

Chapter 3: Preliminary screening of metabolites

3.1 Introduction

Epidermal Growth Factor Receptor (EGFR) is a tyrosine kinase family's 170 kDa transmembrane glycoprotein. It is a cell surface receptor containing three parts: an intracellular tyrosine kinase, an extracellular ligand-binding domain, and a transmembrane anchoring region (Sepay et al., 2022; P. Singh & Bast, 2014). It binds to the epidermal growth factor, triggering dimerization and autophosphorylation of tyrosine residues, which activate a cascade of intracellular signal transduction, and promote cell proliferation, angiogenesis, resistance to apoptosis, cellular invasion, and metastasis (Harari et al., 2007).

Mutation or overexpression of the EGFR gene causes different cancer types, including breast cancer, prostate cancer, ovarian cancer, and NSCLC (non-small-cell lung cancer) (Di Lorenzo et al., 2002). Five amino acids (codons 746 to codon 750) are responsible for the point mutation of EGFR at the L858 position (catalytic domain of EGFR) and cause autophosphorylation and unregulated cell growth. Gefitinib and erlotinib (first-generation EGFR tyrosine kinase inhibitors) are reversible inhibitors that compete with ATP (Adenosine triphosphate) to bind at the catalytic site of EGFR and cause hindrance of autophosphorylation and cell proliferation (X. Lu et al., 2018). The need for inhibitors keeps increasing; consequently, the effectiveness of first-generation inhibitors was hindered due to resistance raised for these inhibitors, which diminished the clinical response over a year. Substitution of threonine 790 with methionine, EGFR^{T790M} (EGFR-L858R/T790M), is one of the most frequent resistance issues against first-generation tyrosine kinase inhibitors and causes a reduction in their inhibitor's activity and prevents their interaction at the catalytic domain (Hata et al., 2012). Approximately 50% of NSCLC-affected patients with mutated EGFR^{T790M} (EGFR-T790M) were resistant to

first-generation tyrosine kinase inhibitors. Thus, the alteration in EGFR at T790M established the development of resistance against first and second-generation tyrosine kinase inhibitors and provoked the finding of inhibitors of the third-generation kinase. Neratinib, Dacomitinib, and Afatinib are second-generation kinase inhibitors that interact covalently with the Cys₇₉₇ residue at the binding pocket of ATP, and their clinical application is restricted because of side effects such as gastrointestinal problems and skin diseases (Sos et al., 2010). Third-generation kinase inhibitors showed significant efficiency in NSCLC-affected patients possessing resistance to first/second-generation inhibitors. Osimertinib is a Food and Drug Administration (FDA)-approved third-generation inhibitor, expressing effective results in NSCLC patients (A. Singh et al., 2022). Further, the development of fourth-generation kinase inhibitors comes into the limelight with the emergence of resistance to third-generation kinase inhibitors. The third-generation EGFR tyrosine kinase inhibitors, i.e., EGFR^{CSTMLR} (EGFR-L858R/T790M/C797S), resistance developed by converting cysteine to serine at 797, which is less reactive and inhibits the interaction between inhibitors and cysteine amino acid (Nagano et al., 2018). Numerous side effects are also associated with third-generation inhibitors, so an alternative drug based on natural sources, having fewer side effects and low cost, needs to be discovered (Nagasaka et al., 2021).

In this regard, *Streptomyces* can be a perfect alternative for discovering EGFR tyrosine kinase inhibitors due to their greater bioavailability, bountiful sources of secondary metabolites, biotic friendliness, lesser side effects, non-toxic and highly effectiveness (Michels et al., 1998). Today, the genus *Streptomyces* is responsible for developing 80% of antibiotics, making it the most important genus in medication discovery against various diseases (Kapadia et al., 2007). They also create complex secondary metabolites such as enzymes, protein kinase inhibitors, apoptosis inducers, and caspase-3 activators, which

are helpful in cancer treatment (Procópio et al., 2012). This chapter aimed to identify new EGFR^{CSTMLR} tyrosine kinase inhibitors from the StreptomeDB database. Initial screening of metabolites was conducted to identify promising candidates by utilizing molecular docking and a dynamics simulation study.

3.2 Materials and Methods

3.2.1 Retrieval and processing of ligand database

StreptomeDB is a curated database of Isolated or mutasynthesized Natural Products (NPs) from *Streptomyces* (Lucas et al., 2012). The StreptomeDB 3.0 is an updated database with 6524 NPs from 3302 actinomyces species (Moumbock et al., 2021). The 3-Dimensional structures of these 6524 NPs were downloaded in SDF file format from <http://www.pharmbioinf.uni-freiburg.de/streptomedb>. All the NPs and reference molecules were converted into mae file format for easy processing by utilizing the LigPrep module of Schrödinger (Schrödinger, LLC, New York, USA, 2022). LigPrep module windows were set to OPLSe (Optimized Potentials for Liquid Simulations) forcefield, ion neutralization, and single isomers generation per ligand (Madhavi Sastry et al., 2013). Reference structure CH7 (CH7233163) and co-crystal ligand Osimertinib of the target were used as the standard reference for the study (Kashima et al., 2020; Soria et al., 2018).

3.2.2 Retrieval and Processing of Target

A 3-D Crystal Structure of EGFR^{CSTMLR} was downloaded from the RCSB-PDB (RCSB Protein Data Bank) with PDB id: 6LUD (Kashima et al., 2020). The receptor was already co-crystallized with a well-known anti-Tyrosine Kinase inhibitor: Osimertinib. The structure was then imported into the Protein preparation wizard of Schrödinger to rectify the problems in protein structure. The protein was further refined by optimizing hydrogen bonds and minimizing its structural energy with the OPLSe force field (Madhavi Sastry et al., 2013; Shelley et al., 2007). The receptor grid generation module in Glide-

Schrödinger creates a grid box over the ligand with a Van-der-Waals scaling factor set to 1.0, and a partial charge cutoff is set to 0.25.

3.2.3 ADME analysis of ligand library

ADME is a test that establishes the absorption, distribution, metabolism, and excretion characteristics of a potential principal compound and helps study its disposition within an organism (Gleeson, 2008). ADME was performed by incorporating the QikProp program of Schrödinger running in standard default settings. This program helps generate the relevant descriptor (set of 46 molecular descriptors) and uses selected descriptors to carry out ADMET predictions. Drug-likeness of a lead depends on various parameters, and they were used to estimate the pharmacokinetic analysis of the StreptomeDB ligand library. The descriptors used for compound evaluation include several physicochemical and pharmacokinetic parameters. Rule of Five and Rule of Three are guidelines for drug-likeness based on molecular weight, hydrogen bond donors/acceptors, and lipophilicity. PSA (Polar Surface Area), WPSA (Water-accessible PSA), FISA, FOSA, and PISA represent surface area components relevant to solubility and permeability. %Human Oral Absorption predicts oral bioavailability. QLogPo/w, QLogP, and QLogPoct denote log partition coefficients indicating lipophilicity across different phases, while QLogPw and QLogS indicate water solubility and aqueous solubility, respectively. QLogBB estimates blood-brain barrier penetration, QLogHERG estimates the potential for cardiotoxicity, and QLogKhasa reflects serum protein binding affinity. EA(eV) (Electron Affinity) and IP(eV) (Ionization Potential) reflect electronic properties influencing reactivity. #metab predicts the number of metabolic reactions. donorHB and acceptHB indicate hydrogen bond donors and acceptors. volume, mol_MW (Molecular Weight), and dipole are basic physicochemical descriptors, while #rotor indicates molecular flexibility through rotatable bonds. glob relates to molecular globularity, and Qpolrz represents

polarizability. QLogPC16 and Qpolrz estimate partitioning in specific media or electronic response. SASA (Solvent Accessible Surface Area) and Total Compounds are structural coverage indicators, while Final Selected likely denotes compounds retained after all filtering steps. Selected 26 descriptors that play a significant role in most drug-likeness criteria were chosen on the ligand library, making them match 95% of known drug ranges

3.2.4 Molecular Docking Simulations

Molecular docking was performed using Glide software from the Schrödinger suite with Standard-Precision protocol (SP), with ten poses included in post-docking minimization, keeping all other parameters default. Re-docking, the co-crystallized ligand with a suitable RMSD (Root Mean Square Deviation) value (less than 2 Å), enabled to verify the docking procedure. The re-docked poses gave away similar interactions compared to the active site with the native pose. The above experiment leads to the inference that the docking protocol is reliable and sustainable for predicting plausible inhibitors in the StreptomeDB database. For further evaluation, the compounds were selected based on the obtained glide score, binding energy, and analysis of the binding pattern of the ligand-protein complex. The PAINS (pan assay interference compounds) was also applied to the selected compound using the Knime Analytics Platform with PAINS, PAINS-A, PAINS-B, and PAINS-C filter criteria of RDKit Molecule catalog filter node (Saini et al., 2023a).

