In-silico Design of Inhibitors Targeting Polymerase IV of *E. coli* Bacteria: Virtual Screening, Docking and Molecular Dynamics

Masters Project Final Stage Report 2021 (CH 594)

Submitted by

Malaya Ranjan Patra

Roll: 195030037

Supervisor: Prof. Pradeepkumar P.I

Department of Chemistry

Indian Institute of Technology Bombay

Mumbai- 400076

Certificate

I certify that the work presented in this report titled "**In-silico Design of Inhibitors Targeting Polymerase IV of *E. coli* Bacteria: Virtual Screening, Docking & Molecular Dynamics**" has been carried out by Malaya ranjan Patra under my supervision.

Date: 12-June-2021

Prof. Pradeepkumar P.I.

Declaration

I declare here that this report is written in my own ideas and own words. Wherever other's ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have followed to all principles of academic honesty and integrity and have not misrepresented or fabricated any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

05-06-2021

Malaya Ranjan Patra

195030037

Acknowledgements

I want to thank Prof. Pradeepkumar for guiding me and giving me good directions throughout my project. It has been a great experience to work under his guidance. Also, I would like to thank my senior Ph.D. student Sruthi Sudhakar and Siddharam Shivappa Bagale, who helped me in computational work and clearing my doubts about the project. Also, I want to thank Ramanathan R. and Akhil Sudarshan for their valuable support and suggestions throughout the project. Also, I want to express my gratitude to the HPC facility of IIT Bombay for providing faster computation.

INDEX

1.INTRODUCTION

Antibiotic resistance occurs when the bacteria is able to defeat the drug which was designed to kill it. E. coli is a rod-shaped bacterium that generally resides in the lower intestine of warm-blooded organisms. Most of the *E. coli* strains are harmless, but pathogenic varieties cause serious food poisoning, septic shock, meningitis or urinary tract infections, kidney failure and fever in humans[1]. Unlike normal flora *E. coli*, the pathogenic varieties produce toxins and other virulence factors that enable them to reside in parts of the body normally not inhabited by *E. coli* and to damage host cells[1]. Antibiotics are becoming less helpful in treating *E. coli* infections because they possess a unique repair mechanism of their damaged DNA[2]. Drug resistance in *E. coli* bacteria is a serious concern[3]. Therefore, we need to find ways to tackle the issue and discover new antibiotics to inhibit the activity of the same. So in this project we have used in-silico methods like virtual screening, ADME filtration, molecular docking and molecular dynamics to tackle the antibiotic resistance problem in E. coli bacteria.

1.1 LESS SENSITIVITY OF E. COLI BACTERIA

Nitrofurans are a broad spectrum of antibiotics used to treat *E. coli* infections. Nitrofurans react with the DNA of *E. coli* bacteria to form N2-adduct with the guanine nucleotide[2]. Nitrofurans are becoming less sensitive to treat *E. coli* infections because they possess a unique repair polymerase that has the ability to bypass the N2-adducts[4]. The structure of Nitrofurazone and N2- furfuryl (fdG) adduct is shown below in Figure 1.

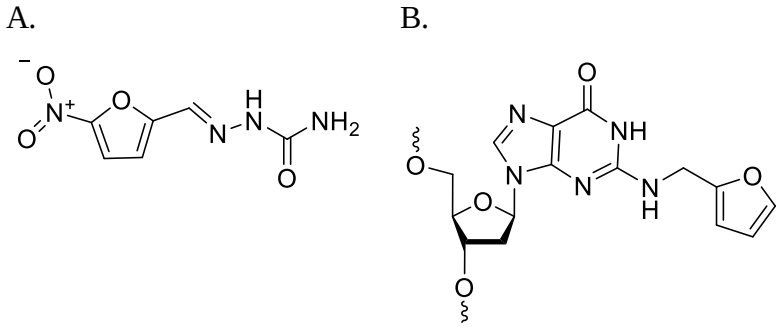


Figure 1. Structures of (A) Nitrofurazone (NFZ) and (B) N^2-furfuryl dG (fdG)

1.2 TRANSLESION SYNTHESIS (TLS)

TLS is a DNA Damaged Tolerance (DDT) pathway that helps in cell survival by replicating even the damaged DNA[4]. This pathway operates during the S-phase of the cell cycle. In this process, the cell recruits low fidelity DNA polymerases (Y-family or TLS polymerases) to bypass the damaged site in DNA in an error-prone or error-free pathway. TLS polymerases include pol η, κ, τ, Rev1, and mitochondrial primase polymerase PrimPol from eukaryotes,

DNA Pol IV (DinB) and Pol V (UmuD′2C) from prokaryotes, and Pol IV (Dpo4), Dbh (DinB homolog) from archaea. Pol ζ is also involved in TLS process, although it belongs to B family polymerases[4]. These TLS polymerases possess low fidelity and processivity due to the absence of exonuclease domain which carries out the proof-reading activity and the presence of a wider active site compared to high fidelity DNA polymerases[5]. The polymerase switch model[4] is shown in Figure 2. According to this model, when replicative DNA polymerase comes across a damaged site, then replication gets hindered or stopped, and high-fidelity DNA polymerase gets replaced with low fidelity polymerase to carry out the error-prone or error-free replication process. Polymerase switching is a key event in this process. It also causes ubiquitinylation of proliferating cell nuclear antigen (PCNA)[4]. This is a very complicated process that might lead to serious miscoding by inserting the wrong nucleotide across the damaged site to rescue the stalled replication. Finally, the high-fidelity polymerase switches back to carry out the normal replication process.

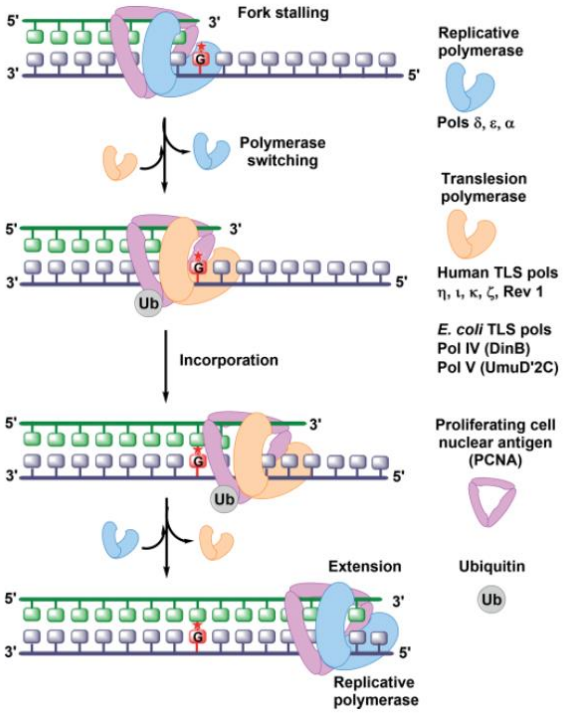


Figure 2. Picture of Polymerase switch model[4].

1.3 TLS POLYMERASE: POLYMERASE IV

Polymerase IV belongs to the Y family polymerase. Y family polymerases are capable of bypassing the adducts formed in the nucleotide efficiently and accurately. Y family polymerases play a very important role in rescuing the stalled replication. Generally, Y family polymerases are characterized by low catalytic efficiency, low processivity, and low fidelity

on normal DNA[6]. Pol IV has four different domains, namely PAD, finger, palm, thumb. PAD domain is also called the little finger domain[7]. The refined crystal structure of Pol IV[7] (PDB-4IRC) is shown in Figure 3. Pol IV is a homolog to the human polymerase κ[4]. Palm domain includes residues 1-10 and 74-165, finger domain includes residues 11-73, thumb includes residues 166-230, and PAD or little finger contains residues 241-341[7].

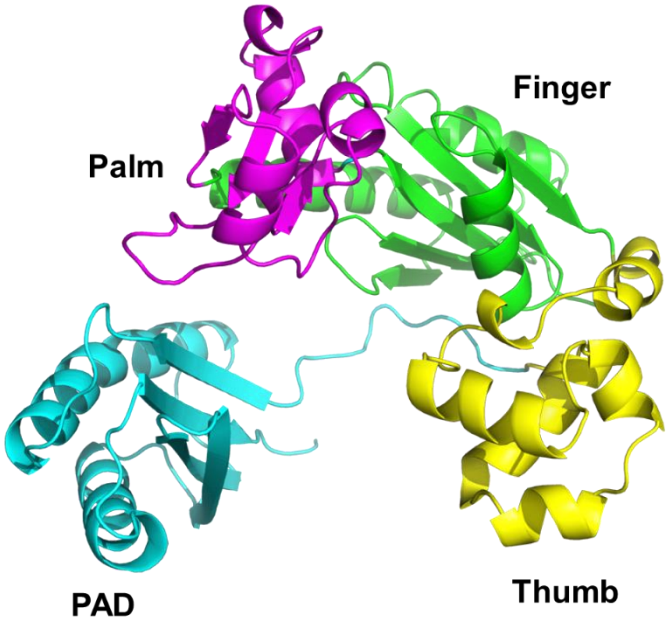


Figure 3. Refined crystal structure of pol iv generated using PyMOL[8] 4.6.

2. OBJECTIVES

The specific objective of our project is to discover inhibitors targeting Polymerase IV(Dpo4) enzyme from *E. coli*. We can achieve this objective by following structure-based drug design route. We need to look for lead compounds with high binding energy towards Pol IV having the drug-likeness properties. Binding energy results can be obtained from virtual screening, molecular docking and molecular dynamic studies. The drug-likeness can be predicted by investigating ADME (Absorption, Distribution, Metabolism, Excretion) properties of compounds.

2.1 SUMMARY OF PROJECT STAGE 1

The active site of Pol iv (PDB-4IRC) was already identified in stage-1 of this project. Active site residues include Asp8, Met9, Asp10, Cys11, Phe12, Phe13, Ile31, Gly32, Gly33, Arg38, Val40, Ser42, Thr43, Arg49, Ser55, Ala56, Leu71, Leu72, Pro73, Gly74, Phe76, Tyr79, Asp 103, Glu104, And Lys 157, Thr248, Phe295, Arg330. The reason for selecting a bigger active site region around the incoming nucleotide is because no experimental studies have been found

to assess the proper active site. Therefore, selecting a larger active site region of protein provides us more scope for understanding the active site region. In stage 1 of the project, the virtual screening of the Maybridge[9] library containing ~14000 compounds was completed. Though some compounds with good drug-like properties were found, the off-target effect was significant.

2.1 PROJECT STAGE 2

In the 2nd stage of this project, in order to minimize this off-target effect, we have used targeted libraries Asinex[10], Enamine[11], ChemDiv[12], Life chemicals[13] with Antibacterial properties . The efficiency in finding a hit compound is more in the case of the targeted library as compared to the normal library. Pol IV of bacteria is the target, so we chose the Antibacterial library. These libraries are the unique compound library for antibacterial research based on the proprietary natural product-like scaffolds that provide great skeletal diversity combined with the presence of polar functional groups and multiple stereogenic centers.

3. RESULTS AND DISCUSSION

3.1 HIGH THROUGHPUT VIRTUAL SCREENING

We got 25 hit compounds that have a considerable affinity towards the Pol IV enzyme. We found that the antibacterial libraries of Asinex[10], Enamine[11] were giving hit compounds; on the other hand, ChemDiv[12] and Life Chemicals[13] library were not giving appropriate hit compound. The top five compounds were selected on the basis of the Schrodinger's glide score[14] (Appendix 1).

