

Pharmacoinformatics Analysis Reveal Flavonoids and Diterpenoids from *Andrographis Paniculata* and *Thespesia Populnea* as Potential Modulators of Hepatitis B Virus Induced Hepatocellular Carcinoma

Vishal S. Patil

KLE Academy of Higher Education and Research: KLE University

Darasaguppe R. Harish (✉ harish.dr@icmr.gov.in)

ICMR National Institute of Traditional Medicine <https://orcid.org/0000-0002-3609-549X>

Umashankar Vetrivel

ICMR-National Institute of Traditional Medicine

Subarna Roy

National Institute of Epidemiology

Sanjay Deshpande

Regional Centre for Biotechnology

Sunil S. Jalalpure

KLE Academy of Higher Education and Research: KLE University

Harsha V. Hegde

ICMR-National Institute of Traditional Medicine

Research Article

Keywords: *Andrographis paniculata*, Hepatitis B, Hepatocellular carcinoma, Molecular docking, Molecular dynamics, Network pharmacology, *Thespesia populnea*

Posted Date: September 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-735128/v1>

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

[Read Full License](#)

Abstract

Herbs are widely utilized in the Western Ghats region of India to treat liver diseases and viral infections. However, such practices lack scientific evidence at the molecular level and may often pose adverse drug reactions. Thus, by this study we intend to identify phytochemicals having druggability and non-toxic profiles with potential activity against HBV-induced HCC. To start with, the details of phytochemicals in traditionally utilized herbs in Western Ghats region were collated from chemical databases and publications. The druggability and toxicity of these compounds were predicted using MolSoft and ADVERpred, respectively. The probable targets of these phytochemicals were predicted using BindingDB. Moreover, compound-gene set pathways, cellular processes, and functional enrichment analysis were also performed using STRING and KEGG pathway databases. Subsequently, Herb-compound-target-disease pathway networks were constructed using Cytoscape 3.6.1. The potential hub protein was virtually screened against ligand dataset using POAP pipeline. Finally, molecular dynamics (MD) simulations of the most potential protein-ligand complexes were performed in triplicate using Desmond 6.1v. Amongst 274 compounds from 16 herbs studied, 36 showed drug likeness with nontoxic properties, and were also predicted to modulate 16 potential targets involved in the pathogenesis of HBV-induced HCC. Among all the molecules screened, flavonoids and diterpenoids from *Andrographis paniculata* and *Thespesia populnea* scored the highest edge count *via* modulating multiple targets and pathways. Moreover, molecular docking and MD simulation (100ns) also inferred that the top ranking Andrographin and Gossypetin to exhibit stable intermolecular interactions with EGFR protein, which was identified as highly connected hub protein in the constructed network. All these findings are suggestive of these moieties as potential therapeutics for targeting HBV-associated HCC sans adverse drug reactions.

Introduction

Hepatitis B, caused by the hepatitis B virus (HBV) is a life-threatening chronic liver infection and is one of the major risk factors for the development of hepatocellular carcinoma (HCC) which accounts for 70–90% of all primary liver cancers. HCC is the sixth most common cancer across worldwide with significant mortality rate and is projected to increase by ~ 40% by 2030 [1, 2]. Currently, more than 248 million individuals worldwide are living with chronic HBV infection. Africa has the most elevated endemicity with 8.83% of HBsAg prevalence. About 95 million people in China have chronic HBV infection [3]. India is meso-endemic for HBV and is the second leading country after China to have about 50 million individuals with chronic HBV infection with an overall prevalence rate of 2.4% (2.2–2.7%) and a much higher of 15.9% (11.4–20.4%) among tribal population [4].

Multiple studies like accumulation of genetic damage due to the immune-mediated hepatic inflammation and chronic hepatic oxidative stress established a robust correlation between chronic HBV infection and the progression to HCC [5–7]. Moreover, HCC mediated by HBV is a polygenic disease that involves multi-protein, multi-pathway pathogenesis. HBV viral proteins alter the expression of endogenous genes within the host tissue. In the hepatocytes, HBV mediated cytokine production by pro-inflammatory reactions and T-cell dysfunction, and inflammation-mediated up and down-regulation of signalling pathways enhances

the development of viral carcinogenesis and ultimately leads to HCC. Further, Signal Transducer and Activator of Transcription 3 (STAT3) and nuclear factor (NF)- κ B play key roles in the regulation of communication between cancer cells and inflammatory cells which further aids in the regulation of tumour angiogenesis. However, HBV-mediated activation of these two pathways promotes immune escapes, production of interleukin IL-6, and IL-22 and ultimately progress to the development of HCC [8, 9]. The HBV X Interacting Protein (HBx or HBXIP) is a transcriptional transactivator that controls the transcription of viral nucleic acid and cellular promotion i.e., it mediates cell cycle regulatory proteins including p16, p21, p27, cyclin A, D1, and B1, promotes CDK2 activity [10–15]. HBx regulates HBV replication and plays a crucial role in the development of HCC. The correlation between the HBV mediated HCC is well established, however the molecular mechanism is not well understood, because on post HBV infection, multiple cascades that are involved for HCC development. Also, numerous scientific reports suggest that the major cause of HCC is due to continuous hepatocyte destruction and regeneration [16].

Since ancient times, herbs and herbal products have been widely utilized as therapeutics for various infectious diseases. Herbs are known to prevent viral infections, delay or arrest the progression or extend the latency period by acting as an agonist, antagonistic, and synergistic *via* modulating multi-protein, multi-pathway associated with the disease pathogenesis [17]. Although the use of herbals against specific etiologies such as HBV infection and its associated HCC is not adequately documented, attempts are being continuously made to evaluate if traditionally used herbs could be used against these specific etiologies by screening of its use against signs and symptoms resembling the disease. In the current study, we identified 16 herbs from the Western Ghats region which is traditionally utilized by the local traditional practitioners against conditions that represent viral infection, fever, inflammation, and jaundice that are documented in earlier studies by the ICMR-National Institute of Traditional Medicine, Belagavi, India. These herbs include *Achyranthes aspera* L., *Aerva lanata* (L.) Juss. ex Schult., *Aloe vera* (L.) Burm.f., *Andrographis paniculata* (Burm.f.) Nees, *Cassia auriculata* (L.) Roxb., *Centella asiatica* (L.) Urban, *Diospyros Montana* Roxb., *Leucas aspera* (Willd.) Link, *Mangifera indica* L., *Phyllanthus virgatus* G. Forst., *Pongamia pinnata* (L.) Panigrahi, *Ricinus communis* L., *Solanum nigrum* L., *Strychnus nux-vomica* L., *Thespesia populnea* (L.) Sol. ex Corrêa, and *Vernonia cinerea* (Carl Linnaeus) H. Rob.

