Chapter 4

In silico Analysis, Isolation, and Cytotoxicity Evaluation of the Coumestans from *Psoralea corylifolia* (L.)

4 *In silico* Analysis, Isolation, and Cytotoxicity Evaluation of the Coumestans from *Psoralea corylifolia* (L.)

4.1 Introduction

Coumestans or Phytoestrogens are polycyclic aromatic compounds consisting of a four-ring oxygen heterocycle system linked to a coumarin and a benzofuran unit through a C=C bond [123]. Natural coumestans are reported to exhibit estrogenic, antimicrobial, antifungal, antioxidant, anti-osteoporotic, anti-inflammatory, antihemorrhagic, hepatoprotective, antifibrotic, anti-proteolytic, anti-diabetic, anticancer, neuroprotective, and immunomodulatory properties [123, 124]. Legumes are notably rich in coumestans, and many coumestans have been successfully extracted from the Leguminosae/Fabaceae family [123]. Within this family, *Psoralea corylifolia* L. (syn. *Cullen corylifolium* (L.) Medik., Fabaceae) is a rich source of coumestans [125] that are concentrated primarily in seeds. Their estrogenic properties support the potential role of *P. corylifolia* in osteoporosis and cardiovascular diseases [126, 127]. The coumestan of *P. corylifolia* includes psoralidin, psoralidin-2',3'-oxide, psoracoumestan, plicadin, etc. (Figure 4.1) and they still remain largely unexplored for their pharmacological potential [128].

EGFR is ubiquitously expressed multifunctional signal transducer. It orchestrates various cellular processes including migration, proliferation, cell fate determination, and apoptosis [129]. It binds to various growth factor ligands like EGF, betacellulin, amphiregulin, epiregulin, epigen etc. and trigger responses [130]. Upon ligand binding, EGFR undergoes dimerization leading to autophosphorylation and triggers resultant signalling pathways like PI3 kinase, Ras-Raf-MAPK, JNK, and PLCγ [131]. ErbB1, the EGFR family representative, is linked to cancer development through mutations or overexpression. EGFR overexpression are common in various cancers, notably breast and lung cancers, making it a focal point for several cancer treatments [129].

EGFR overexpression is a key therapeutic target in breast cancer, causing poor tumor differentiation, larger tumor sizes, and disruption of EGFR pathways [131]. Across all breast cancer subtypes, its overexpression is more prevalent in triple-negative breast cancer (TNBC) and hence, its negative impact is more apparent in TNBC. Hence, EGFR a promising target for TNBC treatment [131].

EGFR growth factors play a crucial role in lung cancer initiation, progression, and metastasis affecting both non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). Mutations in EGFR damages signalling pathways culminating into aggressive and metastatic lung adenocarcinoma. EGF upregulation is observed in lung cancer, while amphiregulin upregulation is linked to poor prognosis and reduced survival rates in NSCLC [132].

In this study, *in silico* approaches were employed to assess anticancer capabilities of coumestans of *P. corylifolia* against EGFR. It involved rigorous molecular docking and MD simulations, followed by *in silico* pharmacokinetic estimations. The results were validated by focussed isolation of identified lead from *P. corylifolia* and assessing its cytotoxic effects on TNBC and NSCLC cell lines.

