Chapter 3

Virtual Screening and Repurposing of FDA-Approved Drugs from Zinc Database to Identify Potential Autophagy Inhibitors Exploiting ATG4A Cysteine Peptidase as a Target: Potential as Novel Anti-Cancer Molecule*

Abstract

Cancer cells utilize extensive autophagy in effort to adapt to high metabolic stress. This indicates that impairing the high autophagic flux might be an attractive target for cancer therapy. Autophagy related gene 4A (ATG4A) is a key player for autophagy and its inhibition may help in tumor clearance. The present study aims to screen candidate drugs from FDA-approved drugs, a subset of Zinc database, to identify potential ATG4A inhibitors that may have anti-cancer activity. Computer aided drug design (CADD) approach was applied for the study using the virtual screening tools Raccoon and MGLTools-1.5.6. We have identified the drug Lumacaftor as a potent inhibitor of ATG4A on the basis of computational approaches viz. molecular docking, MD simulation and MM/PBSA method. The drug is likely to be a potent regimen candidate to be used as an anti-cancer molecule. However, this potent inhibitor against ATG4A as anti-cancer molecule needs further investigation and validation.

* Part of the work is published in Umesh#, Prerna K.# and Dubey V. K. (2021) Virtual screening and repurposing of FDA-approved drugs from ZINC database: an approach for identifying potential autophagy-inhibitors exploiting Autophagy Related 4A Cysteine Peptidase as a target. J. Biomol. Struct. Dyn. 1-17. #Equal contribution.

3.1 Introduction

Autophagy is a catabolic process found in eukaryotes. The process of autophagy starts with the formation of a double membrane-bound structure, termed as autophagosomes. This double membrane bound structure consists of the key enzymes for the catalysis of undesirable structural molecules such as unfolded proteins, protein aggregates and dysfunctional organelles. These, if not degraded, could lead to imbalance in the cellular homeostasis of the cell. Therefore, dysregulation of autophagy leads to the development of many critical diseases including cancer and metabolic and neurodegeneration disorders (Sánchez-Wandelmer and Reggiori 2017; Kundu *et al.* 2021). Consequently, penetration of the autophagic flux has become a novel strategy for the search of novel therapies for various diseases. For targeting the autophagic flux-mediated modulation, we need a healthier sightedness on the process of formation of the autophagosomes, which will help in finding new molecular drug targets for the development of novel anti-cancer molecules (Mizushima and Levine 2010; Gil *et al.* 2018).

Moreover, autophagy process is involved in several physiological processes including cell development, immune defence, tumor suppression, and the stress response mechanisms (Mizushima and Levine 2010; Klionsky *et al.* 2011). The process of autophagy is regulated by core autophagy proteins which are programmed by >30 autophagy related genes (*Atgs*). Initiation of autophagy starts with the formation of a curve-shaped membranous structure termed as phagophore. As the phagophore starts elongating, the curvature part becomes the vesicle luminal side which is termed as autophagosome (Yu *et al.* 2012). Further, the autophagosomes start maturing and finally get fused with the lysosomes, wherein the fillings of autophagosmes are degraded by acidic hydrolyses residing within the lysosomes (Klionsky *et al.* 2011; Gil *et al.* 2018). During elongation of autophagosomes, a key protein, a mammalian homolog of the yeast ATG8 protein family, plays a crucial role. The ATG8

protein comprises of two subfamilies: the γ-aminobutyric acid receptor-associated protein (GABARAP) and the microtubule-associated protein 1 light chain 3 (MAP1LC3) with its three members; LC3a, LC3b and LC3c (Bortnik and Gorski 2017; Wild *et al.* 2014; Schaaf *et al.* 2016). Autophagosome-associated proteins contain highly similar and conserved amino acid sequences from yeast to mammals and play significant roles in living cells. Apart from autophagy, the atg8 protein family orthologs play important role in several intracellular trafficking processes (Grand *et al.* 2011; Shpilka *et al.* 2011). Moreover, proteins belonging to MAP1LC3 and GABARAP families contain a conserved c-terminal glycine which exhibits an important role during autophagosome formation. In autophagy process, the MAP1LC3 family is required for the elongation of phagophore, whereas, the GABARAP family is involved in the formation of the late stages of the autophagosome (Grand *et al.* 2011).

Significant roles of ATG8 homologs in the de-lipidation and lipidation processes are mainly accompanied by the family of ATG4 cysteine proteases. Thus, ATG4 activity needs to be systematically regulated. This regulation is performed by at least four different human homolog members of ATG4 protein family which are known to participate in autophagy: (i) Autophagin-2 (ATG4a), (ii) Autophagin-1 (ATG4b), (iii) Autophagin-3 (ATG4c), and (iv) Autophagin-4 (ATG4d). These Autophagins are accountable for the catalysis or cleavage of several substances of human ATG8 homologs (Wild *et al.* 2014; Zhang *et al.* 2016; Sugawara *et al.* 2005). Among all the members of ATG4 protease family, only humans (*Homo sapiens*) ATG4a and ATG4b have been studied for their crystallographic structures and denoted as hsATG4a and hsATG4b. Among these two, the crystallographic structure for hsATG4b has been reported, which is composed of 2 domains: a conserved domain characteristic of papain family cysteine proteases and another is short finger domain that is implanted into the catalytic domain and is unique to the ATG4 family proteins (Sugawara *et al.* 2005; Kumanomidou *et al.* 2006). Moreover, among the ATG4 protein family, ATG4c and ATG4d

show slightly less activity towards all types of mammalian ATG8 homologous, whereas, ATG4c and ATG4b, but not ATG4a and ATG4d, integrate with the yeast ATG4 in autophagy. All of the homologs of ATG4 show similar binding affinity towards the ATG8 homologs, despite of their tremendously different catalytic efficiency (Li *et al.* 2010). Loss of function of the cysteine peptidase ATG4a leads to impairment of the autophagy process. Therefore, hsATG4a can be a potent target for developing the autophagy-inhibitors (Fernández and López-Otín 2015). Rising tumor cells activate autophagy in order to meet the intensive nutritional requirement for the cell survival. Therefore, we have selected the ATG4a protein as a potential therapeutic target for developing anti-cancer drugs.

Recently, a computational along with the experimental background, a study has identified the ATG4a inhibitor(s) using the FDA-approved drugs from the MedChem Express (MCE) database (Liu *et al.* 2018b). In our current study, we have used the FDA-approved drugs from the ZINC database. Although, there may be some overlap in the drug compounds in these two databases, the database which we are using in our study is a bigger one consisting of ~1600 molecules. In this computational study, we have taken atg4a cysteine peptidase (ATG4a; PDB-ID 2P82) as a target molecule (Umesh *et al.* 2021). The tetramer structure of ATG4a consists of four chains; chain-a, -b, -c, and -d. For molecular docking study, we have selected chain-c for making active site grid. The active site of ATG4a is made up of Cys77, Asp279, and His281, the geometry of which corresponds to the canonical catalytic triad of cysteine protease (Maruyama and Noda 2018).

3.2 Methodology

3.2.1 Preparation of target protein: A total of ~1600 structural molecules were retrieved from the FDA-approved drug database, a subset of ZINC database (Irwin *et al.* 2012). The structure files for these molecules were downloaded in a sdf format from the ZINC database. Then after, the sdf structure files were converted into mol2 structure files using Open Babel

(O'Boyle *et al.* 2011). The procedure was followed by converting the mol2 structure files into the pdbqt format using AutoDockTools-1.5.6 (ADT) and graphical user interface (GUI) version of Raccoon. In this study, we used a computer aided drug design (CADD) approach for the molecular screening purpose using the Raccoon and MGLTools-1.5.6 virtual screening tools. Then after, the online virtual screening tool Raccoon was used to prepare the molecules with an aid of AutoDockTools-1.5.6 using python as a programming language (Morris *et al.* 2009).

Further, the structure of the protein ATG4A cysteine peptidase (ATG4A; PBD-ID: 2P82) was downloaded from the RCSB protein database. Here, it is worth mentioning that for our current study, we have used the crystal structure of ATG4A as a target protein, however, in an earlier study (Liu *et al.* 2018b), the SWISS-MODEL-based generated model of ATG4A was used as a target protein to perform molecular docking, suggesting a significant reliable difference in our computational approach for identifying the ATG4A inhibitor(s). Furthermore, the atomic co-ordinates of the active site of the protein ATG4A were retrieved from the reported literature, in which the catalytic residues were identified as Cys77C, Asp279C, and His281C; the geometry of which corresponds to the canonical triad of the enzyme cysteine protease. Further, energy minimization of the protein (2P82; ATG4A) was done using Swiss PDB viewer (Guex and Peitsch 1997). Prior to docking, the protein was assigned the charge, solvation parameters and fragmental volumes using the AutoDock tool. The PDB structure of the ATG4A protein was further optimized using the AutoDock tool through an established procedure (Morris *et al.* 2009) and converted the sdf structure files into the mol2 structure files using Open Babel.

3.2.2 Drug molecule screening through molecular docking: Screening of the drug molecules from the FDA-approved drug database was performed by computer aided drug design (CADD) approach using Raccoon and MGLTools-1.5.6 virtual screening tools.

Raccoon was used for preparing the molecules using AutoDockTools-1.5.6, as a virtual screening tool using python as a programming language(Morris *et al.* 2009). During docking, the small molecules were considered as a flexible entity and the ATG4A protein as a rigid entity. The most important step in the docking is to prepare grid parameter file (GPF) and docking parameter file (DPF), which was accomplished through the software AutoDock4. With the aid of this software, grid maps for each ligand were generated. The dimensions of the grid box were defined in such a way so that it accommodates one ligand at a time in the active site of the protein. The -X, -Y, and -Z grid coordinates of the protein ATG4A were set as 35.682, 93.845, and 9.468. Then after, the screening of the drug molecules was done by performing molecular docking using the default docking parameter settings of the Lamarckian genetic algorithm, which were as follows: 100 GA runs, 150 individual's population size, maximum of 2.5 million numbers of energy evaluations, and random starting positions and conformation. The positional root-mean-square deviation (RMSD) value less than 1Å was considered to be ideal and clustered together for finding the favourable interaction. The highest negative binding energy (B.E.) of the compounds was considered to show maximum binding affinity with the protein. Earlier identified ATG4A inhibitor, Tioconazole (Liu *et al.* 2018b) was used as a control for our current study.

3.2.3 ADME analysis: The top 25 hits were evaluated for their properties such as absorption, distribution, metabolism, and excretion (ADME) in order to predict their pharmacokinetic properties using an online server SwissADME (Guex and Peitsch 1997). The drugs were also passed through the Lipinski's filter to assess their ability of drug likeness by following the Lipinski's rule of five (Lipinski 2004; Jayaram et al. 2012) using the same online server.