3.2.5 Molecular Dynamics Simulations

The enzyme-inhibitor complexes that had been docked were considered as the initial framework for future simulation runs. The Schrödinger Suite's Desmond simulation tool was used to run the MD (Molecular Dynamics) simulations (Schrödinger, LLC, New York, USA, 2022). The orthorhombic box was solvated using the TIP3P (transferable intermolecular potential with 3 points) solvation/aqueous system, and the solvent buffer was expanded around the protein by 10 Å using the OPLS4 force field (C. Lu et al., 2021). The system was then neutralized and maintained at 0.15 M salt concentration by placing

the required NaCl counter ions. Before employing the simulation study, Desmond's pre-defined system relaxation algorithm was used to equilibrate the system, consisting of six phases in the protocol. The first two were set for the NVT simulation at 10 K temperature, where the first phase was carried out for 100 ps (picosecond) using the Brownian dynamics method to restrict the motion of heavy atoms present in the protein molecules, while in the second stage, the restriction was applied on the large atom of solute for next 12 ps. After that, the NPT condition was applied in the third step by restricting the atom of solute again at 10 K for 12 ps, and the constant pressure was almost one bar. The temperature was likely to be enhanced up to 310 K in the fourth stage with 10 ps time at NPT conditions for relaxing the protein molecules and possessing no restriction on the system. The final simulation time was 1,000 ns (nanoseconds), with a 1 ns recording interval. All simulations used the OPLS4 force field parameters. To calculate the MM/GBSA (Molecular Mechanics and Generalized Born-Surface Area), representing the following equation, the prime module of Schrödinger was used at the last 100 ns of the MD trajectories:

$$\Delta G_{bind} = \Delta G_{comp} - [\Delta G_{protein} + \Delta G_{lig}]$$

Where ΔG_{bind} represents the free binding energy, $\Delta G_{protein}$ is the binding energy of the target protein, and ΔG_{lig} represents the binding energy of the ligand in the protein-ligand complex.

3.3 Results and Discussion

3.3.1 ADMET Properties Predictions

It has been suggested that the compounds used as human therapeutic agents should have drug-likeness properties or exhibit good ADMET profiles (Yousuf et al., 2017). The ADME properties of 6524 compounds were ascertained using the QikProp web service against selected 25 molecular descriptors criteria. QikProp program predicted that the

potential to penetrate the blood/brain barrier and bind with human serum albumin was portrayed by 71.8% and 85% virtual hits. Ro5 (Lipinski's rule of five) for screening drug-like compounds showed that 76% of molecules retained physicochemical properties and followed Lipinski's parameters. In addition, the Ro3 (Ghose rule of three) was applied to enhance the ADME prediction and observed that 67% of molecules were passed through it.

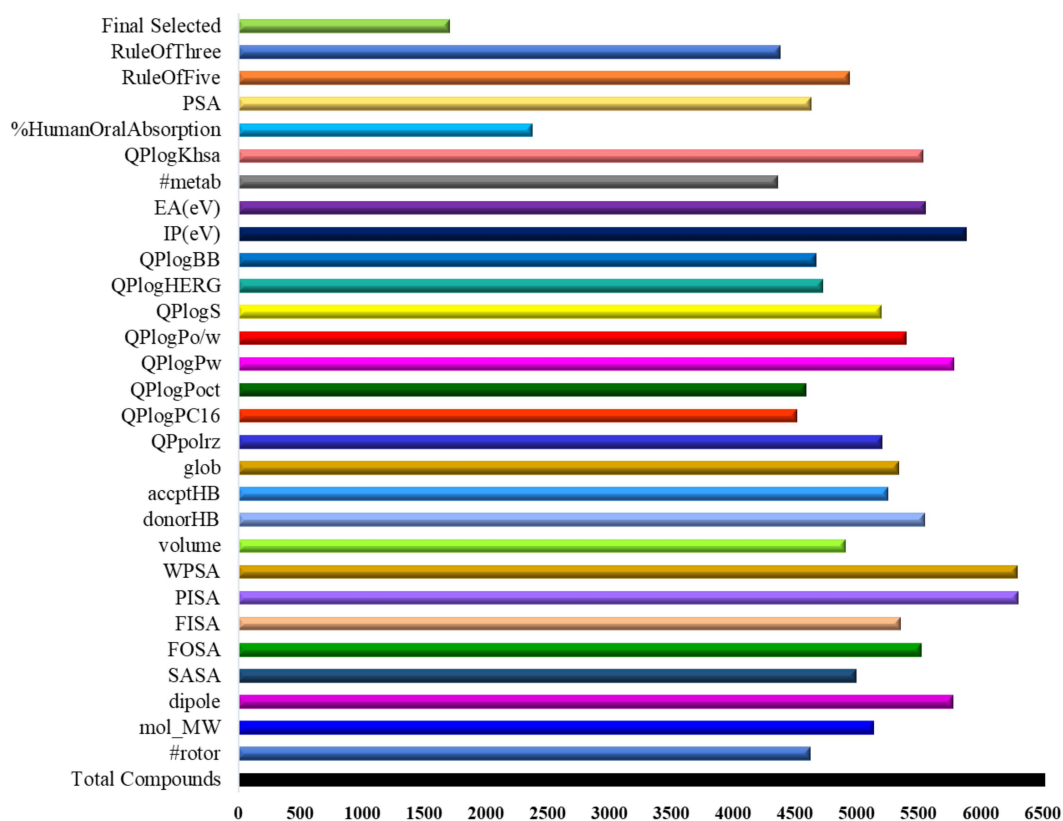


Figure 3.1: ADMET profile of 6524 compounds from the StreptomeDB 3.0 database.

Further, the potassium ion (K^+) channel linked to the Human ether-a-go-go-related gene (HERG) is responsible for long QT syndrome (Hedley et al., 2009). This channel plays a vital role in regulating heartbeats, making it a vulnerable target to the potential therapeutic drug for cardiac toxicity (Vandenberg et al., 2001). So, the value of predicted IC_{50} was calculated for the inhibitory action on the HERG K^+ channel by using drugs to model the

effect of drug toxicity, and it was found that both the standard drugs lie far outside the safe threshold. 1708 (26.2%) compounds depicted a remarkable result by effectively passing through all the ADMET parameters and were selected for anticancer activity analysis by being subjected to the molecular docking study with the target protein EGFR-tyrosine kinase. The ADME properties of all 6524 compounds are presented in Figure 3.1. All the compounds passed through the above criteria lie in the zero-violation category of Ro3 and Ro5 (except C_5024 with only one violation in Ro3), whereas both Osimertinib and CH7 have one violation in Ro3, 1 and 3 violations in Ro5, respectively.

3.3.2 Molecular Docking Study of triple mutant EGFR^{CSTMLR}

Molecular docking of 1708 compounds at the active site of mutant EGFR^{CSTMLR} leads to the identification of a novel potential drug for anticancer. Table 3.1 shows the hit compound's binding score, hydrogen, and hydrophobic bond-forming amino-acid residues.

Out of 1708 compounds, five were screened out, which are C_42: Steptocarbazole A, C_4299: Limazepine H, C_4300: 3'-N-Formylholyrine A, C_5024: CHEMBL1159781, and C_5702: 3'-N-Acetylholyrine A, revealing a better binding score than reference inhibitor CH7: CH7233163 and co-crystallized ligand Osi: Osimertinib (well-known Tyrosine inhibitor). Duration of the interaction, it was found that these selected molecules have also passed all criteria of the PAINS filter as well (De Vivo et al., 2016). The analysis of triple mutant EGFR^{CSTMLR} protein with interacting ligands showed that MET_793 amino acid residue forms a hydrogen bond with ligands in all five hits ligands- EGFR complex, like co-crystallized ligand Osimertinib (Figure 3.2.1 to Figure 3.2.3). Thus, MET_793 interaction can be referred to as "vital" as it was conserved in all protein-ligand complexes (Balogun et al., 2021; Mahajan et al., 2017).

Table 3.1: Binding energy and molecular interaction analysis (Hydrogen and Hydrophobic interaction) ligands with triple mutant EGFR^{L858R/T790M/C797S} and their production source. The bold residues are the conserved one with control compounds. (†: name retrieved from StreptomeDB 3.0 database).

