utilizes the information obtained from the already docked subset of compounds to analyze and predict the docking scores of the remaining compounds in the data set using a partial least-squares regression-based quantitative structure−activity relationship (QSAR) model. Thereby, the computational hour required for the intensive docking of large data sets is drastically reduced. It is generally based on the DNN model, which iteratively trains with small subgroups of samples explicitly docked to deduce the ranking of the remaining compounds in the data set that are yet to be screened. Therefore, eliminating poorly performing molecules beforehand reduces the expense and time spent on docking them. Deep docking utilizes a sequence of iterative processes to enrich and enhance model training, and the utilization of simple 2D QSAR (quantitative structure−activity relationship) descriptors like Morgan fingerprints makes DD an ideal protocol for the high-throughput virtual screening of ultra-large data sets exceeding billions of compounds with limited computational resources.

For the first iteration, random sampling of the molecules was performed, and 844,489 molecules were sampled and split into a validation set (264,120 molecules), test set (280,510 molecules), and initial training set (299,859 molecules). The LigPrepper library was used to convert the sampled molecules into SDF. OpenBabel was then used to convert them into the PDBQT format for docking. The validation and test sets were generated only for the first iteration. For the

second iteration, the training set (299,661 molecules) was iteratively augmented and randomly sampled from the remaining predicted compounds, which were classified as virtual hits obtained from the first iterative run. The third iteration was our final run for the deep docking protocol, for which the training set (297,971 molecules) was iteratively augmented from the virtual hits predicted from the second iterative run.

The iterative nature of the DD process is crucial for enhancing the accuracy, robustness, and efficiency of the virtual screening. Initially, the DNN model is trained on a limited subset of the chemical library, which may not fully capture the diversity and complexity of the entire chemical space, thereby constraining its predictive power. Through iterative sampling and docking of new subsets of molecules, the DD process progressively refines the DNN model, incorporating additional data with each iteration. This ensures a continuous improvement in the model's ability to generalize across a broader chemical space. The iterative approach allows for the systematic exploration of different regions of the chemical space, exposing the model to a wider variety of molecular features and improving its predictive accuracy. The iterative process also optimizes resource usage by focusing computational efforts on a manageable subset of molecules in each iteration, making it feasible to handle an ultra-large library. Each iteration provides an opportunity to validate the model's predictions against the actual docking results, creating a continuous feedback loop that ensures that the model remains reliable and accurate.

6.4.2. MMPBSA Rescoring of the Docking Poses

In order to rescore the top-performing 215 compounds obtained from the deep docking protocol for accessing the binding affinity, we performed Molecular Mechanics Poisson−Boltzmann Surface Area (MMPBSA) on the first docking pose complex using the iPBSA tool, which minimizes the docked receptor−ligand conformations in implicit solvent to calculate the binding free energy by utilizing AmbertTools18 (16, 17).

6.4.3. In Silico Pharmacokinetic Study

The top-scoring 25 compounds obtained from MMPBSA rescoring were further evaluated for their in silico ADME (absorption distribution metabolism excretion) properties. We used the DruMap (18) web server (https://drumap.nibiohn.go. jp/) to predict the ADME properties of the top 25 hits obtained based on the MMPBSA scores of the 215 compounds. The brain-toplasma partition coefficient (Kp,brain) and unbound brain-to-plasma partition coefficient (kp_uu_brain) are the two major properties calculated using DruMap, which indicates probable BBB permeability of the compounds; typically, a value above 1 for a compound is considered to have good BBB permeability. Solubility at pH 7.4 (d_sol), intrinsic clearance (CLint), apparent permeability coefficient (Papp_human_caco2) for human intestinal absorption with Caco-2, and the fraction of a dose absorbed (Fa).

6.4.4. Molecular Dynamics Simulation

The methods for molecular dynamics simulation as detailed in the **Chapter 5** (**Section 5.4.3**).