In the present study, we utilized the knowledge of polypharmacology and also utilized chemoinformatics approaches such as compound-gene set enrichment analysis, network pharmacology, molecular docking, and dynamics studies to identify potential classes of bioactive phytochemicals from herbs having druggable characteristics, non-toxic properties, and potent beneficial effect in the management of HCC induced by HBV. Polypharmacology is the effect of a single compound on two or multiple protein targets and pathways, which involves a network-based regulation of numerous physiological responses [18]. Understanding the effect of multi-component drug therapy using traditional knowledge provides new insight for the identification of potent therapeutics against complex diseases like HCC. The workflow of polypharmacology and chemoinformatics approaches implemented in this study is illustrated in Fig. 1.

Materials And Methods

Selection of plants and retrieval of phytochemicals

Sixteen medicinal plants utilized in the region of Western Ghats and North Karnataka for the treatment of fever, jaundice, inflammation, and viral infection were selected based on online resource documented by ICMR- National Institute of Traditional Medicine. The selected medicinal plants with Ayurveda/common name and list of references for their above claimed uses are provided in Supplementary Table 1. Reported phytochemicals from each plant were retrieved from Dr. Duke's DB (Ref), Phytochemical Interaction DB (<https://www.genome.jp/db/pcidb/>), ChEBI (<https://www.ebi.ac.uk/chebi/>), Database of Ethnomedicinal Plants of Western Ghats (<http://nitmmedplantsdb.in/>) and other public repositories.

Druggability and toxicity profile

Molecular formula (MF), molecular weight (MW), hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), logP value, and canonical SMILES of each phytochemical were retrieved from the PubChem chemical database (<https://pubchem.ncbi.nlm.nih.gov/>). The canonical smile data of individual compounds was queried in MolSoft online server (<https://molsoft.com/mprop/>) to predict the druggability of each phytochemical based on the Lipinski rule of five. The server provides structural information viz., MF, MW, HBA, HBD, logP, and predicts probable drug-likeness score (DLS). Further, phytochemicals having positive DLS were chosen and their toxicity was predicted using the ADVERpred server [19]. We opted this server, as it predicts Hepatotoxicity, Nephrotoxicity, Arrhythmia, and Myocardial infarction of unknown molecules based on the probable activity (Pa) and Probable inactivity (Pi) and also predicts probable side effects. Phytochemicals having Pa value of ≤ 0.5 were proceeded for further studies.

Target identification

Canonical SMILES of the shortlisted phytochemicals were queried into BindingDB [20] for probable target protein identification ($p \geq 0.7$). The Gene IDs of the predicted protein targets were retrieved manually from the UniProt web server (<https://www.uniprot.org/>). Further, the reported targets involved in the pathogenesis of HBV-induced HCC were retrieved from the KEGG database (<https://www.genome.jp/kegg/>) and public repositories.

Gene ontology and pathway enrichment analysis

To understand the molecular functions, processes of probable protein targets modulated by the phytochemicals, and to retrieve the protein-protein interactions systematically, a set of retrieved gene IDs were provided as input to STRING 11.0 (Search Tool for the Retrieval of Interacting Genes/Proteins (<https://string-db.org>)). Gene Ontology (GO) based functional and process enrichment analyses were carried out using the STRING GO tool to retrieve the molecular functions and processes of protein targets. The KEGG pathway (<https://www.genome.jp/kegg/pathway.html>) was utilized to identify the enriched pathways involved in HBV-induced HCC. In this study, in order to obtain significant functions, process, and pathway enrichment, the default false discovery rate (FDR) p-value cut-off of < 0.05 was considered. The overall pathways involved in HBV-mediated HCC are shown in Fig. 2.

Network construction

The compound-protein-pathway network was constructed using Cytoscape v3.6.1 software, an open-source software utilized to visualize biological pathways and complex molecular interaction networks [21]. During the construction of the network, “Network Analyzer” command was used and treated as directed graph. To express the degree of connections among the compound-protein-pathways, edge count calculation was applied. Further, “low values to small sizes” for node size and “low values to bright colors” for node color were applied to design the network [22].

Homology modeling and assessment of active site residues

Among the total targets predicted by bioactive phytochemicals, EGFR was found to score the highest edge count within the network (found to be involved in MAPK, PI3K-Akt, Ras, and Jak-STAT signaling pathways). Hence, EGFR, a highly connected target within the network and also a known therapeutic target for hepatocellular carcinoma was chosen to virtually screen the bioactive phytochemicals intermolecular to infer significant intermolecular interactions at active site residues. The wild type EGFR kinase domain in complex with dacomitinib (PDB id: 4I23 (2.8Å resolution) expressed in *Spodoptera frugiperda* expression system) was selected from the Protein Data Bank (PDB) (<https://www.rcsb.org/structure/4i23>). Discovery Studio Visualizer version 2019 (DSV 2019) was used to visualize the missing amino acid residues. ClustalW (<https://www.genome.jp/tools-bin/clustalw>) multi sequence alignment online server was used to align the sequence of crystal structure (PDB 4I23) with the complete sequence of EGFR from Uniprot (UniProt accession No. P00533). This identified, twenty amino acid residues data to be missing in the PDB id: 4I23 crystal structure. Further, these missing amino acid residues were fixed using Modeller9.10v software [23]. While modelling the ligand unbound form, the crystal structure (PDB id: 4I23) was used as a template, and the sequence of UniProt accession No. P00533 entry was used as query. By using Modeller9.10v, 100 model structures were generated and the model having the least Discrete optimized protein energy (DOPE) score was chosen and checked for structural homology with the template using superimpose RMSD option in PyMOL. The stereochemical check for the modelled structure was assessed using PROCHECK web server. Further, the active site residues were mapped based on the crystal structure of EGFR in complex with dacomitinib.