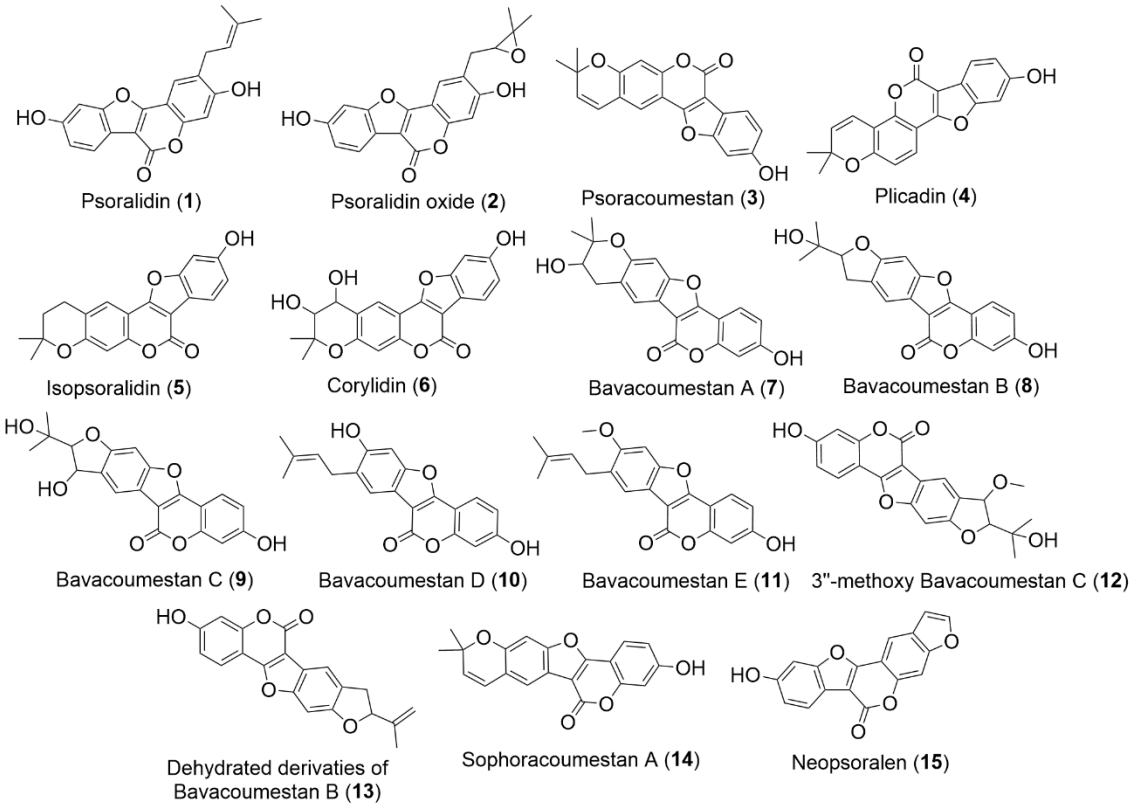

Figure 4.1 Chemical structures of coumestans present in *P. corylifolia*.

4.2 Experimental section

4.2.1 Preparation of ligands

As mentioned in Section 3.2.5.

4.2.2 Protein selection, protein energy minimization, and protein preparation

The crystal structure of EGFR (PDB: 6DUK) was retrieved from the protein data bank (www.rcsb.org) and evaluated for various parameters like resolution, *R*-value free, and *R*-value work. The protein structure was checked for missing residues using UCSF Chimera 1.14 rc [104]. The protein was then subjected to energy minimization to eliminate bad contacts and unfavourable torsion angles. The PDB2PQR server employing PROPKA was used to assign the correct protonation state of residues at pH 7.4 (https://server. poissonboltzmann.org/pdb2pqr). The protein energy minimization was done using the amber plugin on Chimera 1.14. The backbone refinement was done with amber ff14SB force field and the incomplete side chains were replaced via Durnbrack

2010 rotamer library [133]. The energy minimization was carried out at 1000 steps of conjugate gradient and 4000 steps of steepest descent using a step size of 0.02 Å. The minimized protein was then processed in MGL tools 1.5.7.rc to add polar hydrogens only, merge non-polar hydrogens, add Gastieger charge, and subsequently converted to PDBQT format.

4.2.3 Molecular docking studies

4.2.3.1 Grid generation and validation

The EGFR active site was identified using PLIP server. The residues involved in the interaction were used as reference points to build a grid box around the active site. AutoDock uses autogrid 4.0 to calculate the grid maps of interaction energies with various types of atoms viz. A, C, HD, NA, N, OA, S, Br, Cl, and I present in the ligands. The grid size for the study was set to 56 X 80 X 58, xyz points for the minimized protein with a grid spacing of 0.375 Å. The grid centre was placed at coordinates (x, y, and z) 40.849, 90.902, and -62.662, respectively. The grid validation was performed by redocking the co-crystallized ligand and calculating the RMSD between experimentally determined docked pose and co-crystallized ligand using Discovery Studio Visualizer 2021.

4.2.3.2 Molecular docking

The docking study was performed on Autodock 4.2 using LGA. The binding free energy of the ligand-receptor complex was used for scoring various conformations. The scoring function is based on a semi-empirical free energy force field to evaluate conformations obtained by molecular docking. The docking was performed with 100 runs, 150 population size, 27000 number of generations and 2500000 number of energy evaluation. It employs ‘semiemperical free energy force field’ to evaluate conformations at the time of docking simulation. Gefitinib and erlotinib were used as standards for comparing docking results. The compounds with binding energy less than -10 kcal/mol were selected

for the further studies. The post-docking analysis and visualization were performed on Discovery Studio 2021 visualizer.

4.2.4 *In silico* ADMET prediction

As mentioned in Section 3.2.8.

4.2.5 MD simulation

The protocol for MD simulation studies is same as mentioned in Section 3.2.9.

4.2.6 Binding free energy and per residue decomposition studies

The protocol for binding free energy calculation is same as mentioned in Section 3.2.10. In order to determine the energy contribution of the amino acids implicated in inhibitor binding, per-residue breakdown analysis was also carried out. Using the g-mmpbsa tool, binding free energy decomposition was carried out. The overall binding energy of the protein-ligand complex was broken down using this method [134]. For MM-PBSA computations and individual contributions of various amino acids, the Python scripts "MmPbSaStat.py" and "MmPbSaDecomp.py" were used. Met 766, Cys 775, Leu 777, Leu 788, and Leu 858 were taken into consideration for compound **1** in this study, whereas Phe 723, Lys 745, Leu 747, Leu 858, and Lys 875 were taken into consideration for compound **5**. With the majority of the inhibitors, these amino acids consistently established interactions.