Ligands	Docking Score (Kcal/mol)	Hydrogen bond interaction	Hydrophobic interaction	Microbial source of production [†]	Compound ID
C_42 (Steptocarbazole A)	-9.3	GLN_791, MET_793, ASP_800	LEU_718, PHE_723, VAL_726, ALA_743, MET_790, LEU792, PRO_794, LEU_844	<i>Streptomyces sanyensis</i> FMA	PubChem 72501071
C_4299 (Limazepine H)	-9.164	LYS_745, GLU762, MET_793	LEU_718, VAL_726, ALA_743, MET766, MET_790, LEU_792, PRO_794, LEU_844, PHE856	<i>Streptomyces seoulensis</i> IFB-A01	PubChem 102449830
C_4300 (3'-N-Formylholyrine A)	-9.218	MET_793, SER_797, ASP_800	LEU_718, PHE_723, VAL_726, ALA_743, MET_790, LEU792, PRO_794, LEU_844	<i>Streptomyces sp.</i> NB-A13	Pmid 30268972
C_5024 (CHEMBL159781)	-9.074	LEU_718, MET_793, SER_797, ASP_800	PHE723, VAL_726, ALA_743, MET_790, LEU792, PRO_794, LEU_844, LEU1001, MET1002	<i>Streptomyces pulveraceus</i>	PubChem 44271362
C_5702 (3'-N-Acetylholyrine A)	-9.183	MET_793, SER_797, ASP_800	LEU_718, PHE_723, VAL_726, ALA_743, MET_790, LEU792, PRO_794, LEU_844	<i>Streptomyces coelicolor</i> M1146	PubChem 145973966
CH7 (CH7233163) Reference	-6.1148	-	VAL_717, LEU_718, PHE_723, VAL_726, ALA_743, MET_790, LEU_792, MET_793, PRO_794, PHE_795, LEU_844	-	Pmid 32943545
OSI (Osimertinib) Co-crystallized	-8.073	MET_793, SER_797	LEU_718, PHE_723, VAL_726, ALA_743, MET_790, LEU_792, PRO_794, PHE_795, TYR801, LEU_844, LEU1001	-	PubChem 71496458

In addition, all the protein-ligand complexes also shared the common hydrophobic bond-forming amino acid residues VAL_726, ALA_743, MET_790, LEU_792, PRO_794, and LEU_844. Out of five screened compounds from the StreptomeDB 3.0 database, ligand C42 exhibited the best docking score of -9.3 Kcal/mol. It is a derivative of staurosporine containing a carbazole ring, which made a hydrogen bond with MET_793 and Gln_791 in the connecting region of the C and N terminal of the protein. It formed another

hydrogen bond via the hydroxymethyl group of the pyrrole ring with ASP_800. The drug C_42 (Steptocarbazole A or Streptocarbazole A) was investigated for cytotoxic activity by Fu et al. (2012) and found effective against the HL-60, A-549, P388, and HeLa cell lines with IC₅₀ values of 1.4, 5.0, 18.9, and 34.5 μM, respectively. Additionally, its inhibitory effects on the cell cycle and protein kinase were also reported. The researcher also found that Steptocarbazole A stopped the cell cycle of the HeLa cell line in the G2/M phase at a 10 μM concentration (Fu et al., 2012). Ligands C4300 and C5702 have the same binding affinities and share a common bond interaction with the EGFR receptor because of their structural similarity. Only the presence of the formyl group in C4300 and the acetyl group in C5702 distinguishes them. Both comprise a carbazole ring bonded to MET_793 and an oxane ring that interacts with the SER_797 and ASP_800 amino acid residues of the EGFR receptor through its hydroxyl group. Carbazole moiety in C_42, C_4300, and C_5702 is a large-sized, tricyclic, strong pharmacophoric moiety that provides the best binding affinity with the receptor, which leads to the downregulation of protein kinase and eventually apoptosis of cancerous cells (Müller et al., 2021).

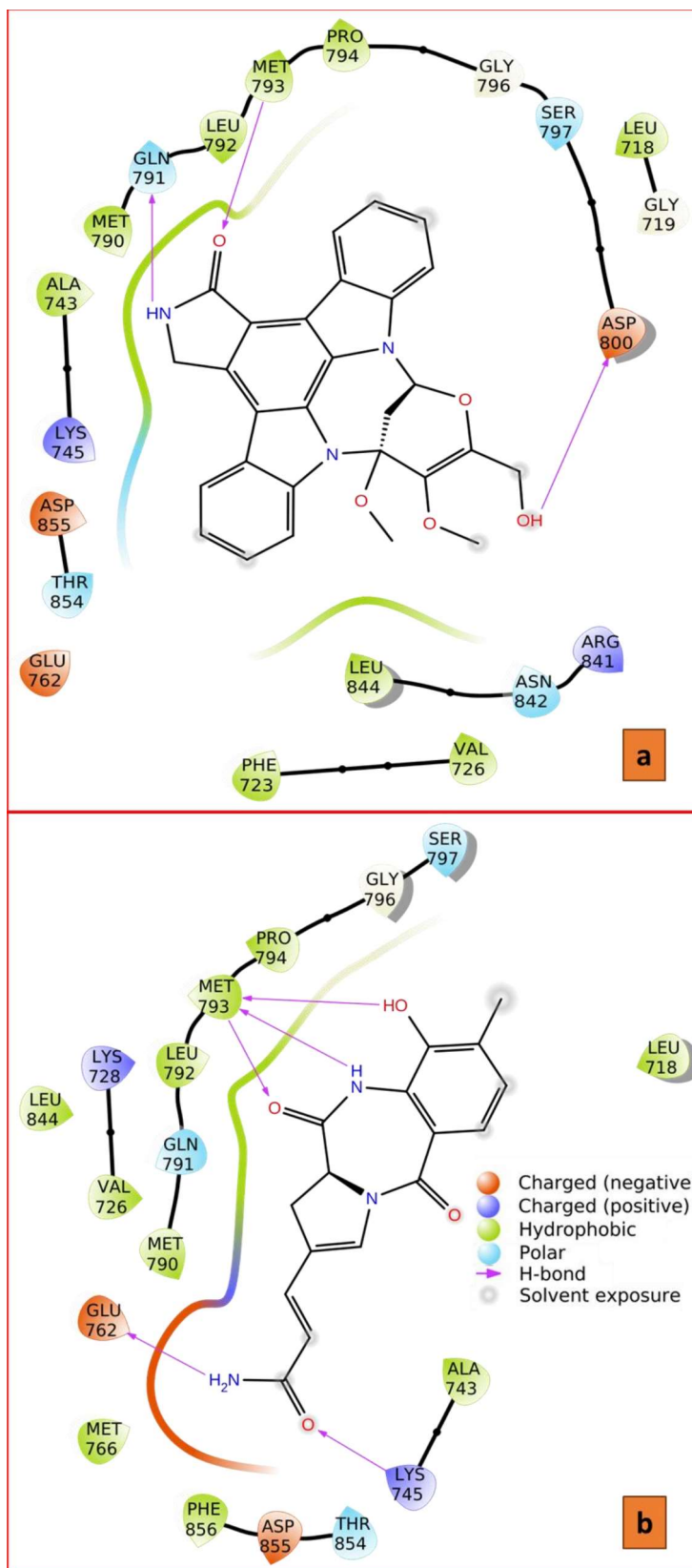


Figure 3.2.1: 2D plot showing the network of the interaction of all compounds representing a: C_42, and b: C_4299 with active site residues of EGFR^(L858R/T790M/C797S) mutant during molecular docking study.