Least potential energy conformation by molecular dynamics (MD) simulation

MD simulation has become an important method in computational biology for understanding the dynamic behaviour of molecular complexes and intermolecular interactions in biologically simulated conditions [24]. To obtain the lowest potential energy (PE in kcal/mol) conformation of the modelled structure, MD simulation was performed for 50 ns using Schrodinger Desmond v6.5 software [25]. After the molecular docking step, MD was also performed in triplicates for protein-ligand complexes to infer the stability of complex formation during 100ns of production run using OPLS force field, to ensure the plausibility. During MD simulation, the entire system was constructed with a predefined Simple Point Charge (SPC) water model, as solvent in a cubic box having 10Å x 10Å x 10Å dimensions as the periodic

boundary. The system was neutralized by adding 6 Na⁺ counterions. Further, the SHAKE algorithm was applied to restrain the geometry of water molecules, bond lengths, and bond angles of heavy atoms. The Particle Mesh Ewald method was applied to estimate the long-range interactions that exist between the molecules. The Lennard-Jones interactions cut-off was set to 10Å. The system was minimized for a 100.0ps production run using OPLS force field. Finally, the NPT ensemble was utilized, implementing Thermostat "Nose-Hoover chain" method with 1.0ps relaxation time and the Barostat "Martyna-Tobias-Klein" method (with 2.0ps relaxation time) were used to maintain pressure of 1.01325bar and 300K temperature. The Coulombic short-range cut-off radius was set to 9.0 Å. The trajectory sampling for the entire run was recorded at an interval of 10.0ps. The residue-wise interaction fluctuations were analyzed by the RMSD, root mean square fluctuation (RMSF), and radius of gyration (rGyr).

Molecular Docking

Preparation of ligand and protein target

Based on the network analysis and polypharmacology approach, 11 phytocompounds from *Andrographis paniculata* and 3 from *Thespesia populnea* were also found to interact with EGFR (a highly connected target within the network). Thus, the 3D structures of 14 phytocompounds were downloaded from the PubChem database. Further, each ligand was energy minimized by mmff94 force field using the steepest descent algorithm at 5000 energy minimization steps by executing the "prepare_ligand4.py" script [26]. The energy minimization step eliminates the clashes within ligand atoms and produces a reasonable starting pose. Further, the convergence criteria were set to default (1e-6) with van der Waals (VDW) cut-off distance to 6.0 Å. Finally, the pose having the lowest energy conformation was chosen and converted into AutoDock vina file (.pdbqt) using AutoDock vina POAP pipeline [26]. The EGFR structure with the lowest potential energy conformation from the MD simulation was chosen for subsequent virtual screening studies.

Ligand- Protein molecular docking

Molecular docking was carried out using AutoDock vina by POAP pipeline [26]. The grid box was created [center x = 4.00245, y = 10.9442, z = -4.3899; and size x, y, z = 26; spacing 1Å] based on the active site residues inferred from an earlier ligand co-crystallized complex (PDB: 4I23) of EGFR protein. The exhaustiveness was set to 100 in vina. The protein-ligand complex having the lowest RMSD value was considered as the best pose and visualized for key intermolecular interactions using Discovery Studio Visualizer 2019 v. The compound having the lowest BE and maximal number of interactions with the active site residues was chosen, and were examined for complex stability by MD simulation (100 ns).

Results

Selection of plants and mining of phytocompounds

Sixteen plants that are utilized by the traditional practitioners/local healers in the region of Western Ghats of India for the treatment of fever, jaundice, inflammation, and viral infection were selected. Initially, 629 phytocompounds from the 16 plants were retrieved from various herbal databases and documented literature. Further, 274 phytocompounds were shortlisted by removing the duplicate and aliphatic chain compounds.

Druglikeness and toxicity profile

Among 274 compounds, 145 compounds showed positive drug-likeness properties. Among these 145 compounds, 36 compounds showed $Pa \leq 0.5$ and nontoxic properties (Table 1). The drug-likeness property and toxicity profile of 274 compounds are shown in Supplementary Table 2 and Supplementary Table 3, respectively.

Target identification

Potential protein targets of 36 phytocompounds were predicted using BindingDB. All these phytocompounds were predicted to modulate 206 protein targets (Supplementary Table 4). Among which 16 targets were found to be potentially involved in the pathogenesis of the HBV infection-associated HCC (Table 2).

Table 2

Phytocompounds and their probable predicted protein targets involved in hepatitis B infection pathways and its associated hepatocellular carcinoma

Plant name	Phytocompounds	Plant part	Compound type	Gene count	Protein targets involved in Hepatitis B infection and HCC pathways
<i>Aerva lanata</i>	Isorhamnetin	Aerial	Flavonoid	7	CDK1, CDK2, CDK6, EGFR, HGF, IGF1R, TERT
	Tiliroside	Aerial	Flavonoid	1	TNF
<i>Aloe vera</i>	2-Methyl-2-Phytyl-6-Chromanol	Whole plant	Phylloquinone	1	AKT1
	Aloe-Emodin	Leaf, Juice	Anthraquinone	1	BCL2
	Aloenin	Whole plant	Glucoside	1	TNF
	Feralolide	Leaf	Coumarin	2	MAP2K1, MAPK1
<i>Andrographis paniculata</i>	5-Hydroxy-7,8,2',3'-Tetramethoxyflavone	Whole plant, Root	Flavonoid	6	CDK1, CDK6, EGFR, HGF, IGF1R, TERT
	5-Hydroxy-7,8,2'-Trimethoxyflavone	Whole plant, Root	Flavonoid	6	CDK1, CDK6, EGFR, HGF, IGF1R, TERT
	Neoandrographolide	Whole plant, Root	Diterpenoid	1	STAT3
	Paniculide-B	Whole plant, Stem	Diterpenoid	3	PRKCA, PRKCB, PRKCG
	Andrographiside	Whole plant, Stem	Flavonoid	4	PRKCB, PRKCA, PRKCG, STAT3
	Altisin	Whole plant, Stem	Flavonoid	6	CDK1, CDK6, EGFR, HGF, IGF1R, TERT
	2'-hydroxy-5,7,8-trimethoxyflavone	Whole plant, Stem	Flavonoid	6	CDK1, CDK6, EGFR, HGF, IGF1R, TERT
	5-hydroxy-7,8,2',3',4'-pentamethoxyflavone	Whole plant, Stem	Flavonoid	6	CDK1, CDK6, EGFR, HGF, IGF1R, TERT
Wightin	Whole plant,	Flavonoid	6	CDK1, CDK6, EGFR, HGF,	