4.2.7 Extraction and isolation of psoralidin (1) from *P. corylifolia*

4.2.7.1 General experimental procedures

The chemicals were obtained from Sigma-Aldrich Company and used as received. The bulk solvents were used only after distillation. The 1H and 13C NMR spectra were recorded on Bruker-Avance III HD 500 MHz NMR spectrometer using internal standard tetramethylsilane (TMS). The TMS was referenced to the residual proton/carbon in the NMR solvent ($CDCl_3$ 7.26/77.1 ppm, CD_3OD 3.31/49.0 ppm, and DMSO-d_6 2.50/39.5

ppm). The chromatographic purifications were performed using silica gel (#60–120 or #100–200 from E. Merck, Germany). The thin-layer chromatography (TLC) was performed on pre-coated silica gel 60 GF254 aluminum sheets from Merck (Germany). The spots of compounds were visualized under short-wavelength (254nm) ultraviolet light. HRMS spectra of the compounds were recorded on HRMS-6540-UHD machines. The human triple-negative breast cancer cell lines MDA-MB-231 and non-small cell lung cancer cell line A549 were acquired from Chittaranjan National Cancer Institute, Kolkata, India and National Centre for Cell Science, Pune, India, respectively. Eppendorf supplied the 96-well plates and T-25 flasks, while Genetix (Genetix Biotech Asia Pvt. Ltd.) supplied the DMEM (Dulbecco's Modified Eagle Medium). FBS (Foetal Bovine Serum), Trypsin-EDTA, and Penicillin-Streptomycin were purchased from Gibco. In order to make PBS (Phosphate Buffer Saline), analytical chemicals were used.

4.2.7.2 Plant material

The seeds of *P. corylifolia* were obtained from the local market in Varanasi. The seeds were authenticated by Prof. N.K. Dubey, Department of Botany, Banaras Hindu University, Varanasi, India. The specimen sample of *P. corylifolia* seeds was also preserved (Voucher specimen no. Faba.2023/01) in the Department of Botany, Banaras Hindu University, Varanasi, India.

4.2.7.3 Extraction and isolation of plant material

The seeds (500 g) were coarsely grounded and extracted via cold maceration with chloroform: methanol mixture in a 1:1 ratio. The extraction cycle was repeated three times and finally, the extract was dried under reduced pressure to get a viscous crude extract. The extract was then dissolved in the minimum amount of solvent and it was adsorbed over silica gel to make the slurry. The slurry was loaded over the silica-gel column chromatography for the separation of primarily psoralidin. The column was run using

increasing concentrations of ethyl acetate in hexane starting from 100% hexane to 100% ethyl acetate. The fraction obtained at 40% ethyl acetate in hexane was repeatedly purified using silica gel column chromatography to get psoralidin. The obtained compound was characterized via NMR spectroscopy, mass spectrometry, and the structures was confirmed by comparing NMR data with the reported literature (Appendix, Figure A.11-A.14) [135].

Psoralidin (**1**). 1H NMR (500 MHz, $CDCl_3$) δ 8.07 (d, J = 8.4 Hz, 1H), 7.97 (s, 1H), 7.94 (s, 1H), 7.45 (d, J = 2.1 Hz, 1H), 7.26 (dd, J = 8.4, 2.1 Hz, 1H), 5.70 (t, J = 7.4 Hz, 1H), 3.69 (d, J = 7.6 Hz, 2H), 2.12 (s, 3H), 2.07 (s, 3H). 13C NMR (126 MHz, $CDCl_3$) δ 159.5, 158.3, 158.1, 156.2, 155.6, 152.5, 132.4, 126.1, 120.9, 120.4, 120.1, 114.6, 113.1, 103.7, 102.1, 101.7, 98.1, 27.1, 25.1, 17.1. HRMS *m/z*: $[M+Na]^+$ calc. 337.1071, obs. 337.1019 [136].

4.2.8 Cytotoxicity screening

4.2.8.1 Cell culture

A549 and MDA-MB-231 were used for validation of identified hit from molecular modeling studies. The cell lines A549 and MDA-MB-231 were grown using vented T25 cell culture flask (Corning, USA) in Ham's F-12 Nutrient Mix (Thermofisher, USA) and Minimum Essential Medium (MEM) (Thermofisher, USA), respectively. The medium was supplied with 10% of fetal bovine serum (Thermofisher, USA), Antibiotic-Antimycotic solution (Thermofisher, USA) containing 10,000 units/mL of penicillin, 10,000 µg/mL of streptomycin, 25 µg/mL Gibco Amphotericin B, and 1.25 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (Thermofisher, USA). The cell lines were incubated at 37 °C temperature with 95% humidity and 5% pressure of CO_2 [137, 138].