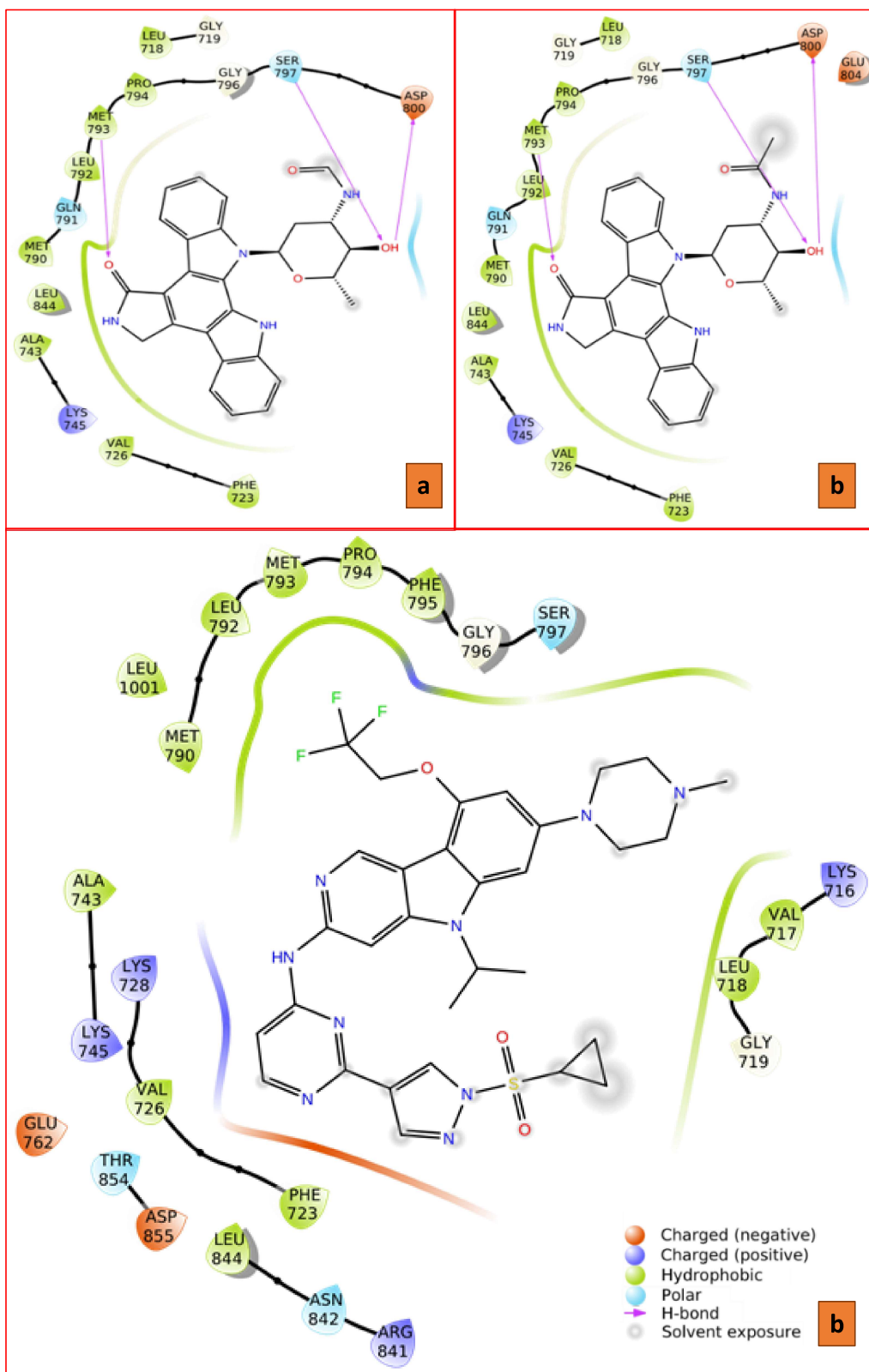


Figure 3.2.2: 2D plot showing the network of the interaction of all compounds representing a: C_4300, b: C_5702, and c: CH7, with active site residues of EGFR^(L858R/T790M/C797S) mutant during molecular docking study.

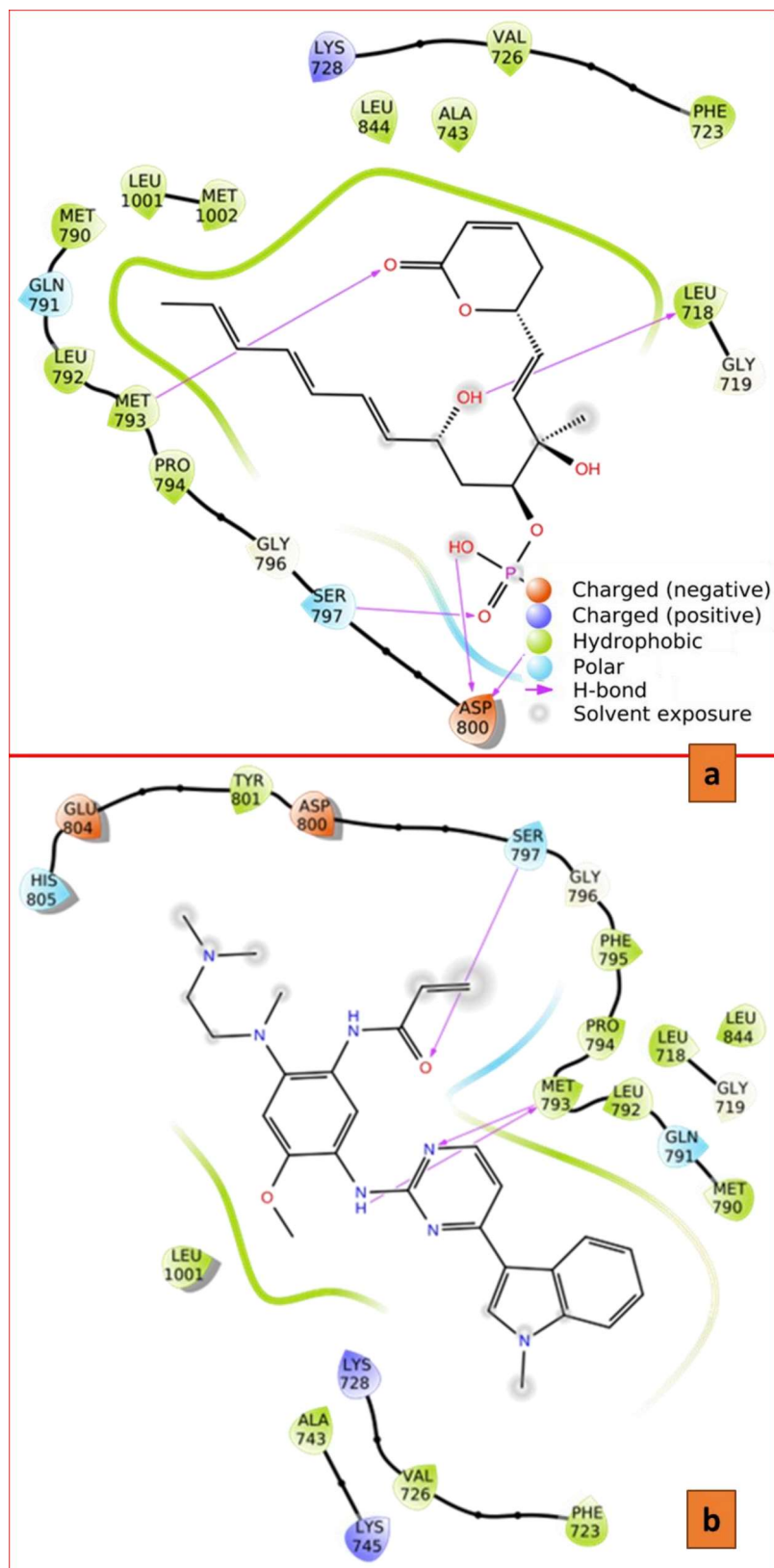


Figure 3.2.3: 2D plot showing the network of the interaction of all compounds representing a: C_5024, and b: Osi with active site residues of EGFR^(L858R/T790M/C797S) mutant during molecular docking study.

Zhou et al. investigated the cytotoxic activities of C_4300 (3'-N-Formylholyrine A) on the growth of PC-3 and SW-620 cell lines, with the cytotoxic activity of IC₅₀ values 2.50 and 0.73 μ M, respectively (Zhou et al., 2019). Xiao et al.(2018) found C_5702 (3'-N-Acetylholyrine A) compound to have cytotoxicity activity against the tumor cell lines HCT-116, K562, and Huh 7.5 as well as the normal hepatic cell line LO2 with IC₅₀ values as >100, 32.4, 35.2, and 41.1 μ M, respectively (Xiao et al., 2018). Wang et al. also tested the same compound for inhibitory action against protein kinase C enzymes θ (PKC θ) and displayed inhibitory activity with an IC₅₀ value less than 2.5 μ M (J. Wang et al., 2020). C5024 showed good interaction with protein by forming the maximum number of hydrogen bonds (four). It contains a phosphate group that interacts with SER_797 and ASP_800 residues in the ATP binding pocket, while its unsaturated lactone group makes interactions in the hinge region with MET_793, which helps in forming a strong bond with the protein. Its anticancer activity is accomplished by the presence of unsaturated lactone and phosphate ester moieties (Leopold et al., 1984). It is an analog of the anticancer drug Fostriecin, which is widely known for inhibiting protein phosphatase (McCluskey et al., 2002; Swingle et al., 2009).

This implies that the reported compound C5024 has both phosphatase and kinase inhibitory properties, suggesting dual action. The anti-tumor properties of C5024 (ChEMBL1159781) have been observed by McCluskey et al. on L1210 and HCT-8 Cell Lines, with GI₅₀ concentrations of 0.22 and 1.5 μ g/mL, respectively (McCluskey et al., 2002). Ligand C4299 possesses amide moieties which formed two hydrogen bonds with Glu_762 and Lys_745 amino acids and improved the binding affinity. Its benzodiazepine ring is oriented in the back of the ATP binding pocket, where it interacts with MET_793 via oxo and hydroxy groups, which act as donors and acceptors and lie in the hinge region of the protein kinase. Molecular docking analysis provides only a tentative result

regarding protein-ligand interaction and binding score, so before concluding, molecular dynamics simulation needs to be carried out to evaluate other factors like protein stability and compactness, solvent effects, and time

3.3.3 Molecular Dynamics Studies of the mutant protein

MD simulation was performed for ligands (C_42, C_4299, C_4300, 5024, and C_5702), reference drug CH7233163, and co-crystallized ligand Osimertinib to confirm their structural stability at the active site of the triple mutant protein EGFR^{CSTMLR} during 1,000 ns MD.