3.2.4 Molecular Dynamics (MD) Simulation study: After molecular docking and ADME analysis, we subjected the best six hit compounds to the MD simulation studies. For this, we used GROMACS v2018.8 simulation package with the GROMOS96 as a force field for all

the calculation purposes(Durrant and McCammon 2011; Lemkul 2019; Oostenbrink et al. 2004). As a part of the process, the PRODRG server was used for generating the parameters file and for building the topology files for the ligands (Schüttelkopf and Van Aalten 2004). Further, during the simulation process, we used the apoprotein as a rigid entity whereas the ligands as a flexible entitities. Furthermore, a cubic dimension box was used for centering the biomolecular complexes with a distance of 1.2 nm from the edge of the box in effort to satisfy the periodic boundary conditions. The cubic box along with biomolecular complexes within it, was allowed to be solvated using the SPC water model system. Further, an ionization step was performed wherein, the charge within the system was made electroneutral with the help of adding the required amount of Na+ or Cl- ions into it. The ionization step was followed by the energy minimization step, for which, we used the steepest descent method with a convergence threshold of 1000 kJ/mol/nm for 50,000 steps for all the drug complexes along with the apoprotein. Moreover, we equilibrated the solvent and ions around the biomolecular complex under an isothermal-isochoric ensemble (NVT ensemble or constant Number of particles, Volume and Temperature) using the modified Berendsen thermostat coupling method (Bussi *et al.* 2007) at a temperature of 300 K with the coupling constant set at 0.1 ps. In addition to this, we further equilibrated the solvent and ions around the biomolecular complex under an isothermal-isobaric ensemble (NPT ensemble or constant Number of particles, Pressure and Temperature) using the Parrinello-Rahman pressure coupling method (Martoňák *et al.* 2003) at pressure of 1.0 bar with the coupling constant set at 2.0 ps. For long-range electrostatic interactions, we subjected the simulation process to the Particle Mesh Ewald (PME) algorithm (Kawata and Nagashima 2001). For both, short range electrostatic and Vander Waal interaction, the cut-off distance was set at 1.0 nm. The parameters for bond length were constrained using the LINear Constraint Solver (LINCS) algorithm (Hess *et al.* 1997). Each of the equilibration phases were subjected for 1000 ps at

the time step (dt) set to 0.002 ps. After the completion of both the equilibration phases, we carried out a final production MD run for the apoprotein along with all six protein-ligand complexes for 100 ns with the time step set at 0.002 ps. The generated trajectories were stored for further analyses.

3.2.5 Calculation of average binding energy through MM/PBSA method: Post-simulation, all the six protein-ligand complexes were further assessed for their average binding energy along with the residues responsible for imparting their maximum contribution to the overall average binding energy. This calculation of average binding energy was performed through molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA), a highly efficient computational method which is widely used in an integration with the MD simulation process to compute the free energy interaction between the biomolecules using the procedure explained in chapter 2.

3.2.6 Analyses and visualization: The docking result was analysed using the AutoDock tool (Morris et al., 1998). The docked protein-ligand complexes were visualised using Pymol-2.3.3 and Ligplot (Laskowski and Swindells 2011). After the successful completion of the MD run production, the analyses of the generated trajectories for evaluating the structural properties; such as, root-mean-square deviation (RMSD), root mean square fluctuation (RMSF) and radius of gyration (Rg) of the protein and its bound ligand complexes were performed using the 'g_rms', 'g_rmsf', and 'g_gyrate' modules respectively, which are an in-built function of GROMACS 18.8 simulation package. The number of intermolecular hydrogen bonds between the biomolecular complexes were assessed, which is based on bond length smaller than 3.6 Å distance between the protein and the ligand and the angle between the donor-hydrogen and acceptor larger than 90° through the g_hbond utility of GROMACS 18.8. The data were plotted using the QtGrace version of GRaphing Advanced Computation and Exploration (GRACE) program.

3.3 Results and Discussion

3.3.1 Molecular Docking: We have retrieved the sdf structure of ~1600 drug molecules from the FDA-approved drug database, a subset of ZINC database. Among them, we have selected the top 25 drug molecules based on their binding affinity with the protein ATG4A (Table 3.1).

Table 3.1. Details of interacted amino acid residue present near the active site of ATG4A protein along with their respective drug molecule with their binding energy (B.E.) and inhibition constant (Ki). The residues of active site are represented as in bold format.

Complex	ZINC ID	B.E. (kcal/mol)	Ki	Active Site Residues
Samsca	ZINC000000538658	-7.1	6.28 µM	**His281**, Tyr57, Gly74, Asn263 Glu144, Trp145, Gly143, Trp75, Asn262
Lumacaftor	ZINC000064033452	-6.27	25.37 µM	Asn263, **His281**, Try265, Pro261, Asn262
Dolutegravir	ZINC000058581064	-5.97	41.86 µM	**Asp279**, Try265, **His281,** Asn263, Pro261, Try57, Asn262
Imatinib	ZINC000019632618	-5.96	43.12 µM	Asn263, **His281,** Try57, Asn262, Gly74, Pro261
Mepron	ZINC000100017856	-5.94	44.44 µM	Gly74, Asn262, Try57, Asn263, **His281**
Benzoyl	ZINC000000001016	-5.93	44.64 µM	Asn262, Trp75, Trp145, Pro261, Glu144, Gly143, Gly74
Caprelsa	ZINC000053683345	-5.91	46.46 µM	Tyr265, Asn263, Asn262, Gly74, **His281, Asp279,** Tyr57
Azelastine	ZINC000000601229	-5.9	47.17 µM	Trp75, Gly74, Pro261, Trp145, Asn262, Gly143
Zytiga	ZINC000003797541	-5.82	54.0 µM	Pro261, Asn263, Asn262, **His281**, Tyr57
Panobinostat	ZINC000022010649	-5.82	53.9 µM	Glu144, Pro261, Asn262, Trp145, Gly143, Asn263, Try57, **His281**
Vorapaxar	ZINC000003925861	-5.77	58.64 µM	**Asp279**, Tyr265, **His281**, Asn263, Pro261, Asn262
Raltegravir	ZINC000013831130	-5.77	59.22 µM	Ser127, Asp72, Arg80, Gly83, Ile128, Met131, Gly76, Leu79, Trp145, Phe146, Val132, Met178, Glu144, Trp75
Axitinib	ZINC000003816287	-5.75	61.05 µM	**Asp279**, Thr282, **His281**, Try265, Asn263, Pro261

Tasmar	ZINC000035342789	-5.75	60.54 µM	Asn262, Asn263, Tyr265, **Asp279**, Tyr57, **His281**
Inspra	ZINC000003985982	-5.72	64.16 µM	Asn263, **His281**, Asn262, Tyr57, Gly74
Belinostat	ZINC000003818726	-5.7	66.4 µM	Gly143, Glu144, Trp75, Gly74, Tyr57, Pro261, Trp145, Asn262
Thalitone	ZINC000000057255	-5.69	67.2 µM	**Asp279**, Tyr265, **His281**, Tyr57, Asn262, Asn263, Gly74
Differin	ZINC000003784182	-5.67	70.26 µM	Gly74, Asn262, Asn263, **His281**
Anzemet	ZINC000000002688	-5.62	76.37 µM	Tyr265, **Asp279, His281**, Asn263, Asn262, Tyr57, Gly74
Meloxicam	ZINC000013129998	-5.61	77.44 µM	Gly74, Trp75, Gly143, Glu144 Asn262, Pro261, Trp145, Gly147
Cefazolin	ZINC000003830405	-5.6	79.02 µM	Glu144, Trp145, Asn262, Gly74, Pro261, Asn263, **His281**, Try57
Pletal	ZINC000001552174	-5.58	81.24 µM	Trp145, Glu144, Trp75, Asn262, Tyr265, Asn263, **His281**, Tyr57, Gly74, Gly143
Idarubicin	ZINC000003920266	-5.46	100.16 µM	Tyr265, **His281, Asp279**, Tyr57, Gly74, Asn262, Asn263
Avage	ZINC000001542199	-5.42	106.11 µM	Asn262, Gly74, **His281, Asp279**, Tyr265
Dolasetron	ZINC000103105084	-5.37	116.51 µM	Tyr265, **Asp279, His281**, Tyr57, Gly74, Asn262, Asn263

The top 25 molecules include Samsca, Lumacaftor, Dolutegravir, Imatinib, Mepron, Benzoyl, Caprelsa, Azelastine, Zytiga, Panobinostat, Vorapaxar, Raltegravir, Axitinib, Tasmar, Inspra, Belinostat, Thalitone, Differin, Anzemet, Meloxicam, Cefazolin, Pletal, Idarubicin, Avage, and Dolasetron. The further insight into the binding interaction of these drug molecules with ATG4A has been analysed using the ligplot program as shown in the Figure 3.1 to Figure 3.7. The active site coordinates of the ATG4A protein (PDB-ID: 2P82) were defined as Cys77, Asp279, and His281 (Maruyama and Noda 2018), and its unliganded structure was used for molecular docking. The c-chain of the protein ATG4A has been found

to have the active site residues involving Cys77, Asp279 and His28, the geometry as active site of cysteine protease. The top hit drug Samsca (-7.1 kcal/mol) was found to form hydrogen bond with Asn262, and hydrophobic interactions with His281, Tyr57, Gly74, Asn263, Glu144, Trp145, Gly143, and Trp75 (Figure 3.1a). Lumacaftor shows binding energy as -6.27 kcal/mol and forms hydrogen bond with Tyr265 and hydrophobic interactions with Asn263, His281, Pro261 and Asn262 (Figure 3.1b).