3.2 ADME FILTRATION

Lipinski "Rule of Five" is the most authentic and well-known rule-based filter of drug-likeness. This rule helps to predict whether a compound is orally well absorbed or not. The rule of five includes;

- Molecular weight (MW) ≤ 500
- Octanol/water partition coefficient (iLOGP = A log P) ≤ 5
- Number of hydrogen bond donors ≤ 5
- Number of hydrogen bond acceptors ≤ 10

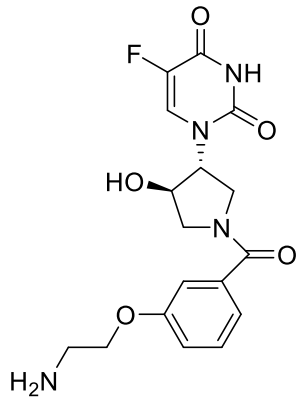
In addition to Lipinski "Rule of Five," several other properties must be obeyed by the ligands; some important ADME properties and their acceptable ranges are given below in Table 2.

Property name	**Acceptable range**
Molecular weight (MW)	50-500
Octanol/water partition coefficient (iLOGP)	–2-10
Topological Polar Surface Area(TPSA)	20-130
Number of H-Bond acceptors	0-10
Number of H-bond Donors	0-5
Rotatable bonds	0-5
Number of heavy atoms	15-50
Lipophilicity of the compound	–0.7-5.0
Molar refractivity(MR)	40-130

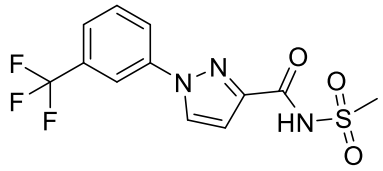
Table 2. The acceptable range of Important Pharmacokinetic properties.

Using SwissADME[15] tool to analyze the predicting the ADME properties, we found that not all the compounds were satisfying the default criteria to be a good drug candidate. Lead compounds with minimum off-targets were found. Only five compounds out of 25 compounds were satisfying the drug-likeness criteria to a maximum extent with respect to their ADME properties (Appendix 2). These Compounds were also subjected to an off-target prediction by SwissTarget web tool. We found that none of these five compounds had known affinity towards any of the human proteins. The structures of the top 5 ligands- Ligand A (Asinex ID-BDH 33911495), Ligand B (Enamine ID- Z1245134172), Ligand C (Enamine ID- Z1497243699), Ligand D (Enamine ID-Z1980497295), Ligand E (Asinex ID- BDI 34017421) are shown below in Figure 4.

A



B



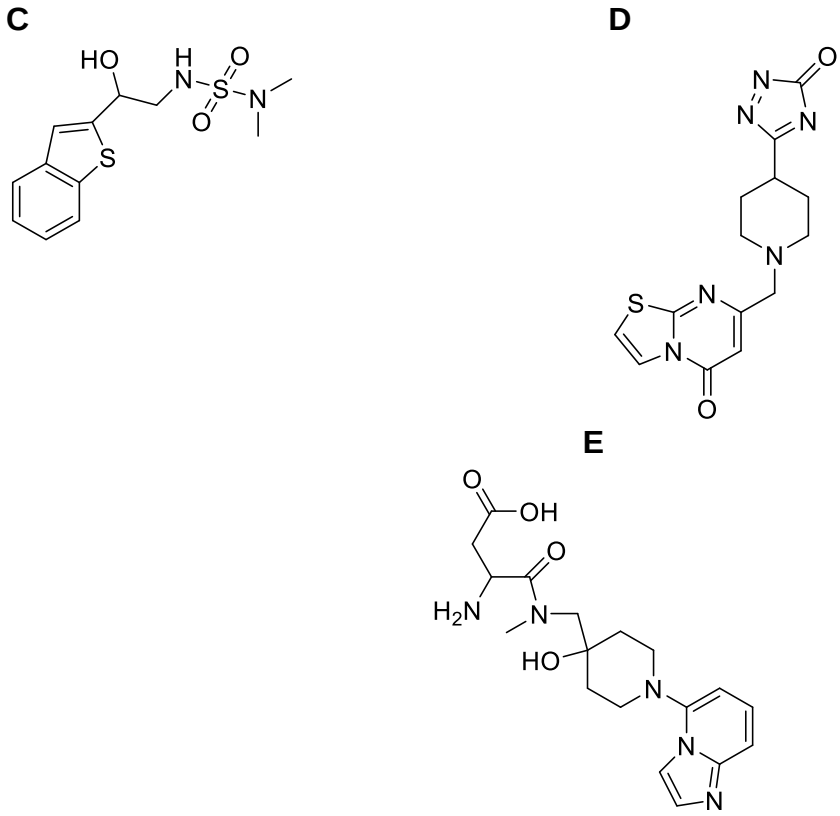
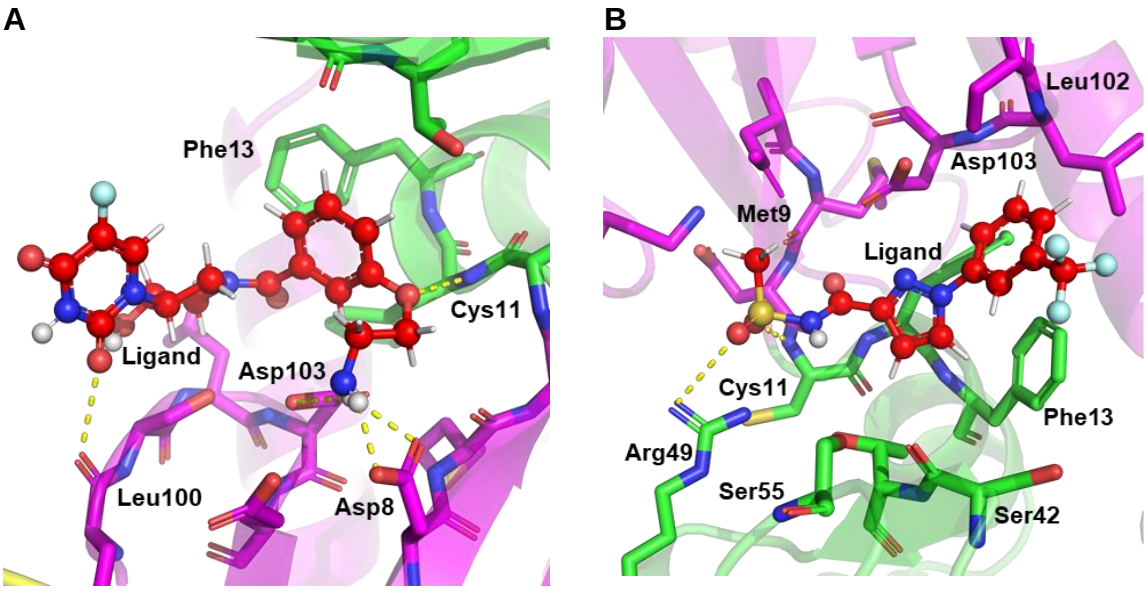


Figure 4. Top 5 most promising compounds (A) Ligand A; (B) Ligand B; (C) Ligand C; (D) Ligand D; (E) Ligand E.

3.3 MOLECULAR DOCKING

Autodock 4.2[16] was used for the molecular docking of the most promising conformer of the top 5 ligands obtained from HTVS with Pol IV (Appendix 1) and 150 conformers were generated for each of the ligands. In molecular docking we have generated more conformers from the best conformer already found from HTVS result and which made the binding energy result more precise and reliable. The site of binding of the ligand and the interaction of protein residues is shown in Figure 5. 2D interaction diagram of the ligands and protein generated using Discovery studio 2020 client[17] is shown in Appendix 3. All the five ligands include van der Waals, conventional hydrogen bond, carbon-hydrogen bond with the protein. Ligand A was docked to a site where within 5 Å, it was surrounded by residues Asp 8, Met9, Asp10, Phe12, Phe13, Thr43, Leu100, Ser101, Glu104. Detailed 2D interaction diagram with specific residues within 5 Å of the ligand can be found (Appendix 3). For Ligand A the main bonding interaction were van der Waals, conventional hydrogen bond, carbon-hydrogen bond (Figure A, Appendix 3). Ligand B was docked to a site where within 5 Å, it was surrounded by residues Asp8, Met9, Asp10, Phe12, Phe13, Thr43, Cys11, Ala14, Ser42, Asp103, Leu102, Lys157 and had π-π stacking, π-π T-shaped stacking and π-alkyl interaction with protein residues (Figure B,

Appendix 3). Ligand C was docked to a site where within 5 Å it was surrounded by residues Asp8, Met9, Asp10, Phe12, Phe13, Thr43, Cys11, Ala14, Ser42, Asp103, Leu102, Lys157, Arg49 and had unfavorable acceptor-acceptor interaction between oxygen atom of sulfonyl group and Met9 residue, π-sulfur, π-π stacking and π-alkyl interaction (Figure C, Appendix 3). Ligand D was surrounded by Asp8, Met9, Asp10, Phe12, Phe13, Thr43, Cys11, Ala14, Ser42, Asp103, Leu102, Lys157, Arg49, Ser42, Ser55, Phe76 residues and the main bonding interactions were, π-σ, π-sulfur, π-alkyl (Figure D, Appendix 3). Ligand E was docked to a site where it was surrounded by the Asp8, Met9, Asp10, Phe12, Phe13, Thr43, Cys11, Ala14, Ser42, Asp103, Leu102, Lys157, Arg49, Ser42, Ser55, Phe76, Tyr46, Leu100 with van der Waals interaction, conventional hydrogen bond, carbon-hydrogen bond, π-σ and π-alkyl interactions (Figure E, Appendix 3). The highest binding energy was observed for Ligand E, which has an Autodock score of -7.39 kcal/mol. The least binding energy among the top 5 selected ligands was observed for Ligand B, which has the Autodock[16] score of -5.83 kcal/mol.