RMSD analysis of the EGFR^{CSTMLR} backbone atom confirmed that all the complexes revealed good stability (less RMSD) compared to the reference inhibitor CH7233163, as shown in Figure 3.3. In addition, the RMSD value for ligands C_42, C_4299, and C_5024 was comparable to co-crystallized ligand Osimertinib, while ligands C_4300 and C_5702 exhibited a lower value than Osimertinib. The average RMSD value for triple mutant EGFR + C_42, EGFR + C_4299, EGFR + C_4300, EGFR + C_5024, EGFR + C_5702, EGFR + CH7233163, and EGFR + Osimertinib was 3.25 Å, 3.57 Å, 2.18 Å, 3.22 Å, 2.88 Å, 4.09 Å, and 3.07 Å, respectively.

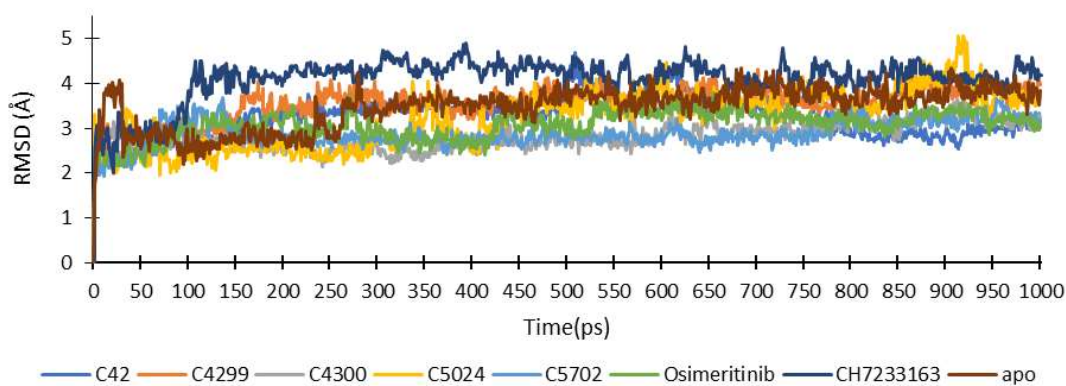


Figure 3.3: RMSD profile of triple mutant EGFR(L858R/T790M/C797S) protein complex with ligands during MD simulation of 1,000 ns.

Root mean square fluctuations (RMSF) denote the local changes in the protein chain around the ligand. The RMSF value of all the interacting residues during the simulation is presented in Figure 3.4. All the complexes revealed a comparable RMSF value with the co-crystallized ligand Osimertinib. It was found that the amino acids surrounding the ligand, GLN_791, MET_793, SER_797, and ASP_800, showed significantly less RMS fluctuations and remained stable during interactions. The average RMSF value of these amino acid residues, GLN_791, MET_793, SER_797, and ASP_800, was 0.66 Å, 0.68 Å, 0.74 Å, and 0.93 Å, respectively. The average RMSF value of ligand C_42 (1.54 Å) was very similar to the co-crystallized ligand (1.54 Å) and smaller than the reference inhibitor CH7233163 (1.71 Å). It proved that protein secondary structures like alpha-helix and beta sheets remain more rigid during simulation.

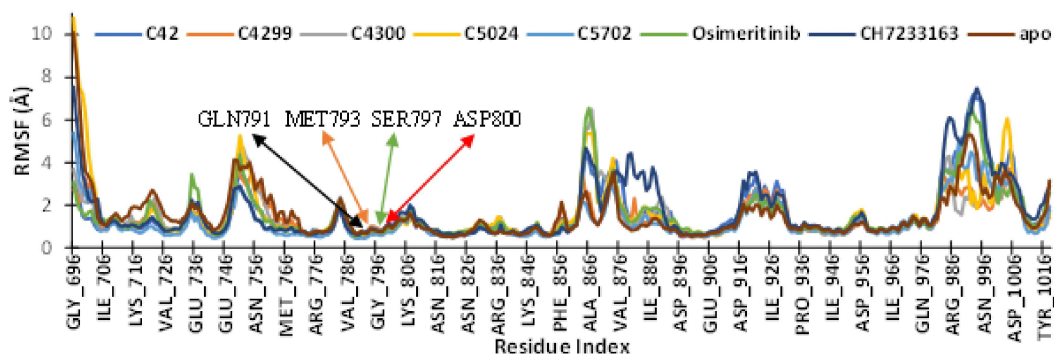


Figure 3.4: RMSF profile of triple mutant EGFR(L858R/T790M/C797S) protein's interacting residues with ligands during MD simulation of 1,000 ns.

The Y axis in Figure 3.5 depicts the R_g (Radius of Gyration) of ligands tells the extendedness of ligands during simulation time. It was assessed that ligand C_42 comprised the lowest radius of gyration value (3.87 Å) among all ligands, co-crystallized ligand, and reference drug, as shown in Figure 3.5, which illustrates that C_42 remained in a folded and rigid state during simulation because of its aromatic structure. R_g of ligand C_4299 (4.12 Å), C_4300 (4.17 Å), C_5024 (4.69 Å), and C_5702 (4.36 Å) was very comparable to co-crystallized inhibitor Osimertinib (4.49 Å). It was also observed that all

ligands exhibited a lower R_g value than the reference drug CH7233163 (5.88 Å) (Figure 3.5).

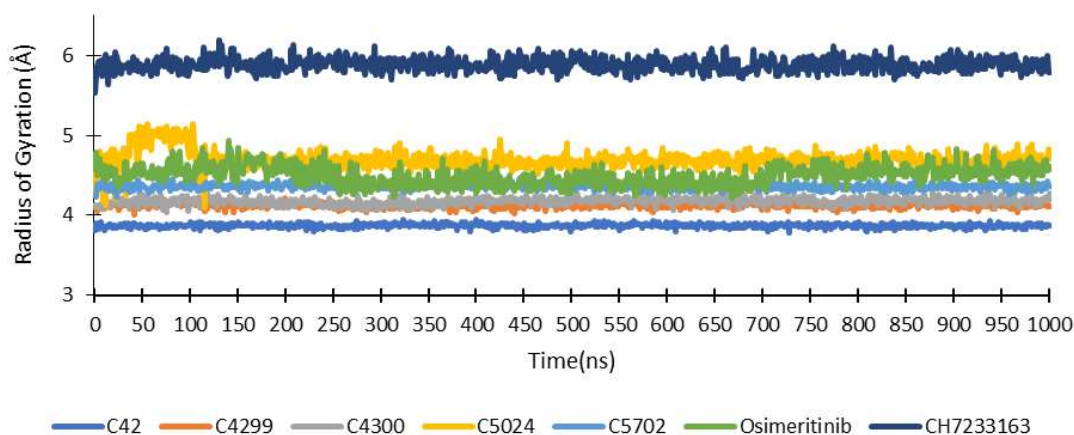


Figure 3.5: The radius of gyration of ligands during MD simulation of triple mutant EGFR(L858R/T790M/C797S) protein-receptor complex.

Solvent Accessible Surface Area (SASA) evaluates the part of the ligand surface area accessible by its surrounding water molecules or solvent. The plot of the SASA of all ligands during the MD simulation is presented in Figure 3.6. The average SASA value of ligand C_42 was the lowest (82.9 Å²), indicating that only a tiny part of the ligand is accessible to water. Ligand C_4299 (105.79 Å²), C_4300 (153.44 Å²), and C_5702 (123.0 Å²) comprised less SASA value than reference drug CH7233163 (222.38 Å²) and co-crystallized ligand Osimertinib (178.20 Å²), while ligand C_5024 had higher value (208.60 Å²) but were comparable to the reference drug (Figure 3.6).

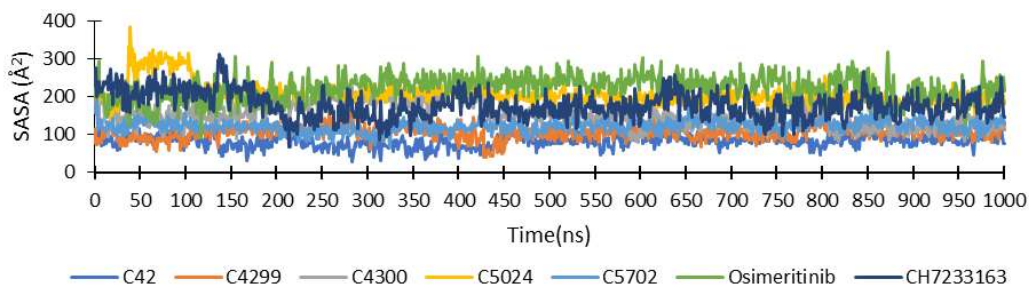


Figure 3.6: Solvent accessible surface area (SASA) of five-hit ligands, reference drug CH7233163, and co-crystallized ligand Osimertinib during MD simulation of 1,000 ns.

The number of hydrogen bonds formed by amino acid residues with all ligands during the simulation was assessed. On average, ligands C_42, C_4299, C_4300, and C_5702 formed six hydrogen bonds till 680 ns; after that, the value was enhanced up to eleven. However, ligand C_5024 shows an average of five hydrogen bonds throughout the simulation, similar to reference drug CH7233163 and co-crystallized ligand Osimertinib, as presented in Figure 3.7.