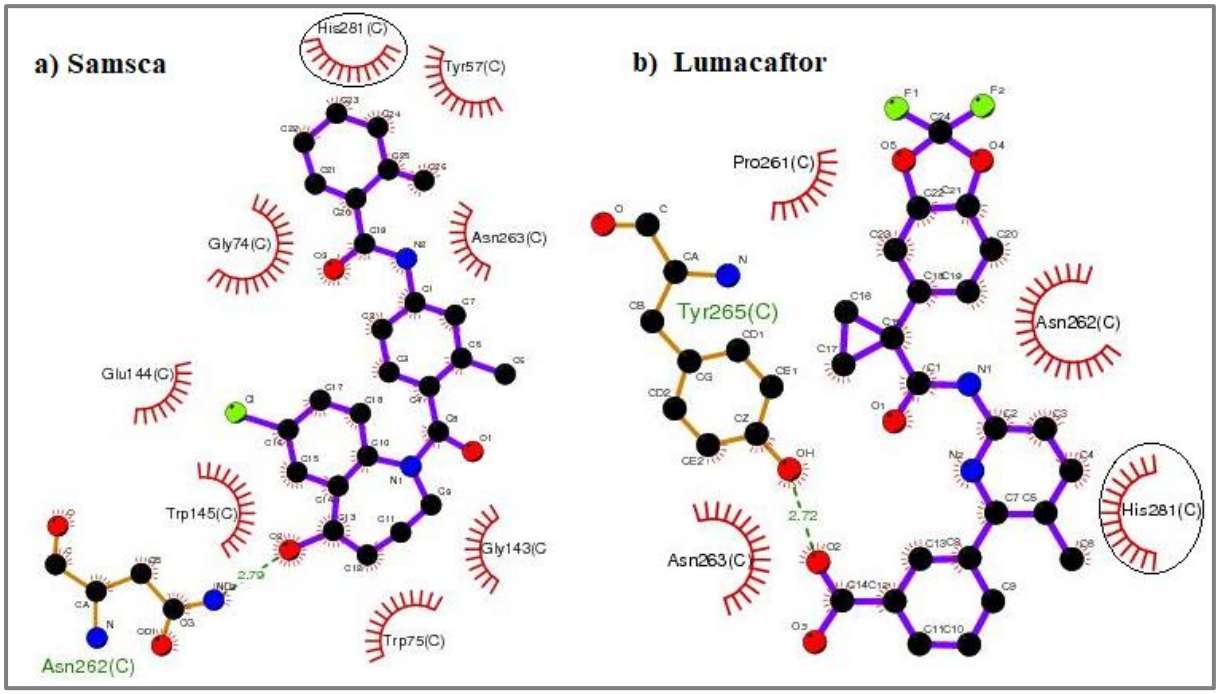

Figure 3.1. The image of Ligplot displaying both type of interaction; hydrophobic and hydrogen interactions. (a) Samsca and **(b)** Lumacaftor with ATG4A. Encircled amino acid residues represent interaction with ATG4A active site residues. His281 forms H-bonding with both the compounds.

Dolutegravir (-5.97 kcal/mol) forms hydrogen bonds with Tyr57 and Asn263 and hydrophobic interactions with Asp279, Try265, His281, Pro261, and Asn262 (Figure 3.2a). Imatinib shows binding energy of -5.96 kcal/mol, and forms H-bonds interaction with Asn262 and Asn263 and hydrophobic interactions with His281, Try57, Gly74, and Pro261 (Figure 3.2b). Mepron shows binding energy of -5.94 kcal/mol, also forms hydrophobic interactions with Gly74, Asn262, Try57, Asn263, and His281 (Figure 3.2c).

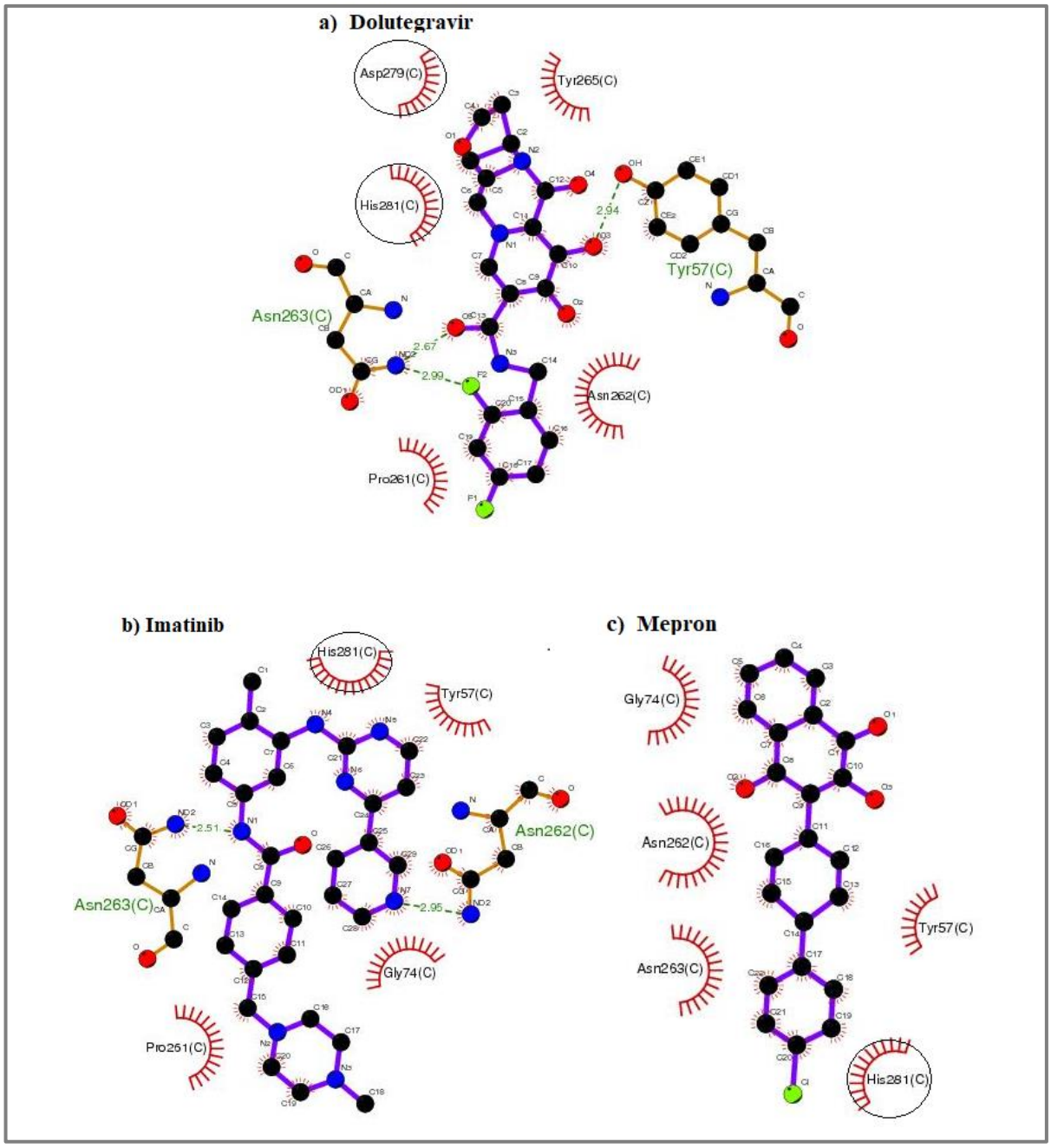

Figure 3.2. The image of Ligplot displaying both type of interaction; hydrophobic and hydrogen interactions. (a) Dolutegravir, **(b)** Imatinib and **(c)** Mepron with ATG4A. Encircled amino acid residues represent interaction with ATG4A active site residues. His281 forms hydrogen bonding with Imatinib and Mepron.

Benzoyl shows binding energy of -5.93 kcal/mol, and forms H-bond with Asn262 and hydrophobic interactions with Trp75, Trp145, Pro261, Glu144, Gly143, and Gly74 (Figure 3.3a). Caprelsa (-5.91 kcal/mol) shows hydrophobic interactions with Tyr265, Asn263, Asn262, His281 and Asp279 and hydrogen bonds with Gly74 and Tyr57 (Figure 3.3b).

Azelastine shows binding energy of -5.9 kcal/mol, and hydrophobic interactions with Trp75, Gly74, Pro261, Trp145, Asn262, and Gly143 (Figure 3.3c). Zytiga (-5.82 kcal/mol) shows hydrogen bonding with Tyr57 and hydrophobic interactions with Pro261, Asn263, Asn262, and His281 (Figure 3.3d).

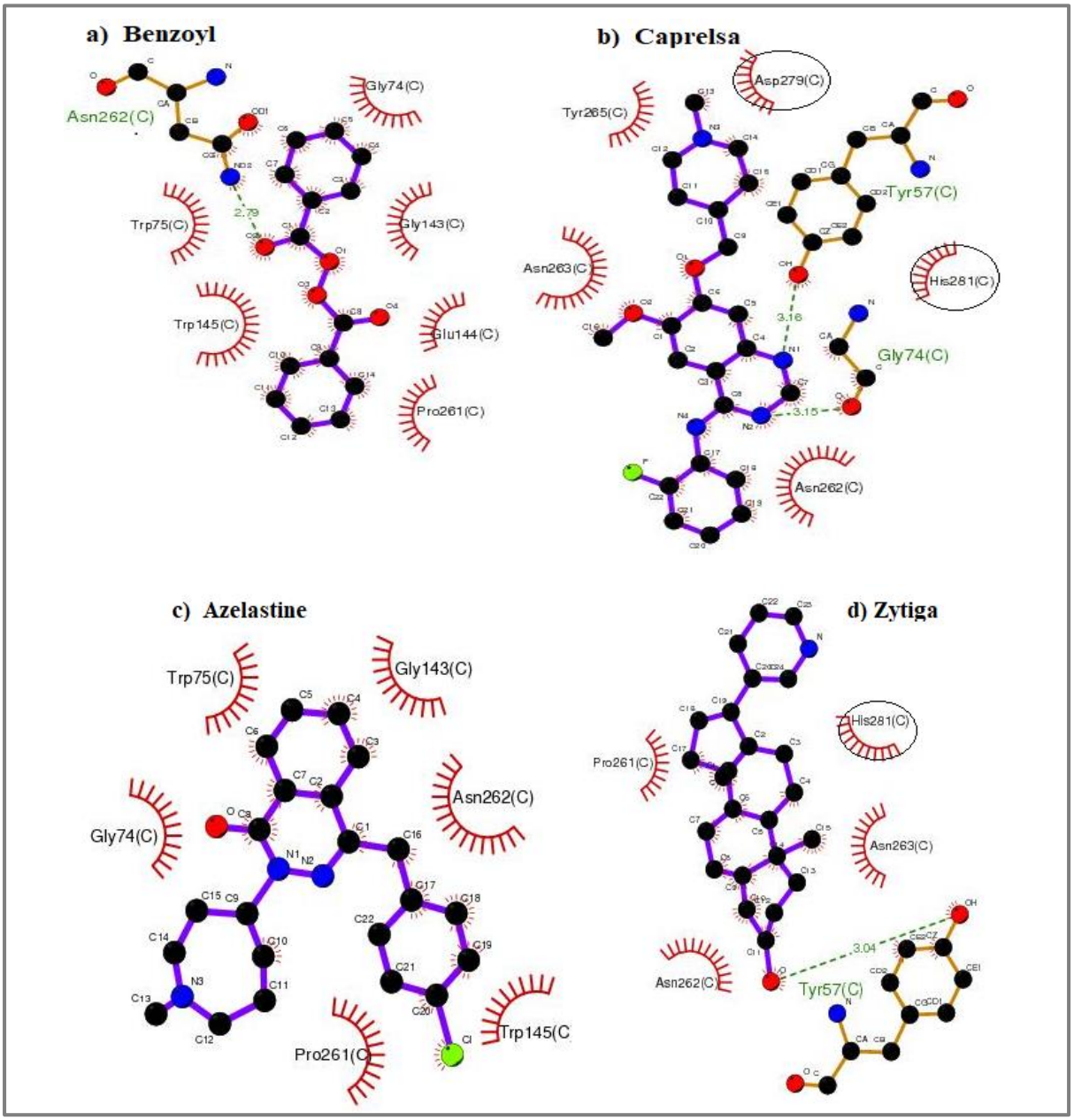

Figure 3.3. The image of Ligplot displaying both type of interaction; hydrophobic and hydrogen interactions. (a) Benzoyl **(b)** Caprelsa, **(c)** Azelastine and **(d)** Zytiga with ATG4A. Encircled amino acid residues represent interaction with ATG4A active site residues.