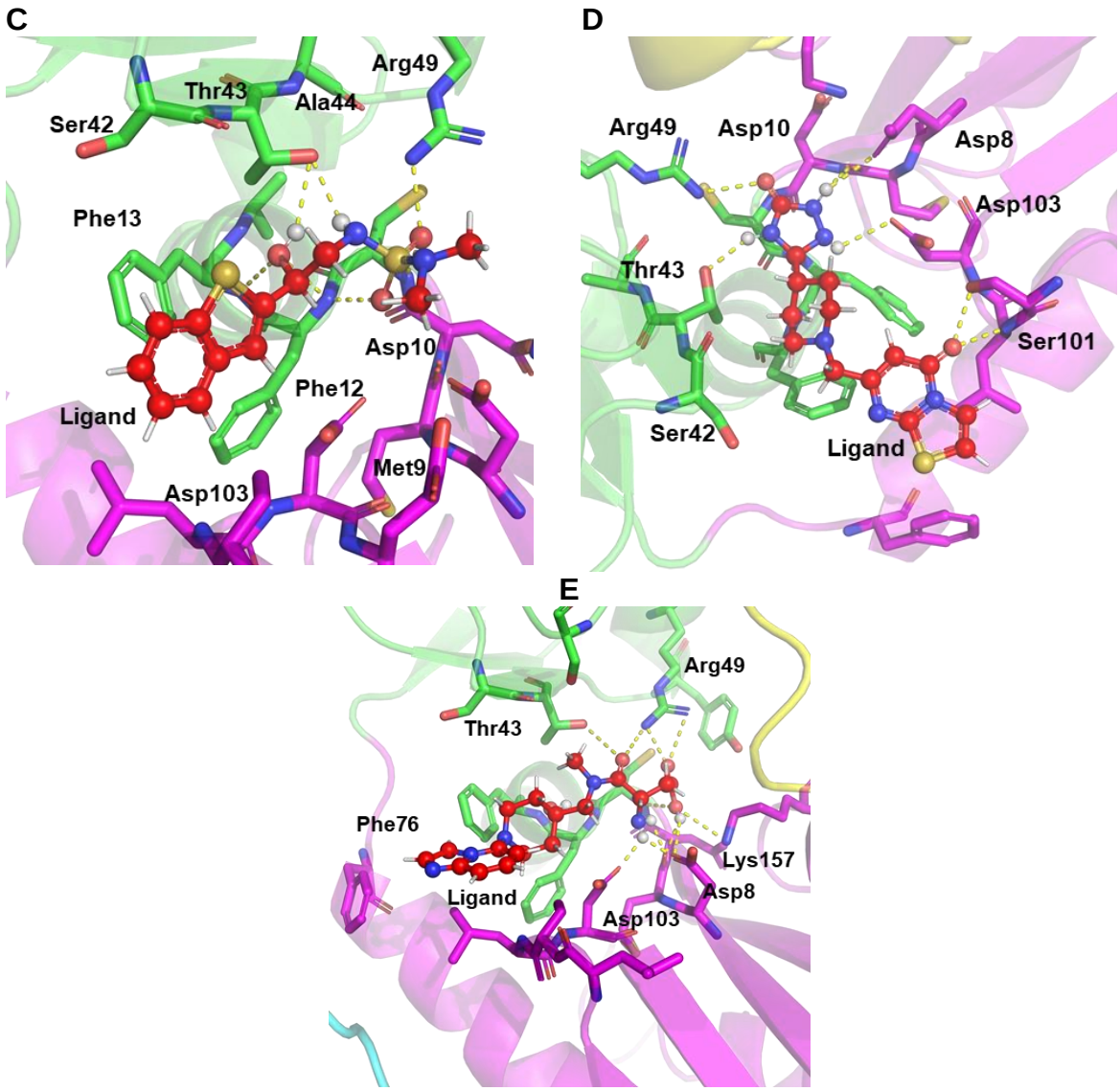
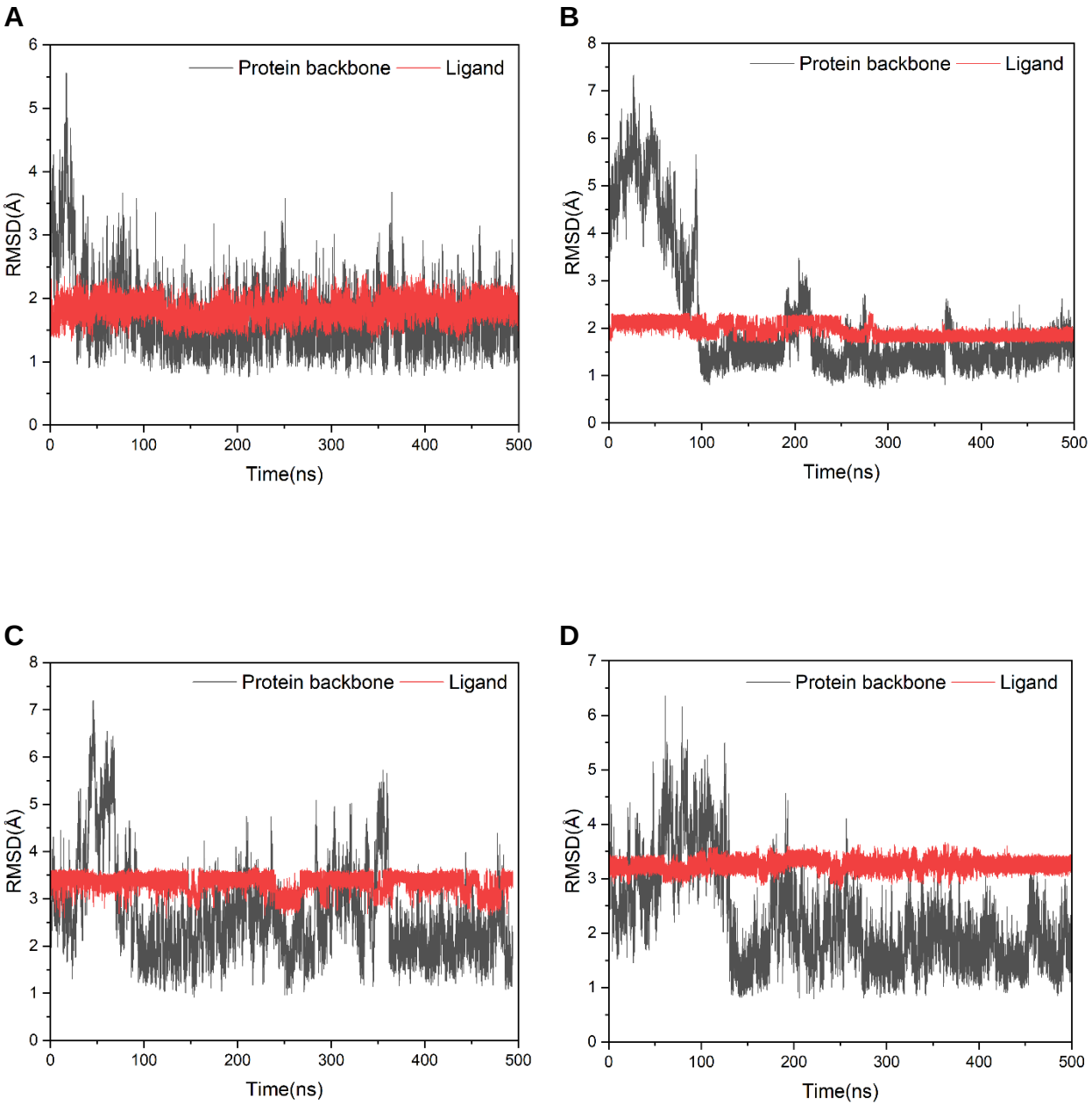


Figure 5. Position of the docked ligand's conformation with highest binding energy and interaction with active site residues of the protein (PDB-4IRC) with (A) Ligand A; (B) Ligand B; (C) Ligand C; (D) Ligand D; (E) Ligand E. Visualization images generated using PyMOL 4.6.

3.4 MD SIMULATION STUDIES

To investigate the stability of our simulated complexes, we have performed Root Mean Square Deviation (RMSD) analysis using cpptraj module of AMBER 18 program[18]. The dynamic stability of simulated complexes was monitored by studying the RMSD for C_α atoms of the protein backbones and all ligand atoms for each complex as a function of time by taking the equilibrated structure as a reference. Generally, smaller fluctuations of RMSDs indicate greater stability of the complex. The RMSD of all four ligands was quite stable (Figure 6). Though the protein initially showed a higher fluctuation in the RMSD value, it eventually became minimal, indicating the stability of the complex. The protein attained a stable state at 100ns, 200ns, 350ns, 150ns in complex A, B, C, D, respectively. In the protein- Ligand E complex, the

backbone RMSD showed a drastic increase from 5 Å to greater than 10 Å around 350ns (Figure 6). This indicates that the protein-Ligand E complex was not stable. On extending the simulation further for 500ns, we have found that the ligand detached from the protein after first 200ns simulation. From the RMSD analysis results, we have found that 4 out of 5 complexes are stable, but Complex E needs further investigation to check its stability.



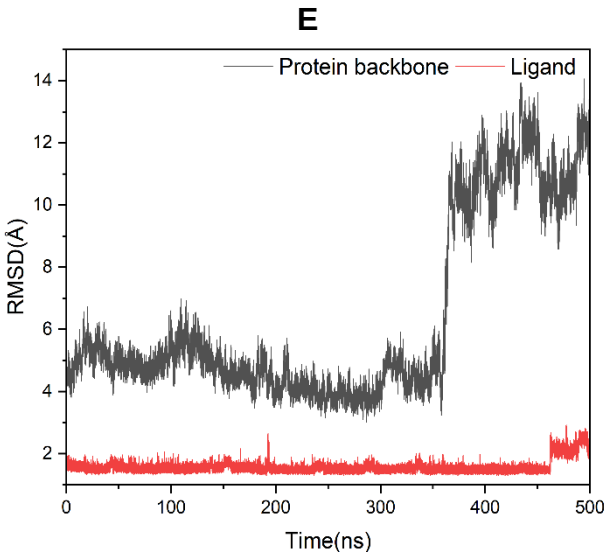
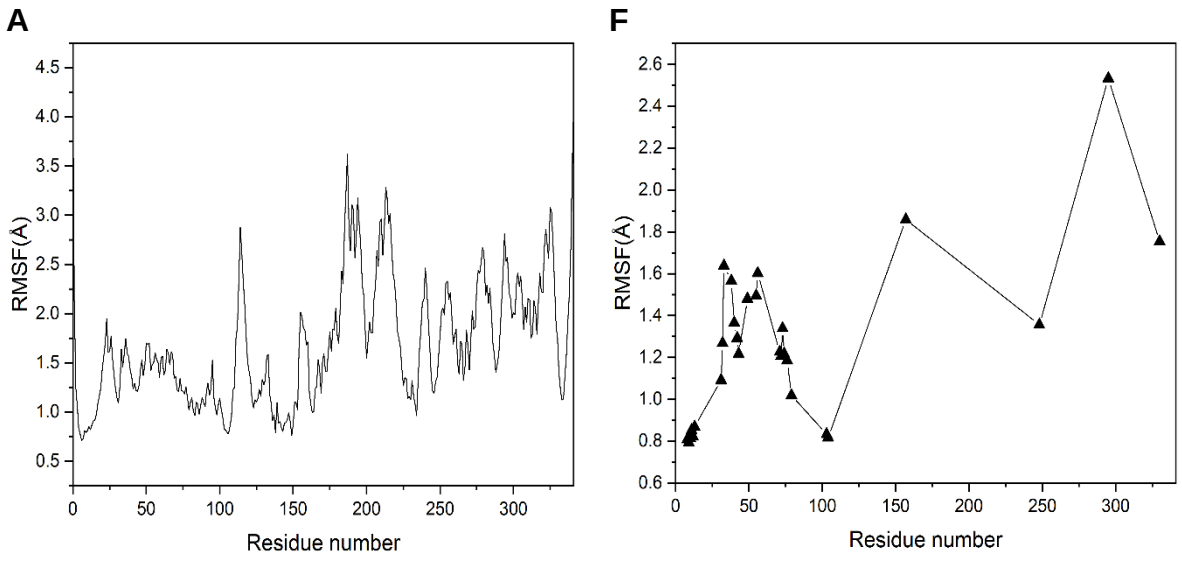
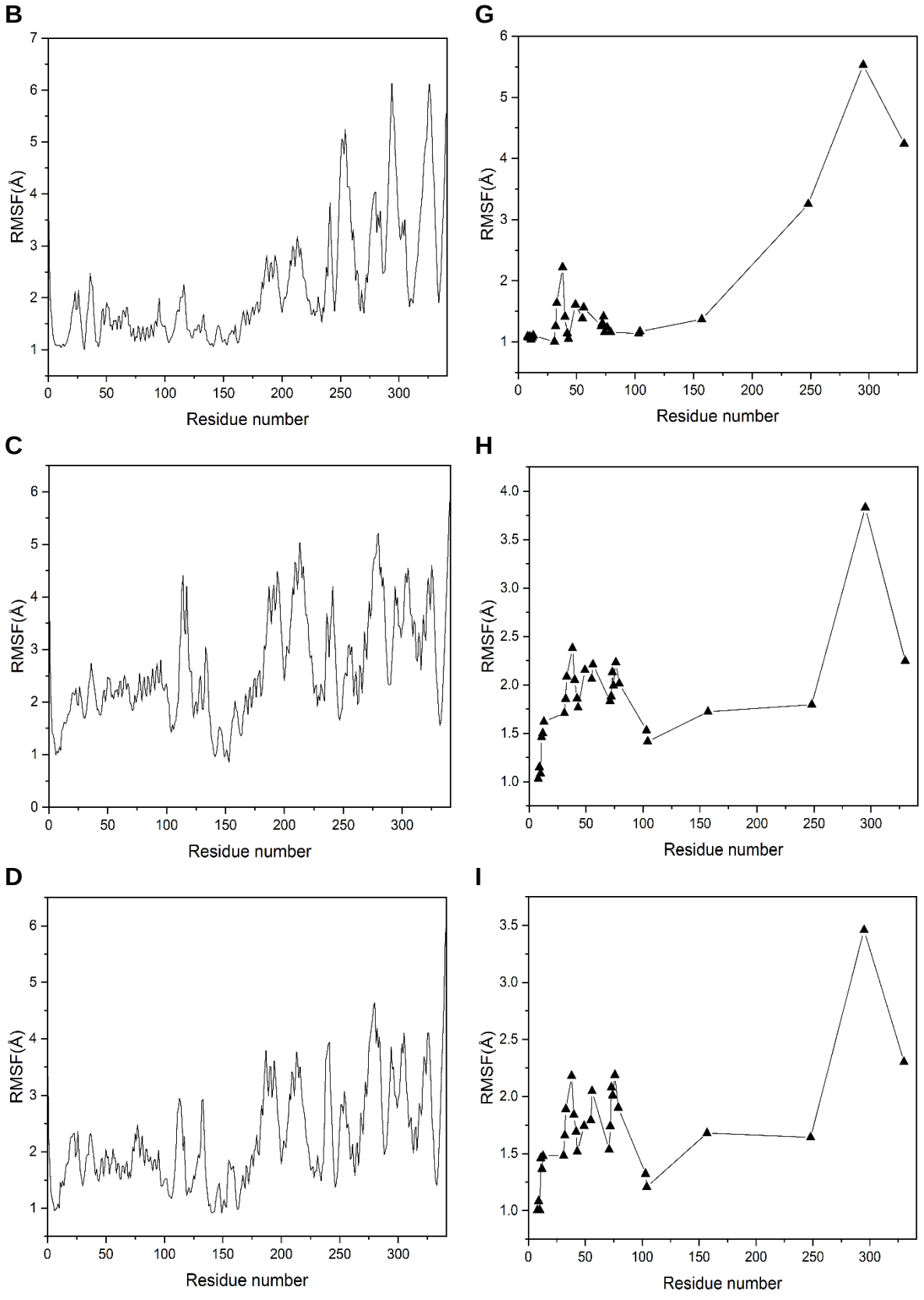