4.2.8.2 Cell viability assay

Cell viability assay of isolated natural coumarins was performed on both A549 and MDA-MB-231 cell lines by MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] assay. Briefly, 1×104 cells were seeded in each well of 96 well tissue culture-treated microplates (Corning, USA) followed by the incubation for 24 h at 37 ℃ with 5% CO_2. On the second day, the cells were treated with the compounds (dissolved in DMSO) at different concentrations ranging from 1 µM to 100 µM (1 µM, 5 µM, 10 µM, 25 µM, 50 µM, 75 µM, 100 µM) and further incubated at 37 ℃ for 24 h. After incubation, 25 µL of MTT solution (5 mg/mL in PBS) was added in each well containing 300 µL of media for the incubation period of 3h. Finally, the crystal formazan formed in each well was dissolved by the addition of 100 µL of DMSO (Merck, Germany) and read at 570 nm on Microplate Spectrophotometer (Epoch BioTek, Agilent) [137, 138]. The IC_{50}, values were obtained by nonlinear regression analysis using Graph pad prism (version 8.0) software.

4.3 Result & Discussion

4.3.1 Ligand preparation, protein energy minimization and protein preparation

The energy minimization of ligands resulted in optimization of bond length and bond angle before docking study. The GAFF in Open Babel 3.1.1 employs simple harmonic form for optimizing angles and bonds of the molecules and it covers most of organic chemical space. Protein energy minimization is crucial for global minimization in docking and simulation studies and it was done using Amber 20 module. This step resulted in potential energy minimization of protein from -137060 kcal/mol to -204320 kcal/mol (Appendix, Figure A.15).

4.3.2 Molecular docking studies

The molecular docking approach has remained an important tool in drug discovery process during the initial screening of potential candidates [139]. The co-crystallized ligand displayed various interactions with Phe 723, Lys 745, Glu 749, Arg 776, Leu 788, Met 790, Asp 855, and Leu 858 (Appendix, Figure A.16). These residues were used to ascertain grid box size [109]. The grid validation between docked ligand and co-crystalized ligand showed RMSD of 0.3409 Å that shows less deviation in selected grid and thus, can be used for docking studies (Appendix, Figure A.17).

After completion of docking studies, the interactions of all ligands were visualized in Discovery Studio Visualizer 2021. Total 15 coumestans of *P. corylifolia* along with Gefitinib and Erlotinib were docked to EGFR and all possessed binding energy less than -8 kcal/mol. The binding energy cut off was set to -10 kcal/mol and two ligands with energy less than this cut off were selected for further studies. The binding energy of ligand **1** and **5** was better than Gefitinib and Erlotinib (Table 4.1). The 2D and 3D interaction diagrams of ligands 1 and ligand 5 with EGFR has been shown in Figure 4.2. The binding energies, ligand efficiency, and interaction diagram of rest of the ligands has been provided in Appendix (Table A.3 and Figure A.18).

Table 4.1 Binding energy, ligand efficiency, and docking interactions of coumestan with EGFR protein (PDB: 6DUK).

Ligands	Binding Energy (kcal/mol)	Ligand Efficiency (kcal/mol)	Ligand-Protein Interactions (PDB Id. 6DUK)
Gefitinib	-9.49	-0.306	Met 766 (Pi-Sulfur), Leu 777 (Pi-Alkyl), Leu 788 (Pi-Sigma, Pi-Alkyl), Met 790 (Pi-Sulfur), Arg 841 (Halogen), Asp 855 (H-Bond)
Erlotinib	-8.71	-0.3	Val 726 (Pi-Sigma, Pi-Alkyl), Lys 745 (H-Bond), Met 766 (Pi-Sulfur, Pi-Alkyl), Met 793 (H-Bond), Met 790 (Pi-Sulfur), Asp 855 (H-Bond)
1	-10.02	-0.401	Met 766 (Pi-Sulfur), Cys 775 (H-Bond), Leu 777 (H-Bond), Leu 788 (Pi-Sigma, Pi-Alkyl), Met 790 (Pi-Sulfur), Leu 858 (Pi-Sigma)
5	-10.00	-0.4	Lys 745 (H-Bond), Met 766 (Pi-Sulfur), Leu 788 (Pi-Sigma, Pi-Alkyl), Thr 854 (H-Bond), Asp 855 (H-Bond), Leu 858 (Pi-Sigma, Pi-Alkyl)

