Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Uncovering Anti-Cancer Potential of Clausena anisata Bioactives via Network Pharmacology-Based Target Prediction, Virtual Screening, Docking, ADMET Analysis, and Molecular Dynamics Simulation

Priyanka Nath Mazumdar
Assam down town University
Kaushik Kalita
Assam down town University
Partha Pratim Dutta
Assam down town University
Dhrubajyoti Gogoi
Assam down town University
Nikhita Nina Bora
Assam down town University
Manash Pratim Sarma
sarmamanash.p.1111@gmail.com
Assam down town University

Research Article

Keywords: Clausena anisata, Network pharmacology, EGFR, Molecular docking, Molecular dynamics simulation, MM-PBSA, Anisolactone, Triphasiol, Anticancer.

Posted Date: September 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11047499/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Introduction:

Despite advances in oncological interventions, multidrug resistance and severe toxicity drive the need for novel, cost-effective therapeutic agents. *Clausena anisata* is a phytochemically rich plant with antineoplastic properties; however, its targets and mechanisms remain poorly understood. To bridge this gap, we utilized integrated network pharmacology and computer-aided drug design (CADD) to elucidate the multitarget anticancer potential of *Clausena anisata* bioactives.

Materials and Methods

For this study, an integrated computational method (network pharmacology, molecular docking, ADMET profiling, molecular dynamics (MD) simulation, and MM-PBSA analysis) screened lead compounds of *Clausena anisata* for cancer treatment. 69 phytochemicals were screened using SwissADME. SwissTargetPrediction predicted protein targets, compared with cancer-associated genes from GeneCards. STRING and Cytoscape analyzed protein-protein interaction networks.

Results

ADMET screening identified seven pharmacokinetic, drug-like compounds adhering to Lipinski, Ghose, Veber, Egan, and Muegge rules, violation-free. Target prediction identified 247 targets; network analysis revealed EGFR as hub. Anisolactone (IMPHY000449) and Triphasiol (IMPHY001745) docked to EGFR, with LibDock scores 116.89 and 116.23, near AG4 (121.43). Complexes remained stable during 100 ns dynamics simulations, with RMSD, RMSF, and free energy landscape plots.MM-PBSA calculated energies of − 28.49 ± 1.25 kcal/mol and − 24.10 ± 1.18 kcal/mol. Compounds showed gut absorption and no PAINS flags.

Discussion

These data collectively indicate that Anisolactone and Triphasiol are promising lead compounds with anticancer properties. By effectively targeting the EGFR, these compounds may in turn inhibit the growth of other proteins involved in cancer progression.

Conclusion

The computational findings provide a robust foundational mechanism, thereby justifying additional in vitro and in vivo studies on *Clausena anisata*.

INTRODUCTION

Cancer represents a major global health challenge, accounting for over 20 million cases and approximately 9.7 million deaths in 2022. According to the WHO (2024), it is projected to reach 35 million by 2050. Although surgery, chemotherapy, immunotherapy, and targeted therapeutics have come a long way, multidrug resistance, off-target toxicity, and the high cost of treatment prevent satisfactory outcomes for patients (Zafar et al. 2025). *Clausena anisata* (wild) is widely used in traditional medicine for its medicinal properties, which are associated with antineoplastic effects. In traditional medicine, the medicinal properties of *Clausena anisata* (Wild.) are attributed to its anti-neoplastic effects, which is attributed to it, making it widely searched for novel, cost-effective, and less toxic anti-cancer agents around the world (Asma et al. 2022). *Clausena anisata* (Wild.) Hook.f. ex Benth. is a deciduous tropical shrub that is widely distributed throughout Africa, the Indian subcontinent, and some parts of Southeast Asia and has been traditionally used for its medicinal properties against fever, rheumatism, hypertension, malaria, and inflammatory diseases (Omara et al. 2022). The stems, roots, bark, and leaves are phytochemically rich in structurally diverse secondary metabolites, such as carbazole alkaloids (girinimbine, murrayamine A, ekeberginine, and clausamine H), furanocoumarins (imperatorin, osthol, bergapten, and xanthotoxin), clausine-series alkaloids (clausine B and clausine C), and phenylpropanoids (trans-anethole and estragole) (Songue et al. 2012).

Clausena species have a wide range of therapeutic potentials, with most of their constituents showing cytotoxic and antiproliferative activities. For example, the pyranomethyl-carbazole alkaloid girinimbine has been reported to induce apoptosis and cell cycle arrest in cancer models and has demonstrated potent acetylcholinesterase inhibitory activity with significant binding affinity, emphasizing its multi-target pharmacological potential (Lu et al. 2011; Mazumder et al. 2020). Similarly, imperatorin, a linear furanocoumarin, has been shown to inhibit the PI3K/AKT signalling pathway, while clausine B exhibited selective cytotoxicity against HeLa cancer cell lines (Wan Mohd Zain et al. 2009; Wang et al. 2017). In addition to these anti-cancer properties, it was recently discovered that extracts of Clausena species have anti-quorum sensing properties, indicating other mechanisms of action in the treatment of bacterial pathogenic diseases and their potential application in cancer treatment (Bouyahya et al. 2022). Although these phytoconstituents have been reported to have promising anti-cancer activities, the exact molecular targets, signalling pathways, and underlying mechanisms of action are not well known. To overcome this limitation, network pharmacology was first proposed by Andrew L. Hopkins to provide a broad method to elucidate how multiple compounds can work on multiple targets simultaneously (Hopkins 2008). It integrates cheminformatics, bioinformatics, and systems biology to allow the construction of compound-target-disease networks that can be analyzed to identify the most relevant targets in cancer development and response to therapy (Federico et al. 2022). Emerging technologies in machine learning-driven network pharmacology have recently improved the prediction of compound-target interactions and disease-associated genes, enabling more thorough and systematic insights into the therapeutic potential of Clausena-derived compounds (Kim et al. 2022). Furthermore, network pharmacology combined with other in silico computational methods, such as molecular docking, ADMET profiling, and

molecular dynamics simulations, has proven to be a powerful and increasingly widely used approach for the prioritization of lead bioactive agents against cancer from natural sources (Zhang 2024).

Clausena anisata has been used for a long time as an anticancer agent, and there have been some initial cytotoxic studies; however, the detailed molecular mechanism underlying its anticancer activity has not been investigated. In the present study, we sought to address this by systematically exploring the bioactive phytocompounds of *Clausena anisata* using an integrated computer-aided drug design (CADD) and network pharmacology approach. The results will be beneficial in providing scientific support for the traditional use of *Clausena anisata* and provide a foundation for future experiments to confirm or develop a drug.

MATERIALS AND METHODS:

This study employed an integrative computational approach to investigate the anti-cancer potential of the phytochemicals of *Clausena anisata*. The workflow included compound retrieval, ADME and drug-likeness screening, target prediction, and network pharmacology analysis.

2.1 Selection, Retrieval of Compounds and ADME Profiling

The study of *Clausena anisata* first started with the collection of phytochemicals present in the plant. These bioactive compounds were obtained from the Indian Medicinal Plants, Phytochemistry, and Therapeutics (IMPPAT) database, which is a carefully selected resource for specialized metabolites of Indian medicinal flora (Mohanraj et al. 2018). For each phytochemical, there is a unique IMPPAT identifier, and its corresponding SMILES string was extracted to facilitate downstream computational analysis. To evaluate the efficacy of these plant chemicals as medicines, all compounds were first subjected to complete Absorption, Distribution, Metabolism, and Excretion (ADME) profiling using the SwissADME web server (Daina et al. 2017). This tool predicts a broad dataset of physicochemical and pharmacokinetic properties for the entire plant-derived- library. Following this initial screening, a list of chemicals was filtered using strict rules to identify “drug-like” molecules. The library was filtered based on established pharmacological criteria, primarily Lipinski’s Rule of Five, as well as the Ghose, Veber, Egan, and Muegge rules (Egan et al. 2000; Ghose et al. 1999; Lipinski et al. 1997; Muegge et al. 2001; Veber et al. 2002). Specific constraints were applied to ensure clinical viability: Lipinski violation = 0, Ghose violation ≤ 1, Veber violation ≤ 1, Egan violation ≤ 1, and Muegge violation ≤ 1.

2.2 Target Prediction

The drug-like plant molecules identified in the previous step were submitted to the SwissTargetPrediction server to determine their most probable biological targets (Gfeller et al. 2014). This server predicts the targets of bioactive molecules by combining 2D and 3D similarity measures with a library of known active compounds. For this analysis, the study was restricted to *Homo sapiens*. The SMILES string of each molecule was examined to determine how closely these chemicals might target cancer proteins, enzymes, or receptors that could potentially be influenced by the phytochemicals. The

results were exported as CSV files and organized by their respective IMPPAT IDs for systematic mapping.

2.3 Network Pharmacology

Network pharmacology was employed to investigate the potential anticancer properties of this plant.. First, the known disease-target (D-T) associations for cancer were retrieved from two databases, the DisGeNET database and GeneCards database, while compound-target (C-T) interactions of *Clausena anisata* were obtained from target prediction analyses (Piñero et al. 2019; Stelzer et al. 2016). Common targets between the D-T and C-T datasets were identified using a Venn diagram generated using Venny 2.0, enabling the construction of a compound-target-disease (C-T-D) network to identify potential therapeutic targets. The network was constructed and analyzed using the NetworkAnalyzer tool in Cytoscape v3.10.4 (Assenov et al. 2008; Shannon et al. 2003). A GIFtS score < 60 was considered for disease-related pathway selection in GeneCards. The overlapping targets from the C-T-D network were further submitted to the STRING database with a high-confidence interaction score threshold (≥ 0.95) to construct a protein-protein interaction (PPI) network (Szklarczyk et al. 2023). The PPI network was visualized and analyzed using Cytoscape v3.10.4, and Gene Ontology (GO) enrichment analysis, including biological process, cellular component, and molecular function categories, was performed using STRING. Hub targets were identified based on the maximum degree of centrality using the cytoHubba plugin in Cytoscape software (Chin et al. 2014). This network pharmacology approach enables the systematic exploration of compound-target, disease-target, and compound-target-disease interactions to identify key targets and pathways associated with the anticancer potential of the studied plant.

2.4 Molecular Docking

The three-dimensional structures of nine phytocompounds (Table S1, Supplementary) from *Clausena anisata* were obtained from the IMPPAT database and prepared using the “Prepare Ligands” protocol in Discovery Studio 2022 (Mohanraj et al. 2018). The three-dimensional structure of the target *Clausena anisate* protein crystal (PDB ID: 2.69 Å) was obtained from the Protein Data Bank and prepared by removing water and heteroatoms, adding hydrogens, and assigning the CHARMm force field. The active site was defined as a sphere with a radius of 9.5 Å from the geometric centroid of the co-crystallized ligand. The LibDock protocol was set to generate 10 distinct poses for each ligand, with a docking tolerance of 0.25 (Rao et al. 2007). Conformational sampling was performed using the CAESAR method, generating a maximum of 255 conformers per ligand within an energy window of 20 kcal mol^{-1} above the global energy minimum (Li et al. 2007). For each docking run, 100 receptor hotspots were generated by placing polar and apolar probes (Pal et al. 2019a). To validate the docking protocol, the co-crystallized ligand was extracted and re-docked; an RMSD value of < 2.0 Å was considered acceptable, confirming the protocol’s ability to reliably reproduce the native binding mode (Diller and Merz 2001). The Docking Score for each pose was calculated using the following formula: Docking Score = − ((receptor interaction energy/ligand) + Ligand Internal Energy) (Pal et al. 2019b). The chemicals that showed strong

interactions with key parts of the active site and had high docking scores were chosen as the best candidates for further testing.