		Stem			IGF1R, TERT
	5-hydroxy-7,8-dimethoxyflavanone	Whole plant, Stem	Flavonoid	3	AKT1, HGF, TERT
	5-hydroxy-7,8-dimethoxyflavone	Whole plant, Stem	Flavonoid	6	CDK1, CDK6, EGFR, HGF, IGF1R, TERT
	Skullcapflavone I	Whole plant, Stem	Flavonoid	6	CDK1, CDK6, EGFR, HGF, IGF1R, TERT
	5-hydroxy-7,8,2',5'-tetramethoxyflavone	Whole plant, Stem	Flavonoid	6	CDK1, CDK6, EGFR, HGF, IGF1R, TERT
	5,4'-dihydroxy-7,8,2',3'-tetramethoxyflavone	Whole plant, Stem	Flavonoid	6	CDK1, CDK6, EGFR, HGF, IGF1R, TERT
	Andrographidine C	Whole plant, Stem	Flavonoid	2	EGFR, TNF
<i>Cassia auriculata</i>	Pelargonidin 5-galactoside	Leaves	Anthocyanidin	2	EGFR, TNF
<i>Centella asiatica</i>	3-Glucosylkaempferol	Leaves	Flavonoid	2	EGFR, TNF
	3-Glucosylquercetin	Leaves	Flavonoid	2	EGFR, TNF
<i>Mangifera indica</i>	Cyanidin 3-arabinoside	Peel	Anthocyanin	1	TNF
	Myrtilin	-	Phenolic compound	1	TNF
	Peonidin galactoside	Peel	Anthocyanin	1	TNF
	Petunidin 3-glucoside	Peel	Anthocyanin	1	TNF
<i>Phyllanthus virgatus</i>	Kaempferol 8-C-sulfate	Whole plant	Flavonoid	1	EGFR
<i>Pongamia pinnata</i>	Pinnatin	Seed	Flavonoid	6	CDK1, CDK6, EGFR, HGF, IGF1R, TERT
	Pongamone A	Stem	Flavonoid	1	EGFR
<i>Thespesia populnea</i>	Gossypetin	Flower, Stem, Root	Flavonoid	7	CDK1, CDK2, CDK6, EGFR, HGF, IGF1R, TERT
	Herbacetin	Flower	Flavonoid	7	CDK1, CDK2, CDK6, EGFR,

					HGF, IGF1R, TERT
	Kaempferol-7- Glucoside	Flowers	Flavonoid	2	EGFR, TNF
<i>Vernonia cinerea</i>	Luteolin-7-Mono- Beta-D-Glucoside	Seed	Flavonoid	2	EGFR, TNF

Gene set enrichment analysis and network construction

Gene ID's of the 16 targets were fetched and provided as input to STRING 11.0 search. The GO functional and process enrichment showed featured 46 molecular functions and 599 molecular processes to be potentially modulated by these 16 targets (Supplementary Table 5 and Supplementary Table 6 respectively). Further, the KEGG pathway enrichment analysis inferred those 139 pathways to be modulated by 36 compounds. Among these 139 pathways, 14 were found to be involved in the pathogenesis of HBV-induced HCC. The 16 targets involved in 139 pathways were predicted to be potentially modulated by the phytochemicals are shown in Supplementary Table 7. Pathways that are solely involved in the HBV-induced HCC modulated by the phytochemicals are shown in Table 3.

Table 3

Pathways involved in the HBV infection and its associated hepatocellular carcinoma modulated by the phytochemicals

KEGG ID	Pathway name	Gene count	False Discovery Rate	Modulated gene within a pathway
hsa05161	Hepatitis B	11	1.07E-18	AKT1, BCL2, CDK2, CDK6, MAP2K1, MAPK1, PRKCA, PRKCB, PRKCG, STAT3, TNF
hsa05225	Hepatocellular carcinoma	11	3.43E-18	AKT1, CDK6, EGFR, HGF, IGF1R, MAP2K1, MAPK1, PRKCA, PRKCB, PRKCG, TERT
hsa01521	EGFR tyrosine kinase inhibitor resistance	11	6.37E-21	AKT1, BCL2, EGFR, HGF, IGF1R, MAP2K1, MAPK1, PRKCA, PRKCB, PRKCG, STAT3
hsa05206	MicroRNAs in cancer	9	2.07E-14	BCL2, CDK6, EGFR, MAP2K1, MAPK1, PRKCA, PRKCB, PRKCG, STAT3
hsa04010	MAPK signaling pathway	10	7.51E-14	AKT1, EGFR, HGF, IGF1R, MAP2K1, MAPK1, PRKCA, PRKCB, PRKCG, TNF
hsa04151	PI3K-Akt signaling pathway	10	3.22E-13	AKT1, BCL2, CDK2, CDK6, EGFR, HGF, IGF1R, MAP2K1, MAPK1, PRKCA
hsa04014	Ras signaling pathway	9	5.60E-13	AKT1, EGFR, HGF, IGF1R, MAP2K1, MAPK1, PRKCA, PRKCB, PRKCG
hsa04150	mTOR signaling pathway	8	1.52E-12	AKT1, IGF1R, MAP2K1, MAPK1, PRKCA, PRKCB, PRKCG, TNF
hsa04210	Apoptosis	5	2.18E-07	AKT1, BCL2, MAP2K1, MAPK1, TNF
hsa05203	Viral carcinogenesis	5	8.12E-07	CDK1, CDK2, CDK6, MAPK1, STAT3
hsa04668	TNF signaling pathway	4	3.85E-06	AKT1, MAP2K1, MAPK1, TNF
hsa04630	Jak-STAT signaling pathway	4	1.57E-05	AKT1, BCL2, EGFR, STAT3
hsa04115	p53 signaling pathway	3	4.58E-05	CDK1, CDK2, CDK6
hsa04110	Cell cycle	3	0.00021	CDK1, CDK2, CDK6

Network analysis

Network analysis identified Skullcapflavone I, Wightin, 2'-hydroxy-5,7,8- trimethoxyflavone, Altisin, Andrographiside, 5-Hydroxy-7,8,2'-Trimethoxyflavone, 5-hydroxy-7,8,2',3',4'- pentamethoxyflavone, 5-hydroxy-7,8,2',5'- tetramethoxyflavone, 5,4'-dihydroxy-7,8,2',3'- tetramethoxyflavone, Andrographidine C, Wightin, 5-hydroxy-7,8-dimethoxyflavanone, 5-hydroxy-7,8-dimethoxyflavone, 5-Hydroxy-7,8,2',3'- Tetramethoxyflavone and Paniculide-B from *Andrographis paniculata* (whole plant); Herbacetin, Gossypetin, Kaempferol-7-Glucoside from *Thespesia populnea* (Flowers) ; Isorhamnetin from *Aerva lanata* (Aerial part); and Pinnatin from *Pongamia pinnata* (Seed) to have highest edge count. Further,

among 14 pathways, EGFR tyrosine kinase inhibitor resistance, Hepatitis B, and Hepatocellular carcinoma pathways were identified as enriched pathways *via* modulating EGFR, AKT1, BCL2, HGF, IGF1R, MAP2K1, MAPK1, PRKCA, PRKCB, PRKCG, STAT3, CDK2, CDK6, STAT3, TNF, and TERT as a common target protein (Table 2 and Table 3). However, the above-selected compounds were considered based on polypharmacology therapy due to its probable affinity towards multiple targets and pathways. The herb-compound-protein-pathway network is shown in Fig. 3. For further docking studies, based on the polypharmacology analysis, phytochemicals that are predicted to modulate multiple protein targets and pathways were selected, therefore 11 phytochemicals from *Andrographis paniculata* were predicted to target 13 protein molecules and 3 phytochemicals from *Thespesia populnea* were predicted to target 8 protein molecules involved in HBV induced HCC and are represented in Venn diagram (Fig. 4).