Figure 6. RMSD of the protein backbone and ligand atoms from the 500 ns MD simulation of protein with (A) Ligand A; (B) Ligand B; (C) Ligand C; (D) Ligand D; (E) Ligand E.

Root Mean Square Fluctuation (RMSF) of protein residues indicate the time-independent shift in the position of the residues after the simulation. RMSF analysis was also performed using cpptraj module of AMBER 18 program (Figure 7). More fluctuations were seen for residues in the palm, thumb, and PAD as compared to the finger domain in protein-Ligand A complex. In protein-Ligand B complex, the fluctuations were quite high in the PAD region as compared to the rest of the domains. In protein-Ligand C complex, significant fluctuations were observed for residues in the PAD and thumb domain only. In complex D, the fluctuations were more concentrated to the PAD and the thumb domain only. In complex E, more fluctuations were observed for the residues in the finger and the PAD domain. Therefore, finger and PAD or little finger domain residues fluctuations indicates the interaction with the ligand in all the complexes.




B
RMSF(Å)
Residue number
G
RMSF(Å)
Residue number
C
RMSF(Å)
Residue number
H
RMSF(Å)
Residue number
D
RMSF(Å)
Residue number
I
RMSF(Å)
Residue number

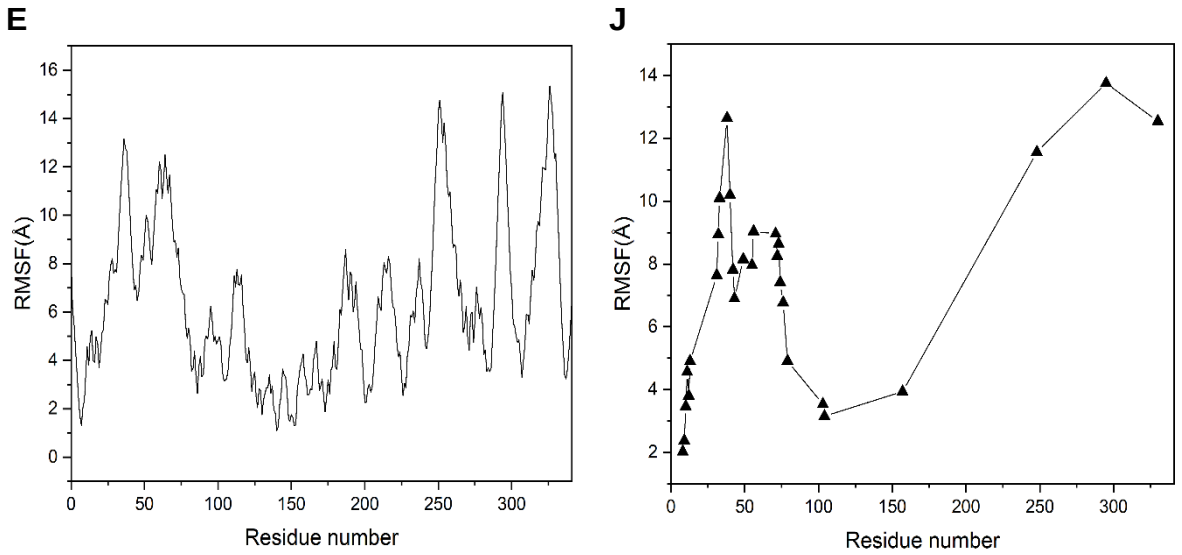
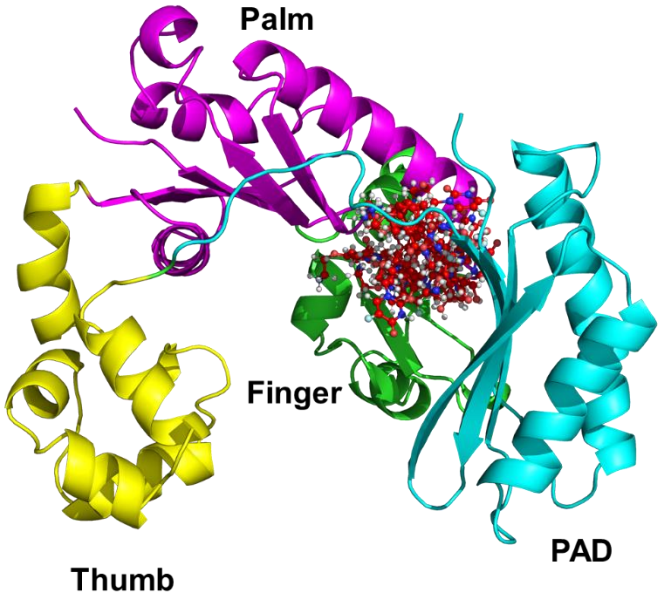


Figure 7. RMSF of all protein residues from the 500 ns MD simulation of protein with (A) Ligand A; (B) Ligand B; (C) Ligand C; (D) Ligand D; (E) Ligand E. RMSF of only active site residues of protein with (F) Ligand A; (G) Ligand B; (H) Ligand C; (I) Ligand D; (J) Ligand E.

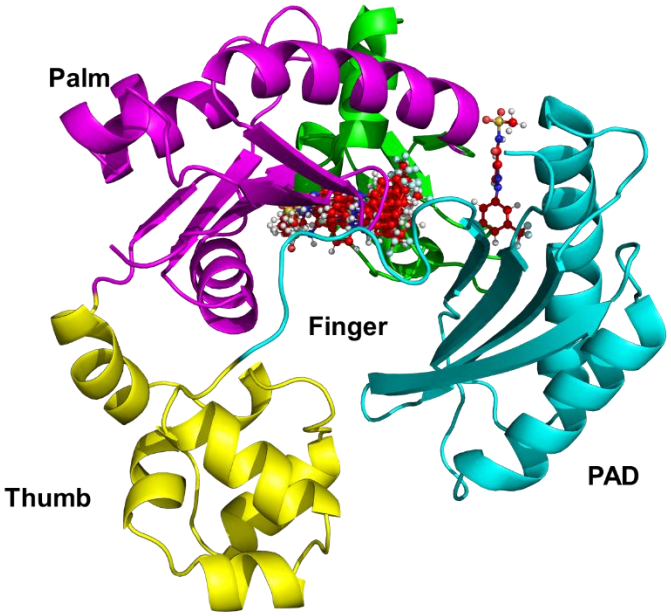
The H-bond analysis was performed for all the 4 complexes (Appendix 4). For Ligand A, the major H-bonding interactions was observed with Gly244, Glu246 residues (Table A, Appendix 4). For Ligand B, the major H-bonding interaction was observed with Thr43, Arg49, Asp77 residues (Table B, Appendix 4). For Ligand D, the major H-bonding interaction was observed with Asp103, Phe13, Asp154, Lys180, Ser55 residues (Table D, Appendix 4). For Ligand E, there was negligible H-bonding interaction with Gly244, Asn84, Arg87 (Table E, Appendix 4).

Hierarchical Clustering[19] analysis was also performed on the 500ns trajectory of the protein-Ligand systems (Appendix 5). Complex with Ligand A forms two major clusters, cluster 1 and cluster 2 with ~ 41% and ~ 38% population respectively (Figure A, Appendix 5). The complex with Ligand B forms one major cluster with ~ 80% population (Figure B, Appendix 5). The complex with Ligand C forms many small clusters. The complex with Ligand D forms one major cluster with ~ 70% population (Figure C, Appendix 5). The complex with Ligand E forms many small clusters. The position of ligand in representative structures from all clusters for each complex is shown in Figure 8.