Panobinostat shows binding energy of -5.82 and hydrogen bonds with Try57, Trp145, and Asn262 and hydrophobic interactions with Glu144, Pro261, Gly143, Asn263, and His281 (Figure 3.4a). Vorapaxar shows binding energy of -5.77 kcal/mol and H-bonding with Asn263.Vorapaxar shows hydrophobic interactions with Asp279, Tyr265, His281, Pro261, and Asn262 (Figure 3.4b). Raltegravir shows binding energy -5.77 kcal/mol, with H-bonding with Met78, Ile128, Met131, and Arg80 with additional hydrophobic interaction Ser127, Asp72, Gly83, Gly76, Leu79, Trp145, Phe146, Val132, Glu144, and Trp75 (Figure 3.4c). Axitinib shows binding energy of -5.75 kcal/mol, and H-bonding with Asn263 and hydrophobic interactions with Asp279, Thr282, His281, Try265, and Pro261 (Figure 3.4d).

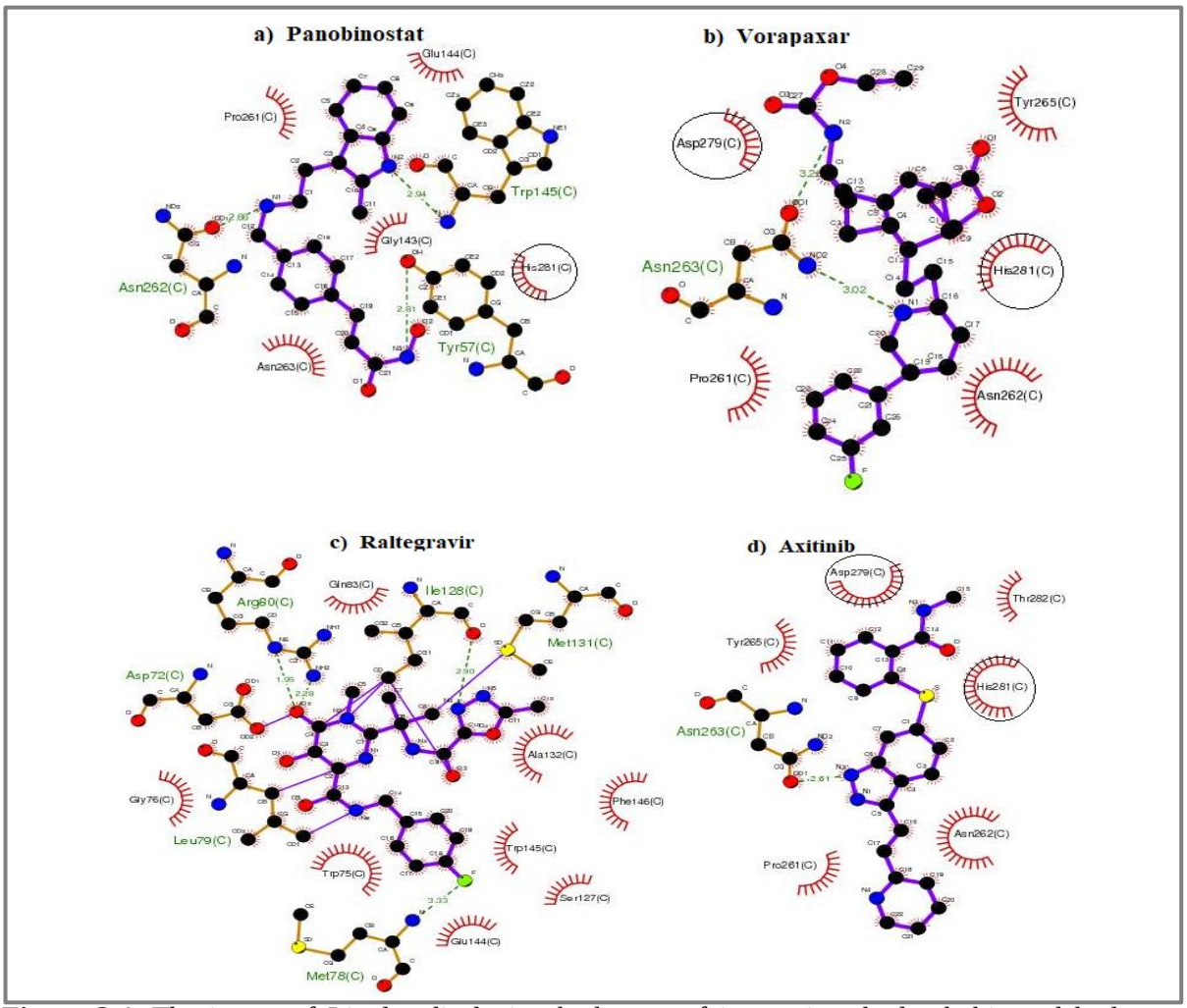

Figure 3.4. The image of Ligplot displaying both type of interaction; hydrophobic and hydrogen interactions. (a) Panobinostat (b) Vorapaxar (c) Raltegravir, and (d) Axitinib with ATG4A. **Encircl**ed amino acid residues represent interaction with ATG4A active site residues.

Tasmar shows binding energy of -5.75 kcal/mol and hydrogen bonding with Tyr57 and hydrophobic interactions with Asn262, Asn263, Tyr265, Asp279, and His281 (Figure 3.5a). Inspra shows binding energy of -5.72 kcal/mol, and H-bonding with Tyr57 and hydrophobic interactions with Asn263, His281, Asn262, and Gly74 (Figure 3.5b). Belinostat shows binding energy as -5.7 kcal/mol and hydrogen bonding with Tyr57 and Asn262 and hydrophobic interactions with Gly143, Glu144, Trp75, Gly74, Pro261, and Trp145 (Figure 3.5c). Thalitone shows binding energy as -5.69 kcal/mol and hydrogen bonding with Asp279 and Asn263 and also, it shows hydrophobic interactions with Tyr265, His281, Tyr57, Asn262, and Gly74 (Figure 3.5d).

Differin (-5.67 kcal/mol) forms hydrophobic interactions with Gly74, Asn262, Asn263, and His281 (Figure 3.6a). Anzemet (-5.62 kcal/mol) forms hydrogen bond with Tyr57 and hydrophobic interactions with Tyr265, Asp279, His281, Asn263, Asn262, and Gly74 (Figure 3.6b). Meloxicam shows binding energy of -5.61 kcal/mol, and forms H-bond with Trp145 and hydrophobic interactions with Gly74, Trp75, Gly143, Glu144, Asn262, Pro261, and Gly147 (Figure 3.6c). Cefazolin shows binding energy (-5.6 kcal/mol) forms H-bonds with Tyr57, Trp145, and Asn263 and hydrophobic interactions with Glu144, Asn262, Gly74, Pro261, and His281 (Figure 3.6d).

Pletal shows binding energy of -5.58 kcal/mol, and is seen to form H-bonds with Trp145 and Tyr57 and other hydrophobic interactions with Glu144, Trp75, Asn262, Tyr265, Asn263, His281, Gly74, and Gly143 (Figure 3.7a). Idarubicin (-5.46 kcal/mol) forms H-bonds with Tyr57 and Asn263 and hydrophobic interactions with Tyr265, His281, Asp279, Gly74, and Asn262 (Figure 3.7b). Avage (-5.42 kcal/mol) forms hydrophobic interactions with Asn262, Gly74, His281, Asp279, and Tyr265 (Figure 3.7c). Dolasetron shows binding energy of -5.37 kcal/mol, and forms H-bond with Tyr57 and hydrophobic interactions with Tyr265, Asp279, His281, Gly74, Asn262, and Asn263 (Figure 3.7d).

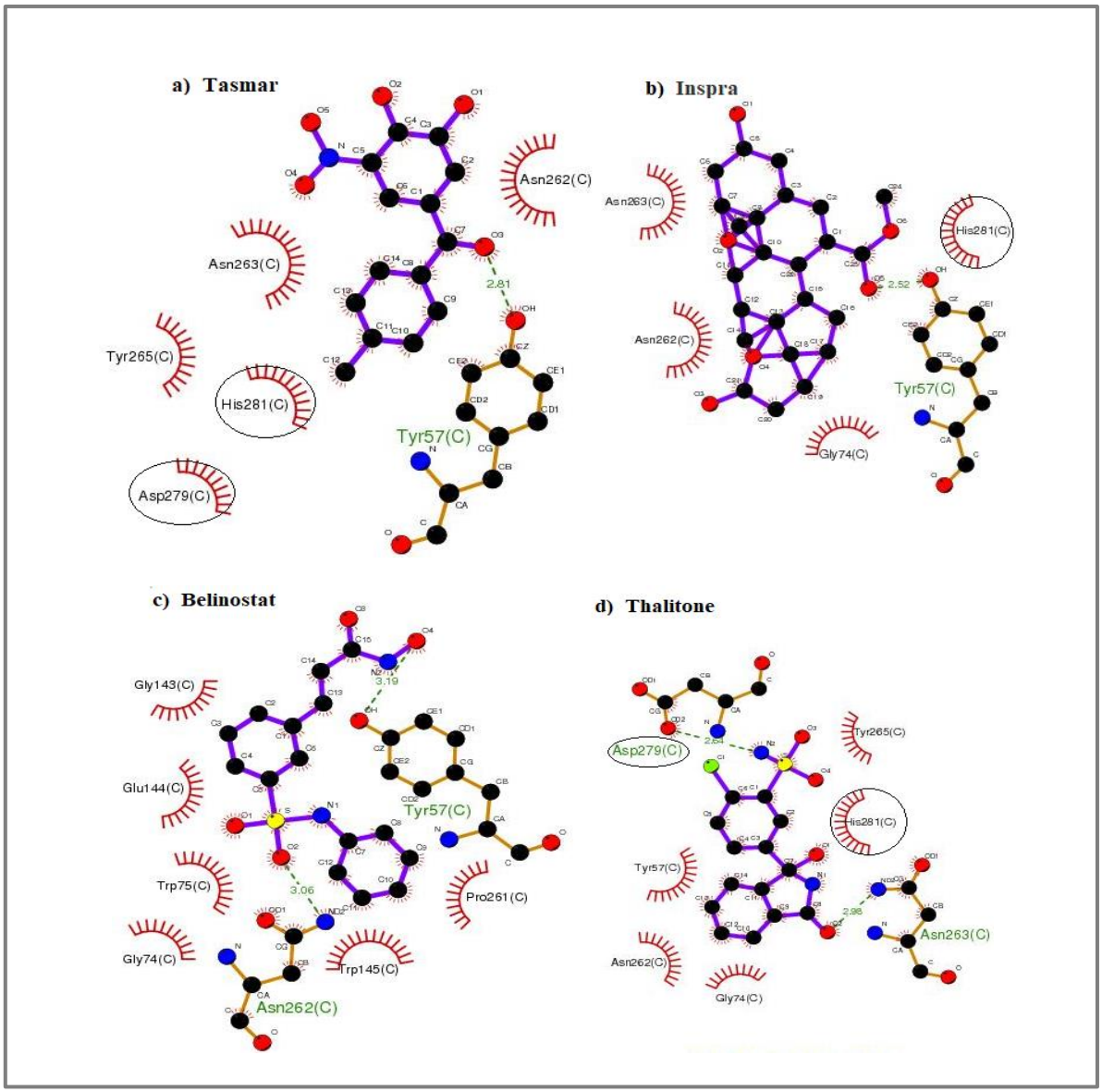

Figure 3.5. The image of Ligplot displaying both type of interaction; hydrophobic and hydrogen interactions. (a) Tasmar **(b)** Inspra **(c)** Belinostat and **(d)** Thalitone with ATG4A. Encircled amino acid residues represent interaction with ATG4A active site residues.