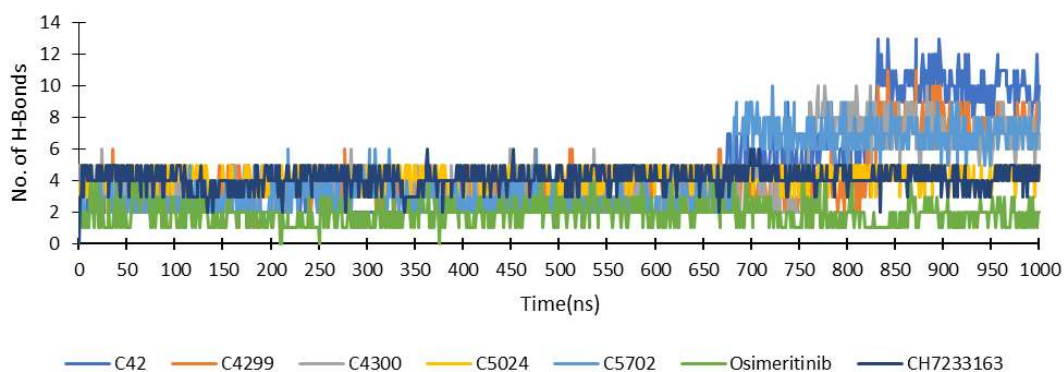


Figure 3.7: Hydrogen bonds formed between protein EGFR^{L858R/T790M/C797S} and ligands during MD simulation of 1,000 ns.

MD simulation results also revealed that all the ligands form a hydrogen bond with protein residue MET_793, like reference ligand CH7233163 and co-crystallized ligand Osimertinib with 100% occupancy; in addition, ligand C_4299 remains interacted with MET_793 for the maximum period (Figure 3.8).

Ligand C_42 shows hydrophobic interaction with LEU_718, PHE_723, VAL_726, ALA_743, MET_790, and LEU_844 residues, which was also seen in co-crystallized ligand Osimertinib. C_42 also showed a water bridge with a SER_797, like in the co-crystallized ligand (Figure 3.8). Ligand C_42 makes hydrogen bonds with MET_793 and Gln791 with a 100% interaction percentage, which existed in the reference ligand as well, with 96% and 98% interaction probability, as presented in Figure 3.9.1(a). co-crystallized ligand Osimertinib interacted with MET_793 and SER_797 residues with 99% and 55% interaction percentages, respectively (Figure 3.9.3(b)).

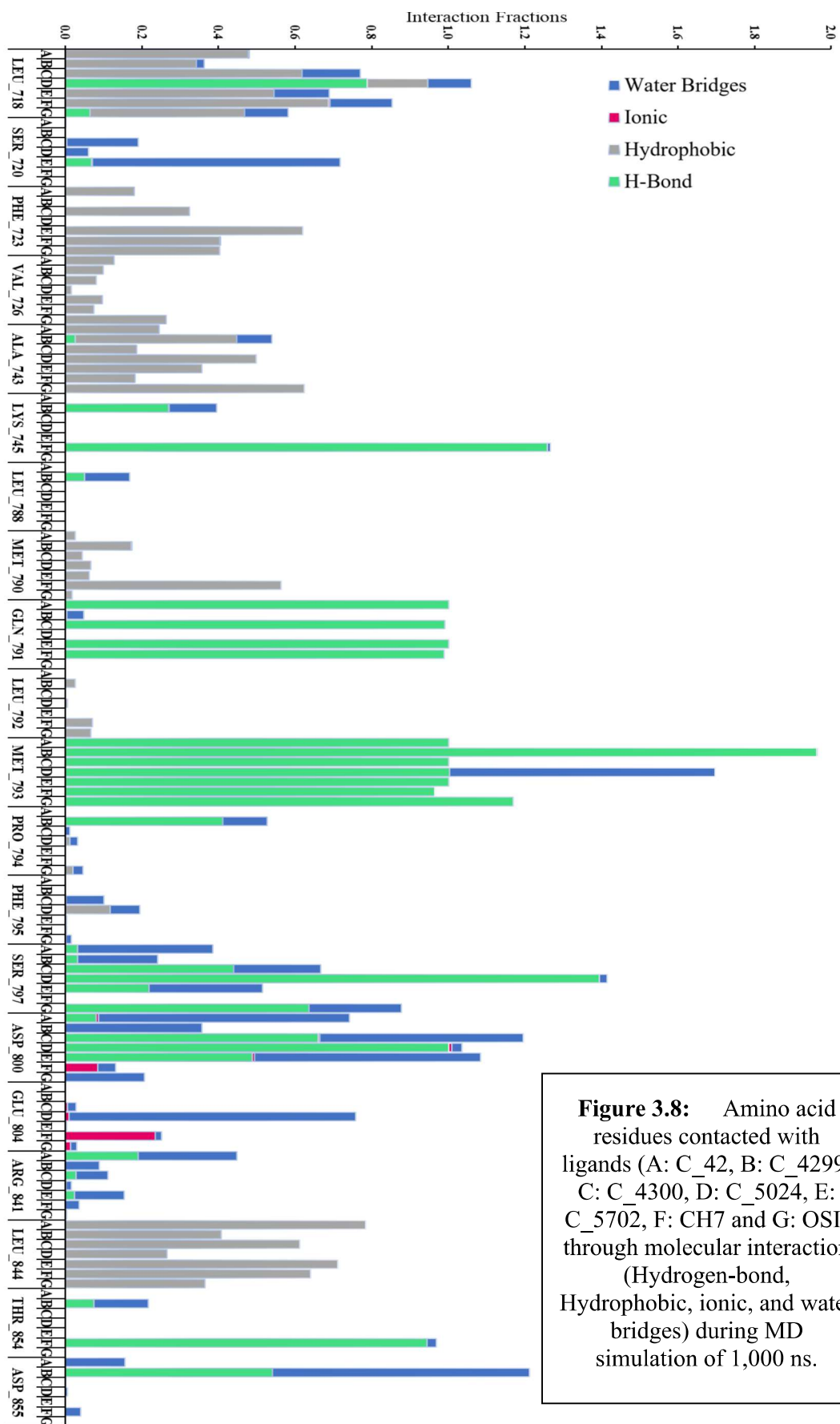


Figure 3.8: Amino acid residues contacted with ligands (A: C_42, B: C_4299, C: C_4300, D: C_5024, E: C_5702, F: CH7 and G: OSI) through molecular interaction (Hydrogen-bond, Hydrophobic, ionic, and water bridges) during MD simulation of 1,000 ns.

Ligands C_5702, C_4299, C_4300, and C_5024 also show strong interaction with MET_793, and the interaction percentage was 97% to 100%. They also interact with ASP_800 with 34%, 35%, 57 %, and 98% probability, respectively. Water-mediated interaction was also revealed in ligand C_4299 with ASP_855 and LEU_718, Ligand C_5024 with MET_793, and Ligand 5720 with Ser720 (Figure 3.9.1 – Figure 3.9.3).

Thus, based on molecular docking and molecular dynamics simulation results, the screened ligand C_42 exhibited the best interaction with the triple mutant EGFR^{CSTMLR} protein. It comprised better binding energy, the R_g, and SASA value than the reference ligand CH7233163 and the co-crystallized ligand Osimertinib. Its RMSD and RMSF values were comparable to the co-crystallized ligand and lower than the reference drug. In the beginning, it displayed four hydrogen bonds, but towards the end, it had increased to eleven. On the other hand, the co-crystallized ligand and reference drug revealed four hydrogen bonds throughout the simulation (Figure 3.7). Additionally, compared to the co-crystallized ligand, it also demonstrated conserved hydrogen bond interactions with the amino acids MET_793; water bridges with SER_797 residue; and hydrophobic interactions with LEU_718, PHE_723, VAL_726, ALA_743, MET_790, and LEU_844 residues. Thus, the overall analysis proved that ligand C_42 comprises better interaction and can be considered as an anticancer drug against triple mutant EGFR^{L858R/T790M/C797S}.

3.3.4 MM/GBSA calculations and per-residue free energy decomposition

With the MMGBSA technique, the free energies of protein-ligand complexes were estimated. One hundred frames were collected from the last 100 ns of MD simulation trajectories to explore the binding capabilities of ligands with the EGFR^{L858R/T790M/C797S}.