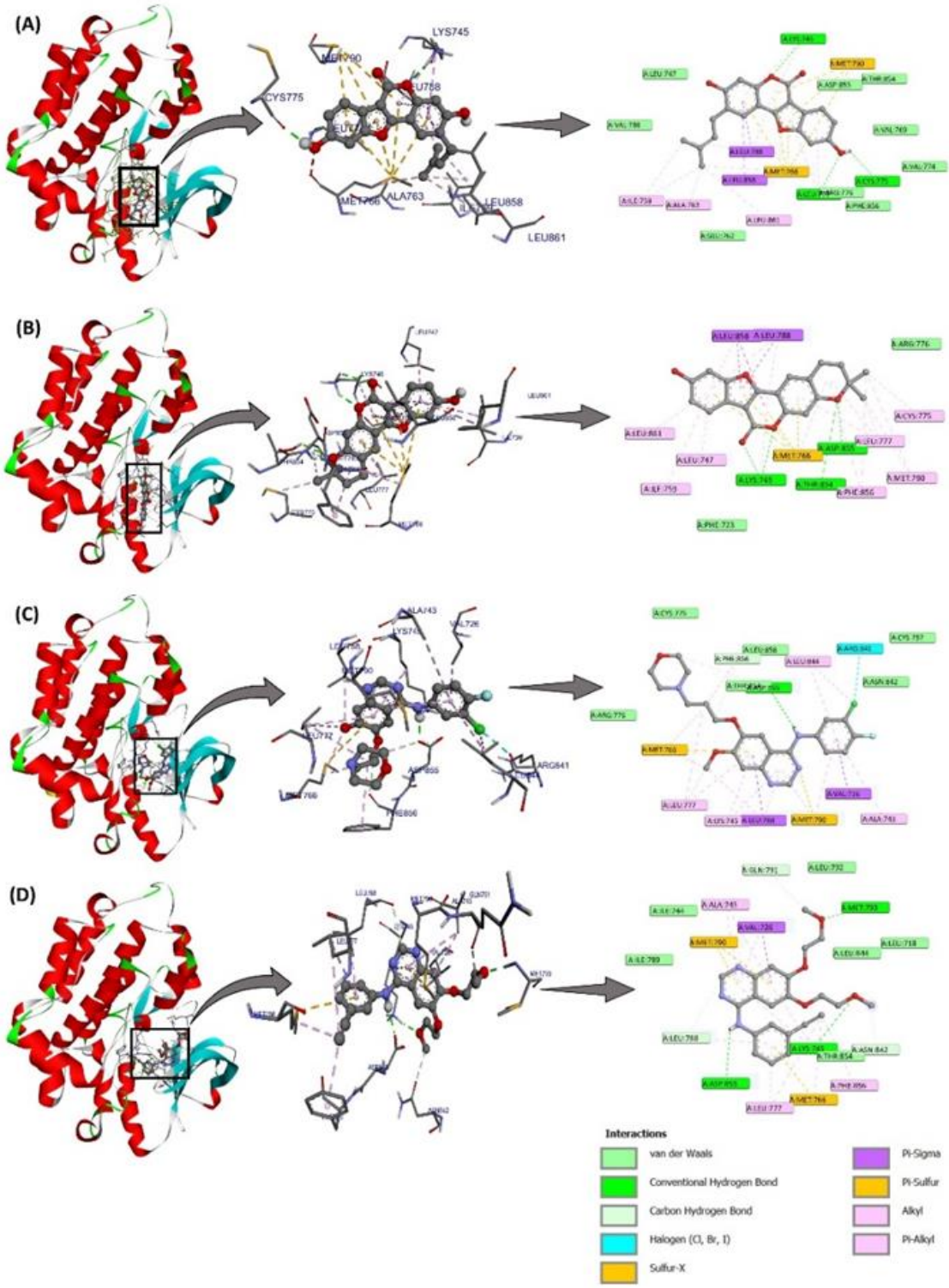

Figure 4.2 2D and 3D interaction diagram of (A) Compound 1, (B) Compound 5, (C) Gefitinib, and (D) Erlotinib with EGFR protein (PDB: 6DUK).

4.3.3 *In silico* ADMET prediction

As the bioavailability and drug response depend on absorption thus, drug must have HIA above 70% for good absorption. The BBB permeability in range of 0.1 to 2.0 for predicted

compound is an indicator of moderate penetration through BBB [121]. Carcinogenic potential is an undesirable trait for drug-like molecules. hERG inhibition is also an undesirable trait for drug candidates as it could lead to cardiac dysfunctions [140]. The other desirable trait of drug candidate is that it should not inhibit liver microenzymes. Compound **1** and **5** possessed good *in silico* ADMET properties summarized in Table 4.2. The ADMET properties of rest of the compounds is also provided in Appendix, Table A.4.

Table 4.2 Predicted ADMET properties of best hits.