2.5 MD Simulation

The Molecular Dynamics Simulations (MDS) were conducted using the GROMACS 2024.4 simulation package on a Linux workstation equipped with an Intel Core i9-14900 processor and an NVIDIA GeForce RTX 4090 GPU. GROMACS is extensively employed for high-performance molecular simulations because of its efficient parallel computation capabilities on modern computing architectures (Abraham et al. 2015; Berendsen et al. 1995; Dutta et al. 2026). The ligand topologies were derived from SwissParam, based on the MMFF94 force field compatible with CHARMM (Zoete V et al., 2011), whereas the protein and ions were modelled using the CHARMM36m force field. Each system was placed in an orthorhombic simulation box filled with TIP3P water molecules and neutralized with appropriate numbers of Na^+ and Cl^- ions to maintain electrostatic neutrality (Jorgensen et al. 1983; Zoete et al. 2011). Energy minimization was performed using the Steepest Descent algorithm for 5000 steps to eliminate steric clashes and unfavourable contacts. Following minimization, the system underwent equilibration in two phases: an initial 200 ps NVT (constant number of Particles, Volume and Temperature) equilibration at 300.15 K, followed by 200 ps NPT (constant number of Particles, Pressure and Temperature) equilibration at 1 atm. The production MD simulation was executed for 100 ns using the leap-frog integrator with a time step of 2 fs, corresponding to 50,000,000 simulation steps. The system temperature was maintained at 300 K using a velocity-rescale (V-rescale) thermostat to ensure proper canonical ensemble sampling, whereas the pressure was controlled at 1 bar using a Parrinello-Rahman barostat with isotropic pressure coupling (Bussi et al. 2007; Parrinello and Rahman 1981). Long-range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method, which provides an accurate treatment of electrostatic forces under periodic boundary conditions. A cutoff distance of 1.2 nm was applied for both electrostatic and van der Waals interactions using the Verlet cutoff scheme for neighbour searching. All covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm, which enabled stable simulations with larger integration time steps. Additionally, center-of-mass motion removal was applied linearly throughout the simulation to prevent the system drift (Essmann et al. 1995; Hess et al. 1997). The coordinates and energy values were recorded at regular intervals of 10 ps during production runs. The resulting trajectories were subsequently analyzed to evaluate the structural stability and dynamic behaviour of the protein-ligand complexes using standard parameters, including the Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration (Rg), solvent-accessible surface area (SASA), and Hydrogen Bond (HB) occupancy analysis.

The conformational dynamics and binding energetics of the ligand-PIK3CA complexes were characterized using PCA, which was performed by analyzing the Cα fluctuations of the protein trajectories in the complexes (Castelli et al. 2024). The resulting Gibbs free energy was then plotted as $\Delta G = -RT \ln P(q)$ against the first two principal components (PC1 and PC2) to create a free energy landscape (FEL) to reveal the most stable conformational states (Valdés-Tresanco et al. 2021). Finally,

binding free energies were computed using the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) methodology as part of the gmx mmpbsa toolset, which consists of two components: a gas-phase energy term (ΔEMM) and a solvation energy term (ΔGsol) - TΔS was taken as the average of the two (Xu et al. 2025).

RESULTS

3.1 ADME Profiling and Drug-Likeness Assessment

The pharmacokinetic and drug-likeness properties of the 69 *Clausena anisata* phytocompounds were examined for their suitability as potential drug candidates (Tables S2-S5). Physicochemical analysis revealed that all the compounds satisfied the major drug-likeness criteria with minimal violations, such as Lipinski, Ghose, Veber, Egan, and Muegge criteria, indicating their good oral bioavailability potential. The molecular weights (MWs) of the compounds were in the range from to 162.14-470.52 g/mol, which falls within the typical range for drug-like molecules. The topological polar surface area (TPSA) ranged from 12.53 to 118.73 Å², which showed a suitable balance of polarity and permeability. The compounds with lower TPSA values (< 25 Å²), such as IMPHY007840, IMPHY011542, IMPHY011890, IMPHY012586, and IMPHY014906, have relatively better permeabilities through the membranes, whereas the compounds with relatively higher TPSA values (> 100 Å²), such as IMPHY006381 and IMPHY010516, may have relatively poorer permeabilities but relatively better aqueous interactions. The partition coefficient and solubility profile also suggested good lipophilicity for the compounds, with consensus logP values between 1.51 and 4.41, below the threshold generally considered to be poor absorption. Most phytocompounds showed moderate solubility, while IMPHY007161 and IMPHY005587 were soluble, indicating better absorption and systemic availability. However, compounds such as IMPHY000449, IMPHY003757, IMPHY007590, IMPHY002840, and IMPHY013007 were poorly soluble and may require formulation optimization to improve bioavailability, despite their good permeability. All compounds showed high gastrointestinal absorption (GI), suggesting good oral drug potential in pharmacokinetic prediction. Most low-molecular-weight and lipophilic compounds, including IMPHY007161, IMPHY008913, IMPHY004940, IMPHY003993, IMPHY013007, and IMPHY016921, were predicted to cross the BBB. BBB non-permeant predictions were made for highly oxygenated compounds (IMPHY003373, IMPHY006381, and IMPHY010516), perhaps because they are more polar and complex. Moreover, very few compounds (IMPHY006381, IMPHY010516, IMPHY007590, IMPHY002840, and IMPHY001745) were predicted to be substrates of P-glycoprotein (P-gp), which is a possible efflux pathway that could decrease the intracellular concentration. Cytochrome P450 (CYP450) inhibition profiling revealed differential metabolic liabilities among the phytocompounds. IMPHY000449, IMPHY003757, IMPHY013007, and IMPHY004940 were predicted to inhibit multiple CYP isoforms (CYP1A2, CYP2C19, CYP2C9, and CYP3A4), which may be a concern regarding drug-drug interactions during metabolism. In contrast, compounds such as IMPHY003373, IMPHY010516, IMPHY007421, and IMPHY011647 exhibited minimal CYP inhibition, indicating a better metabolic profile. The log Kp values of skin permeability ranged from − 4.26 to − 8.29 cm/s and indicated that this product had generally low

dermal permeability, which was favourable for oral administration instead of dermal. Nearly all compounds had no Lipinski violations, a favourable bioavailability score (0.55), no PAINS alerts, and low Brenk alerts, suggesting minimal concerns for nonspecific assay interference and toxicophoric features. The synthetic accessibility (SA) scores ranged from 2.39 to 6.51, indicating that the majority of these compounds are moderately accessible for chemical synthesis, except for structurally complex compounds such as IMPHY003373, IMPHY006381, and IMPHY010516, which likely require more complex synthetic routes. The ADME assessment identified several phytocompounds with balanced physicochemical properties, acceptable solubility, high GI absorption, and good drug-likeness, which were further evaluated using molecular docking and molecular dynamics simulation studies and experimental validation studies.

3.2 Network Pharmacology

The multi-component, multi-target, and multi-pathway interactions of the selected phytocompounds with cancer-associated targets were explored using network pharmacology analysis with Cytoscape software. A set of compounds with desirable drug-like and synthetic accessibility properties were screened (Table S5), and seven phytocompounds were identified: IMPHY000449, IMPHY003757, IMPHY005428, IMPHY001745, IMPHY008465, IMPHY000879, and IMPHY009650. Lipinski, Ghose, Veber, Egan and Muegge filters were used to check the selected compounds and none of them showed any violation thereby indicating their potential as orally bioavailable lead molecules. All of the selected compounds also had an acceptable synthetic accessibility score (SA: 2.90–4.14) and were without PAINS alerts, and with a common bioavailability score of 0.55.

Genes associated with the compounds were gathered and overlapping with the cancer-associated genes from the GeneCards database with a GIFtS score > 60 were selected as therapeutic targets that could be related to cancer. The overlapping genes were identified and visualized using the Venny tool, and 247 common targets were identified between the genes associated with phytocompounds and cancer genes (Figure c). These genes, which were present in multiple copies, were then analyzed for other protein-protein interactions and pathway enrichment analyses and were selected as putative therapeutic targets.

The compound-target (C-T) interaction network, disease-target (D-T) interaction network, and compound-target-disease (C-T-D) interaction network were built and analyzed in Cytoscape using the Network Analyzer plugin to search for important topological features (Figure S1 a-c). The C-T-D network integrated 839 nodes with an average of 7.17 neighbors, indicating significant connectedness between phytocompounds, target proteins, and disease-associated pathways (Table S6). The network had two connected components, with a clustering coefficient of 0, network diameter and radius of 1, and a characteristic path length of 1, suggesting that the nodes were tightly interconnected and communication was efficient. The integrated network exhibited higher connectivity and complexity than the individual networks based on C-T and D-T, indicating the multi-target pharmacological effects of the selected phytocompounds.

The common genes were then submitted to the STRING database with a high confidence interaction score of > 0.95 to further investigate the interaction of proteins and enriched signaling pathways with strong and reliable protein-protein interactions. The cancer-associated pathways were enriched, as revealed by Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis (Figure b), such as pathways in cancer, breast cancer, prostate cancer, proteoglycans in cancer, central carbon metabolism in cancer, microRNAs in cancer, PD-L1 expression in cancer, and PD-1 checkpoint pathway in cancer. Gene Ontology (GO)-based biological process enrichment also showed significant enrichment in events related to phosphorylation, such as protein phosphorylation, tyrosine phosphorylation, serine phosphorylation, receptor protein kinase signaling pathways, amino acid modification, and cellular response to chemical stress (Figure d). Based on these results, it can be inferred that the pick phytocompounds could have anticancer activity by regulating the kinase-mediated signaling cascades and oncogenic regulatory networks (Fig. 1).

The CytoHubba plugin was then used in Cytoscape to identify hub genes based on topological parameters, such as degree, betweenness centrality, closeness centrality, radiality, and shortest path length (Table S7). Several phytocompounds had the highest degree values (degree = 39), indicating that they were well connected with several targets and signaling pathways. In the protein target list, EGFR was the main hub gene with the highest degree among the disease-associated genes (degree = 31) and with relatively high betweenness centrality (0.020693) and closeness centrality (0.829787) values, indicating that this protein plays a crucial role in cancer progression and signaling regulation. Epidermal Growth Factor Receptor (EGFR) was identified as the target of choice for molecular docking analysis because of its high degree-based connectivity, high centrality in the network, and known involvement in tumorigenesis. EGFR plays a vital role in cell proliferation, migration, survival, resistance signaling, and angiogenesis in various types of cancers. Its high level of association with phosphorylation-related biological processes and its high enrichment in oncogenic pathways further validated its selection as the main molecular target. Thus, molecular docking and molecular dynamics simulation studies were performed against the EGFR to verify the binding affinity and stability of the selected phytocompounds and understand how they could possibly inhibit EGFR.