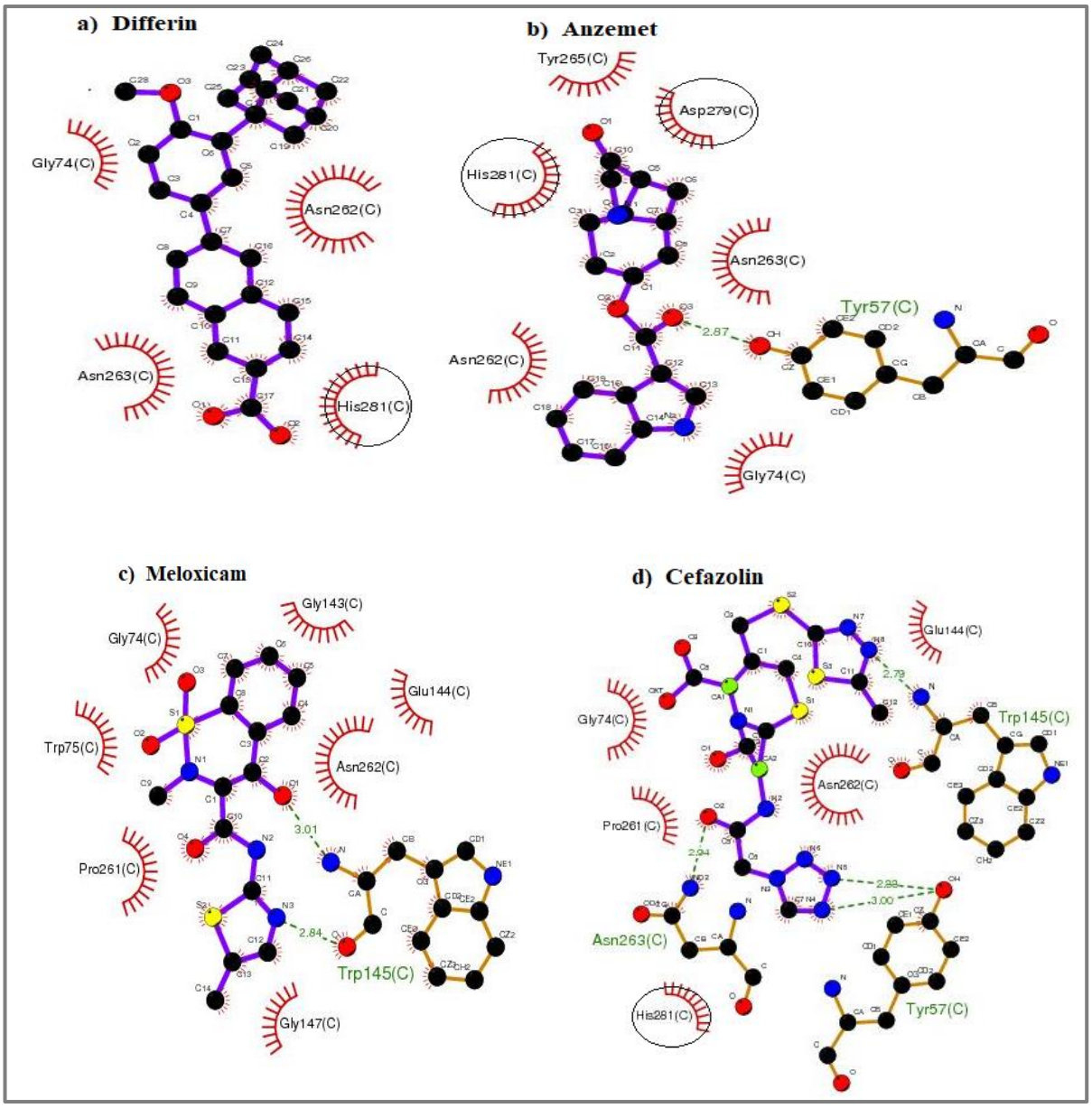

Figure 3.6. The image of Ligplot displaying both type of interaction; hydrophobic and hydrogen interactions. (a) Differin **(b)** Anzemet **(c)** Meloxicam and **(d)** Cefazolin with ATG4A. Encircled amino acid residues represent interaction with ATG4A active site residues.

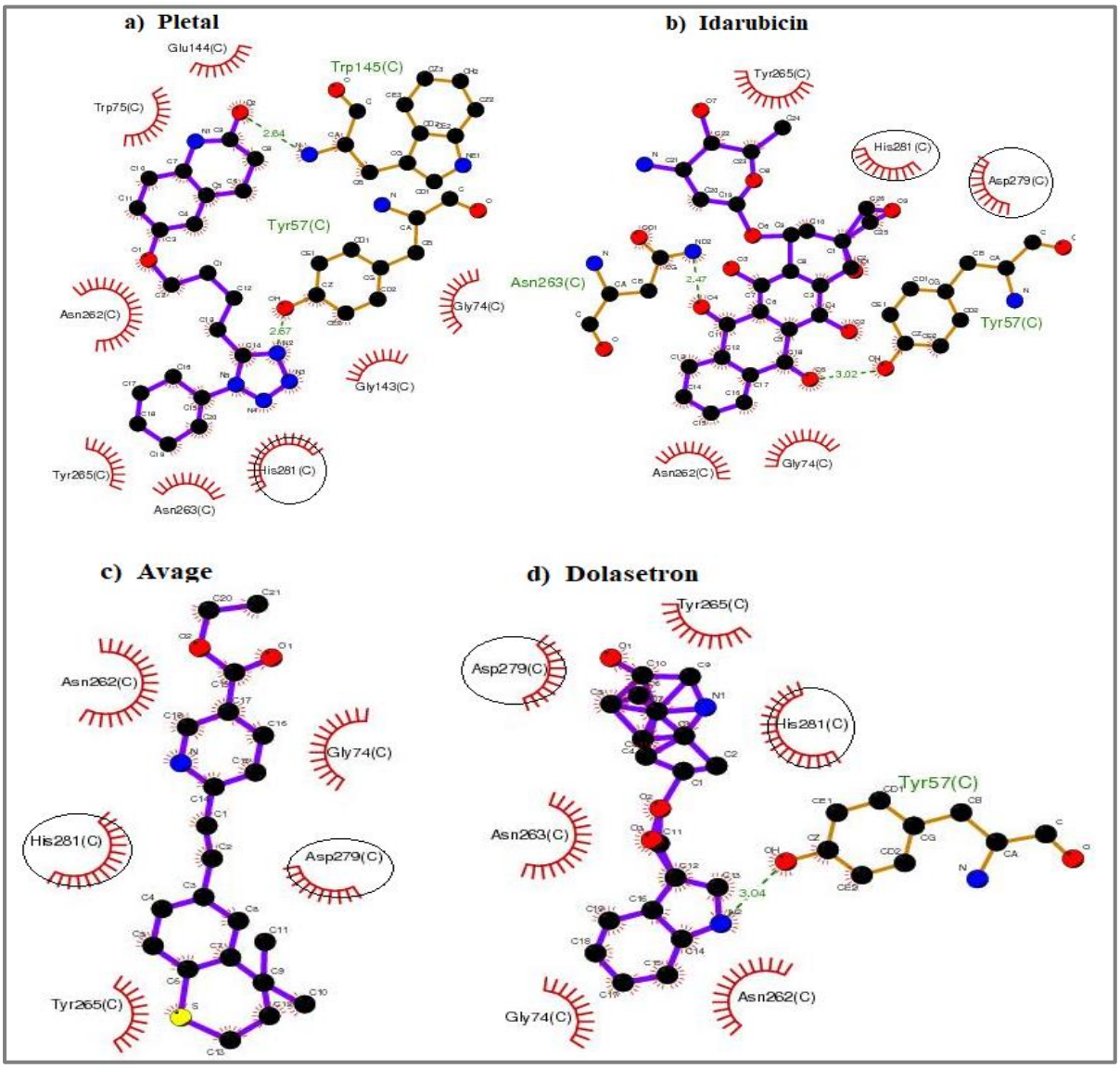

Figure 3.7. The image ofLigplot displaying both type of interaction; hydrophobic and hydrogen interactions. (a) Pletal **(b)** Idarubicin **(c)** Avage and **(d)** Dolasetron with ATG4A. Encircled amino acid residues represent interaction with ATG4A active site residues.

The overall molecular docking result reveals that the residues viz. Tyr57, Trp145, Asn262, Asn263, and Tyr265 are consistent throughout most of the drug complexes in making intermolecular hydrogen bond. The control drug Tioconazole has been reported to form a favourable binding at the active site of the ATG4A protein, however, it is interesting to note that all the top 25 drug complexes exhibit higher binding energy than that of the control drug Tioconazole. Moreover, our docking result shows that the residues viz. Tyr57,

Trp145, Asn262, Asn263, and Tyr265 of almost all of the top 25 drugs have also been found to be present within the active site of the protein ATG4A. Therefore, it may be suggested that these residues behave as the key contributor residues which impart their role in making favourable intermolecular hydrogen bond during the process of molecular docking and thus, providing the overall stability to the docked complexes.