Ligands	**1**	**5**
BBB (C_{brain}/C_{plasma})	1.415	0.073
Aqueous solubility (mg/ml)	4.54	4.96
HIA (%)	94.29	96.12
hERG inhibition	Medium risk	Medium risk
Rodent carcinogenicity	Non-carcinogen	Non-carcinogen
CYP2D6 inhibition	Non-inhibitor	Non-inhibitor

4.3.4 MD simulation study

MD simulation is an advanced technique simulating systems with biological relevance and involves studying the trajectories and motion of the molecules in presence of other complexes over time, thus aids in studying the conformational changes occurring in the molecules [141]. It also gives information regarding structural features of protein within the system and ligand-protein interactions.

Here two ligand-protein complexes- compound **1**-EGFR and compound **5**-EGFR were studied through MD simulations using Pmemd.cuda module of Amber 20 [142]. The RMSD plot of protein-ligand complexes is shown in Figure 4.3. The ligand showed protein backbone stability during the simulation run with an average protein RMSD of 1.34 Å and 1.23 Å for compound **1**-EGFR and compound **5**-EGFR, respectively. The

RMSD values of heavy atoms present in the ligands demonstrated that the ligand remained stable throughout the simulation time (Figure 4.3). The average ligand RMSD for compound **1**-EGFR and compound **5**-EGFR was 0.68 Å and 0.71 Å, respectively.

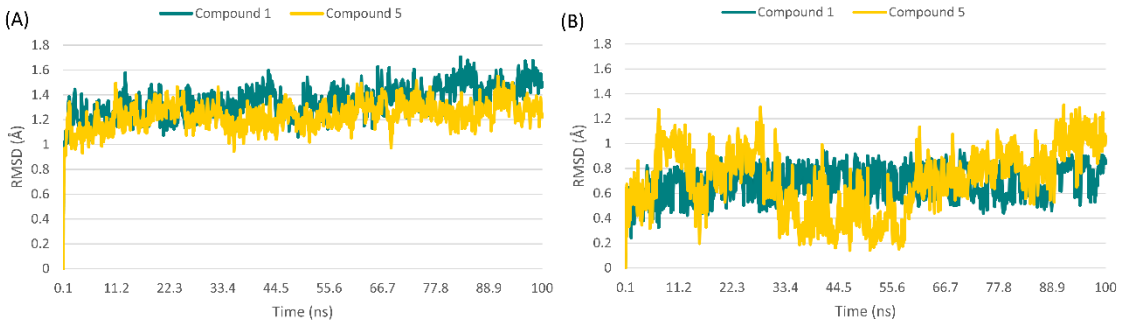

Figure 4.3 RMSD of the protein-ligand complexes. (A) Protein RMSD and (B) Ligand RMSD.

The RMSF analysis was done for C-α atoms of residues and the RMSF plot showed stability in the secondary conformation of protein at the time of simulation (Figure 4.4). The average RMSF values for compound **1**-EGFR and compound **5**-EGFR were 0.92 Å and 0.91 Å, respectively. The high fluctuations at the C-terminal and N-terminal region are inevitable and hence, could be seen in the current plot also. The active site residues showed apparently lower values of RMSF and so possessed good stability during the run. The RMSF analysis showed both complexes were stable throughout the MD simulation.

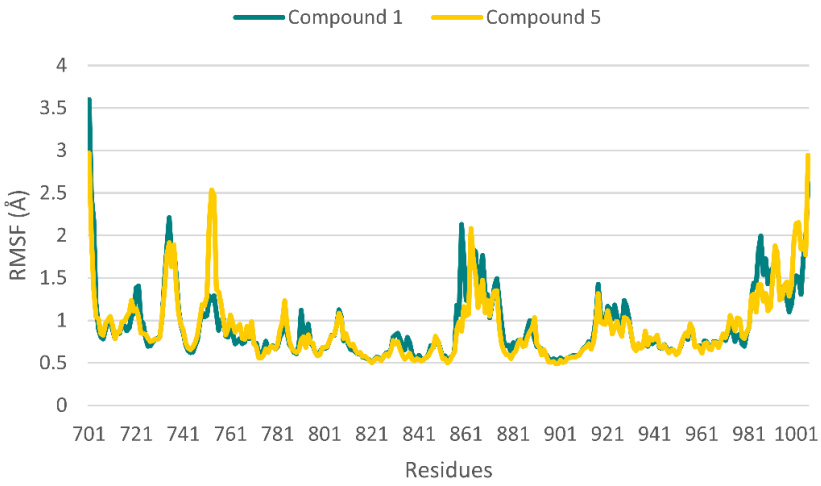

Figure 4.4 Residue wise RMSF deviations in the protein-ligand complexes.

The apparent protein folding or unfolding at the time of simulation run was determined by radius of gyration (RoG) analysis. The average values of RoG were 19.73 ± 0.06 Å for compound **1**-EGFR and 19.62 ± 0.07 Å for compound **5**-EGFR (Figure 4.5).