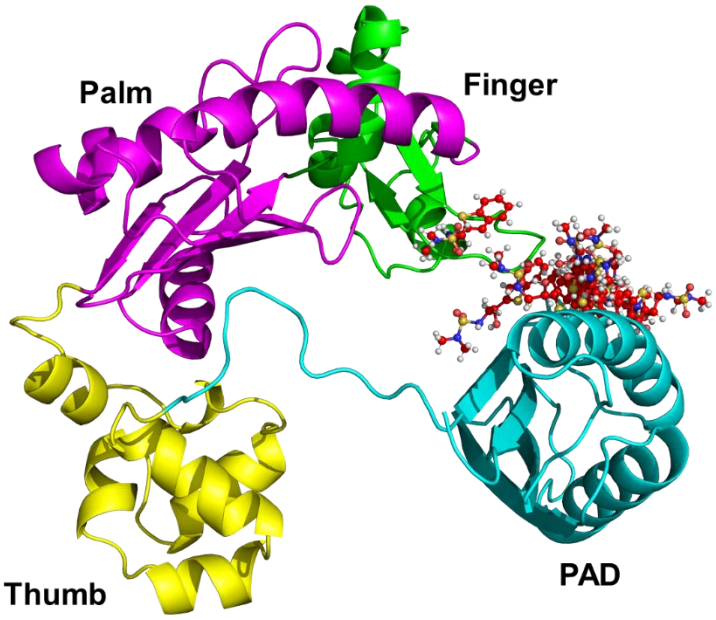
A



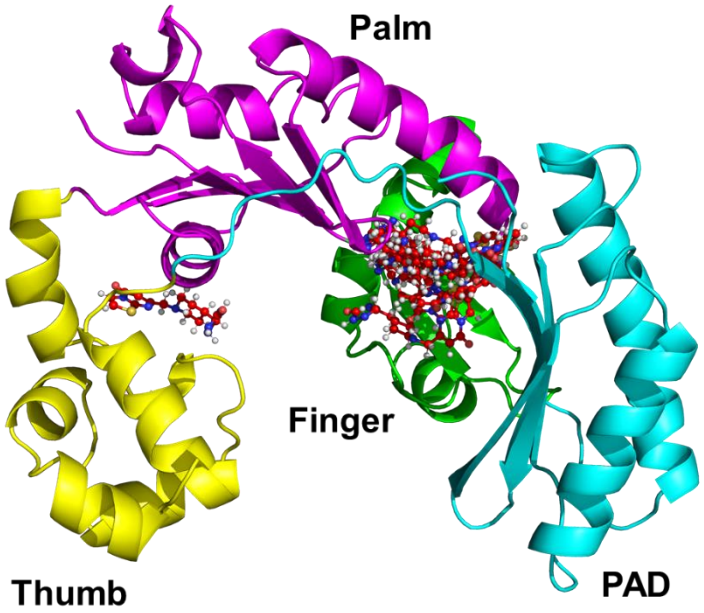
B



C



D



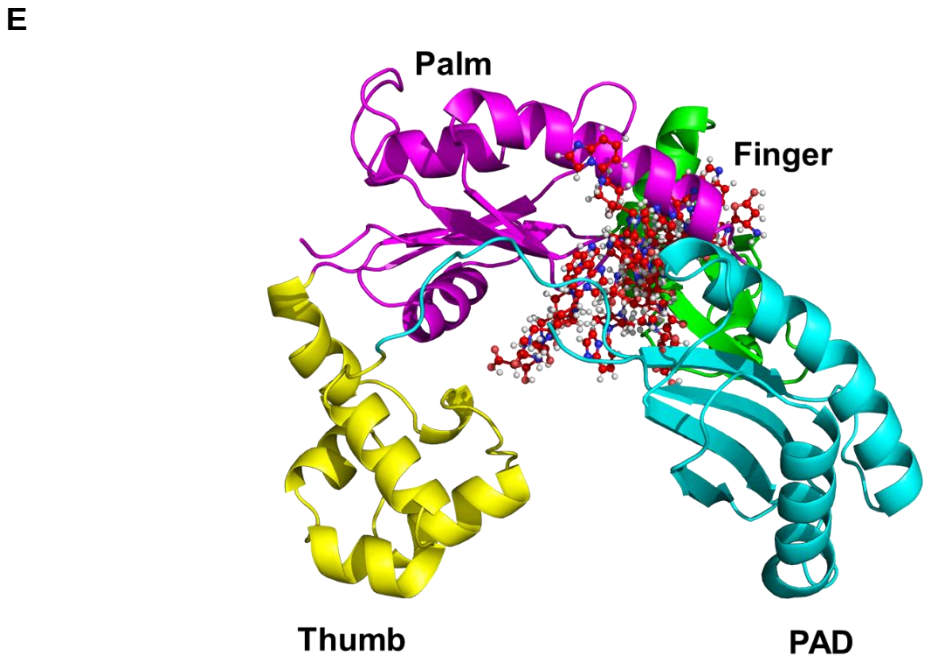


Figure 8. Hierarchical Clustering analysis from the 500ns MD simulation of protein with (A) Ligand A; (B) Ligand B; (C) Ligand C; (D) Ligand D; (E) Ligand E. Visualization images showing the position of the ligand from all representative structures generated using PyMOL 4.6.

The representative structure from the major clusters of all complexes is shown in Figure 9. Ligand A was surrounded by active site residues Phe12, Phe13, Ser42 from the finger domain, Leu102, Asp103 from palm domain and Val243, Gly244, Glu246 from PAD domain (Figure A, Appendix 6). Ligand B was interacting with residues Pro73 from finger domain, Phe76, Asp77 from palm domain, Glu246, Arg247, Leu266, Arg272, Arg273 from the PAD domain (Figure B, Appendix 6). Ligand C was mainly interacting with residues Arg247, Met249, Ile262, Arg265, Glu269, Arg272 of the PAD region (Figure C, Appendix 6). Ligand D was found to show interactions with residues Lys150, Asp154, Met155 from palm domain, Phe172, Thr175, Pro177, Lys180 from thumb domain (Figure D, Appendix 6). Ligand E was found to interact with residues Arg87, Leu100, Tyr106 residues from palm, Pro182 of thumb domain as well as the Arg232, Arg238, Arg240 residues from the PAD domain (Figure E, Appendix 6).

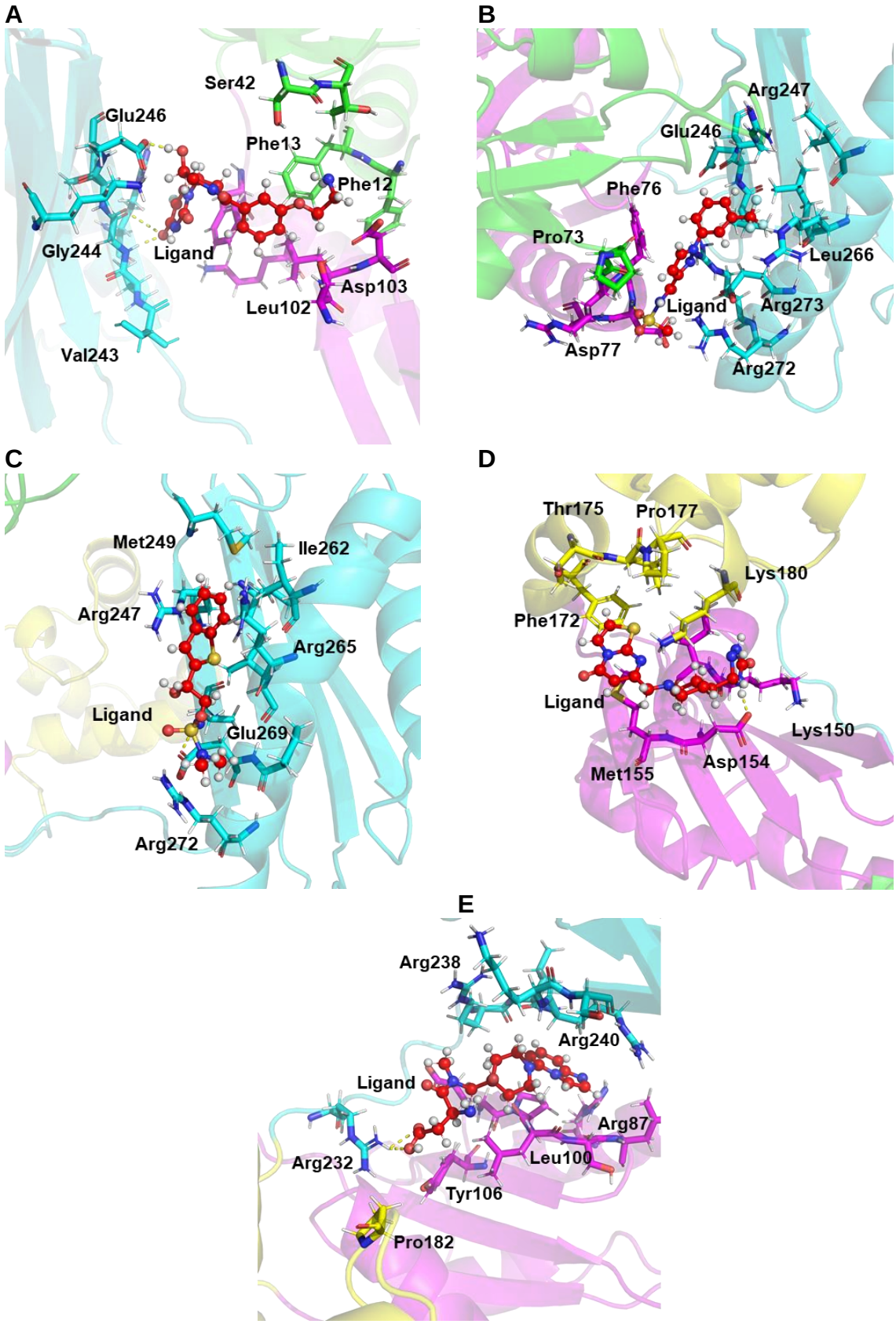


Figure 9. Representative structure from major cluster from 500ns simulation of the protein with (A) Ligand A; (B) Ligand B; (C) Ligand C; (D) Ligand D; (E) Ligand E. The position of ligand and its interaction with residues of the protein within 5 Å distance.

MMGBSA[20] (Molecular Mechanics/Generalized Born Surface Area) analysis was done by using the mmpbsa.py module of AMBER 18. The molecular-mechanical energy calculations were performed using MM/GBSA and ignoring the entropy contribution. The entropy contribution could be found by performing normal mode analysis, but in practice, entropy contributions could be neglected if only a comparison of states of similar entropy is desired, such as five ligands binding to the same protein. The reason for this is that normal mode analysis calculations are computationally expensive and tend to have a large margin of error that introduces significant uncertainty in the result[21]. The highest binding free energy among the five ligands was observed for Ligand A, i.e. ~ –21 kcal/mol. The lowest binding free energy among the five ligands was observed for ligand E, i.e. ~ –14 kcal/mol. The detailed result of this analysis is shown in Table 1.

Energy component (kcal/mol)	**Complex A**	**Complex B**	**Complex C**	**Complex D**	**Complex E**
ΔE_{VDW}	–29.88±0.07	–30.31±0.06	–22.74±0.07	–26.55±0.07	–27.31±0.07
ΔE_{ELEC}	–28.79±0.19	–25.05±0.21	–17.79±0.13	–26.29±0.19	–12.58±0.17
ΔE_{GB}	46.05±0.15	38.05±0.16	27.61±0.12	37.77±0.17	29.15±0.17
ΔE_{SURF}	–4.45	–4.41	–3.13	–3.33±0.01	–3.24
ΔG_{GAS}	–58.68±0.20	–55.36±0.20	–40.53±0.17	–52.84±0.23	–39.89±0.19
ΔG_{SOLV}	41.59±0.15	33.64±0.16	24.47±0.12	34.43±0.16	25.90±0.16
ΔG_{BIND}	–17.08±0.08	–21.72±0.07	–16.06±0.07	–18.41±0.08	–13.99±0.06

Table 1. Binding free energy components of Pol iv and ligand A, B, C, D, E has been calculated by considering every 100th frame from a total of 500000 frames. ΔE_{VDW} is the Vander Waals contribution. ΔE_{ELEC} is the electrostatic contribution. ΔE_{GB} is the energy of Generalized Born. ΔE_{SURF} is the energy of the surface. ΔG_{GAS} is the binding free energy in the gas phase complex. ΔG_{SOLV} is the binding free energy in the solvated complex. ΔG_{BIND} is the effective binding free energy, i.e., the sum of ΔG_{GAS} and ΔG_{BIND}.