3.3.2 Assessment of the Druglikeness. After molecular docking, we subjected the top 25 drugs to pass through the Lipinski's filter and assessed their ability to follow the Lipinski's rule of five, thereby to show the druglikeness. According to the Lipinski's rule of five, the molecule which follows the four out of five parameters i.e., molecular weight less than or equal to 500 g/mol, MLogP less than or equal to 4.15, number of H-bond acceptors (N or O) less than or equal to 10 and number of H-bond donors (NH or OH) less than or equal to 5, is considered as an orally consumable drug and is potential enough to be used in the biological system. The comparative detailed assessment given in Table 3.2 revealed that most of the drugs possess the ability to show the druglikeness with no violations, whereas, a few of the drugs viz. Samsca, Azelastine, Zytiga, and Differin possess the druglikeness ability with a single violation of MLogP greater than 4.15, which suggests that these compounds may have a little less lipophilicity. Altogether, these results support all the drugs in favour of possessing the pharmacokinetic properties. Further, the top six drugs viz. Samsca, Lumacaftor, Dolutegravir, Imatinib, Mepron, and Benzoyl were selected for further studies. Noticeably, we included the drug Samsca in the top six hit drugs, as this drug exhibits the highest negative binding energy (-7.1 kcal/mol) along with the purpose to increase the chemical space for drug discovery. These results together potentiate the ability of the top six drug complexes viz. Samsca, Lumacaftor, Dolutegravir, Imatinib, Mepron, and Benzoyl to be subjected for further studies.

Table 3.2. Detailed list of the ADME properties of selected small molecules along with the Lipinski's parameters. Water solubility has been represented by LogS, whereas, lipophilicity of the small molecule is represented by MLogP. TPSA represents theTotal Polar Solvent Accessibility. MW represents the Molecular Weight.

Compound	MLogP	TPSA (Å^2)	MW (g/mol)	H-Donor	H-Acceptor	LogS	Druglikeness (Violations)
Samsca	4.16	69.64	448.94	2	3	-5.71	Yes, (1)
Lumacaftor	2.69	97.75	452.41	2	8	-5.45	Yes, (0)
Dolutegravir	0.79	100.87	419.38	2	7	-4.01	Yes, (0)
Imatinib	2.15	86.28	493.60	2	6	-5.07	Yes, (0)
Mepron	3.28	54.37	366.84	1	3	-5.90	Yes, (0)
Benzoyl	3.50	52.60	242.23	0	4	-3.68	Yes, (0)
Caprelsa	3.79	62.74	475.35	1	6	-5.41	Yes, (0)
Azelastine	4.28	38.13	381.90	0	3	-5.20	Yes, (1)
Zytiga	4.42	33.12	349.51	1	2	-5.03	Yes, (1)
Panobinostat	2.31	77.15	349.43	4	3	-3.79	Yes, (0)
Vorapaxar	4.13	77.52	492.58	1	6	-6.00	Yes, (0)
Raltegravir	1.03	155.73	444.42	3	10	-3.55	Yes, (0)
Axitinib	3.12	95.97	386.47	2	3	-5.02	Yes, (0)
Tasmar	0.97	103.35	273.24	2	5	-3.86	Yes, (0)
Inspra	2.75	82.20	414.49	0	6	-3.15	Yes, (0)
Belinostat	1.55	103.88	318.35	3	4	-2.87	Yes, (0)
Thalitone	1.63	117.87	338.77	3	5	-2.75	Yes, (0)
Differin	5.34	46.53	412.52	1	3	-7.37	Yes, (1)
Anzemet	1.95	62.40	324.37	1	4	-3.63	Yes, (0)
Meloxicam	-0.26	136.22	351.40	2	5	-4.34	Yes, (0)
Cefazolin	0.08	234.93	454.51	2	9	-2.02	Yes, (0)
Pletal	3.04	81.93	369.46	1	5	-3.92	Yes, (0)
Idarubicin	-1.05	176.61	497.49	5	10	-4.14	Yes, (0)
Thalitone	0.77	117.87	338.77	3	5	-2.75	Yes, (0)
Avage	4.00	64.49	351.46	0	3	-5.26	Yes, (0)
Dolasetron	1.95	62.40	324.37	1	4	-3.63	Yes, (0)

3.3.3 MD Simulation Analysis: After docking and assessing the druglikeness, based on the highest negative binding energy, we selected six compounds and subjected them for MD simulation studies along with the apoprotein itself in a triplicate manner. MD simulation-based studies helped us in evaluating the stability of the docked biomolecular complexes for a course of 100 ns simulation with the help of the generated trajectories for the RMSD, the RMSF, the Rg, and the intermolecular hydrogen bonds. For calculating the RMSD, we selected the 'backbone' as well as 'ligand' for least-squares fitting using the 'g_rms' utility of

the GROMACS simulation package. From the figure 3.8A, it is evident that the complex Lumacaftor showed the maximum backbone stability among all the six complexes with a maximum deviation at ~0.3 nm. The RMSD trajectory for the complex Lumacaftor exhibited a steady pattern throughout the 100 ns simulation. Additionally, it is also evident from the graph that only the complex Lumacaftor showed a converged pattern of RMSD with that of the apoprotein (~0.2 nm), however, rest all the complexes showed the RMSD value above than that of the apoprotein (Figure 3.8B; depicts the RMSD value of all the protein-ligand complexes as well as of the apoprotein from 80 ns to 100 ns). This suggests that the drug Lumacaftor did not show any unstable instances upon binding to the apoprotein, however, all the five complexes may exhibit a little instability with respect to their binding with the apoprotein. Moreover, followed by Lumacaftor, the rest of the complexes showed backbone RMSD value ranged between ~0.2-0.35 nm and exhibited a descent stable pattern throughout the course of 100 ns simulation. Furthermore, we also assessed the stability of all the complexes with respect to ligand RMSD (Figure 3.8C), which showed that except for the complex Mepron, all the complexes exhibited a stable ligand RMSD value ranges from ~0.25-0.5 nm throughout the 100 ns course of simulation. In the case of ligand RMSD as well, we found the complex Lumacaftor as showing maximum stability with the RMSD value of ~0.25 nm. Therefore, the overall RMSD value suggests the complex Lumacaftor to be the most stable drug among all the six candidates subjected for MD simulation studies.

Moreover, we analyzed the effect of the drug binding on the dynamic features of residues of the apoprotein. For this, we calculated the RMSF values for the apoprotein as well as for all the six ligand complexes using the 'g_rmsf' utility of GROMACS for 100 ns of time scale. It can be observed from the figure 3.8D that the value of RMSF for all of the residues of apoprotein and its bound complexes ranged from ~0.1-1.0 nm for the entire course of simulation. Among all the complexes, the complex Lumacaftor showed the lowest

fluctuation. However, this result was further assessed through performing a comparative analysis of the effect of binding of drugs to the active site residues of the apoprotein. In this analysis, we also calculated the RMSF value of the active site residues of the apoprotein itself. From the figure 3.8E, it is evident that the RMSF values of the active site residues of the apoprotein and its bound forms show maximum fluctuation of ~0.25 nm for the entire time scale of 100 ns. The analysis of RMSF of the active site residues revealed that the complex Lumacaftor showed lesser fluctuation among all the other complexes as compared to the RMSF value of the active site residues of the apoprotein alone. This analysis brought us to a suggestion that Lumacaftor is stable enough among all the complexes to be analyzed further.

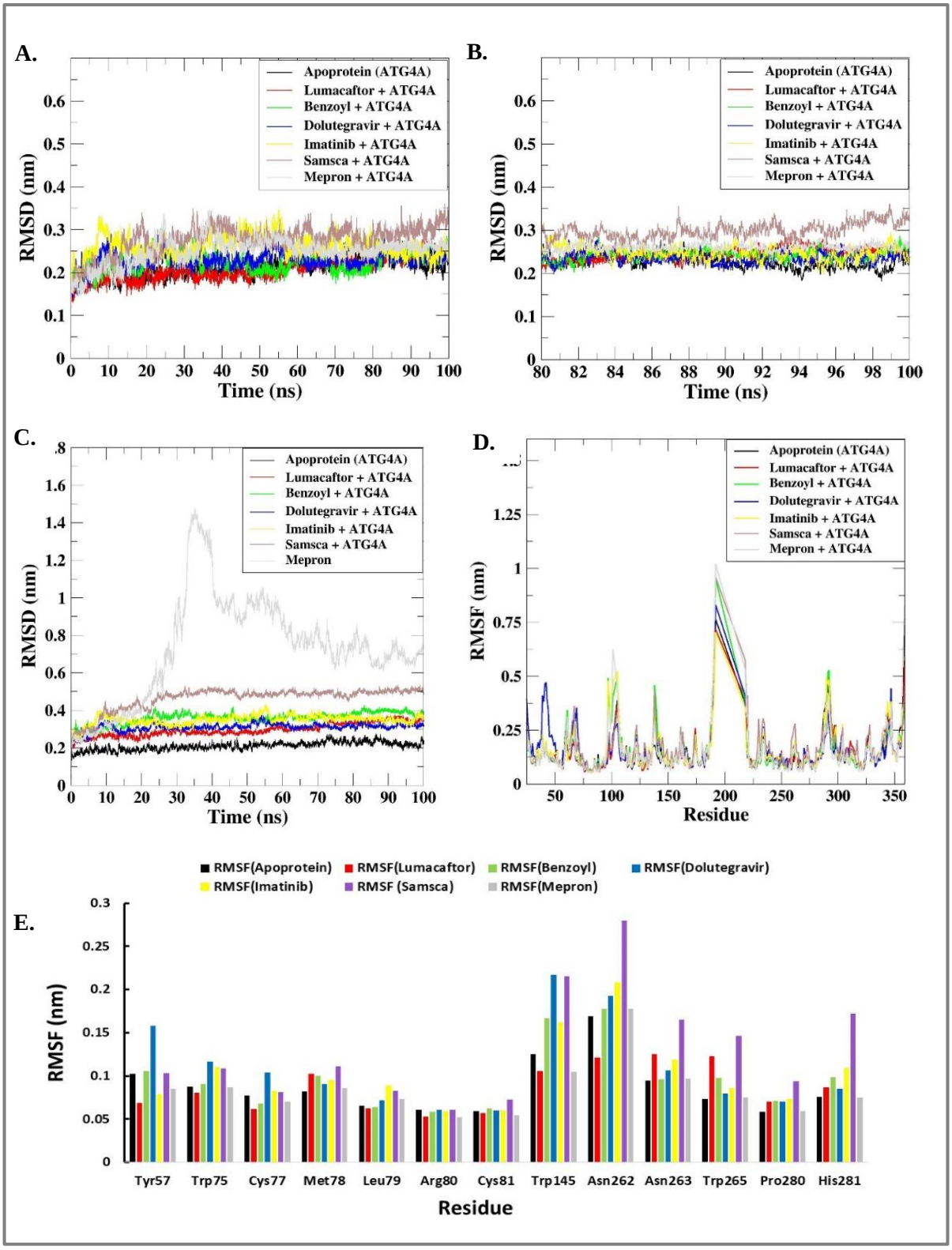

Figure 3.8. Analyses of MD simulation result with respect to RMSD and RMSF. **A.** Comparative analysis of backbone RMSD of all the protein-ligand complexes along with the apoprotein for a 100 ns trajectory. **B.** For an enlarged and clear view of the plot, the backbone RMSD is calculated from 80 ns to 100 ns for the protein as well as for all the protein-drug complexes. **C.** Comparative analysis of Ligand RMSD of all the protein-ligand complexes along with the apoprotein for a 100 ns trajectory. **D.** A comparative analysis of the RMSF values of all the six drug complexes along with the protein alone for a 100 ns trajectory. **E.** A comparative analysis of the RMSF values with respect to the active site of all the six drug complexes along with the protein alone for a 100 ns trajectory.