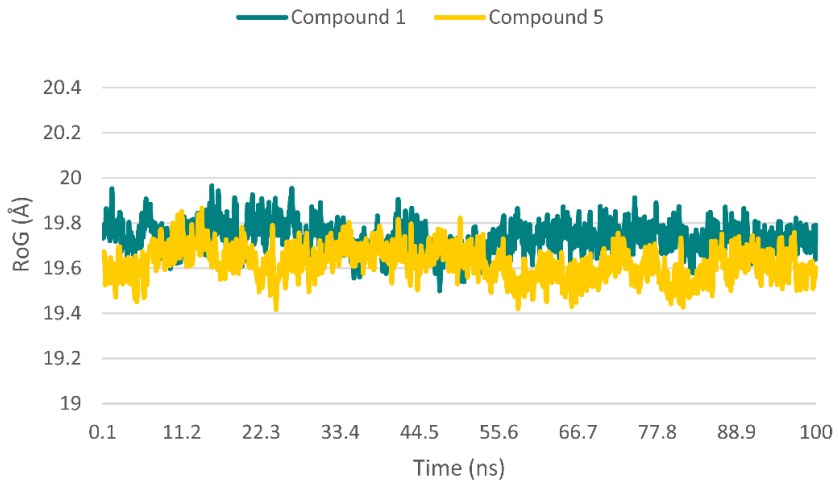

Figure 4.5 Radius of gyration of the protein-ligand complexes.

Understanding the stability of protein-ligand complex requires an understanding of the H-bond interaction analysis as fitting of ligand in the binding site largely depends on H-bonding. Lys 745, Cys 775, and Leu 777 displayed H-bond interaction with ligand **1**. Lys 745, Thr 854, and Asp 855 displayed H-bond interaction with ligand **5** (Figure 4.6).

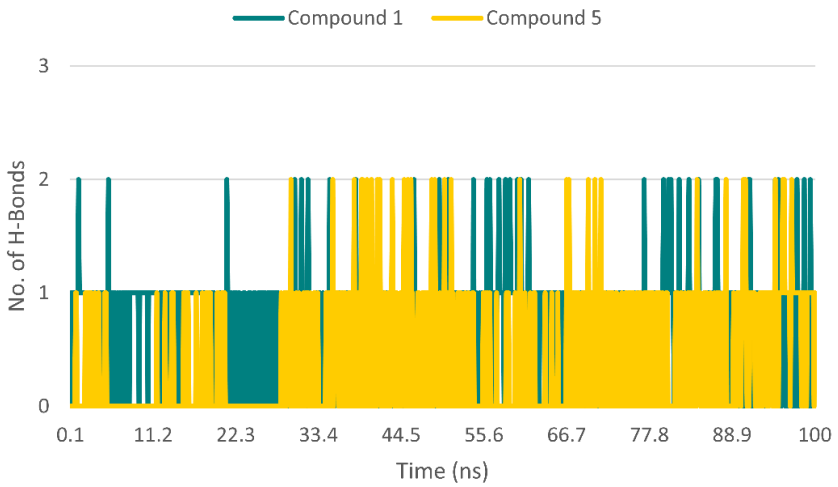

Figure 4.6 Number of H-bond interaction between compound 1-EGFR and compound 5-EGFR complex during MD run.

4.3.5 Binding free energy and per-residue decomposition studies

The energy contribution of non-bonded interaction energies i.e., Vader Waals (ΔE_{vdw}) and electrostatic energy (ΔE_{ele}) for the two complexes was determined by MM-GBSA and MM-PBSA methods. The complexes were found to exhibit high stability as the net binding free energies calculated by both methods were low. In contrast, the gas phase energy contribution (ΔE_{vdw} and ΔE_{ele}) was high that indicates that the stability of the complex was primarily because of ligand conformation with reference to receptor. The ligand **1** and **5** displayed binding energy of -50.92 ± 1.83 kcal/mol and -42.94 ± 1.92 kcal/mol, respectively in the GB solvation model. Both coumestans exhibited good binding free energy in the PB solvation model as well. The contribution of all the energies in MM-GBSA assay and MM-PBSA assay has been shown in Figure 4.7.