Further, the study was extended to molecular docking and dynamics simulations. As discussed earlier, EGFR was chosen as the most potential target, as it was found to be the highly connected protein target within the network and is also a well-known therapeutic target in many types of cancer viz., lung cancer [27], glioblastoma [28], breast carcinoma [29], HCC [30, 31], and so on. Importantly, among 14 pathways prioritize, EGFR was also found to be involved in Hepatocellular carcinoma, MAPK, PI3K-Akt, Jak-STAT, and Ras signaling pathways, and EGFR tyrosine kinase inhibitor resistance pathway. Based on the docking study, 11 phytochemicals from *Andrographis paniculata* and 3 from *Thespesia populnea* were identified to potentially target EGFR protein.

Homology modeling and active sites assessment

As the WT-EGFR crystal structure (PDB id: 4I23) had missing amino acids, the entire structure was rebuilt by homology modelling. In order to fill the missing residues and to obtain a plausible structure, 100 homology structural models of WT EGFR were generated using Modeller9.10v software. Among these, Model number-30 was chosen as best model, as it featured the lowest DOPE score (-37355.91kcal/mol) and it also had an RMSD of 0.292 w.r.t. template PDB id: 4I23 (Fig. 5). The active site residues of WT EGFR reported in the co-crystallized structure are Leu23, Ala48, Lys50, Glu67, Met71, Leu93, Thr95, Gln96, Met98, Pro99, and Gly101 [32].

Model stability and lowest potential energy conformation sampling by MD simulation

The stability of the WT EGFR modelled structure (Fig. 6) was validated by MD run of 50 ns and it was found to be highly stable with less perturbations. The lowest potential energy conformation was found to be -129809.51kcal/mol at 45252.0th ps (45.2ns) and was selected for further docking studies.

Molecular Docking study

Molecular docking of 11 phytochemicals from *Andrographis paniculata* and 3 from *Thespesia populnea* were performed with WT EGFR and the corresponding binding energies (BE) and intermolecular interactions are summarized in Table 4. Firstly, all 11 compounds from, *A. paniculata* were predicted to form interactions with the active site residues. Among these, Andrographin scored the lowest BE of

-8.3kcal/mol via forming two hydrogen bonds i.e. Met98...=O, Met98...OH and forming ten hydrophobic interactions i.e. Thr159, Lys50, Val31 (2), Ala48 (2), Leu149 (2), Leu23 (2) residues. Among these, seven interactions were found to span the active site residues. The intermolecular interactions of Andrographin with WT-EGFR are shown in Fig. 7. Secondly, all the 3 compounds from *T. populnea* were predicted to form stable interactions with active site residues. Among these, Gossypetin scored the lowest BE of -8.5kcal/mol via forming three hydrogen bond interactions i.e. Met98...OH, Thr95...OH (2) and eight hydrophobic bonds i.e. Thr159, Ala48, Val31 (2), Leu149 (2), Leu23 (2). Among these, six interactions were found to span with active site residues. The interactions of Gossypetin with WT EGFR are shown in Fig. 8.

Table 4

Binding energy and molecular interactions of phytochemicals with WT EGFR protein target.

Compound name	BE (kcal/mol)	Hydrogen bond interactions (No. of interactions)	Van der Waals, Pi-Alkyl, CH, Pi-Cation, Pi-Sigma, Pi-Pi stacked, Pi-Pi T-shaped interactions (No. of interactions)	Active sites residues within Interactions	No. of Interactions with Active site residues
Kaempferol-7-Glucoside	-8.8	Asp105...OH, Thr95...OH, Gln96...OH	Val31, Leu149, Leu23 (3), Cys102	Thr95, Gln96	5
Herbacetin	-8	Thr95...OH, Met98...OH (2), Gly101...OH	Leu23 (2), Leu149 (2), Val31 (2), Ala48, Cys80, Thr159	Thr95, Met98, Gly101, Leu23, Ala48	7
Gossypetin	-8.5	Met98...OH, Thr95...OH (2)	Thr159, Ala48, Val31 (2), Leu149 (2), Leu23 (2)	Thr95, Met98, Leu23, Ala48	6
Andrographidine C	-8.3	Thr95...OH (2), Gln96...OH, Gly26	Arg146, Val31, Leu149, Leu23 (2), Cys102 (2), Asp105	Thr95, Gln96, Leu23, Cys102	7
5-Hydroxy-7,8,2'-Trimethoxyflavone (Andrographin)	-8.3	Met98...=O, Met98...OH	Thr159, Lys50, Val31 (2), Ala48 (2), Leu149 (2), Leu23 (2)	Met98, Lys50, Ala48, Leu23	7
Skullcapflavone I	-8.1	Met98...=O, Met98...OH	Thr159, Val31 (2), Lys50, Ala48 (2), Leu23 (2), Leu149, Asp105	Met98, Lys50, Ala48, Leu23	7
2'-hydroxy-5,7,8-trimethoxyflavone	-8.1	Nil	Gly101, Leu23 (2), Leu149 (2), Val31 (2), Gly24	Leu23	2
5,4'-dihydroxy-7,8,2',3'-tetramethoxyflavone	-7.7	Nil	Gln96, Ala48, Thr159, Lys50, Val31 (2), Leu149, Leu23 (2), Cys102 (2), Asp105	Gln96, Ala48, Lys50, Leu23, Cys102	7