3.5 SUMMARY AND FUTURE DIRECTIONS

We have followed structure-based drug design route to find potential inhibitors of poliv enzyme of *E. coli* bacteria. The whole process is represented below in Figure 10. We have discovered five lead compounds with potential inhibiting tendency against poliv. The results so far, we have obtained using In-silico methods should be further verified using experimental methods. These top 5 compounds with potential inhibiting tendency can be further assessed using inhibitory assays.

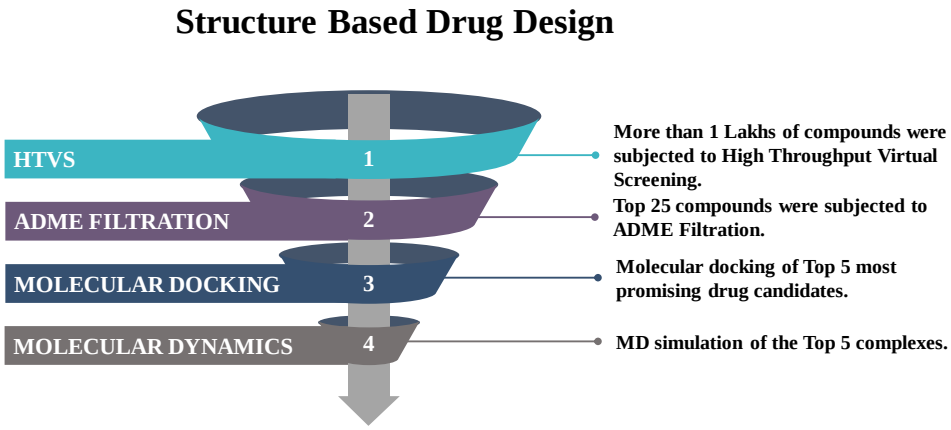


Figure 10. Workflow of whole in-silico process.

4. METHODOLOGY

4.1 HIGH THROUGHPUT VIRTUAL SCREENING

For the HTVS, we have used four different Antibacterial and one Premium library. The Antibacterial libraries are Asinex comprising 5968 compounds, Enamine containing 28831 compounds, Life chemicals comprising 11100 compounds, and ChemDiv containing 15000 compounds. Also, the Enamine premium library containing 45664 compounds has been used for the same purpose. The total number of compounds used for HTVS is 106563. Maestro and glide of Schrodinger suite have been used for the virtual screening. Maestro was the GUI used for carrying out the whole HTVS process. Glide[14] is designed for virtual screening with varying precision like HTVS(High throughput virtual screening), SP(Standard precision), XP(Extra precision). Preprocessing of the refined structure of Pol IV (PDB- 4IRC) was done using protein preparation wizard by assigning bond orders, adding hydrogens, creating zero-order bonds to metal, creating disulfide bonds, converting selenomethionines to methionines, deleting water beyond 5Å from het groups, and generating het states using Epik[22] at p^H of 7±2.0. LigPrep was used for the preparation of ligands imported in sdf format. During the preparation of ligands Ionization was kept neutral, no tautomers were generated, desalting was done and specified chiralities were retained. OPLS[23] force field was used to optimize the atoms of the ligands as well as for protein. In the ligand docking, scaling factor of 0.80 and partial charge cutoff of 0.15 were used for scaling the van der Waals radii. Post docking minimization is enabled, and five poses per ligand were included. The grid was set around the active site residues, namely Asp8, Met9, Asp10, Cys11, Phe12, Phe13, Ile31, Gly32, Gly33, Arg38, Val40, Ser42, Thr43, Arg49, Ser55, Ala56, Leu71, Leu72, Pro73, Gly74, Phe76, Tyr79, Asp 103, Glu104, and Lys157, Thr248, Phe295, Arg330. Following this, high precision docking

was done for selected ligands using Autodock 4.2[16], and during this process, 150 conformers were generated using Lamarckian Genetic Algorithm.

4.2 ASSESSMENT OF ADME PROPERTIES

After the virtual screening, further optimization of the drug candidates was done by subjecting individual compounds to ADME screening. ADME stands for Absorption, Distribution, Metabolism, and Excretion/Elimination. SwissADME[15] web tool was used to serve our purpose of ADME screening and SwissTargetPrediction[15] web tool for off-target prediction. All the essential pharmacokinetic properties (Appendix 2) and off-target effect were taken care of, and five ligands (Figure 4) were selected for further docking studies.

4.3 MOLECULAR DOCKING

AutoDockTools 4.2[16] was utilized for docking of 5 ligands obtained after ADME filtration. Before subjecting the ligands to molecular docking, all the top ligands were optimized at hf/6-31+g(d) level of theory using gaussian 09 software[24]. The crystal structure of Pol IV (PDB-4IRC) was obtained from RCSB website. The protein was prepared before docking, and Kollman charges were employed to represent the protein atoms. The grid was set to around the active site residues covering Asp8, Met9, Asp10, Cys11, Phe12, Phe13, Ile31, Gly32, Gly33, Arg38, Val40, Ser42, Thr43, Arg49, Ser 55, Ala56, Leu71, Leu72, Pro73, Gly74, Phe76, Tyr79, Asp 103, Glu104, Lys 157, Thr248, Phe295, Arg330. Genetic Algorithm (GA) was used as a search parameter, and Lamarckian GA (4.2) is used for generating the docking parameter file. 150 GA runs were carried out for all the ligands with the protein. Clustering of these docked conformations was also done to investigate the most suitable site and conformation of the ligand. The docked conformation of each ligand was extracted from the most populated cluster for further study in Molecular dynamics.

4.4 MOLECULAR DYNAMICS STUDIES

After docking is finished, individual clusters were visualized. The refined structure of Pol IV and conformer with highest binding energy of five promising ligands were selected for the MD simulation studies. The restrained electrostatic potential[25] (RESP) fitting technique was employed to estimate the partial atomic charges of all the five selected ligands. ESP charge calculation was done at HF/6-31G* level of theory. The RESP charge fitting of the ligand was then done using the antechamber module of AmberTools 20. The parameters of protein and ligand were derived from the ff14SB[26] force field and the General AMBER Force Field 2 (GAFF2) in AMBER 18. Each of the prepared complexes was solvated in a cubic box

containing TIP3P water molecules, with a minimum distance of 15Å from any edge of the box to any complex atom. 150 mm salt concentration was maintained by adding 59 Na^+ and Cl^- ions in the system. The system was initially minimized by imposing harmonic constraint of 20 kcal mol^{-1} $Å^{-2}$ on the water and counterions. In the second stage, the restriction was removed to facilitate the free movement of all atoms. During each stage, 7000 steps of the steepest descent method, followed by 7000 steps of the conjugate gradient method was performed for the minimization. Thereafter, the system was heated from 0 K to 300 K over 300ps. During the heating stage, the Langevin thermostat was used to control the temperature of the system, and also a position restraint of 20 kcal mol^{-1} $Å^{-2}$ was imposed. Then, each complex was equilibrated at 300 K over 1000ps simulation time in an isothermal isobaric (NPT) ensemble. Finally, 500ns production classical MD simulation was carried out using GPU accelerated PMEMD[27] for each complex in the NPT ensemble with a time step of 2fs. The Particle mesh Ewald (PME) method was employed to estimate the long-range electrostatic interactions, with the cutoff parameter of 10 Å for the non bonded interactions. SHAKE method was applied to constraint all covalent bonds connecting hydrogen atoms[28]. The coordinates of each complex were saved at an interval of 1ps for subsequent analysis. The RMSDs, RMSF, H-bond, and Clustering analysis of the trajectories were performed using cpptraj module of AMBER 18 program. VMD[29] was used to render the movies from the 500ns simulation of all the complexes.

REFERENCES

(1) E. Coli: Symptoms, Diagnosis, and Treatment https://www.verywellhealth.com/e-coli-symptoms-diagnosis-treatment-4174407.

(2) Kottur, J.; Sharma, A.; Gore, K. R.; Narayanan, N.; Samanta, B.; Pradeepkumar, P. I.; Nair, D. T. Unique Structural Features in DNA Polymerase IV Enable Efficient Bypass of the N2 Adduct Induced by the Nitrofurazone Antibiotic. *Structure* **2015**, *23* (1), 56–67.

(3) Olorunmola, F. O.; Kolawole, D. O.; Lamikanra, A. Antibiotic Resistance and Virulence Properties in Escherichia Coli Strains from Cases of Urinary Tract Infections. *African J. Infect. Dis.* **2013**, *7* (1), 1–7.

(4) Ghodke, P. P.; Pradeepkumar, P. I. Site-Specific N2-DG DNA Adducts: Formation, Synthesis, and TLS Polymerase-Mediated Bypass. *European J. Org. Chem.* **2020**.

(5) Yang, W.; Gao, Y. Translesion and Repair DNA Polymerases: Diverse Structure and

Mechanism. *Annu. Rev. Biochem.* **2018**, *87*, 239–261.

(6) Ghodke, P. P.; Pradeepkumar, P. I. Site-Specific N^2-DG DNA Adducts: Formation, Synthesis, and TLS Polymerase-Mediated Bypass. *European J. Org. Chem.* **2020**, *2020* (44), 6831–6844.

(7) Sharma, A.; Kottur, J.; Narayanan, N.; Nair, D. T. A Strategically Located Serine Residue Is Critical for the Mutator Activity of DNA Polymerase IV from Escherichia Coli. *Nucleic Acids Res.* **2013**, *41* (9), 5104–5114.

(8) The PyMOL Molecular Graphics System, Version 4.6.0 Schrödinger, LLC.

(9) Maybridge.com

(10) Antibacterial – Asinex.com https://www.asinex.com/antibacterial_compound_library-html.

(11) Antibacterial Library - Enamine https://enamine.net/compound-libraries/targeted-libraries/antibacterial-library.

(12) Antibacterial compounds library - 100% quality control. Worldwide delivery. https://www.chemdiv.com/antibacterial-compounds-library.

(13) Antibacterial Screening Compound Library | Targeted and Focused Screening Libraries | Screening Libraries | Life Chemicals https://lifechemicals.com/screening-libraries/targeted-and-focused-screening-libraries/antibacterial-library.

(14) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759.

(15) Daina, A.; Michielin, O.; Zoete, V. SwissADME: A Free Web Tool to Evaluate Pharmacokinetics, Drug-Likeness and Medicinal Chemistry Friendliness of Small Molecules. *Sci. Rep.* **2017**, *7*.

(16) Morris, G. M.; Ruth, H.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. Software News and Updates AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791.

(17) Systèmes, D. Biovia Discovery Studio ® 2016 Comprehensive Modeling and

Simulations. **2016**.

(18) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *Journal of Computational Chemistry*. December 2005, pp 1668–1688.

(19) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3* (6), 2312–2334.

(20) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opinion on Drug Discovery*. Informa Healthcare May 1, 2015, pp 449–461.