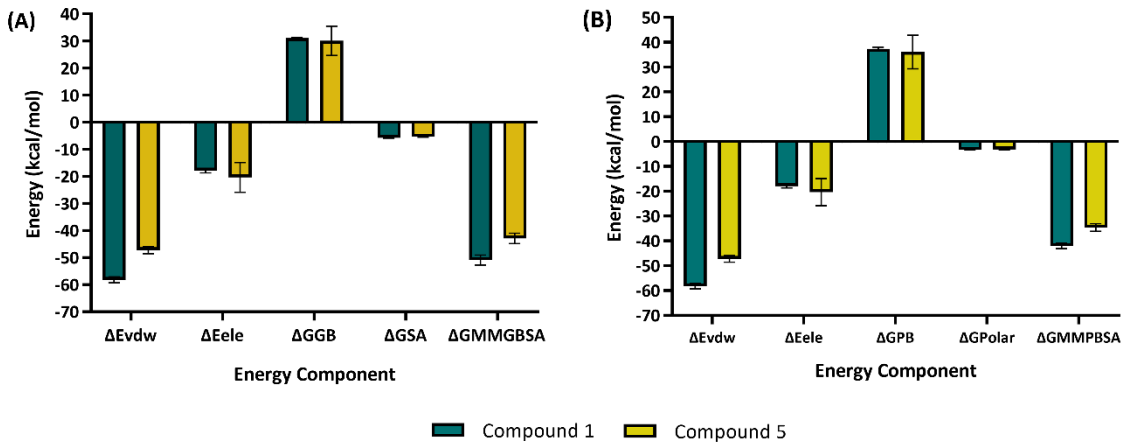

Figure 4.7 Energy contributions of protein-ligand complexes in (A) MM-GBSA and (B) MM-PBSA.

Per-residue binding energy decomposition analysis revealed the contribution of various amino acids towards total binding energy. From the decomposition analysis, the contributions of the consistently interacting amino acids were extracted. In case of compound 1, the decomposition energy for residues Met 766, Cys 775, Leu 858, Leu 788, and Leu 777 were -3.19, -2.54, -2.18, -2.02, and -1.98, respectively. While for compound

5, the decomposition energy for Leu 747, Leu 858, Lys 745, Lys 875, and Phe 723 were -2.65, -1.90, -1.84, -1.72, and -1.71, respectively.

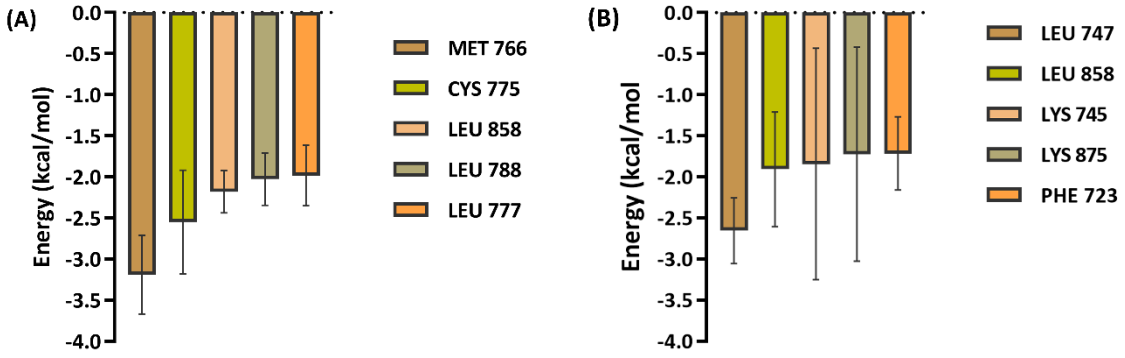

Figure 4.8 Per residue energy decomposition for key residues for (A) Compound 1 and (B) Compound 5.

4.3.6 Isolation and purification of Psoralidin (1)

The seed extract of *P. corylifolia* was subjected to silica-gel column chromatography for the isolation of primarily psoralidin. The fraction containing psoralidin was subjected to repeated column chromatography to obtain pure compound. The structure of the isolated compound was confirmed by NMR spectroscopy, Mass spectrometry, and compared with that of reported literature. The isolated compound 1 was the evaluated for cytotoxicity against MDA-MB-231 and A549 cell lines.

4.3.7 Cytotoxicity screening

The best hit, i.e., compound **1** (Psoralidin) was evaluated for cytotoxic potential against two cancer cell lines viz MDA-MB-231 and A549 using MTT assay for validation of *in silico* results. The findings revealed significant cytotoxicity of compound **1** on the cancer cells with IC_{50} value of 22.21 ± 1.65 µM and 23.64 ± 0.39 µM against MDA-MB-231 cells and A549 cells, respectively. The good cytotoxicity shown by compound **1** in *in vitro* settings validated the robustness of *in silico* analysis.

4.4 Conclusion

The study conducted molecular modeling studies on 15 coumestans from *P. corylifolia* against EGFR, revealing two best hits: psoralidin (**1**) and isopsoralidin (**5**), which showed good inhibitory activity against EGFR. MD simulations showed stability of complexes with EGFR in simulated biological environments. The *in vitro* cytotoxicity performed on the best hit and isolated psoralidin against MDA-MB-231 and A549 cells also validated the credibility of *in silico* method. Psoralidin displayed good cytotoxic potential against MDA-MB-231 and A549 cells with IC_{50} value of 22.21 µM and 23.64 µM, respectively. Since both cell lines overexpress EGFR, it could be inferred that the possible mechanism of action of psoralidin is inhibition of EGFR. The study's findings demonstrate the efficacy and potential of coumestans in inhibiting EGFR receptors, providing prospective leads for cancer therapy.