Compound name	BE (kcal/mol)	Hydrogen bond interactions (No. of interactions)	Van der Waals, Pi-Alkyl, CH, Pi-Cation, Pi-Sigma, Pi-Pi stacked, Pi-Pi T-shaped interactions (No. of interactions)	Active sites residues within Interactions	No. of Interactions with Active site residues
5-hydroxy-7,8-dimethoxyflavone	-8.2	Met98...=O	Leu23 (2), Thr159, Val31 (2), Lys50, Ala48 (2), Lue149 (2), Asp105	Met98, Lys50, Ala48, Leu23	5
5-hydroxy-7,8,2',5'-tetramethoxyflavone	-8.4	Met98...=O	Val31 (2), Lys50, Thr159, Ala48, Leu23 (2), Leu149 (2), Asp105	Met98, Lys50, Ala48, Leu23	5
5-hydroxy-7,8,2',3',4'-pentamethoxyflavone	-8	Nil	Lys50, Thr159 (2), Val31 (2), Gln96, Ala48, Leu149, Leu23 (2), Cys102 (2)	Lys50, Gln96, Ala48, Leu23	5
Wightin	-8.3	Met98...=O, Met98...OH	Val31 (2), Lys50, Thr159, Ala48 (2), Leu23 (2), Leu149 (2), Asp105	Met98, Lys50, Leu23, Ala48,	7
5-Hydroxy-7,8,2',3'-Tetramethoxyflavone	-8.6	Met98...=O	Asp105, Val31 (2), Thr159, Lys50, Leu23 (2), Leu149 (2), Ala48, Gln96	Met98, Lys50, Leu23, Ala48, Gln96	6
Altisin	-7.8	Met98...O-	Val31 (2), Ala48, Leu149 (2), Leu23 (2)	Met98, Ala48, Leu23	4

Molecular dynamics simulation of protein-ligand complexes

Based on the lowest BE and maximum intermolecular interactions with active site residues, two compounds were prioritized for MD simulation i.e. Andrographin from *A. paniculata* and Gossypetin from *T. populnea*. The stability of complex formation WT-EGFR-Andrographin complex and WT-EGFR-Gossypetin complex were validated by MD simulation for 100ns. The simulation for both the complex were carried out in triplicate, to ensure the plausibility.

Andrographin with WT EGFR complex

The Andrographin with WT-EGFR simulation of system consisted of 39399 atoms with 11384 water molecules. The system was neutralized by adding 8 Na⁺ ions (12.777 mM concentration) and was subjected to 100ns run of production run with 10ps recording interval. The ligand RMSD with respect to protein values were within the range of 0.3Å to 3.5Å from 0 to 100ns in all three replicates. The average RMSD of ligand w.r.t. protein was 1.24Å and ligand w.r.t. ligand was 0.44Å. The C- α and backbone RMSD fluctuation was within 0.7 to 3.6Å, and the average RMSD was 2.28Å. The RMSF analysis was performed to infer for residue-wise fluctuations observed during the production run. The average RMSF value of atoms w.r.t. protein was \sim 0.7Å and ligand was found to be \sim 0.2Å. The average RMSF difference of both protein-ligand atoms was found to be \sim 0.5Å. Additionally, the radius of gyration (rGyr) trajectory over the 100ns production run was also analysed. The rGyr deviation of Andrographin with EGFR was found within 3.60Å to 3.68Å, which indicates the higher compactness of Andrographin-EGFR complex.

Further, Andrographin-EGFR contact analysis concluded that Met98 to form a stable hydrogen bonded interactions for around 98–99% and 73–77% of the duration with = O and OH group of Andrographin, respectively in all three replicates. Pro99 and Cys102, and Asp105 were found to be involved in water-bridge interactions and showed around 10–15%, 12–20%, and 19–32% of the time with interaction fraction in all three replicates, respectively. Figures 9, 10, and 11 represents the RMSD, RMSF, rGyr, and contacts of Andrographin with WT EGFR.

Gossypetin and WT EGFR complex

The Gossypetin with WT EGFR simulation of system consisted of 39389 atoms with 11383 water molecules. The system was neutralized by adding 8 Na⁺ ions (12.778 mM concentration) and subjected to 100ns run of production run with 10ps recording interval. The Ligand RMSD values were observed within the range of \sim 0.3Å to 4.5Å from 0 to 100ns. The average RMSD of ligand w.r.t. protein was \sim 1.4Å and ligand w.r.t. ligand was \sim 0.3Å. The ligand w.r.t. ligand RMSD of replica 3, it showed slight fluctuation from 35ns to 50ns (0.3 to 0.7Å) and again attained stability from 50ns at 0.7Å. Further, decrease in RMSD was seen at 75ns (0.7Å to 0.3Å) and found to stable throughout 100ns for all three replicates. The C- α and backbone RMSD fluctuation was to be within 0.699 to 3.53Å and the average RMSD was found to be \sim 2.6Å. The RMSF analysis was performed to infer for residue-wise fluctuations observed during the production run. The average RMSF value of atoms w.r.t. protein was \sim 1.4Å and ligand was 0.220Å. The average RMSF difference of both protein-ligand atoms was \sim 1.2Å. The radius of gyration (rGyr) trajectory over the 100ns production run was also analysed. The rGyr deviation of Gossypetin with EGFR was found within 3.7Å to 3.8Å, which indicates the higher compactness of Gossypetin-EGFR complex.

Further, Gossypetin-EGFR contact analysis concluded that Gln96 to form a stable hydrogen bonded interactions for around \sim 93% and 21% of the duration with OH group of Gossypetin in replica 1 and 3 and around 53% and 27% of water-bridge in replica 2 and 3. Met98, Thr95, and Asp105 residues were found to be a common interactive residues and involved in hydrogen bond, hydrophobic bond, and water-

bridge interactions in all three replicates. Figures 12, 13, and 14 represents the RMSD, RMSF, rGyr, and contacts of Gossypetin with WT- EGFR.

Discussion

Hepatocyte damage due to HBV infection and the pathogenesis of liver disease is not directly associated with the HBV, however it is caused by the immune-mediated host HBV interactions [33]. The complex mechanisms that cause HBV-mediated liver disease are not fully understood due to the involvement of complex protein-protein interactions with multiple host factors [34]. However, the utilization of already existing single drug molecules against a cluster of HBV mediated liver disease pathways may cause adverse drug effects *via* disrupting the host tissue mechanisms. The utilization of chemotherapeutic agents during HCC may alter hepatic metabolism, reduce the efficacy, or may cause liver toxicity by the disposition of the drug in liver tissue [35]. Further, physicians utilize palliative treatments such as systemic chemotherapy, interferon, transarterial chemo-embolization, and hormonotherapy against HCC when there are no curative treatment options. However, the utilization of these therapies for HCC patients is challenging because HCC is highly resistant to systemic therapies [36]. In the current study, we elucidated the molecular mechanisms and probable safe use of traditionally used herbs of Western Ghats, Karnataka against jaundice, fever, inflammation, and viral infections *via* cluster and enrichment analysis of potential HBV disease protein targets that participate in the progression of immune processes and inflammatory responses. The results obtained from the current study predicts that flavonoids and diterpenoids from *A. paniculata* and flavonoids from *T. populnea* could potentially prevent the disease progression and could be effective therapeutics against chronic HBV-induced HCC.