6.4.4.1. H-Bond Occupancy Analysis

The hydrogen-bond occupancy can be defined as the number of times a particular H-bond has occurred between the amino acids of the protein and ligand during the MD simulation. This helps identify the crucial residues involved in the Hbond formation and contributes to the overall structural stability of the formed complex. The occurrence of the per-residue interaction of the amino acids in the active site of the ChAT protein with the ligands throughout the simulation time was determined by utilizing the hydrogen-bond analysis tool in the VMD 1.9.3 program (19). The criteria for H-bond donor−acceptor distance were set to <180° (20, 21).

6.4.4.2. Molecular Mechanics Poisson−Boltzmann and Surface Area (MMPBSA) Analysis

The molecular mechanics Poisson−Boltzmann and surface area (MMPBSA) are used to calculate the binding free energy to measure the internal energy of the system, which is the

primary driving force in biomolecular processes like protein folding, protein−ligand complex formation, and biochemical reactions. Therefore, accurate prediction of the free energy is an important aspect for understanding the protein−ligand binding interactions during its complex formation (22). MMPBSA is a popular methodology to estimate the binding free energy with high precision, reliability, and reproducibility (23). The enthalpy and entropy contributions were calculated for the energy minima frame and its two adjacent frames obtained from the free energy landscape diagram. The trajectories obtained from the molecular dynamic simulations were fitted, and periodic boundary conditions (PBC) were removed before proceeding with the MMPBSA calculations. In this study, we used the single trajectory method for calculating the binding free energy using the gmx_MMPBSA tool that associates the applications of GROMACS and Amber Tools, where it utilizes the output from GROMACS to construct input files for the AmberTools to perform the free energy calculation, which can be easily reproducible.

The process involved three primary steps: (i) establishment of the MMPBSA.py calculation setup given in the gmx_MMPBSA tool to perform the calculations, and the obtained trajectory from the GROMACS was utilized as input by converting it into a suitable format for Amber tools; (ii) calculations were performed for the binding free energy with various solvation models (PB, GB, and 3DRISM), and; (iii) the results obtained from the calculations were analyzed using thegmx_MMPBSA_ana command, a graphical user interface (GUI) (24).

6.4.4. Similarity Index Analysis

Determination of chemical similarity is a crucial aspect in which a large database of small-molecule libraries is to be screened by in silico structure-based virtual screening to identify novel compounds with new scaffolds having similar or better bioactivity with improved overall pharmacokinetic and pharmacodynamic parameters compared to the previously reported

compounds. Thus, once a lead compound has been identified from screening, it is an important aspect to be considered in order to verify the novelty in its structure, thereby giving a completely new compound as a potential hit that will likely have similar or better bioactivities. The Tanimoto algorithm has been widely used in computing for measuring similarity. It is based on fingerprint-based representations, where each individual molecule is encoded into a series of bits that denote the presence (1) or absence (0) of specific fragments within the chemical structure. Here, we have utilized rdkit to compute the Tanimoto similarity of the compounds and plotted the heatmap using matplotlib.

6.5. Conclusion

Artificial Intelligence has the inherent potential to revolutionize the drug discovery process by facilitating researchers for rapid screening of an ultra-large chemicals database, identifying novel scaffolds, and designing new molecules. AI is widely used for the development of predictive models. AI-deep docking utilizes a quantitative structure-activity relationship based on a deep model algorithm trained on the obtained docking scores of the randomly sampled subset of the molecule database to predict the docking outcome for the remaining molecules that are yet to be docked. This removes the low-scoring molecules in each iterative run and enriches the number of probable hits with high docking scores. The discovery of novel molecules as potential ChAT inhibitors with good pharmacokinetic and pharmacodynamic properties hold immense potential, e.g., for the development of potent diagnostic and theranostic agents in the near future for early-stage evaluation of the health of the cholinergic neuronal system in the brain in order to diagnose the onset of AD and assess whether an intervention is able to arrest the progression of AD at an early stage. The involvement and upregulation of non-neuronal ChAT have opened another avenue for utilizing potent inhibitors of ChAT. For instance, ChAT is upregulated in various cancer cells, specifically in colon and

lung cancers. Evidence indicates that cancer cells may use acetylcholine as an autocrine growth factor, as well as an antiapoptotic agent (25-28). Therefore, such novel ChAT inhibitors have immense potential for development as radiolabeled theranostic agents that can be used to track the location of cancer cells in the body using PET imaging. Alternatively, ChAT overexpressed in cancer cells can be targeted by selective ligands with radiation via a radiolabeled moiety, and simultaneous inhibition of ChAT decreases cell access to acetylcholine and thereby aids in killing the targeted cancer cells without affecting the surrounding healthy cells.