3.3 Molecular Docking

The binding potential of the selected phytocompounds to the identified hub target from network pharmacology was further validated by molecular docking using the LibDock module against EGFR. Before docking, the binding site was identified prior to the co-crystallized ligand, AG4, and the docking protocol was validated by selecting the same active binding pocket occupied by AG4 to ensure that the docking was reliable and reproducible. Seven phytocompounds (IMPHY000449, IMPHY001745, IMPHY003757, IMPHY008465, IMPHY009650, IMPHY000879, and IMPHY005428), which were found to be drug-like and suitable for network pharmacology analysis, were docked into the validated active site of EGFR and compared with the reference ligand AG4 (Table S8 and Figure S1). The docking results showed that the phytocompounds exhibited good binding affinities towards the EGFR catalytic pocket, ranging from 81.24 to 116.89, which represents the stability of the ligand-receptor interactions. Of the

screened phytocompounds, one compound (c1 = IMPHY000449) had a libdock score of 116.89, which was very close to the libdock score of the reference compound (AG4 = 121.426), indicating a high binding potential. IMPHY000449 formed a conventional hydrogen bond with PHE832 (1.94 Å) and carbon hydrogen interactions with GLY772 and ASP831, resulting in stable anchoring in the active pocket. Likewise, IMPHY001745 (c2) had a high docking score of 116.229 and formed several conventional hydrogen bonds with important residues such as ASP831, ARG817, and THR766 and carbon hydrogen interactions with residue ASN818, indicating good binding geometry and interaction stability.

Similarly, IMPHY003757 (c3) had a LibDock score of 114.642 and had hydrophobic interactions with MET769, THR766, LYS721 and π-sigma interactions with LEU820, as well as alkyl interactions, which further increased hydrophobic stabilization within the EGFR binding cavity. IMPHY008465 (c4) also had a high binding affinity (LibDock score: 107.151) through multiple conventional hydrogen bond interactions with THR766 and LYS721 and carbon hydrogen bond interactions with ASP831 and ILE765, indicating a good fit in the receptor active site. The LibDock values of 95.40 and 92.43 were obtained for IMPHY009650 (c5) and IMPHY000879 (c6), respectively, which had moderate docking scores. IMPHY009650 formed conventional hydrogen bonds with LYS721 and THR766, carbon hydrogen interactions with ILE720, and hydrophobic alkyl interactions with ALA719, which are indicative of stable receptor association. Likewise, IMPHY000879 formed hydrogen bonds and engaged in carbon hydrogen bonding with ASP831, ILE765, and THR766. IMPHY005428 (c7) had the lowest docking score (81.236); however, it still had stable hydrogen bonding interactions with LYS721 and THR766, carbon hydrogen bonds with GLU738 and ILE765, and π-sigma bonds with LEU820, indicating that it would still be able to bind within the active region.

The co-crystallized ligand AG4 (c8) had the highest LibDock score (121.426) and had key interactions with MET769, CYS751, and the residues on the binding region (AQ4999), which demonstrated the feasibility of the docking approach and ensured the selection of the appropriate active pocket. Importantly, several phytocompounds shared common interacting residues with AG4 (LYS721, THR766, MET769, and ASP831), suggesting similar docking of the interacting residues from the validated binding site and possibly similar EGFR-inhibitory modes.

Overall, the docking analysis indicated that IMPHY000449, IMPHY001745, and IMPHY003757 were more strongly bound to the EGFR active site than the reference ligand AG4, in terms of binding affinity and interaction stability. Their strong hydrogen bonding and hydrophobic interactions with catalytically important residues show their potential as promising EGFR-targeting anticancer agents, which deserve further validation via molecular dynamics simulation and MM-PBSA binding free energy analysis.

Table 1
Result of molecular docking

Ligand ID*	LibDock Score	Bond Type	Amino Acid Involves	Distance (Å)
(c1) MPHY000449	116.89	Conventional Hydrogen Bond	PHE832	1.94276
		Carbon Hydrogen Bond	GLY772	2.00601
		Carbon Hydrogen Bond	ASP831	2.21185
		Pi-Alkyl	LYS721	3.94871
		Pi-Alkyl	LYS721	3.96805
(c2) IMPHY001745	116.229	Carbon Hydrogen Bond	ASP831	1.74562
		Conventional Hydrogen Bond	ASP831	1.93648
		Conventional Hydrogen Bond	ARG817	2.15219
		Conventional Hydrogen Bond	THR766	2.30241
		Carbon Hydrogen Bond	ASN818	2.49347
(c3) IMPHY003757	114.642	Carbon Hydrogen Bond	MET769	2.18034
		Conventional Hydrogen Bond	THR766	2.27383
		Conventional Hydrogen Bond	LYS721	2.72945
		Pi-Sigma	LEU820	2.76597
		Alkyl	LEU820	3.68272
(c4) IMPHY008465	107.151	Conventional Hydrogen Bond	THR766	2.02847
		Carbon Hydrogen Bond	ASP831	2.27778
		Conventional Hydrogen Bond	THR766	2.61779
		Conventional Hydrogen Bond	LYS721	2.9928
		Carbon Hydrogen Bond	ILE765	3.04835

Ligand ID*	LibDock Score	Bond Type	Amino Acid Involves	Distance (Å)
(c5) IMPHY009650	95.4008	Carbon Hydrogen Bond	IMPHY009650	2.40683
		Conventional Hydrogen Bond	LYS721	2.72566
		Conventional Hydrogen Bond	THR766	2.78501
		Carbon Hydrogen Bond	ILE720	2.79983
		Alkyl	ALA719	2.81712
(c6) IMPHY000879	92.4336	Carbon Hydrogen Bond	ASP831	2.21411
		Conventional Hydrogen Bond	THR766	2.47042
		Conventional Hydrogen Bond	LYS721	2.75336
		Conventional Hydrogen Bond	THR766	3.06485
		Carbon Hydrogen Bond	ILE765	3.09308
(c7) IMPHY005428	81.2361	Carbon Hydrogen Bond	GLU738	2.31951
		Conventional Hydrogen Bond	LYS721	2.64034
		Pi-Sigma	LEU820	2.74803
		Carbon Hydrogen Bond	ILE765	2.85668
		Conventional Hydrogen Bond	THR766	2.95035
(c8) AG4 (Reference)	121.426	Conventional Hydrogen Bond	MET769	1.78155
		Carbon Hydrogen Bond	AQ4999	1.78538
		Carbon Hydrogen Bond	AQ4999	2.12338
		Carbon Hydrogen Bond	CYS751	2.51521
		Carbon Hydrogen Bond	AQ4999	2.58773

3.4 Molecular Dynamics Simulation

MDS was performed for all ligand-protein complexes, and the data were compared with the reference lead. MDS parameters such as RMSD, RMSF, ROG, and SASA were estimated and calculated for the best

hits. The outcome of this is now shown below.

The structural stability of the EGFR-phytocompounds (c1-c7) and EGFR-AG4 (c8_ref) complex was assessed by a 100 ns molecular dynamics (MD) simulation of the complexes using RMSD analysis (Fig. 3). The RMSD values of the lowest-energy complex (c8_ref) showed a high degree of stability, with RMSD fluctuations around the active site of less than 0.7 nm during the simulation. A significantly greater RMSD was observed after equilibration for c1 (IMPHY000449), c2 (IMPHY001745), and c6 (IMPHY000879), which have relatively stable trajectories in the binding pocket of EGFR, suggesting that the interactions and conformational stability remain intact. c4 (IMPHY008465) and c7 (IMPHY005428) had moderate deviations with the beginning of a relatively stable conformation at later stages of the simulation (~ 1.5–1.9 nm). In general, the RMSD data suggest that IMPHY000449, IMPHY001745, and IMPHY000879 would form relatively stable complexes with EGFR, with behaviour more similar to that of the reference ligand. Thus, the RMSD data concur with the molecular docking results and indicate that these three compounds would likely possess favourable binding persistence for the length of the simulation.

RMSF: The root mean square fluctuation (RMSF) analysis was performed to assess the flexibility of the residues and local conformational changes that occurred in the EGFR-ligand complexes throughout the 100-ns molecular dynamics simulation (Fig. 4). Overall, there were relatively low fluctuations (< 0.5 nm) in most residues, suggesting that the protein backbone was relatively stable upon ligand binding. The reference complex AG4 (c8_ref) exhibited relatively small amplitude of residue fluctuations throughout the trajectory, reflecting the validated stable interactions in the active site. The phytocompound complexes c1 (IMPHY000449), c2 (IMPHY001745), and c6 (IMPHY000879) exhibited lower RMSF values and fluctuation patterns closer to those of the reference ligand, indicating better stabilization of amino acid residues and reduced local flexibility. The fluctuation profiles of c4 (IMPHY008465) and c7 (IMPHY005428) were intermediate with moderate flexibility, while they displayed comparatively higher fluctuations at several residue regions towards the terminal residues, which might be indicative of increased conformational mobility and decreased rigid-binding interactions. The high variations seen primarily in the terminal loop regions are quite normal, as they are very flexible regions and do not necessarily signify instability of the binding pocket.

ROG: To investigate the compactness and conformational stability of EGFR in complex with the phytocompounds along the 100 ns molecular dynamics simulation, the radius of gyration (Rg) analysis was carried out (Fig. 5). In general, all complexes showed relatively constant Rg trajectories between ~ 2.1 and 2.5 nm, which corresponded to the maintenance of protein compactness during ligand binding. An overall compact structure with moderate fluctuations was observed in the reference complex AG4 (c8_ref). The differences in the Rg profiles were relatively small after equilibration for the phytocompounds c4 (IMPHY008465), c6 (IMPHY000879), and c7 (IMPHY005428), which showed relative stability and good structural compactness with moderate variations over time. In contrast, the fluctuations of c3 (IMPHY003757) and c5 (IMPHY009650) were relatively high in the initial part of the simulation, decreasing to a more stable state later, suggesting that transient conformational

rearrangements occurred prior to reaching the equilibrium state. The results show that most complexes have Rg values that slowly return to their original values, indicating that ligand binding did not cause significant unfolding or destabilization of EGFR. Overall, the RMSD and RMSF results are in agreement with the Rg results, where IMPHY000449, IMPHY001745, and IMPHY000879 are relatively stable complexes with the lowest RMSD, RMSF, and Rg values throughout the simulation.

SASA:

The changes in the solvent-accessible surface area (SASA) of EGFR upon binding to the ligands during the 100 ns molecular dynamics (MD) simulation period were evaluated (Fig. 6). Overall, the fluctuations in the SASA of all EGFR-ligand complexes ranged from moderate (~ 170–210 nm^2), suggesting the integrity of the complexes during the simulation time. The reference complex AG4 (c8_ref) showed relatively low and small fluctuations in the SASA values, indicating the consistency of solvent exposure and binding at the experimentally verified active site. The other phytocompound complexes, c4 (IMPHY008465), c5 (IMPHY001380), and c6 (IMPHY000879), exhibited reasonably stable SASA curve values, with the values for c4 and c6 reaching equilibrium after a while. The other phytocompound complexes, c4 (IMPHY008465), c5 (IMPHY001380), and c6 (IMPHY000879), also showed acceptable SASA curve values stability with moderate fluctuations during simulation, whereas the values for c4 and c6 reached equilibrium after a while. Conversely, c3 (IMPHY003757) and c5 (IMPHY009650) had relatively more SASA variations, especially in the initial and intermediate stages, implying higher conformational changes and temporary surface exposure of the protein surfaces. The overall decrease and stabilization in the SASA values for most complexes suggest that ligand binding favors compact conformations without significant destabilization of the EGFR structure.