Identification of potent candidates from herbal medicines is not only the key focus of new drug discovery against HBV-induced HCC, drug safety is also a key concern. This means that ensuring that the drugs have a positive benefit-risk balance is the goal [37]. Hepatotoxicity induced by herbs is one of the most important factors limiting treatment. The issue has often been responsible for post-marketing warnings and withdrawal [38]. Many researchers have well documented the adverse hepatic reactions, idiosyncratic liver damage caused by herbs. The major hepatotoxic herbs reported so far are, *Teucrium chamaedrys*, *Larrea tridentate*, *Cimicifuga racemosa*, *Scutellaria lateriflora*, and *Scutellaria baicalensis* [39]. In the current study, we screened 274 phytocompounds from 16 herbs for their druggable characteristics using the Lipinski rule of five and selected only compounds having positive drug-like properties. Further, these compounds were screened for major adverse drug reactions/ side effects i.e. Hepatotoxicity, Cardiac toxicity, Nephrotoxicity, and potential to cause Myocardial infraction using the ADVERpred online prediction. To minimize the toxicity, compounds having a probable toxic activity (Pa) with a p-value ≤ 0.5 were considered. Further, compounds were also predicted for their probable affinity towards potential drug targets using BindingDB (p-value ≥ 0.7) and also segregated the target associated with HBV infection and HCC with the aid of published literature, KEGG pathway, and Therapeutic Target Database (TTD).

STRING database was also utilized to identify the molecular functions, biological processes, and pathways regulated by the predicted targets, wherein, the targets were meticulously chosen solely based on potential molecular functions, processes, and pathways with higher significance ($p \leq 0.05$) and those that are only associated with HBV-induced HCC. Based on the results obtained from STRING, we constructed compound-protein-pathway network building to understand the relations between compounds, target proteins, and pathways. It is important to consider that one pathway containing many protein molecules modulated by one compound is more essential than the function of one protein that involves many pathways modulated by one drug molecule. The impact of one target in numerous pathways may be less, and the impact of one pathway having multiple protein targets modulated by the one compound could be large [40]. Based on this theory, we chose 11 phytochemicals viz., Andrographidine C, 5-Hydroxy-7,8,2'-Trimethoxyflavone (Andrographin), Skullcapflavone I, 2'-hydroxy-5,7,8-trimethoxyflavone, 5,4'-dihydroxy-7,8,2',3'-tetramethoxyflavone, 5-hydroxy-7,8-dimethoxyflavone, 5-hydroxy-7,8,2',5'-tetramethoxyflavone, 5-hydroxy-7,8,2',3',4'-pentamethoxyflavone, Wightin, 5-Hydroxy-7,8,2',3'-Tetramethoxyflavone, and Altisin from *A. paniculata* [majorly are of Andrographolide derivatives] contained in *A. paniculata* whole plant and Gossypetin, Herbacetin and Kaempferol-7-Glucoside contained in *T. populnea* flowers [41]. These phytochemicals were identified as flavonoids and diterpenoids and were found to modulate thirteen protein molecules involved in the progression of HBV infection to HCC i.e. CDK1, CDK2, CDK6, EGFR, HGF, IGF1R, TERT, STAT3, PRKCA, PRKCB, PRKCG, AKT1, and TNF. Further, these protein targets were also found to modulate 14 major molecular pathways i.e. Hepatitis B pathway, Hepatocellular carcinoma, MicroRNAs in cancer, MAPK signaling pathway, PI3K-Akt signaling pathway, Ras signaling pathway, mTOR signaling pathway, Apoptosis, Viral carcinogenesis, TNF signaling pathway, Jak-STAT signaling pathway, p53 signaling pathway, and Cell cycle. The role of each protein target associated with the identified pathways involved in the pathogenesis of HBV infection and progression of HCC is shown in Supplementary Table 8.

After gene ontology, enrichment, and network analysis, we further extended the study to examine the intermolecular interactions between phytochemicals and predicted protein targets using docking and MD simulation studies. As a result, we have chosen EGFR, a therapeutic target for cancer (identified to be involved in the MAPK, PI3-Akt, Ras, and Jak-STAT signaling pathways within the network). Earlier studies also have demonstrated that overexpression of EGFR to be common in HCC, meaning that it can play a role in the pathogenesis and treatment of the disease. From early inflammation and hepatocellular proliferation to fibrogenesis and neoplastic transformation, the EGFR signaling mechanisms have been recognized as a central player in all stages of the liver response to injury [30, 31]. Schiffer et al [42] reported Gefitinib, a potent EGFR tyrosine kinase inhibitor to curtail HCC development in rats exposed to diethylnitrosamine. Further, Zheng et al [43] reported that Tropomodulin 3 levels to rise in HCC condition, especially when it spreads outside the liver. Tmod3 expression was also linked to aggressive malignancy and poor patient survival in HCC patients [44]. Tmod3 enhances matrix metalloproteinase-2, -7, and -9 transcription, which is dependent on PI3K-AKT. The interactions between Tmod3 and the EGFR, which promotes EGFR phosphorylation is required for signalling activation of PI3K-Akt. Similarly, VersicanV1 was also found to enhance EGFR-PI3K-Akt signalling pathway in HCC cells [44]. A recent study by Liu et

al [45] demonstrated that, in HCC cells, the EGFR-P38 MAPK axis may up-regulate PD-L1 *via* miR-675-5p and down-regulate HLA-ABC via HK2 and this might be responsible for the immune suppression in HCC, and they suggest that EGFR signalling might be targeted for HCC immunotherapy. Hence, based on the potential role of EGFR in HCC and also key network hub status in the current study, the phytochemicals were screened against EGFR. First, the WT EGFR x-ray crystallographic protein structure was obtained from the PDB and the missing residues were fixed by homology modelling. The structural stability of the best model was confirmed by molecular dynamics for 50ns production run. The least potential energy conformation from MD simulation was retrieved for utilization in docking studies. The molecular docking of 11 compounds from *A. paniculata* and 3 from *T. populnea* showed significant intermolecular interactions with active site residues. Further, based on the BE and highest interactions with active site residues, 5-Hydroxy-7,8,2'-Trimethoxyflavone (Andrographin) from *A. paniculata* and Gossypetin (*T. populnea*) were identified as lead hits against WT EGFR. On post docking studies, the MD method was used to delineate the effects of protein structure stability on ligand binding. To explain the structural modification of the protein upon ligand binding, we run a three replicates of 100ns MD simulation for Andrographin and Gossypetin with WT EGFR, to ensure the plausibility of complex formation. The protein-ligand interactions RMSD was found to be $\sim < 0.8\text{\AA}$ and formed stable interactions throughout 100ns production run in all three replicates.