(21) Amber Advanced Tutorials - Tutorial 3 - MM-PBSA - Introduction http://ambermd.org/tutorials/advanced/tutorial3.

(22) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for PKa Prediction and Protonation State Generation for Drug-like Molecules. *J. Comput. Aided. Mol. Des.* **2007**, *21* (12), 681–691.

(23) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296.

(24) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., Gaussian 16 Rev. B. 01, Wallingford, CT. **2016**.

(25) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280.

(26) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713.

(27) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine

Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878–3888.

(28) Hu, H.; Chen, B.; Zheng, D.; Huang, G. Revealing the Selective Mechanisms of Inhibitors to PARP-1 and PARP-2 via Multiple Computational Methods. *PeerJ* **2020**, *2020* (5), 1–23.

(29) VMD - Visual Molecular Dynamics http://www.ks.uiuc.edu/Research/vmd.

APPENDICES

APPENDIX 1

Ligand	Autodock score	Schrodinger's glide score
A	–7.01	–6.64
B	–5.83	–5.20
C	–6.15	–5.56
D	–7.07	–6.01
E	–7.39	–6.24

Table A1. The results obtained from Autodock 4.2 and Schrodinger's glide 2020. The higher the score, the better the binding tendency of the ligand to the pol iv. Autodock and glide score units are in kcal/mol.

APPENDIX 2

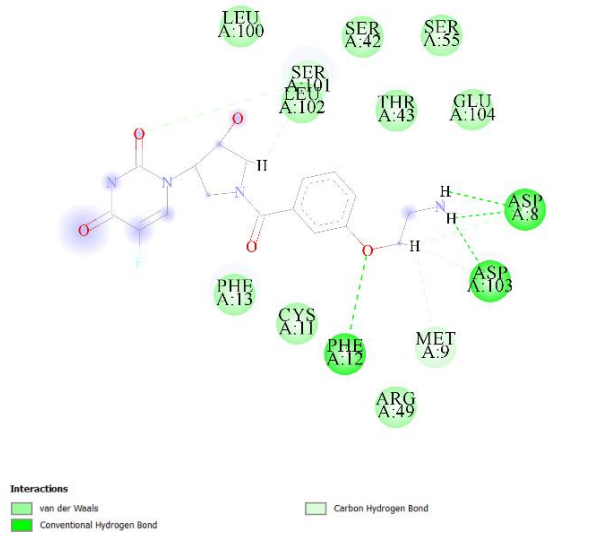
Molecule	Ligand A	Ligand B	Ligand C	Ligand D	Ligand E
Formula	C17H18FN4O5	C12H10F3N3O3S	C12H15N2O3S2	C14H17N6O2S	C18H24N5O4
MW	378.35	333.29	300.4	333.39	375.42
Heavy atoms	27	22	19	23	27
Aromatic heavy atoms	6	11	9	9	9
Fraction Csp3	0.41	0.17	0.33	0.5	0.5
Rotatable bonds	6	5	5	3	7
H-bond acceptors	7	7	5	5	6
H-bond donors	2	1	1	3	2
MR	101.11	70.53	75.5	100.26	100.73
TPSA	127.8	89.44	86.03	119.01	104.17
iLOGP	0	1.53	0	0	0

Consensus Log P	–0.72	1.88	1.22	–0.19	–0.66
ESOL Solubility (mg/ml)	2.40E+01	2.20E–01	1.09E+00	5.89E+00	1.10E+02
ESOL Class	Very soluble	Soluble	Soluble	Very soluble	Very soluble
GI absorption	Low	High	High	High	High
Lipinski violations	0	0	0	0	0
Ghose violations	1	0	0	1	0
Veber violations	0	0	0	0	0
Egan violations	0	0	0	0	0
Muegge violations	0	0	0	0	1
Bioavailability Score	0.55	0.55	0.55	0.55	0.55
PAINS alerts	0	0	0	0	0
Brenk alerts	1	0	0	0	0
Leadlikeness violations	1	0	0	0	1
Synthetic Accessibility	4.51	2.31	3.35	3.68	3.62

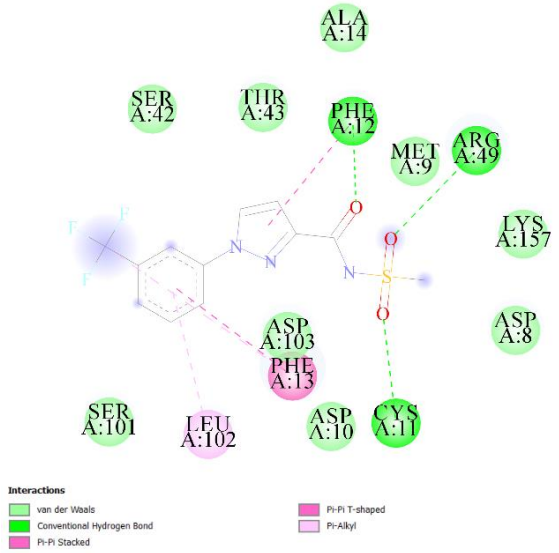
Table A2. Important pharmacokinetic properties of Ligand A, B, C, D, E obtained using SwissADME webtool.

APPENDIX 3

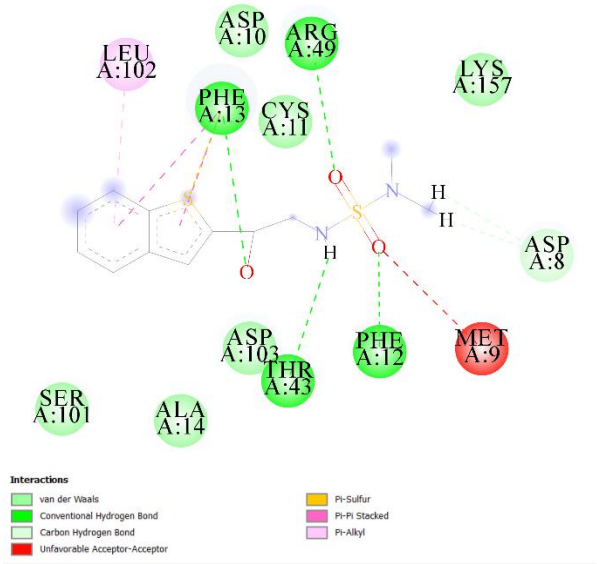
A



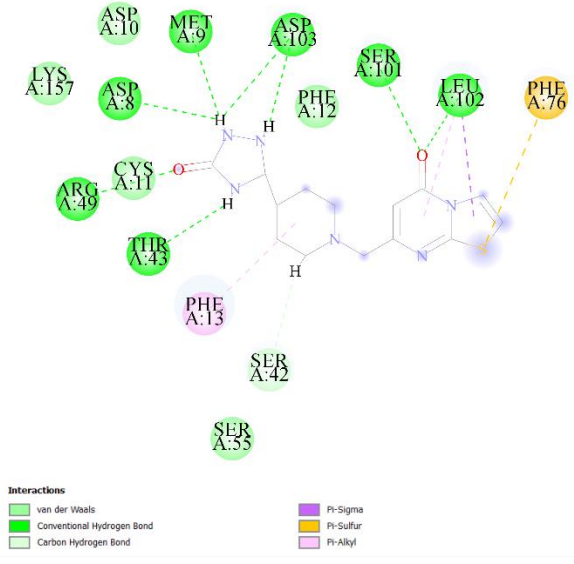
B



C



D



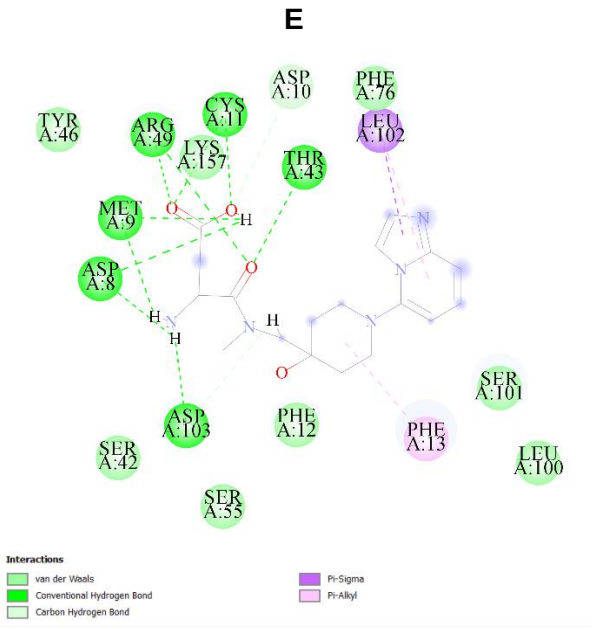


Figure A3. 2D interaction diagrams showing the interaction of the protein residues with the (A) Ligand A; (B) Ligand B; (C) Ligand C; (D) Ligand D; (E) Ligand E from the major clusters after docking in autodock 4.2 . The diagrams were generated using the Discovery studio 2020 client software.

APPENDIX 4

A

Acceptor	Donor H	Donor	Frames	Occupancy(%)	Average Distance	Average Angle
GLU246@OE2	LIG342@H4	LIG342@O5	135095	27	2.67	163.28
GLU246@OE1	LIG342@H4	LIG342@O5	125786	25	2.67	163.04
LIG342@O3	GLY244@H	GLY244@N	48822	9	2.89	157.51
LIG342@N2	ARG240@HH21	ARG240@NH2	43078	8	2.85	160.91
PRO99@O	LIG342@H2	LIG342@N2	20001	4	2.86	159.46

B

Acceptor	Donor H	Donor	Frames	Occupancy(%)	Average Distance	Average Angle

LIG342@O2	ASP77@H	ASP77@N	107652	21	2.89	162.29
LIG342@O1	THR43@HG1	THR43@OG1	49973	9	2.79	156.37
LIG342@O2	ARG49@HH22	ARG49@NH2	49753	9	2.85	151.78
LIG342@O2	ARG49@HH12	ARG49@NH1	45509	9	2.84	151.54
LIG342@O3	CYS11@H	CYS11@N	30372	6	2.89	158.40

C

Acceptor	Donor H	Donor	Frames	Occupancy(%)	Average Distance	Average Angle
LIG342@O2	ALA250@H	ALA150@N	24776	5	2.86	153.47
LIG342@O2	ARG247@HH11	ARG247@NH1	24225	4	2.85	154.97
LIG342@O3	ALA250@H	ALA150@N	20313	4	2.87	158.06