PCA and FEL

MD trajectories were used to perform Principal Component Analysis (PCA) and Gibbs free energy landscape (FEL) analyses to further investigate the conformational dynamics and stability of the EGFR-ligand complexes (Figs. 7–9). The distribution of the variance across the principal motions (Fig. 8) indicated that the first few principal motions had a large variance compared to the variance of the remaining principal motions, which diminished quickly with the principal motion. This indicates that the dominant conformational changes in the protein-ligand systems were dominated by a few collective motions. The 2D projection on PC1 and PC2 (Fig. 7) showed that the complexes exhibited different sampling behaviours. For the reference complex (c8_ref), the conformational space was tightly packed with a low deviation from the reference and high dynamic stability during the simulation. The conformational distributions of c1, c2, c4, and c6 were fairly restrained, indicating relatively stable conformational changes and ligand accommodation in the EGFR binding pocket. In contrast, c3, c5, and c7 were found in wider conformational regions, indicative of a more flexible conformation and more structural changes during the simulation. The Gibbs free energy landscapes (Fig. 9) also supported these observations: low-energy basins (blue) indicated energetically favourable conformational states. For the reference complex and c1 (IMPHY000449), c2 (IMPHY001745), and c4 (IMPHY008465), low-

energy minima were found, and they were well-defined and concentrated, indicating energetically stable equilibrium conformations and favourable protein-ligand stability. In contrast, c3, c5, and c7 had several separate minima along the energy landscape, indicating metastable conformational states and increased flexibility. In general, the conformational analysis revealed that the selected phytocompounds maintained conformational behaviour similar to that of the reference ligand, suggesting their ability to inhibit EGFR with stability.

MM-PBSA:

We performed MM-PBSA calculations to determine the binding free energies of these selected phytocompounds and our reference ligand to determine how well these compounds bind to EGFR compared to our reference ligand (Table 2, Fig. 10, and Supplementary Figure S5).The reference ligand (C8) demonstrated the highest overall binding energy (ΔTOTAL = − 37.77 ± 1.60 kcal/mol), which was mostly due to the strong van der Waals interactions with the protein (ΔVDWAALS = − 47.07 ± 1.53 kcal/mol), further supporting its stable binding to the EGFR active site. The binding energies of all phytocompounds against EGFR are listed in Table 3. C3 (IMPHY003757) and C1 (IMPHY000449) compounds had the most favourable binding energies, namely − 28.73 ± 1.26 kcal/mol and − 28.49 ± 1.25 kcal/mol, respectively, suggesting a stable and energetically favourable interaction with EGFR. Notably, hydrophobic stabilization by van der Waals interactions was stronger in C1 than in C2 (− 36.77 ± 0.61 kcal/mol and − 13.40 ± 0.67 kcal/mol, respectively), indicating the significance of hydrogen bonding and polar contacts in stabilizing the complex. Complexes C6 (IMPHY000879) and C2 also showed very high binding affinities with ΔTOTAL values of − 23.86 ± 1.25 and − 24.10 ± 1.18 kcal/mol respectively, which indicates that the complexes are stable in accommodating the ligand during the simulation process. In contrast, C4 (IMPHY008465), C5 (IMPHY009650), and C7 (IMPHY005428) exhibited relatively less favourable binding energies, indicating relatively weaker interactions with the receptor. The gas-phase interaction energies (ΔGGAS) were favourable for most complexes, but their solvation energies (ΔGSOLV) were positive, representing the energetic cost of exposure to the solvent. The results obtained from the MM-PBSA energy decomposition corroborated the molecular dynamics results (Fig. 10 and Supplementary Figure S5), indicating that the phytocompounds C1, C2, and C3 are the most energetically stable, with binding energies very close to that of the reference inhibitor, justifying the potential of the compounds as candidates to target EGFR.

Table 2
MM-PBSA binding free energy summary of compounds (kcal/mol)

Complex	ΔVDWAALS	ΔEEL	ΔENPOLAR	ΔGGAS	ΔGSOLV	ΔTOTAL
C1	-36.77 ± 0.61	-0.60 ± 0.22	-3.86 ± 0.03	-37.37 ± 0.68	8.88 ± 1.05	-28.49 ± 1.25
C2	-28.93 ± 0.61	-13.40 ± 0.67	-3.55 ± 0.38	-42.33 ± 0.92	18.23 ± 0.74	-24.10 ± 1.18
C3	-31.95 ± 1.14	-6.22 ± 0.47	-3.51 ± 0.03	-38.17 ± 1.25	9.44 ± 0.20	-28.73 ± 1.26
C4	− 23.38 ± 2.79	-3.82 ± 1.41	-2.85 ± 0.54	-27.20 ± 3.13	7.81 ± 1.86	-19.39 ± 3.64
C5	-17.91 ± 1.14	-0.95 ± 0.73	-2.30 ± 0.20	-18.86 ± 1.36	2.75 ± 0.90	-16.11 ± 1.63
C6	-25.39 ± 1.22	-8.68 ± 0.14	-3.18 ± 0.00	-34.07 ± 1.25	10.21 ± 0.13	-23.86 ± 1.25
C7	-23.92 ± 0.60	-5.82 ± 0.30	-2.68 ± 0.01	-29.74 ± 0.68	8.97 ± 0.58	-20.77 ± 0.90
C8	-47.07 ± 1.53	-8.37 ± 0.25	-4.80 ± 0.06	-55.45 ± 1.55	17.67 ± 0.39	-37.77 ± 1.60

DISCUSSION

In the present study, a computationally integrated approach was used to explore the anticancer activity of *Clausena anisata*. The phytocompounds obtained after performing ADMET analysis and drug likeness were found to be suitable for further evaluation as potential drug candidates. They were also not associated with any major rule violations or PAINS alerts.

EGFR emerged as one of the key hub targets identified by network pharmacology analysis that was associated with cancer progression and therefore is a therapeutic target. Molecular docking studies showed that Anisolactone and Triphasiol had a strong binding affinity with the EGFR active site by forming stable hydrogen bonds and hydrophobic interactions with the active site of EGFR, similar to the reference ligand.

In addition, the stability of the protein–ligand complex and the favourable binding free energy of the complexes were validated by performing molecular dynamics simulations and MM-PBSA analysis during the simulation time. The overall results of these findings show that *Clausena anisata* extract contains promising phytochemicals that show potential anticancer activity against EGFR. However, these computational predictions require further in vitro and in vivo studies to verify them and determine their therapeutic efficacy and safety.

Conclusion

The integrated computational workflow, which utilized network pharmacology, molecular docking, molecular dynamics simulation, and MM-PBSA binding free energy analysis, clearly demonstrated the therapeutic potential of the selected phytocompounds in EGFR-associated cancer pathways. EGFR is one of the main hub targets with high network connectivity and pathway relevance, which was selected for molecular docking and further validation using network pharmacology analysis. Docking results showed that the selected phytocompounds had strong binding with the validated binding pocket AG4 of EGFR and favorable LibDock scores, showing favourable interactions with key residues, such as LYS721, THR766, ASP831, MET769, and PHE832 of EGFR. Next, molecular dynamics simulations for 100ns further confirmed the stability of the protein–ligand interaction with RMSD, RMSF, Rg, SASA, PCA, and Gibbs free energy landscape analyses showing that the simulations of three complexes (c1, c2, and c4) had dynamic behaviour close to that of the reference ligand, while the remaining complexes showed poor behaviour. The binding free energy of MM-PBSA corroborated these results, showing that c1 (anisolactone) and c2 (triphasiol) had favorable binding free energies of − 28.49 ± 1.25 and − 24.10 ± 1.18 kcal/mol, respectively, which demonstrated energetically favourable binding to EGFR. Importantly, the analysis of hydrogen bond interactions during the 100 ns molecular dynamics simulation (Figure S4) revealed that some hydrogen bonding patterns persisted and remained high throughout the simulation time, especially for c2 (Triphasiol), which showed a high occupancy of hydrogen bonds for longer periods of the simulation, suggesting a stronger interaction persistence in the EGFR binding cavity. Although c3 had a slightly more favourable binding free energy, it had more extensive conformational fluctuations and multiple metastable states, suggesting lower dynamic stability than the other binders. Together, the docking results, molecular dynamics stability parameters, hydrogen bond persistence, PCA/FEL conformational stability, and MM-PBSA energetics identified two compounds (anisolactone (c1) and triphasiol (c2)) as potential EGFR inhibitory phytocompounds for further in vitro and in vivo studies against EGFR-driven cancer.

Abbreviations

CADD
Computer-Aided Drug Design
IMPPAT
Indian Medicinal Plants, Phytochemistry and Therapeutics Database
SMILES
Simplified Molecular Input Line Entry System
ADME
Absorption, Distribution, Metabolism and Excretion
CSV
Comma-Separated Values
CT
Compound-Target

DT
Disease-Target
CTD
Compound-Target-Disease
PPI
Protein-Protein Interaction
KEGG
Kyoto Encyclopedia of Genes and Genomes
GO
Gene Ontology
PDB
Protein Data Bank
kcal mol^{-1}
Kilocalories per Mole per Molecule (binding free energy unit)
RMSD
Root Mean Square Deviation
PAINS
Pan-Assay Interference Compounds
MM/PBSA
Molecular Mechanics/Poisson-Boltzmann Surface Area
REF
Reference Ligand
ΔGGAS
Gas-Phase Free Energy (Gas-Phase Interaction Energy)
RMSF
Root Mean Square Fluctuation

Declarations

Acknowledgments

The authors are thankful to Assam down town University, Guwahati for providing necessary support in carrying out this work. Biorender.com was used for the design of the figures.

Author contributions

PNM: Conceived and designed the study, performed the experiments, analyzed the data and drafted the manuscript. KK: Conceived and designed the study, analyzed the data, and drafted the manuscript. PPD: Contributed to methodology development, data acquisition, and data analysis. NNB contributed to data curation, literature review, and preparation of figures/tables. DG and MPS: Supervised the project, provided critical intellectual input, and revised the manuscript for important content.

All authors read and approved the final version of the manuscript.

Ethics declarations

The authors confirm the originality of the article. Institutional Review Board approval is not required.

Funding

This research received no external funding.

Data Availability Statement

All data generated and analyzed during this study, including network pharmacology outputs, docking results, ADMET predictions, and molecular dynamics simulation data, are included in this article and its supplementary material.

Declaration of patient consent

Patient's consent was not required as there are no patients in this study

Conflicts of Interest

The authors declare no conflicts of interest.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

The authors declare that no AI-assisted technology was used for content generation or image manipulation. Paperpal was used only for language polishing, including grammar and sentence correction.

References

1. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015 Sep;1–2:19–25.
2. Asma ST, Acaroz U, Imre K, Morar A, Shah SRA, Hussain SZ, et al. Natural Products/Bioactive Compounds as a Source of Anticancer Drugs. Cancers (Basel). 2022 Dec 15;14(24):6203.
3. Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M. Computing topological parameters of biological networks. Bioinformatics. 2008 Jan 15;24(2):282–4.
4. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. 1995 Sep;91(1–3):43–56.
5. Bouyahya A, Chamkhi I, Balahbib A, Rebezov M, Shariati MA, Wilairatana P, et al. Mechanisms, Anti-Quorum-Sensing Actions, and Clinical Trials of Medicinal Plant Bioactive Compounds against Bacteria: A Comprehensive Review. Molecules. 2022 Feb 22;27(5):1484.

6. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J. Chem. Phys. 2007 Jan 7;126(1).
7. Castelli M, Magni A, Bonollo G, Pavoni S, Frigerio F, Oliveira ASF, et al. Molecular mechanisms of chaperone-directed protein folding: Insights from atomistic simulations. Protein Science. 2024 Mar 26;33(3).
8. Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst. Biol. 2014 Dec 8;8(S4):S11.
9. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci. Rep. 2017 Mar 3;7(1):42717.
10. Diller DJ, Merz KM. High throughput docking for library design and library prioritization. Proteins: Structure, Function, and Genetics. 2001 May 1;43(2):113–24.
11. Dutta N, Neog J, Bari A, Gogoi D, Singh SK, Sarma D. Silver catalyzed synthesis of N-arylated quinazolin-4(3H)-ones: Biological screening and in silico insights against Mycobacterium tuberculosis. J. Mol. Struct. 2026 Aug;1366:146045.
12. Egan WJ, Merz , Kenneth M., Baldwin JJ. Prediction of Drug Absorption Using Multivariate Statistics. J. Med. Chem. 2000 Oct 1;43(21):3867–77.
13. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J. Chem. Phys. 1995 Nov 15;103(19):8577–93.
14. Federico A, Fratello M, Scala G, Möbus L, Pavel A, del Giudice G, et al. Integrated Network Pharmacology Approach for Drug Combination Discovery: A Multi-Cancer Case Study. Cancers (Basel). 2022 Apr 18;14(8):2043.
15. Gfeller D, Grosdidier A, Wirth M, Daina A, Michielin O, Zoete V. SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res. 2014 Jul 1;42(W1):W32–8.
16. Ghose AK, Viswanadhan VN, Wendoloski JJ. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. J. Comb. Chem. 1999 Jan 12;1(1):55–68.
17. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 1997 Sep;18(12):1463–72.
18. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat. Chem. Biol. 2008 Nov 20;4(11):682–90.
19. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 1983 Jul 15;79(2):926–35.
20. Kim Y, Park J-H, Cho Y-R. Network-Based Approaches for Disease-Gene Association Prediction Using Protein-Protein Interaction Networks. Int. J. Mol. Sci. 2022 Jul 3;23(13):7411.
21. Li J, Ehlers T, Sutter J, Varma-O'Brien S, Kirchmair J. CAESAR: A New Conformer Generation Algorithm Based on Recursive Buildup and Local Rotational Symmetry Consideration. ChemInform. 2007 Dec 4;38(49).

22. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 1997 Jan;23(1–3):3–25.
23. Lu S-H, Wu JW, Liu H-L, Zhao J-H, Liu K-T, Chuang C-K, et al. The discovery of potential acetylcholinesterase inhibitors: A combination of pharmacophore modeling, virtual screening, and molecular docking studies. J. Biomed. Sci. 2011 Dec 21;18(1):8.
24. Mazumder K, Biswas B, Raja IM, Fukase K. A Review of Cytotoxic Plants of the Indian Subcontinent and a Broad-Spectrum Analysis of Their Bioactive Compounds. Molecules. 2020 Apr 20;25(8):1904.
25. Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Chand RPB, Aparna SR, Mangalapandi P, et al. IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics. Sci. Rep. 2018 Mar 12;8(1):4329.
26. Muegge I, Heald SL, Brittelli D. Simple Selection Criteria for Drug-like Chemical Matter. J. Med. Chem. 2001 Jun 1;44(12):1841–6.
27. Omara T, K. A, J. V, Kagoya S. Clausena anisata (Willd.) Hook.f. ex Benth. (Rutaceae): ethnomedicinal uses, phytochemistry, pharmacological activities, toxicity, and clinical application. Traditional Medicine Research. 2022;7(6):51.
28. Pal S, Kumar V, Kundu B, Bhattacharya D, Preethy N, Reddy MP, et al. Ligand-based Pharmacophore Modeling, Virtual Screening and Molecular Docking Studies for Discovery of Potential Topoisomerase I Inhibitors. Comput. Struct. Biotechnol. J. 2019a Jan 28;17:291–310.
29. Pal S, Kumar V, Kundu B, Bhattacharya D, Preethy N, Reddy MP, et al. Ligand-based Pharmacophore Modeling, Virtual Screening and Molecular Docking Studies for Discovery of Potential Topoisomerase I Inhibitors. Comput. Struct. Biotechnol. J. 2019b Jan 28;17:291–310.
30. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. 1981 Dec 1;52(12):7182–90.
31. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2019 Nov 4;
32. Rao SN, Head MS, Kulkarni A, LaLonde JM. Validation Studies of the Site-Directed Docking Program LibDock. J. Chem. Inf. Model. 2007 Nov 1;47(6):2159–71.
33. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003 Nov;13(11):2498–504.
34. Songue J, Kouam, Dongo E, Mpondo T, White R. Chemical Constituents from Stem Bark and Roots of Clausena anisata. Molecules. 2012 Nov 20;17(11):13673–86.
35. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr. Protoc. Bioinformatics. 2016 Jun 20;54(1).
36. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced

genome of interest. Nucleic Acids Res. 2023 Jan 6;51(D1):D638–46.

37. Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E. gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. J. Chem. Theory Comput. 2021 Oct 12;17(10):6281–91.
38. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. J. Med. Chem. 2002 Jun 1;45(12):2615–23.
39. Wan Mohd Zain WNI, Rahmat A, Othman F, Yap TYH. Antiproliferative Properties of Clausine-B against Cancer Cell Lines. Malays. J. Med. Sci. 2009 Jul;16(3):29–34.
40. Wang KS, Lv Y, Wang Z, Ma J, Mi C, Li X, et al. Imperatorin efficiently blocks TNF-α-mediated activation of ROS/PI3K/Akt/NF-κB pathway. Oncol. Rep. 2017 Jun;37(6):3397–404.
41. Xu X, Zhou F, Zheng L, Wang S, Peng X, Li D. Sampling Challenges of MM/PBSA Binding Energy Calculations. J. Phys. Chem. B. 2025 Nov 13;129(45):11666–78.
42. Zafar A, Khatoon S, Khan MJ, Abu J, Naeem A. Advancements and limitations in traditional anti-cancer therapies: a comprehensive review of surgery, chemotherapy, radiation therapy, and hormonal therapy. Discover Oncology. 2025 Apr 24;16(1):607.
43. Zhang H. Exploring Herbal Compounds as Targeted Therapies for Breast Cancer: Insights from Network Pharmacology, Molecular Docking, MD Simulation, ADME-Toxicity and DFT Profiles. Iranian Journal of Pharmaceutical Research. 2024 Dec 25;23(1).
44. Zoete V, Cuendet MA, Grosdidier A, Michielin O. SwissParam: A fast force field generation tool for small organic molecules. J. Comput. Chem. 2011 Aug 3;32(11):2359–68.
45. Global cancer burden growing, amidst mounting need for services [Internet]. [cited 2026 Jul 28]. Available from: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services

Figures

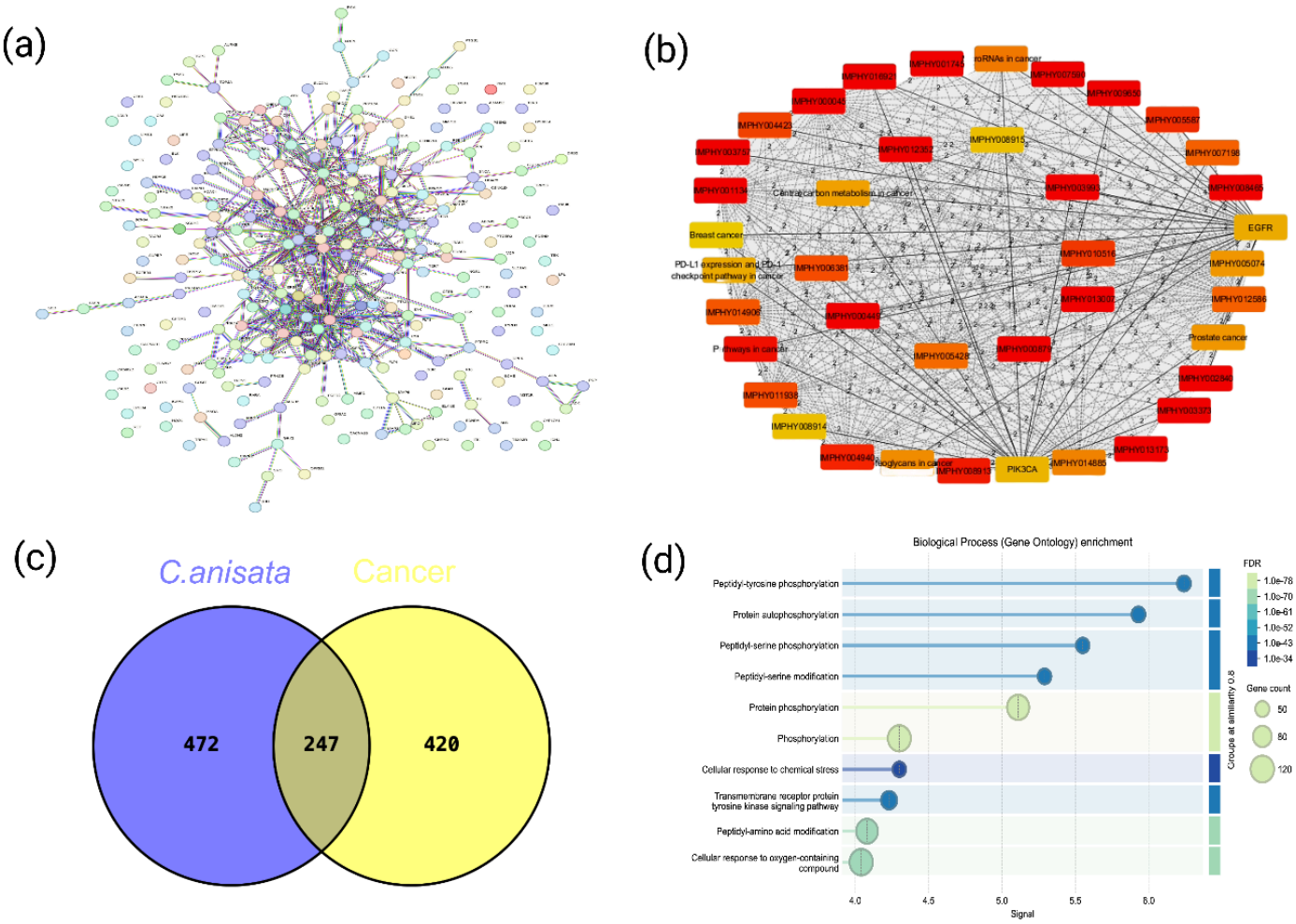


Figure 1

Protein-protein interaction (PPI) network diagram of drug-disease common targets using the STRING database (a), top 20 hub genes identified from the PPI network (b), Venn analysis of cancer-associated genes and plant compound targets (c), and enrichment analysis of the common targets between cancer and plant compounds (d).