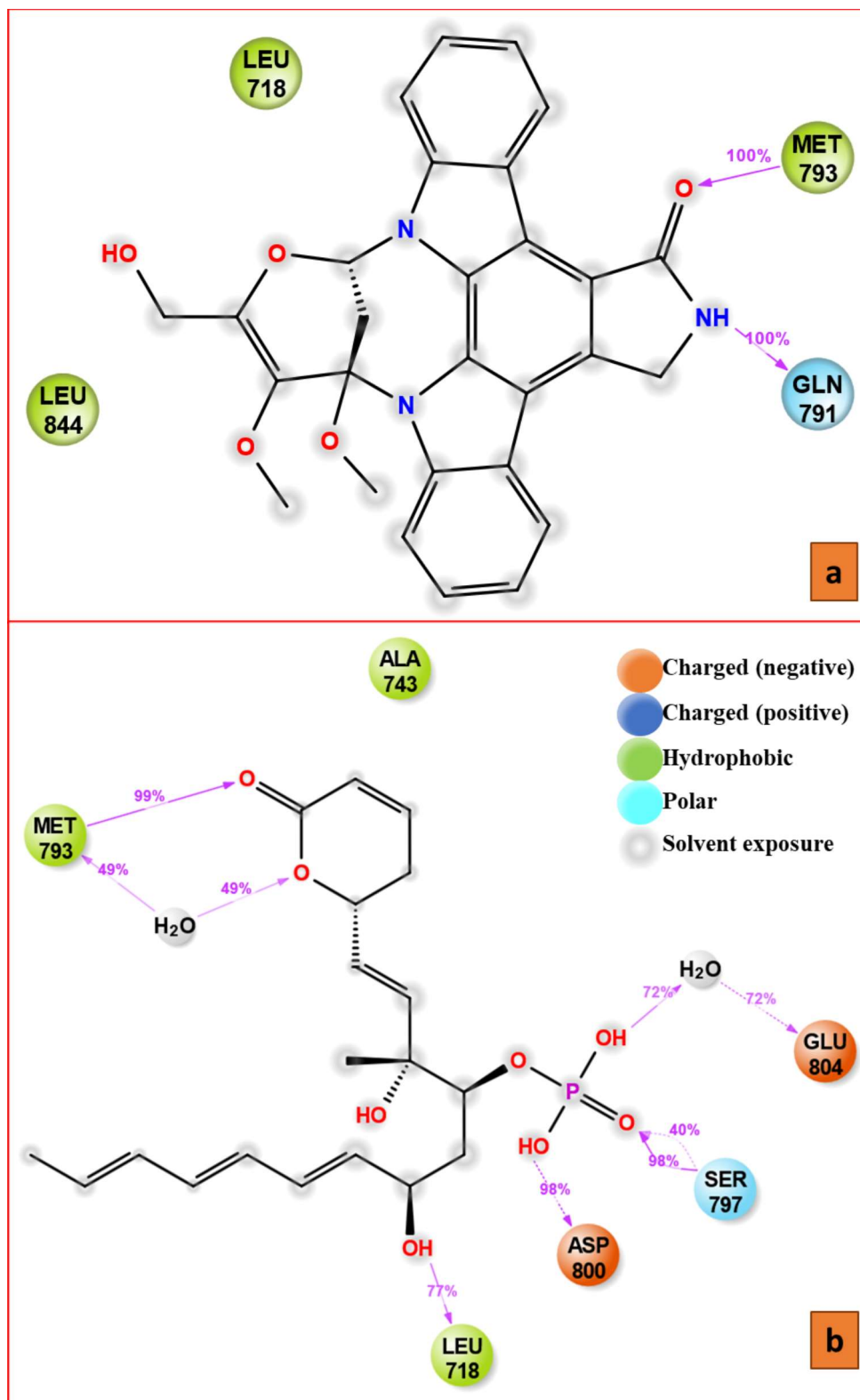


Figure 3.9.1: 2D plot represents the ligands representing a: C_42, and b: C_5024 with interacting residues of EGFR^{L858R/T790M/C797S} and their interaction percentage during MD simulation of 1,000 ns.

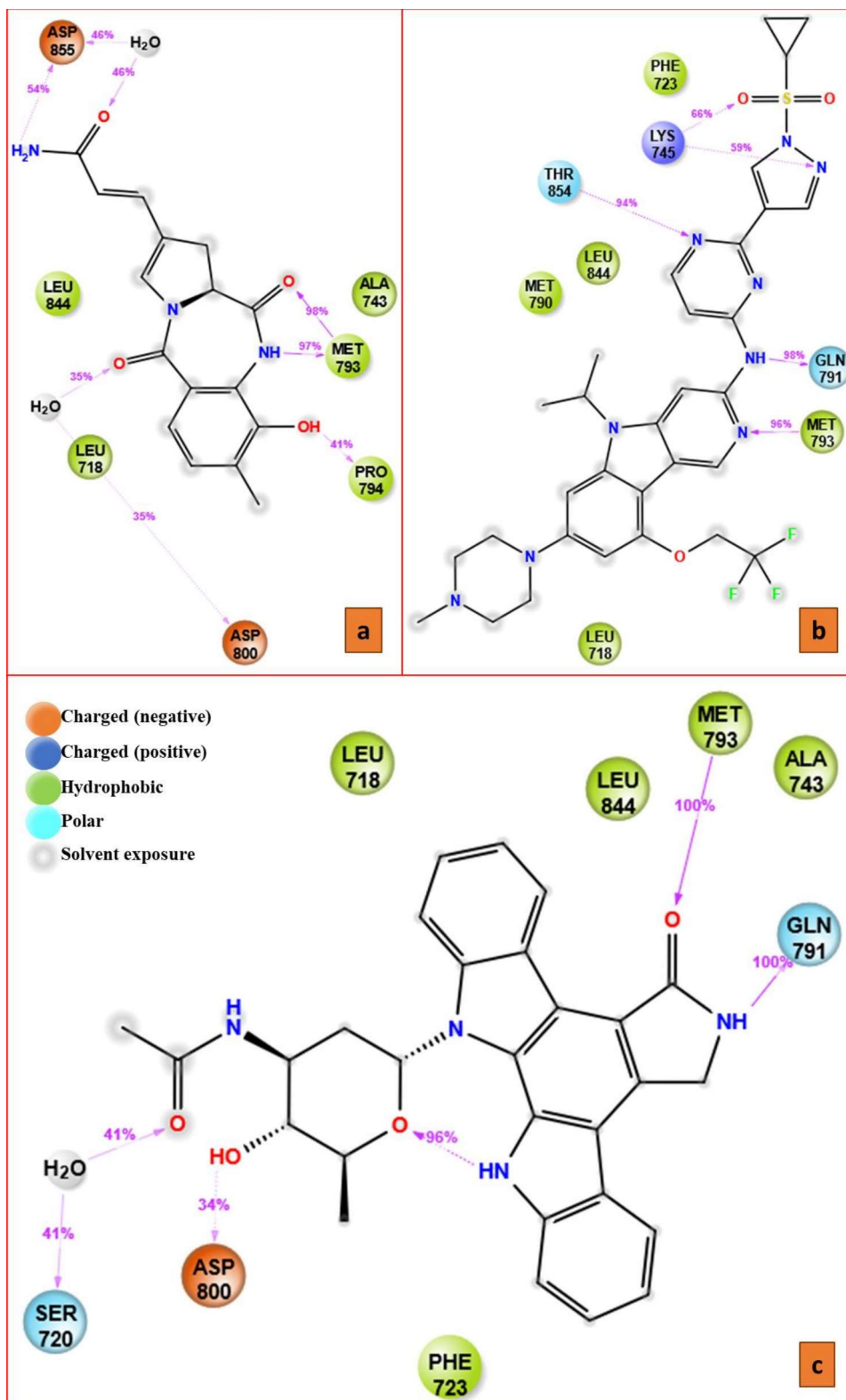


Figure 3.9.2: 2D plot represents the ligands representing a: C_4299, b: CH7, and c: C_5702 with interacting residues of EGFR^{L858R/T790M/C797S} and their interaction percentage during MD simulation of 1,000 ns.