In further analysis, we evaluated the intermolecular hydrogen bond formation for all six complexes in the entire course of simulation of 100 ns. The hydrogen bonds contribute to the overall stability of the structure of the protein. It can be clearly visualized from the Figure 3.9A that on an average two, while a maximum of three intermolecular hydrogen bonds were seen to be formed in the complex Lumacaftor. Moreover, the complexes viz. Benzoyl (Figure 3.9B), Dolutegravir (Figure 3.9C), Imatinib (Figure 3.9D), Samsca (Figure 3.9E), and Mepron (Figure 3.9F) showed an average of two, four, two, three, and three while a maximum of three, five, three, four, and five intermolecular hydrogen bonds respectively.

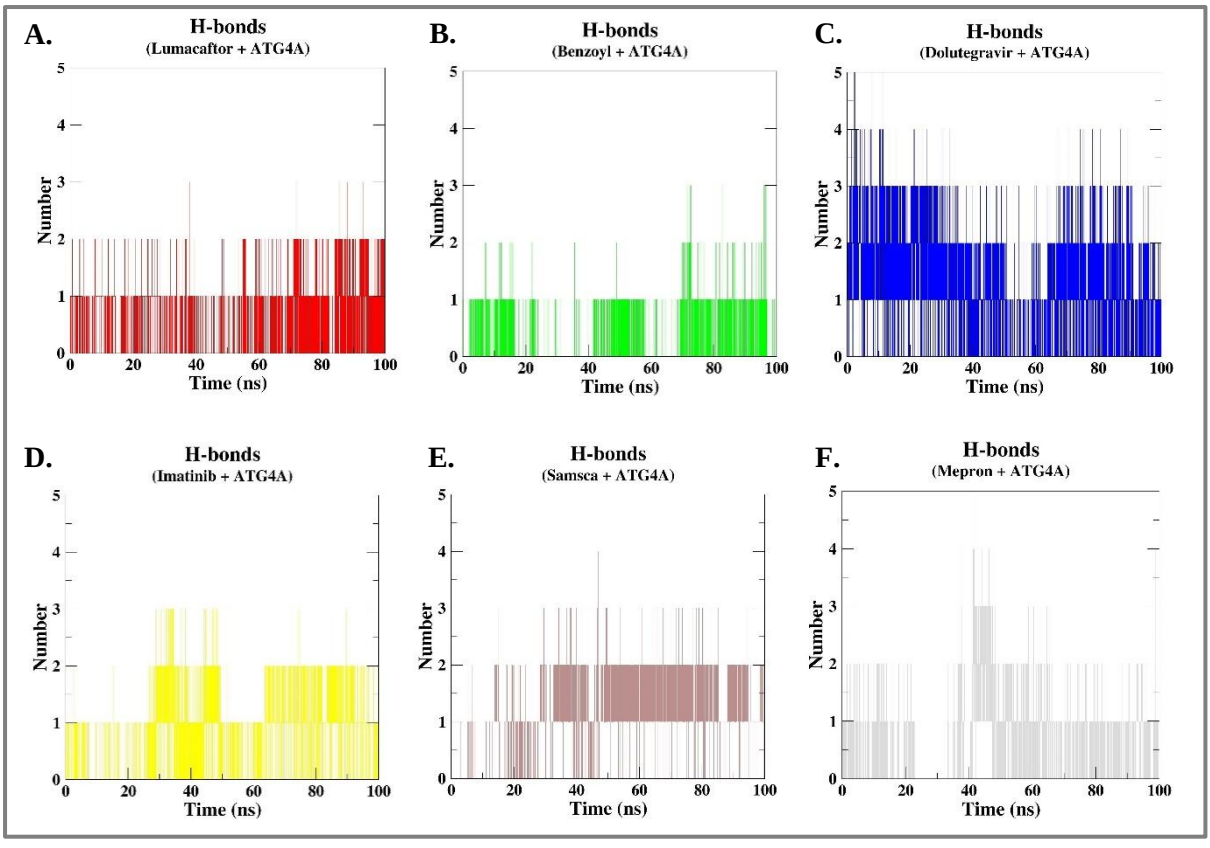

Figure 3.9. Number of H-bonds between the protein-drug complexes. These plots depict the number of hydrogen bond formed between the protein-drug complexes viz. A. Lumacaftor, B. Benzoyl, C. Dolutegravir, D. Imatinib, E. Samsca, and F. Mepron during the time scale of 100 ns.

Interestingly, it was observed that the residues Tyr57, Trp145, Asn262, Asn263, and Tyr265 were the most consistent residues found in almost all of the protein-ligand complexes during the process of molecular docking as well as MD simulation, which were contributing maximum to the intermolecular hydrogen bond formation, suggesting that these are the key residues which imparts the maximum stability to the structure of the protein-ligand complexes. Interestingly, it was also observed that these key residues also fall in the active site of the protein, hence suggesting that these residues are very crucial in making favorable interactions with the drugs, and hence stabilizing the overall structure of the drug complex.

Furthermore, we also analyzed the level of compaction in the structure of apoprotein as well as its all six bound complexes. For this, we calculated the radius of gyration along with the solvent accessible surface area (SASA) using the 'g_gyrate' and 'g_sasa' modules of GROMACS respectively. From the Figure 3.10A, it can be seen that the value of Rg, ranges between ~1.9-1.95 nm, showed a significant stable pattern for the apoprotein as well as for all the protein-ligand complexes throughout the 100 ns course of simulation. With a clearer view (Figure 3.10B), it can be observed that the Rg value of the complex Lumacaftor exhibited a lower value than that of the apoprotein alone, suggesting that upon binding of the complex Lumacaftor, the level of compaction gets increased as compared to unbound state of the apoprotein. Further, Figure 3.10C showed the change in the value of SASA for the apoprotein and its bound compounds with the course of 100 ns simulation. The plot (Figure 3.10D) reveals the value of SASA for the complex Lumacaftor (149.6298 nm^2) to be lesser as compared to that of the apoprotein (152.3086 nm^2) alone, however, rest all of the ligands viz. Benzoyl (153.5445 nm^2), Dolutegravir (155.1425 nm^2), Imatinib (154.751 nm^2), Samsca (154.8778 nm^2), and Mepron (155.1064 nm^2) showed higher SASA value as compared to the that of the value for apoprotein. Therefore, the results generated after Rg analysis and SASA, suggest the drug Lumacaftor to be consistently higher stable than that of all the other

compounds in the ranking. Further, the post-simulation-dependent variations in the overall binding energy of the drug complexes were analysed with the help of the MM-PBSA method.

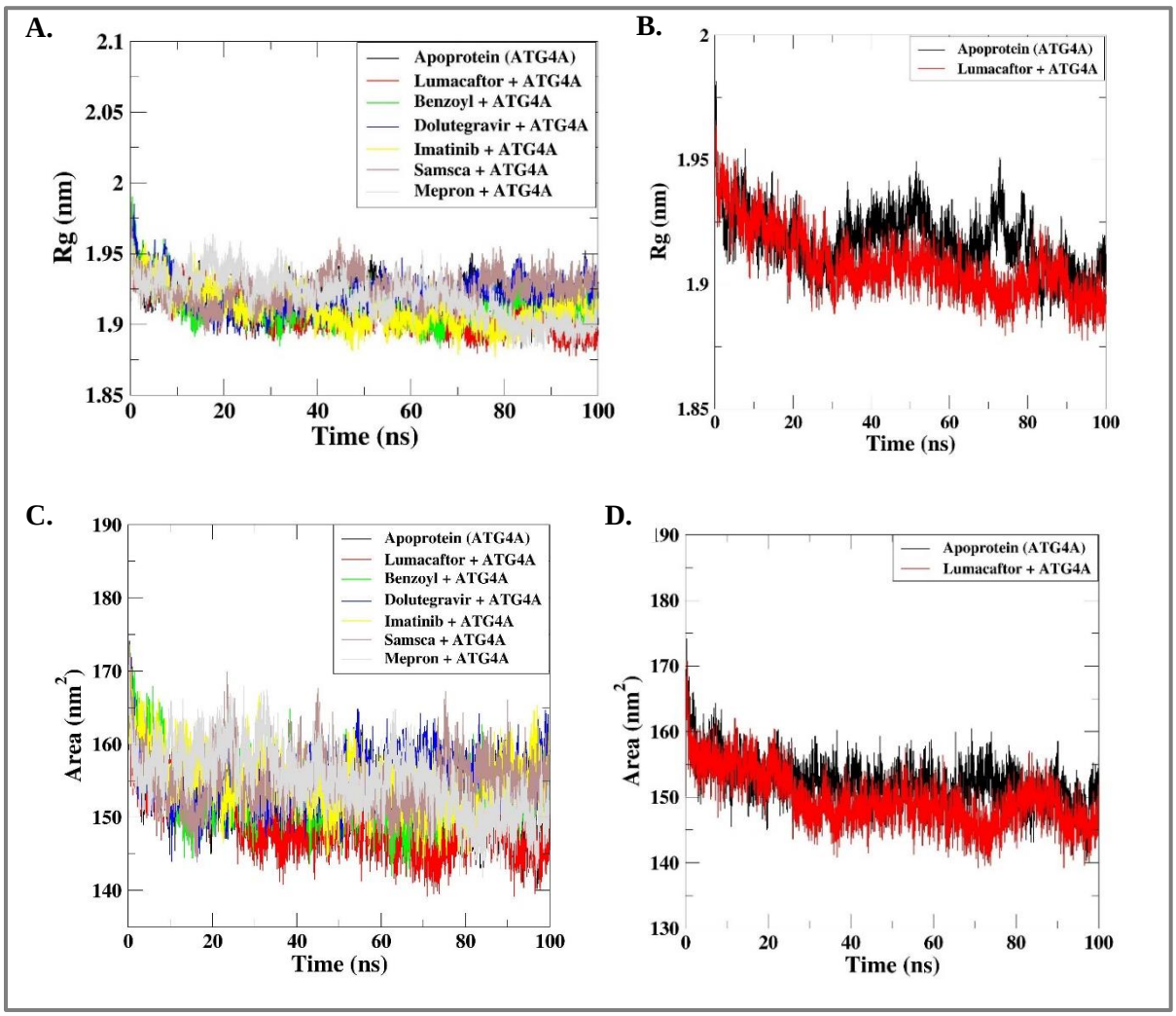

Figure 3.10. Analyses of MD simulation result with respect to Rg and SASA. A. Comparative analysis of Rg values of all the protein-ligand complexes along with the apoprotein for the time scale of 100 ns. **B.** A comparative analysis of the Rg value of the apoprotein with that of the drug complex Lumacaftor. **C.** A comparative analysis of SASA values of all the protein-ligand complexes along with the apoprotein for the time scale of 100 ns. **D.** comparative analysis of the SASA value of the apoprotein with that of the drug complex Lumacaftor.