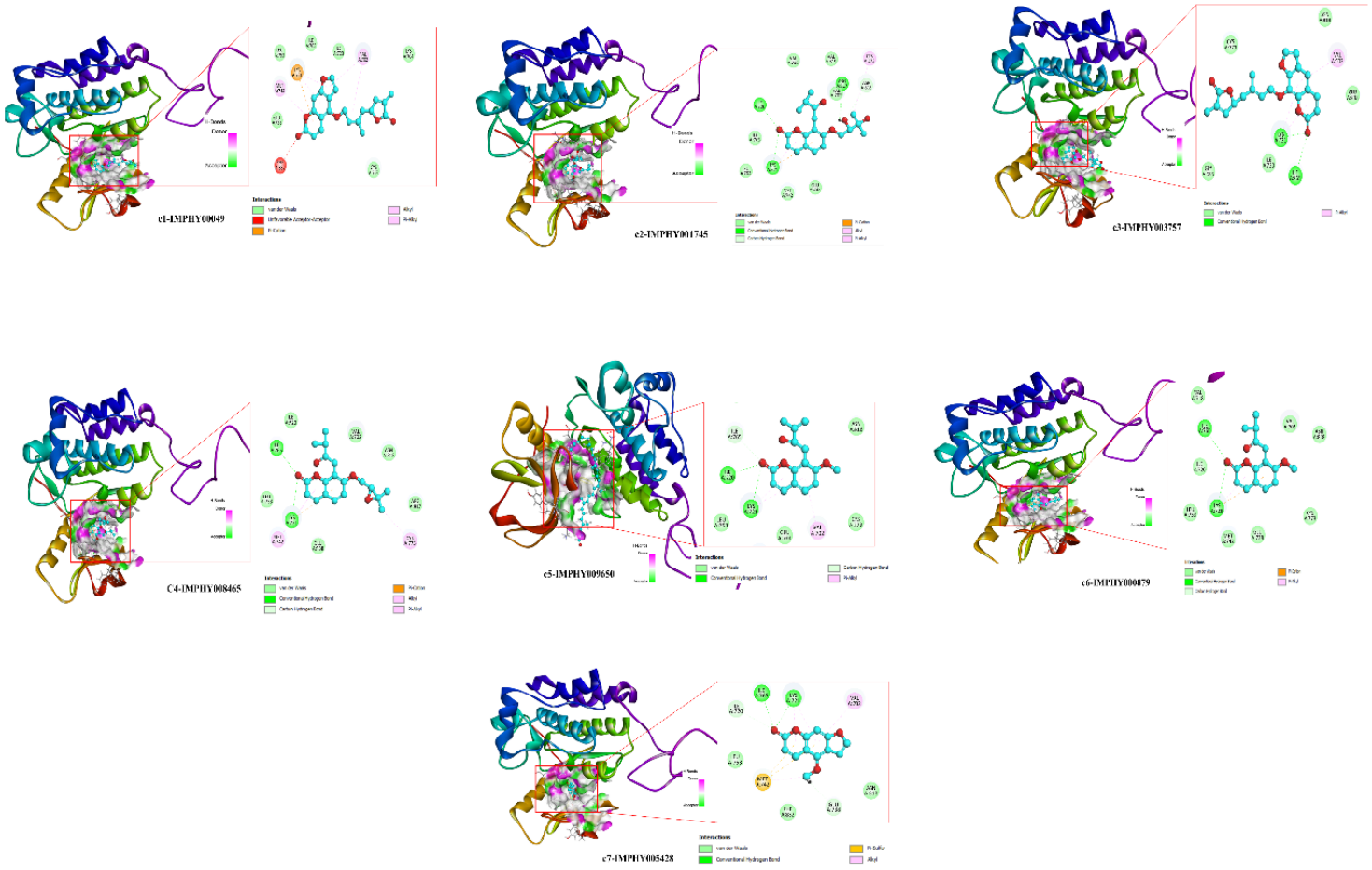


Figure 2

Ligand-protein docking interactions.

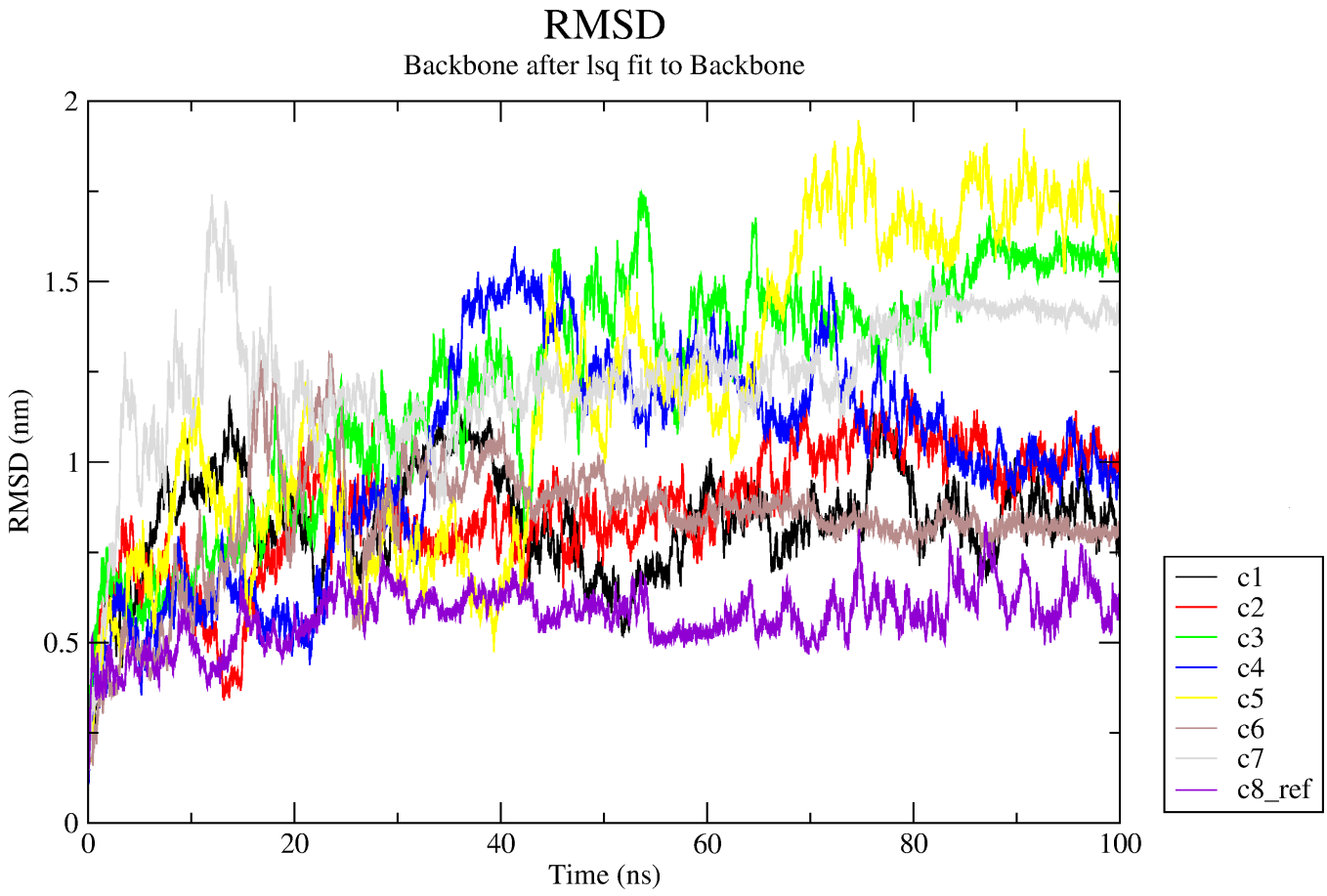


Figure 3

RMSD profiles of Ligand-Protein Complexes.

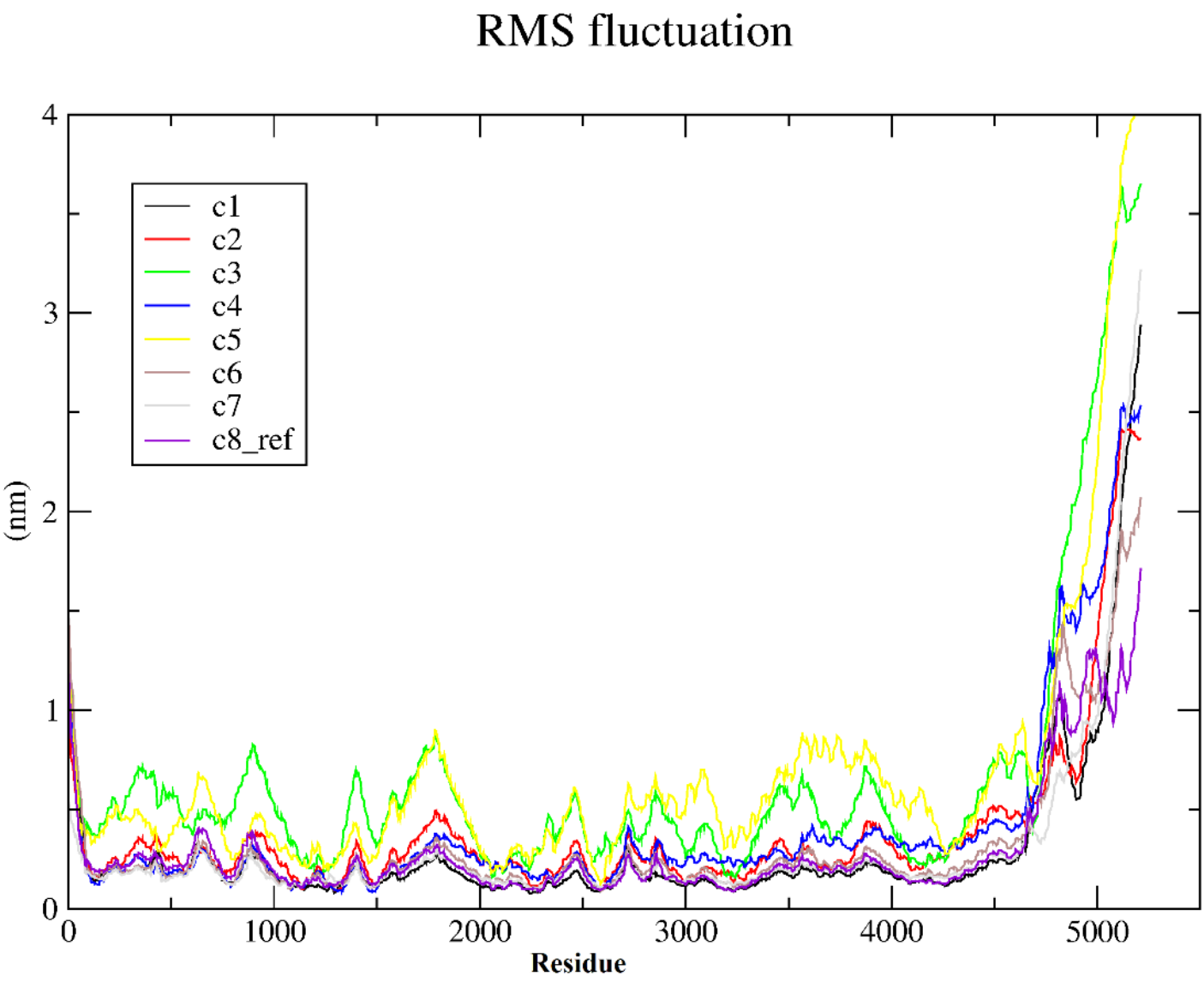


Figure 4

RMSF ligand–protein complexes.