We showed that deep docking methodology could be applied in parallel with MPI-Vina for swift evaluation of docking scores of a database with 1.3 billion small molecules against the active binding cavity of the ChAT protein. The final top-scoring compounds obtained from the last iteration gave us 3049 compounds that were chemically diverse and exhibited superior docking scores. Further, the application of various physicochemical filters to obtain optimum pharmacokinetic parameters narrowed down the number to 215 top hits. Nonetheless, we further refined this number using the MMPBSA approach, which provided the top 25 most promising compounds, followed by in silico pharmacokinetic analysis using the DruMap web server. Based on the pharmacokinetic parameters, the top 5 hit compounds were selected for in-depth analysis by 200 ns classical molecular dynamics simulations, which revealed that all 5 hit molecules displayed excellent complex formation using ChAT. The folding process was gradual without any abnormality during the 200 ns simulations, and the system reached convergence. The FEL analysis revealed the energy minima conformation obtained by the complexes and MMPBSA calculations, which demonstrated excellent binding free energy for all five systems. For instance, CPD4 displayed the best binding free energy with a ΔG value of -41.84 kcal/mol, while CPD1 had a worse ΔG value of -35.02 kcal/mol, which is still quite a strong score. Furthermore, the H-bond occupancy analysis revealed that all the 5 hits have strong H-bond interaction throughout the simulation process which indicates that the

compounds may inhibit ChAT if tested in-vitro. In addition, we compared the top 5 hits with the reported ChAT inhibitors from the literature based on their Tanimoto similarity, which uses fingerprint-based representations where each molecule is divided into a series of bits and then correlates between two sets of fingerprints, with values ranging between 0 and 1, indicating no commonality between two fingerprints to identical fingerprints. Here, we showed that the similarity was less than 0.2 between the compounds. This showed that AI-deep docking structure-based high-throughput virtual screening can be implemented to identify new and novel hit compounds.

6.6. References

1. 2024 Alzheimer's disease facts and figures. Alzheimers Dement. 2024;20(5):3708-821.
2. Maia EHB, Assis LC, De Oliveira TA, Da Silva AM, Taranto AG. Structure-based virtual screening: from classical to artificial intelligence. Frontiers in chemistry. 2020;8:343.
3. Han R, Yoon H, Kim G, Lee H, Lee Y. Revolutionizing medicinal chemistry: the application of artificial intelligence (AI) in early drug discovery. Pharmaceuticals. 2023;16(9):1259.
4. Kalyane D, Sanap G, Paul D, Shenoy S, Anup N, Polaka S, et al. Artificial intelligence in the pharmaceutical sector: current scene and future prospect. The future of pharmaceutical product development and research: Elsevier; 2020. p. 73-107.
5. Gentile F, Agrawal V, Hsing M, Ton A-T, Ban F, Norinder U, et al. Deep docking: a deep learning platform for augmentation of structure based drug discovery. ACS central science. 2020;6(6):939-49.
6. Gentile F, Yaacoub JC, Gleave J, Fernandez M, Ton A-T, Ban F, et al. Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking. Nature Protocols. 2022;17(3):672-97.
7. Mehta N, Musso D, White H. Water soluble choline acetyltransferase inhibitors: SAR studies. European journal of medicinal chemistry. 1985;20(5):443-6.
8. Juba B, Le HS, editors. Precision-recall versus accuracy and the role of large data sets. Proceedings of the AAAI conference on artificial intelligence; 2019.
9. Zhong S, Zhang Y, Xiu Z. Rescoring ligand docking poses. Curr Opin Drug Discov Devel. 2010;13(3):326-34.
10. Gorelov S, Titov A, Tolicheva O, Konevega A, Shvetsov A. Determination of hydrogen bonds in Gromacs: new implementation to overcome the limitation. bioRxiv. 2023:2023.09. 01.555860.
11. Van der Spoel D, van Maaren PJ, Larsson P, Tîmneanu N. Thermodynamics of hydrogen bonding in hydrophilic and hydrophobic media. The Journal of Physical Chemistry B. 2006;110(9):4393-8.
12. Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, et al. ZINC20—a free ultralarge-scale chemical database for ligand discovery. Journal of chemical information and modeling. 2020;60(12):6065-73.