D

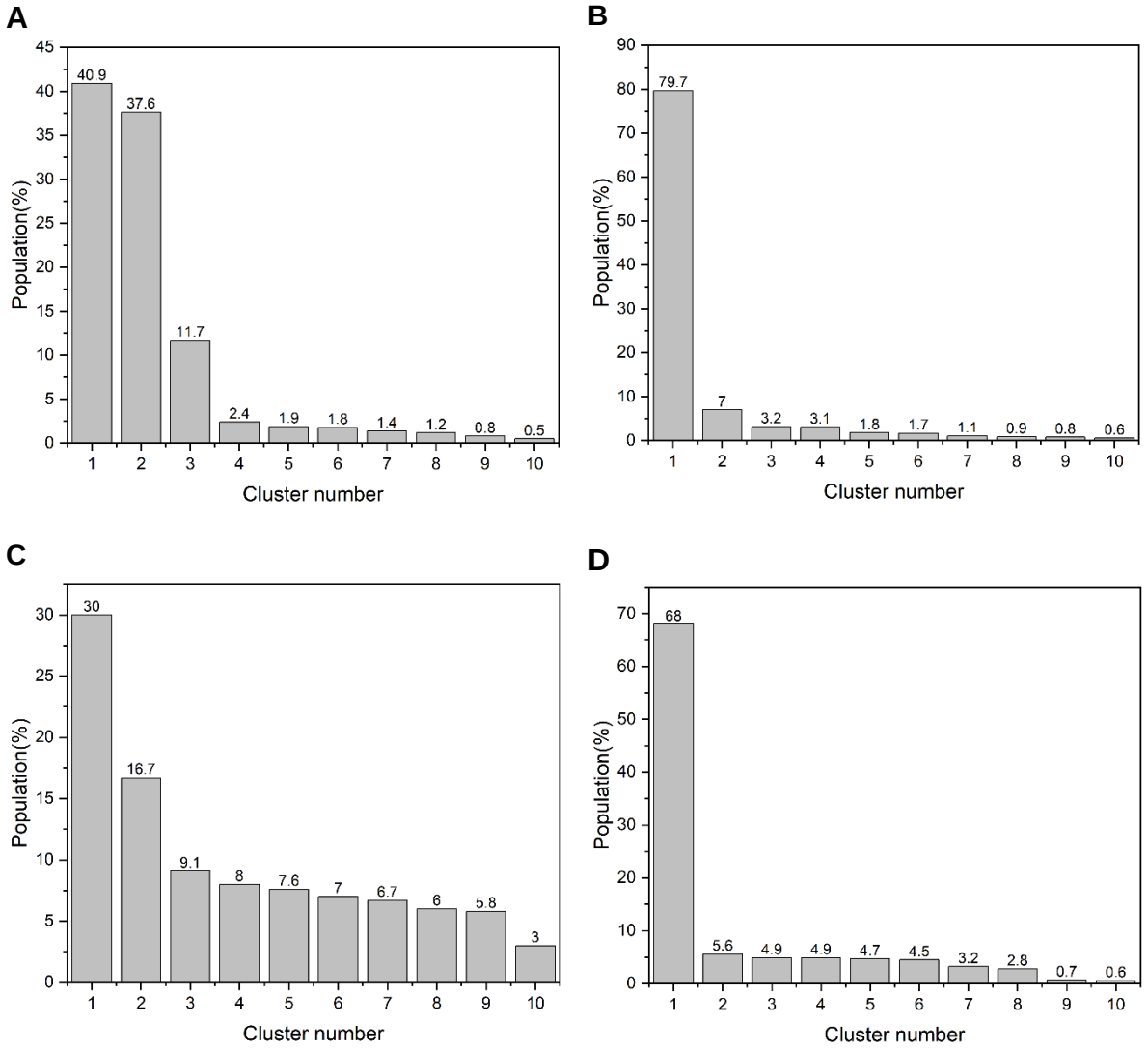
Acceptor	Donor H	Donor	Frames	Occupancy(%)	Average Distance	Average Angle
ASP103@OD1	LIG342@H3	LIG342@N6	110099	22	2.82	160.70
LIG342@O2	PHE13@H	PHE13@N	107582	21	2.87	165.30
ASP154@OD1	LIG342@H1	LIG342@N4	81202	16	2.78	162.11
LYS180@O	LIG342@H2	LIG342@N5	80931	16	2.81	157.22
ASP103@OD2	LIG342@H3	LIG342@N6	32408	6	2.83	162.30

E

Acceptor	Donor H	Donor	Frames	Occupancy(%)	Average Distance	Average Angle
LIG342@O2	GLY244@H	GLY244@N	16220	3	2.88	160.10

Table A4. A, B, C, D, E containing the result obtained from H-bond analysis from 500ns MD simulation of protein with Ligand A, B, C, D, E respectively. LIG342 is the label for the Ligand.

APPENDIX 5



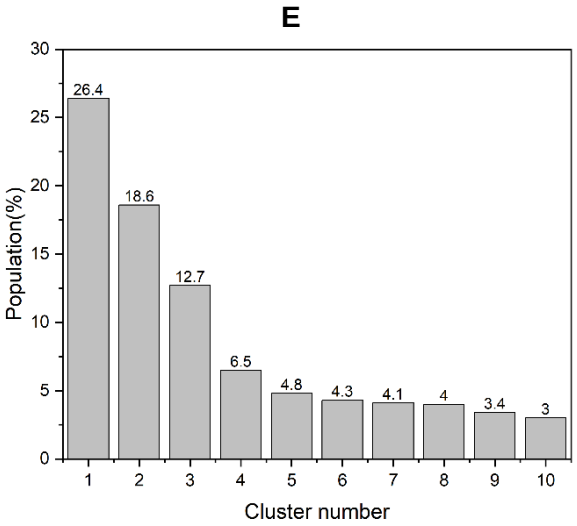
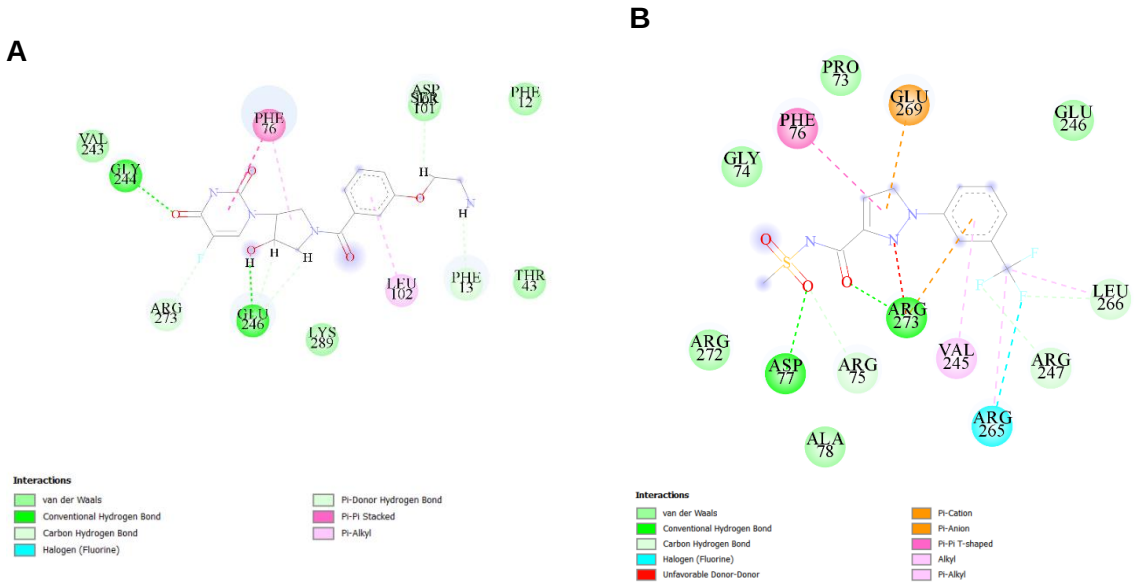


Figure A5. Histograms represent the hierarchical clustering analysis of protein with (A) Ligand A; (B) Ligand B; (C) Ligand C; (D) Ligand D; (E) Ligand E performed using cpptraj module of AMBER 18 Program. Protein-Ligand B and protein-Ligand D complex majorly exist in single conformation. Protein-Ligand A exists majorly in two clusters and many small clusters, whereas the rest of the complexes mostly exist in many small clusters with some change in conformation.

APPENDIX 6



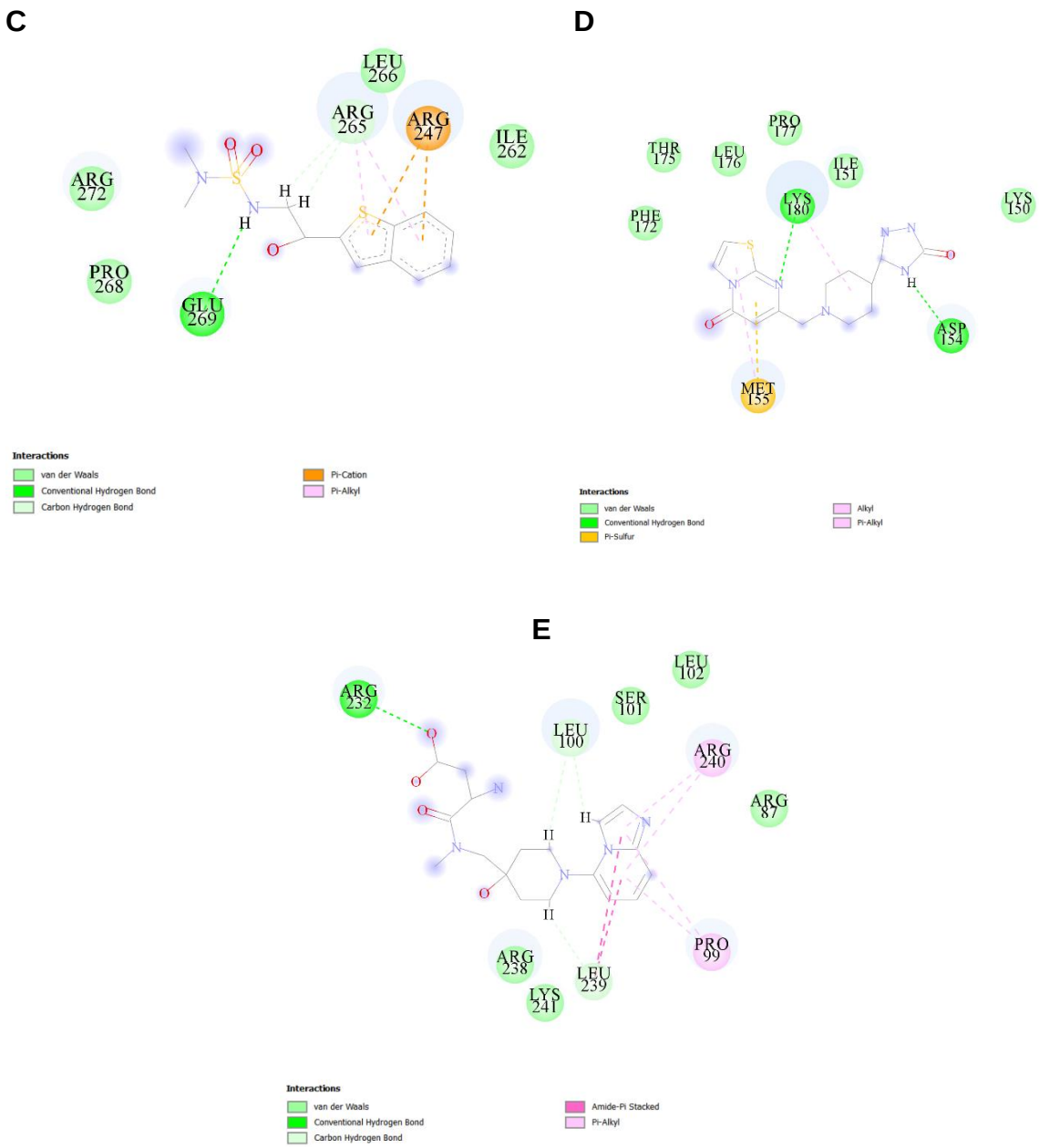


Figure A6. 2D interaction diagrams showing the ligand interaction with residues of the protein within the representative structure of the complex obtained from 500ns of MD simulation of protein with (A) Ligand A; (B) Ligand B; (C) Ligand C; (D) Ligand D; (E) Ligand E. These diagrams are generated using the Discovery studio 2020 client software.