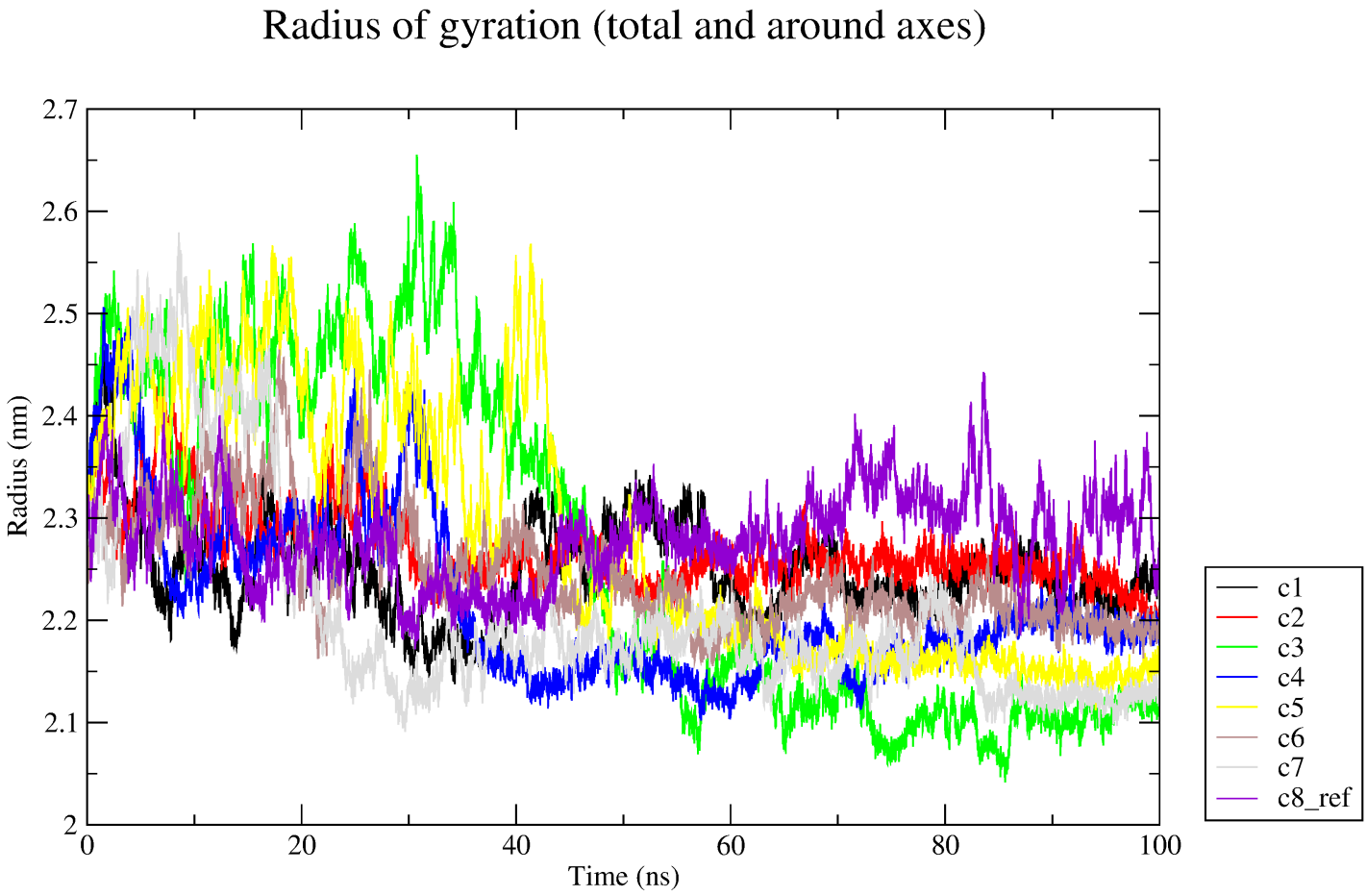


Figure 5

Radius of Gyration Ligand-Protein complexes.

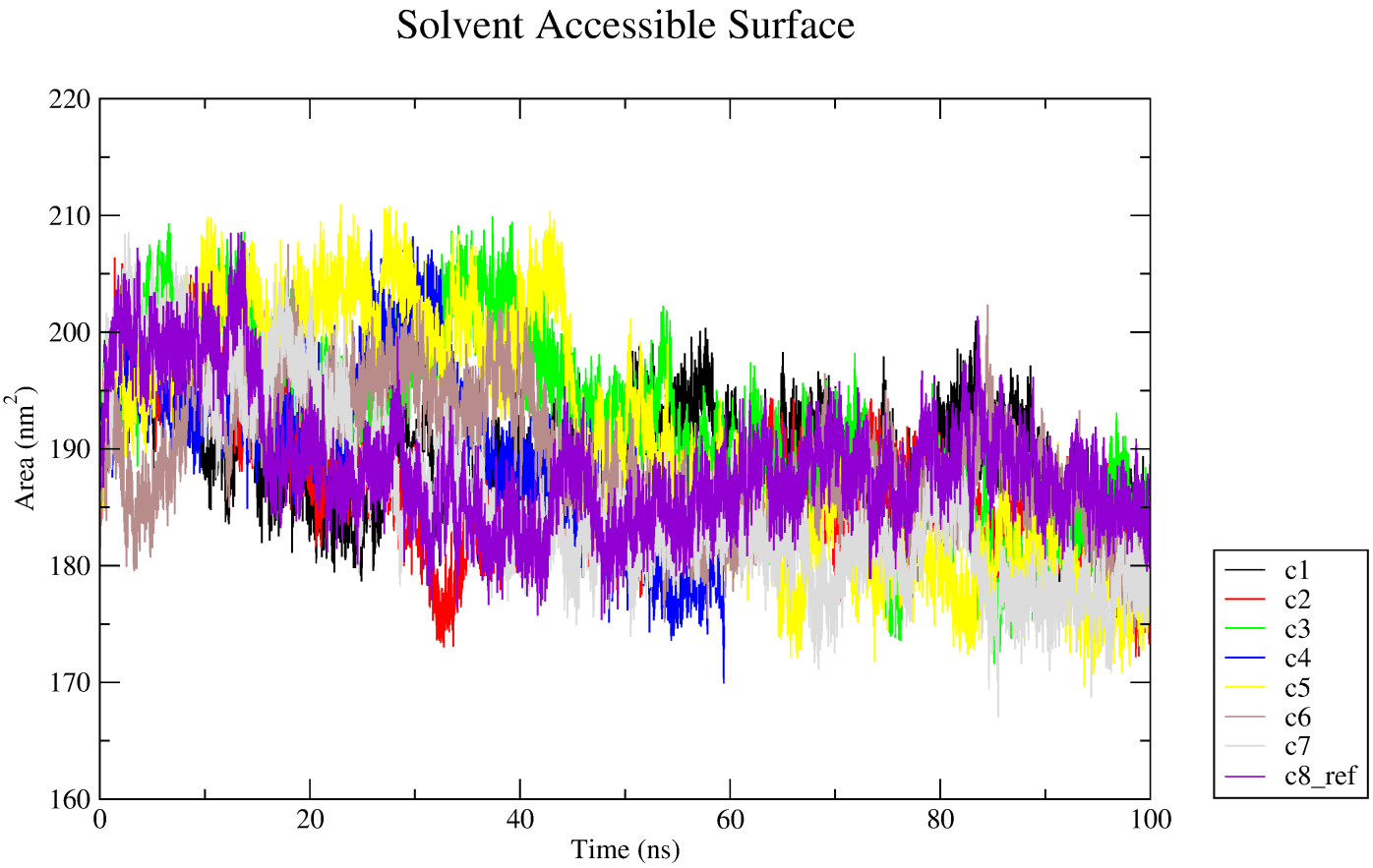


Figure 6

SASA Profile -Ligand-Protein Complexes

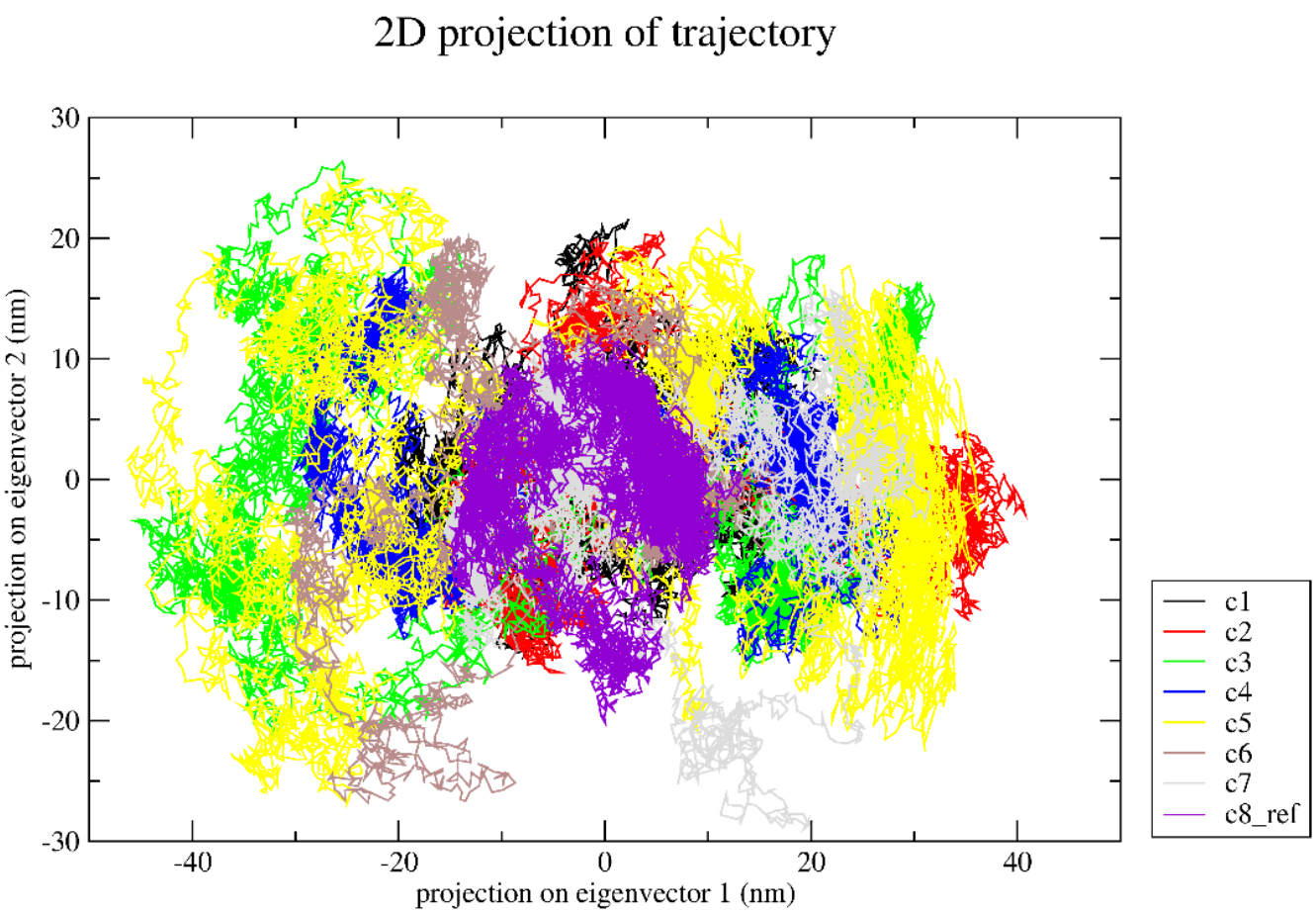

2D projection of trajectory
projection on eigenvector 2 (nm)
projection on eigenvector 1 (nm)
30
20
10
0
-10
-20
-30
-40
-20
0
20
40
c1
c2
c3
c4
c5
c6
c7
c8_ref