13. Kim A-R, Rylett RJ, Shilton BH. Substrate binding and catalytic mechanism of human choline acetyltransferase. Biochemistry. 2006;45(49):14621-31.
14. Burley SK, Berman HM, Kleywegt GJ, Markley JL, Nakamura H, Velankar S. Protein Data Bank (PDB): the single global macromolecular structure archive. Protein crystallography: methods and protocols. 2017:627-41.
15. Svozil D, Kvasnicka V, Pospichal J. Introduction to multi-layer feed-forward neural networks. Chemometrics and intelligent laboratory systems. 1997;39(1):43-62.
16. Sahakyan H. Improving virtual screening results with MM/GBSA and MM/PBSA rescoring. Journal of Computer-Aided Molecular Design. 2021;35(6):731-6.
17. Song LF, Lee T-S, Zhu C, York DM, Merz Jr KM. Using AMBER18 for relative free energy calculations. Journal of chemical information and modeling. 2019;59(7):3128-35.
18. Kawashima H, Watanabe R, Esaki T, Kuroda M, Nagao C, Natsume-Kitatani Y, et al. DruMAP: A novel drug metabolism and pharmacokinetics analysis platform. Journal of Medicinal Chemistry. 2023;66(14):9697-709.
19. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of molecular graphics. 1996;14(1):33-8.
20. Patil R, Das S, Stanley A, Yadav L, Sudhakar A, Varma AK. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. PloS one. 2010;5(8):e12029.
21. Jawad B, Adhikari P, Podgornik R, Ching W-Y. Key interacting residues between RBD of SARS-CoV-2 and ACE2 receptor: combination of molecular dynamics simulation and density functional calculation. Journal of chemical information and modeling. 2021;61(9):4425-41.
22. Baidya AT, Das B, Devi B, Långström B, Ågren H, Darreh-Shori T, et al. Mechanistic insight into the inhibition of choline acetyltransferase by proton pump inhibitors. ACS Chemical Neuroscience. 2023;14(4):749-65.
23. Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E. gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS. Journal of chemical theory and computation. 2021;17(10):6281-91.
24. Miller III BR, McGee Jr TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA. py: an efficient program for end-state free energy calculations. Journal of chemical theory and computation. 2012;8(9):3314-21.
25. Darreh-Shori T, Rezaeianyazdi S, Lana E, Mitra S, Gellerbring A, Karami A, et al. Increased active OMI/HTRA2 serine protease displays a positive correlation with cholinergic alterations in the Alzheimer's disease brain. Molecular neurobiology. 2019;56:4601-19.
26. Song P, Sekhon HS, Lu A, Arredondo J, Sauer D, Gravett C, et al. M3 muscarinic receptor antagonists inhibit small cell lung carcinoma growth and mitogen-activated protein kinase phosphorylation induced by acetylcholine secretion. Cancer research. 2007;67(8):3936-44.
27. Song P, Sekhon HS, Jia Y, Keller JA, Blusztajn JK, Mark GP, et al. Acetylcholine is synthesized by and acts as an autocrine growth factor for small cell lung carcinoma. Cancer research. 2003;63(1):214-21.
28. Resende RR, Adhikari A. Cholinergic receptor pathways involved in apoptosis, cell proliferation and neuronal differentiation. Cell Communication and Signaling. 2009;7:1-20.