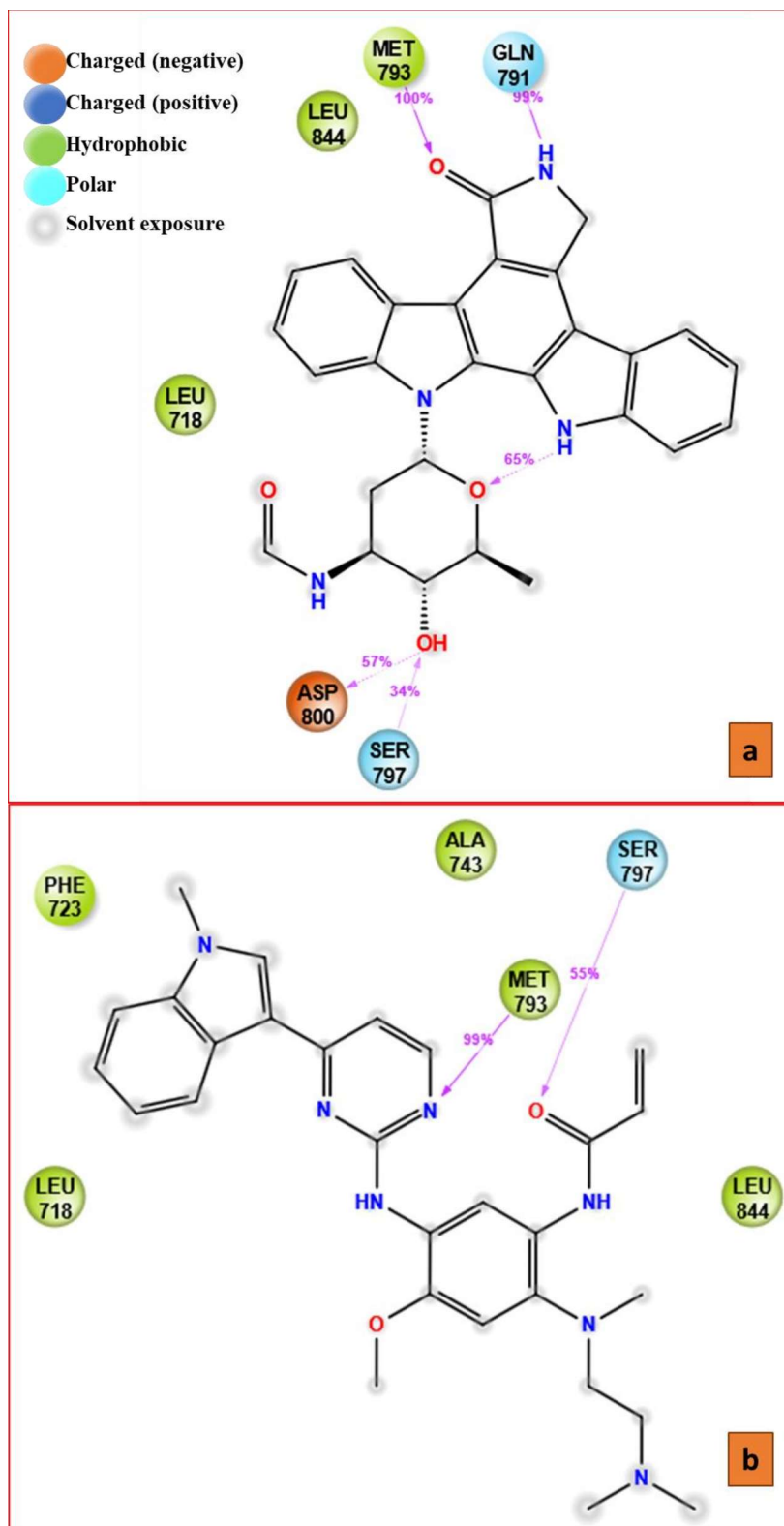


Figure 3.9.3: 2D plot represents the ligands representing a: C₄₃₀₀, and b: Osi with interacting residues of EGFR^{L858R/T790M/C797S} and their interaction percentage during MD simulation of 1,000 ns.

Later on, looking at the components of free energy, it was observed that coulomb energies and electrostatic energies of inhibitors are better than CH7 drugs, with a remarkable difference. In contrast, electrostatic energy values of ligands found akin to Osimertinib, also lipophilic energy contribution, stay equivalent with reference drugs (Table 3.2).

Van-der-waals energy input was found to be lower than CH7 but nearly the same as Osimertinib, but in the calculation of ΔG_{bind} , coulomb and lipo-energy contribution ratios are higher in all cases than native ligands (Table 3.2). Therefore, it concludes that selected ligands have a great affinity toward the active site of the mutant EGFR protein.

The per-residue decomposition was also performed on contributing amino acids to interpretative the energy influence profile in protein-ligand interaction (Figure 3.10). The residues having free energy equal to -1 kcal mol^{-1} or lesser were taken into account to know the depth of involvement in the protein-ligand complex stability, and it found that the residues LEU_718, PHE_723, VAL_726, ALA_743, LYS_745, MET_790, LEU_792, MET_793, GLY_796, SER_797, LEU_844, and THR_854 have played an essential role in the binding of all the seven ligands.

Table 3.2: Free energies (kcal/mol) of complex EGFR^{L858R/T790M/C797S} with C_42, C_4299, C_4300, C_5024, C_5702, CH7 and Osimertinib.

Ligand	Binding free energy (ΔG_{bind})	Coulomb energy ($\Delta E_{coulomb}$)	H-bond	Lipophilic energy (ΔE_{lipo})	Generalized Born electrostatic solvation energy (ΔE_{ete})	Van der Waals energy (ΔE_{vdw})
C_42	-60.06	-18.95	-1.11	-15.61	24.02	-48.60
C_4299	-39.16	-18.40	-1.78	-10.31	20.29	-31.01
C_4300	-59.85	-24.20	-1.42	-16.08	26.84	-45.42
C_5024	-51.24	-28.24	-2.18	-12.48	22.07	-32.24
C_5702	-61.23	-24.95	-1.55	-16.86	28.52	-47.40
CH7	-68.74	-10.17	-1.91	-16.69	21.47	-64.80
Osi	-50.60	-16.33	-0.80	-14.55	27.00	-46.80

From the analysis of Figure 3.10, C_5702, C_4300, and CH7 have shown making interaction with the highest number of residues, followed by C_42, then Osimertinib, and last, C_4299 and C_5024, where all the ligands, including CH7, have shared most of the residues (8 - 12 residues), the ligand Osimertinib found to be gone a bit out of the league with only two residues (Figure 3.10).

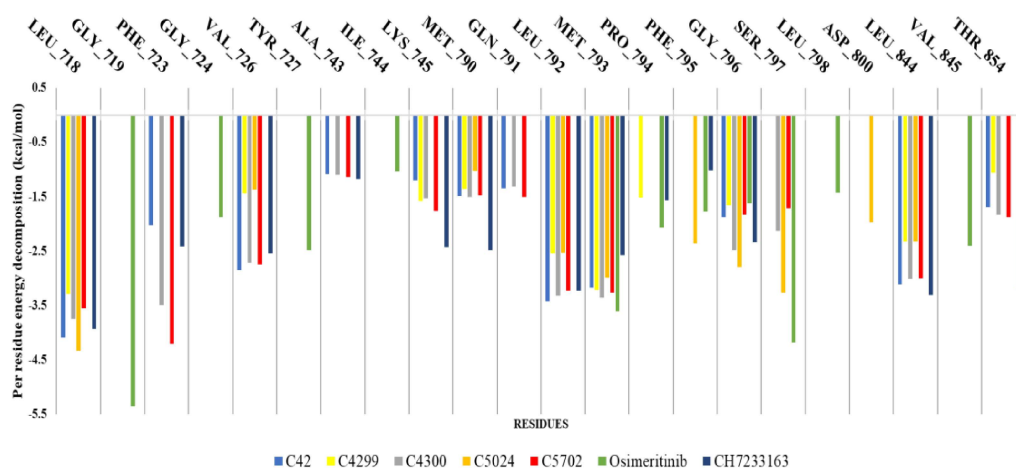


Figure 3.10: Per-residue-energy-decomposition analysis of all ligands against EGFR^{L858R/T790M/C797S} structure.

3.4 Conclusions

In the current chapter, an attempt has been made to identify a novel drug against the triple mutant EGFR^{CSTMLR} as a cancer therapy from *Streptomyces* sources. In order to achieve this, the ADMET profile of all the *Streptomyces*-derived compounds from the StreptomeDB database was assessed, and the screened hits were subjected to a computational study. The docking result revealed five compounds, C_42 (Steptocarbazole A), C_4299 (Limazepine H), C_4300 (3'-N-Formylholyrine A), C_5024 (CHEMBL1159781), and C_5702 (3'-N-Acetylholyrine A), with better binding affinity than both reference drug CH7233163 and co-crystallized ligand Osimertinib. As these drugs are derived from Actinomycetes so, their upscaled production can easily be achieved by the fermentation process. Further, the drugs shortlisted here were already validated experimentally by other researchers against several cancerous cell lines and

protein kinases. Based on the simulation study, ligand Steptocarbazole A binds at the active site of EGFR^{CSTMLR} with outstanding protein stability, compactness, and solvent effects, revealing excellent inhibitory potential. While Osimertinib, a well-known anti-Tyrosine inhibitor, is chemically manufactured and has several adverse effects-developed resistances. On the other hand, C_42 (Steptocarbazole A), derived from natural sources, can be considered a potential inhibitor of triple mutant EGFR in cancer treatment.

ADMET profiling and computational study identified several potent inhibitors against the triple mutant EGFR receptor tyrosine kinase, including Steptocarbazole A, Limazepine H, 3'-N-Formylholyrine A, ChEMBL1159781, and 3'-N-Acetylholyrine A, all derived from *Streptomyces*. This study highlights the potential of the genus *Streptomyces* in developing effective receptor tyrosine kinase inhibitors. Thus, the study progressed by procuring *Streptomyces clavuligerus* and isolating *Streptomyces fragilis* to produce bioactive metabolites. And investigate their potential inhibitory effects on receptor tyrosine kinases and conduct protein kinase inhibitory assays.