Figure 7

PCA

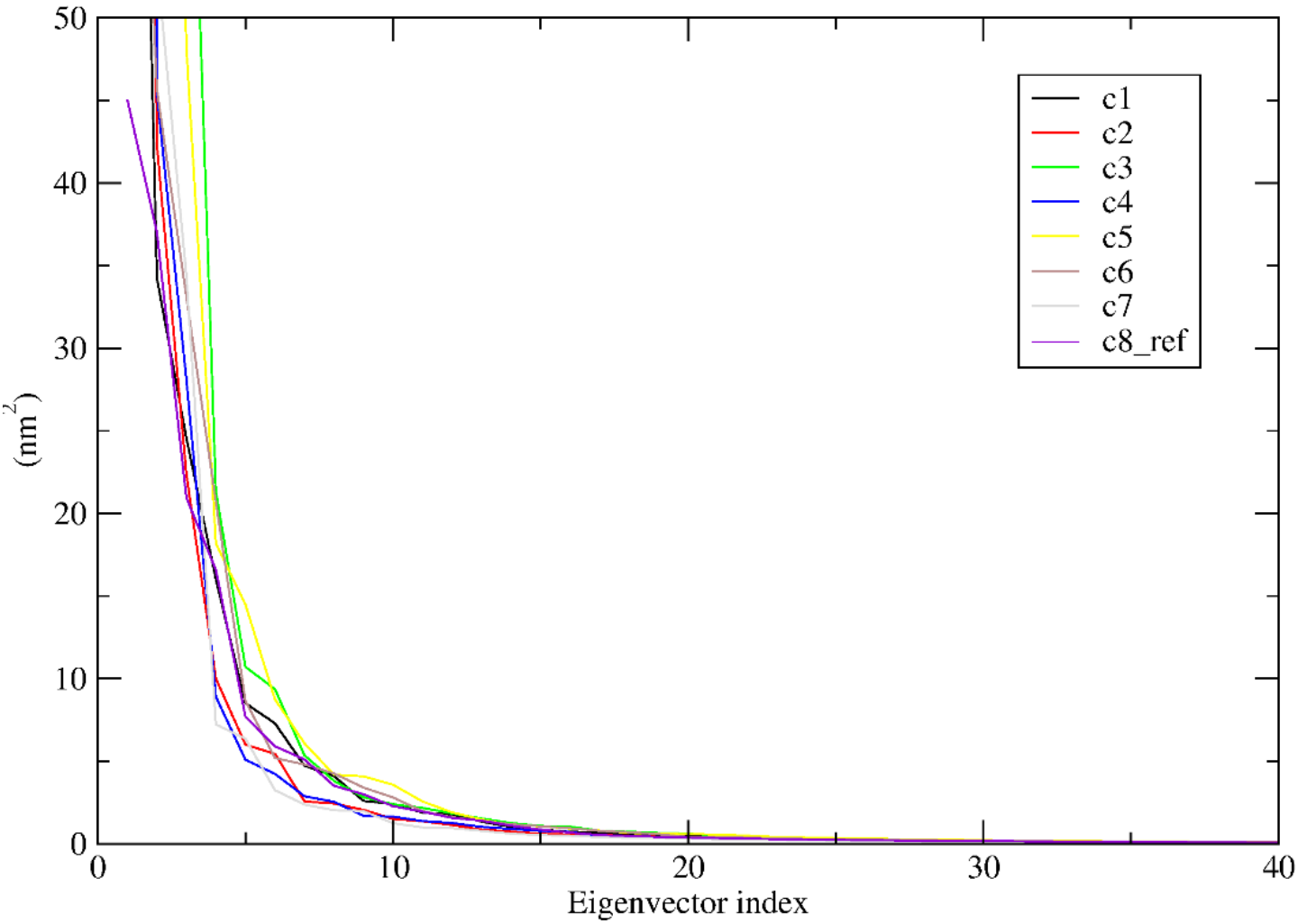


Figure 8

Eigen Vector

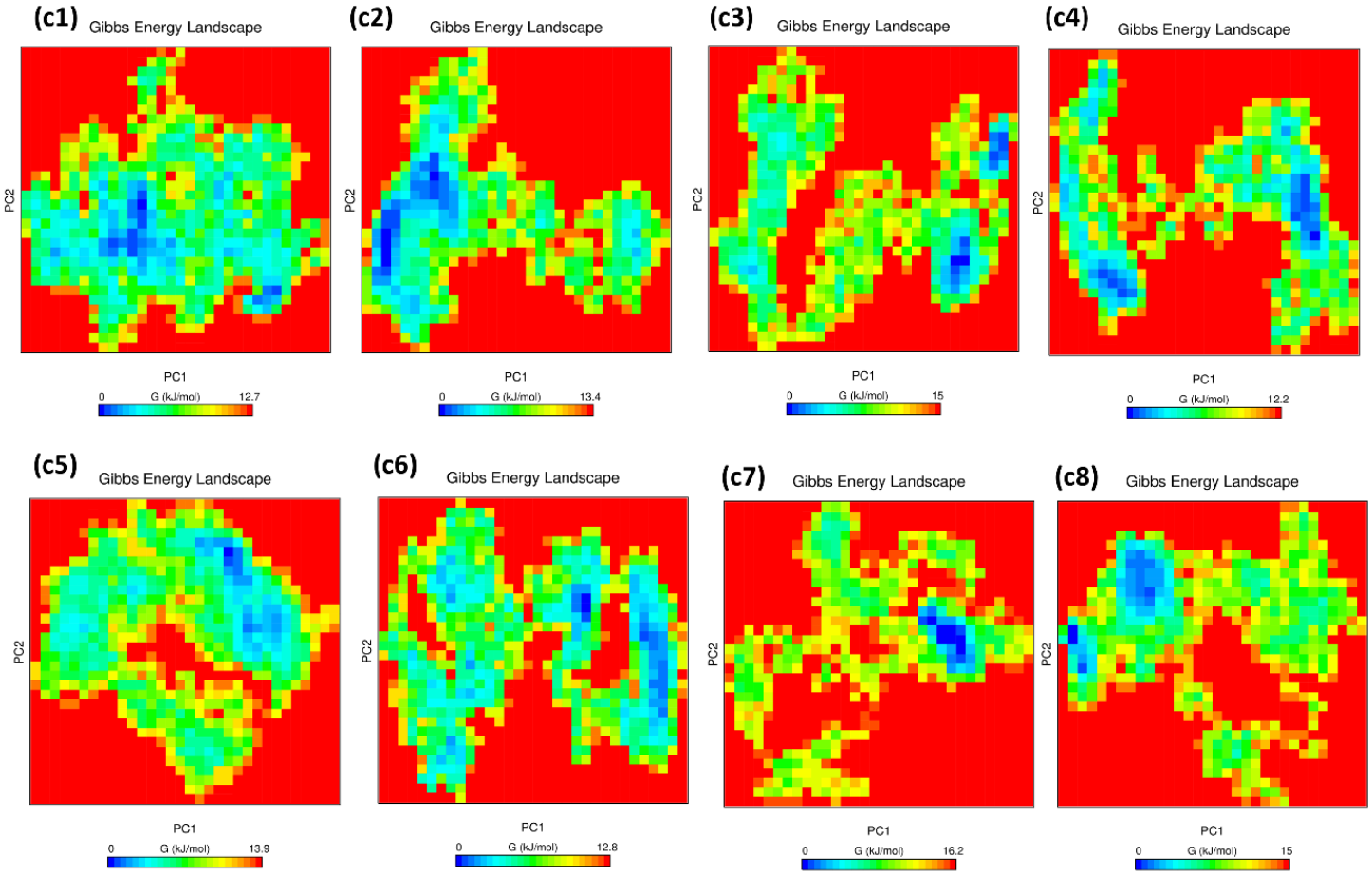


Figure 9

Gibbs Free energy

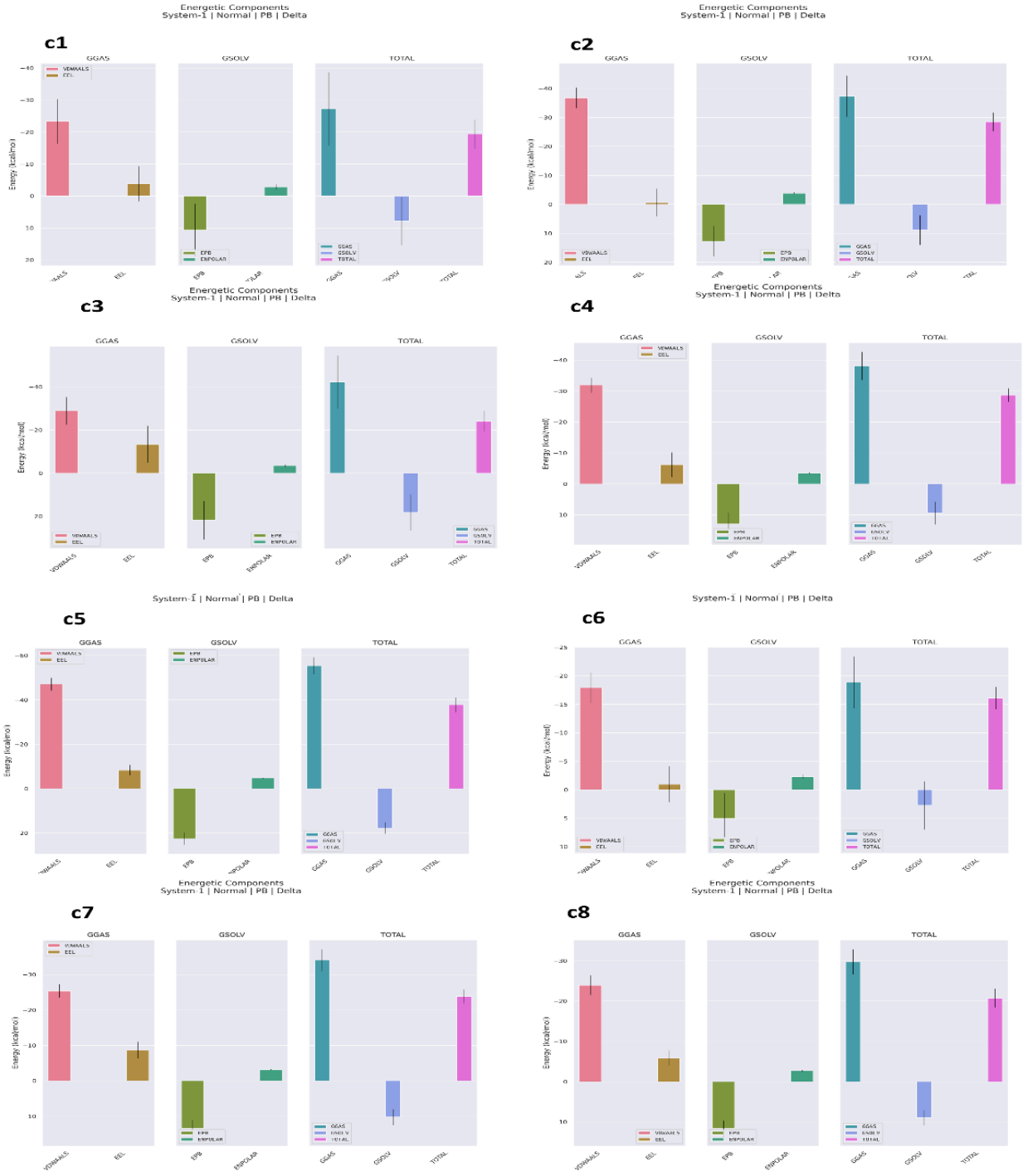


Figure 10

MM/PBSA binding free energies of the complex in 100 ns simulation

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryData.docx