3.3.3 Analysis of Average Binding Energy through MM/PBSA Method: Further, post-simulation, the average binding energy was calculated to evaluate the details of the energetic components of the biomolecular complexes using the 'g_mmpbsa' tool. The energy terms

viz. E_{MM}, G_{polar}, and $G_{nonpolar}$ for all the ligand complexes were analysed with the help of 201 snapshots extracted from the production trajectories at the time scale of every 0.1 ns from 80-100 ns course of simulation (Table 3.3). The comparative details of all the complexes given in Table 3.3 clearly showed the complex Lumacaftor to be acquiring of highest negative binding energy of -99.201 ± 1.156 kJ/mol. Followed by the complex Lumacaftor, the drugs Samsca, Imatinib, Benzoyl, Dolutegravir, and Mepron exhibited the average binding energy of value -70.184 ± 2.529 kJ/mol, -70.139 ± 3.914 kJ/mol, -43.672 ± 2.866 kJ/mol, -41.701 ± 2.657 kJ/mol, and -32.475 ± 1.372 kJ/mol respectively. Other than binding energy component of all thedrug complexes, Van der Waal energy, electrostatic energy, polar solvation and SASA energy are also mentioned in the Table 3.3 for the detailed investigation into the different energetic estimators of the biomolecular complexes. Furthermore, both the energetic components, Van der Waal energy and SASA energy revealed the complex Lumacaftor to be comprising of highest negative energy. Therefore, altogether with the aforementioned factors, we can suggest the complex Lumacaftor to be the most stable drug among all the drug complexes.

Table 3.3. Detailed comparative analyses for all the energetic components formed post-simulation between the protein-drug complexes. The values for various energy terms along with their standard deviation (SD) were calculated by MM/PBSA method.

Complex	**Binding energy (kJ/mol)**	**Vander Waal energy (kJ/mol)**	**Electrostatic energy (kJ/mol)**	**Polar solvation energy (kJ/mol)**	**SASA energy (kJ/mol)**
Samsca	-70.184 ± 2.529	-105.342 ± 3.634	-20.138 ± 1.013	66.444 ± 2.660	-11.141 ± 0.385
Imatinib	-70.139 ± 3.914	-124.034 ± 0.880	-16.066 ± 0.462	82.203 ± 4.001	-12.357 ± 0.161
Lumacaftor	-99.201 ± 1.156	-146.765 ± 1.263	-11.328 ± 0.429	72.509 ± 0.984	-13.634 ± 0.123
Mepron	-32.475 ± 1.372	-50.029 ± 1.663	-16.025 ± 1.151	39.695 ± 1.847	-6.177 ± 0.175
Benzoyl	-43.672 ± 2.866	-58.646 ± 3.940	-2.576 ± 0.235	23.382 ± 1.775	-5.867 ± 0.399
Dolutegravir	-41.701 ± 2.657	-79.140 ± 2.198	-4.600 ± 0.418	51.648 ± 2.487	-9.619 ± 0.210

Moreover, the residues of the drug complexes contributing maximum to the average binding energy were also analysed using the same module of 'g_mmpbsa'. This will give a clearer insight into the binding interaction of the biomolecular complexes. Figure 3.11A represents the contributor residues viz. Tyr265, Met309, and Leu314 of the complex Lumacaftor which impart their maximum to the average binding energy. Similarly, in the complex Benzoyl (Figure 3.11B), the residues Cys125, Phe161, Trp164, and Asn165 impart maximum contribution to the average binding energy. Furthermore, the drug complex Dolutegravir (Figure 3.11C) exhibits the contributor residues Tyr57, Ala73, and Trp145 which had shown the maximum contribution to the average binding energy. The another complex Imatinib (Figure 3.11D) showed Pro69, Trp75, Glu144, and Trp145 as contributor residues in providing maximum contribution to the average binding energy. Similarly, the complex Samsca (Figure 3.11E) exhibited Trp101, Thr299, and Leu303 as the residues which provide their maximum contribution to the average binding energy. The Figure 3.11F represents the complex Mepron showing Gln107, Pro108, and Glu291 as the contributor residues to the average binding energy. The result suggests that there are a few key residues viz. Trp75, Cys125, Trp145, and Tyr265 which are found to be consistent throughout most of the drug complexes in imparting their maximum contribution to the average binding energy of the biomolecular complexes. Moreover, among all the aforementioned key residues, Trp145 and Tyr265 of the drug complex Lumacaftor, Dolutegravir, and Imatinib lie in the active site and have been found constantly present in all the computational approaches used for the current study i.e., molecular docking, post-simulation intermolecular hydrogen bond formation, and MM/PBSA method, suggesting that these key residues may have immense potential to make favourable interactions with the drug.

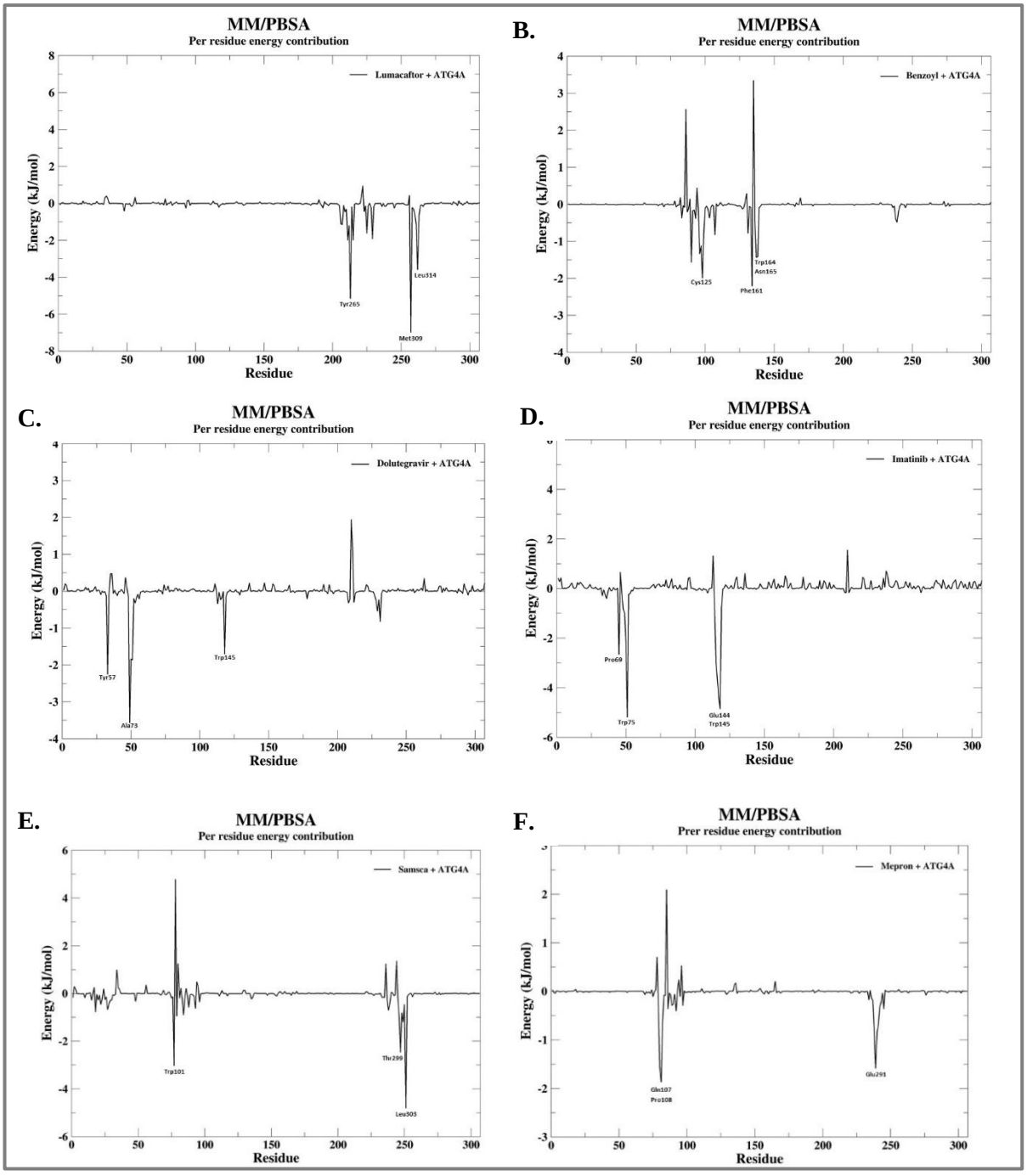

Figure 3.11. The plots depict the contribution of individual residues of each of the six protein-drug complexes in imparting their maximum contribution to the average binding energy. The calculation was done through the MM/PBSA method. The residues contributing maximally to the average binding energy are labelled in each plot.

3.4 Conclusion

Recent computational studies have greatly increased our knowledge and helped to develop novel potential anti-autophagy or anti-cancer drugs through screening the already available FDA-approved drugs. However, earlier studies have already reported a few of the ATG4A inhibitors, but in effort to increase the chemical discovery space in search of novel autophagy inhibitors, our current study focuses on inhibiting the activity of the protein ATG4A using a large pool of FDA-approved drugs from the ZINC database. Among the large number of FDA-approved drugs, we have sorted six drugs viz. Lumacaftor, Benzoyl, Dolutegravir, Imatinib, Samsca, and Mepron. On the basis of considering the parameters RMSD (both, backbone RMSD and ligand RMSD), RMSF (with respect to both, the whole residues of the drug complex and only the active site residues of the drug complex), Rg, and MM/PBSA, the drug complex Lumacaftor have been found constantly better performed and at first position in the drug complex ranking. Hence, our study strongly suggest that the drug Lumacaftor can be the potent regimen candidate to inhibit the function of ATG4A and can be used as a remarkable drug for further studies. Our current study can be marked as a contribution in developing more effective and specific valuable drugs that could potentially inhibit the autophagy associated ATG4A protein, which thereby will potentiate the drug discovery in the field of cancer biology.