Andrographis paniculata and Thespesia populnea

Andrographis paniculata Nees. is known as a “king of bitters”, a member of an Acanthaceae family, and native to India and Sri Lanka. In Ayurveda, the roots and leaves are used for medicinal purposes [46]. *A. paniculata* is a key ingredient of many herbal formulations for the treatment of liver disease, hepatitis, diabetes, cancer, etc [47, 48]. A previous study by Tan et al [49] demonstrated Andrographolide to downregulate EGFR expressed in epidermoid carcinoma (A-431) cells. Andrographolide and Neoandrographolide from *A. paniculata* showed a protective effect against tert-butylhydroperoxide intoxication induced hepatotoxicity in mice [50, 51], and protective activity against galactosamine induced hepatotoxicity in rats (Handa et al. 1990b). The crude extract of *A. paniculata* showed potent inhibitory activity against HBV surface antigen [53]. Dehydroandrographolide and Andrographolide showed potent inhibitory activity against HBV DNA replication [54]. Neoandrographolide is a potent anti-inflammatory compound that acts via inhibition of iNOS and COX-2 expression through inhibiting p38 MAPKs activation [55]. It also suppresses NO production in LPS activated macrophages [53].

Thespesia populnea Soland ex. Correa belongs to the family Malvaceae. It is mainly found in tropical regions and the coastal forests of India. Ayurvedic physicians use bark decoction for the treatment of skin and liver diseases. The bark and flower possess hepatoprotective, antioxidant, anti-inflammatory, memory improving, cholesterol-lowering activities, and have been reported for the management of Alzheimer's disease [56]. In the current study, we identified Gossypetin, Herbacetin, and Kaempferol-7-Glucoside flavonoids from *T. populnea* flower as potent hepatoprotective hits against HBV-induced HCC. It supports earlier reports of these compounds as potent antiviral flavonoids [41, 57, 58]. Arthanari et al

had also demonstrated that methanolic flower extracts of *T. populnea* have a strong antiviral potency and possessed nontoxic properties [59].

The idea of modulating multiple proteins by a single molecule has triggered researcher's interest in finding the first strike on specific targets. Phytocompounds from *A. paniculata* and *T. populnea* have also been shown to have antiviral and hepatoprotective properties in the past. As a result, the current research discovered 11 compounds from *A. paniculata* and three from *T. populnea* that may have effectiveness against hepatocellular carcinoma caused by the HBV virus, and also further prompted us to investigate the binding affinity and intermolecular interactions of these compounds with active site residues of EGFR, a well-known cancer therapeutic target. Although the identified small molecules possess druggability and non-toxic profile, which ensures safety cum efficacy.

Conclusion

Currently, there is no satisfactory therapy for HBV infection and HCC. The present study identified a group of druggable and nontoxic phytocompounds from herbs modulating the therapeutic targets involved in HBV infection and HCC through *in silico* techniques involving compound-protein-pathway enrichment analysis coupled with network pharmacology, molecular docking, and dynamics simulation studies. Results of this study suggest that flavonoid and diterpenoid fraction of *A. paniculata* and flavonoid fraction of *T. populnea* could be potent as a combined therapy for the treatment of patients with HBV infection and its associated HCC without expecting adverse drug reactions. Theoretical and predicted knowledge generated in this study needs to be further confirmed using *in vitro* studies including in viral cell lines and animal models.

Abbreviations

HBV Hepatitis B Virus, HCC Hepatocellular Carcinoma, SMILES Simplified Molecular Input Line-Entry System, MF Molecular Formula, MW Molecular Weight, HBA Hydrogen Bond Acceptor HBD Hydrogen Bond Donor, STRING Search Tool for the Retrieval of Interacting Genes/Proteins, KEGG Kyoto Encyclopedia of Genes and Genomes, GO Gene Ontology, RCSB PDB Research Collaboratory for Structural Bioinformatics Protein Data Bank, TTD Therapeutic Target Database, MD Molecular Dynamics, NO Nitric Oxide, COX-2 Cyclooxygenase-2.

Declarations

Acknowledgments

The authors gratefully acknowledge the extra-mural funding of the Indian Council of Medical Research, Department of Health Research, New Delhi (ICMR Grant IRIS ID 2019-0045). The authors are thankful to

the KLE Academy of Higher Education and Research, Belagavi, India for the support.

Funding

The authors gratefully acknowledge the extra-mural funding of the Indian Council of Medical Research, Department of Health Research, New Delhi (ICMR Grant IRIS ID 2019-0045).

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of data and material

All data generated or analysed during this study are included in this published article (and its supplementary information files).

Code availability

Not applicable

Authors' contribution

VSP main worker, designed and performed the study, writing, and drafting of the manuscript. DRH and SR supervised the study, provided inputs, critically reviewed drafts, and finalized the manuscript. UV supervised VSP on Network pharmacology, molecular docking, MD simulation, provided inputs, and critically revised the manuscript. SD helped VSP in data mining, analysis, and manuscript writing. SSJ and HH co-supervised, provided botanical inputs and critical revision of the manuscript. All the authors read and approved the final.

References

1. Liu C, Yang S, Wang K et al (2019) Alkaloids from Traditional Chinese Medicine against hepatocellular carcinoma. *Biomed Pharmacother* 120:109543. <https://doi.org/10.1016/j.biopha.2019.109543>
2. Torresi J, Tran BM, Christiansen D et al (2019) HBV-related hepatocarcinogenesis: the role of signalling pathways and innovative ex vivo research models. *BMC Cancer* 19:707.

<https://doi.org/10.1186/s12885-019-5916-6>

3. Niu B, Hann H (2017) Hepatitis B Virus – Related Hepatocellular Hepatitis Hepatocellular Carcinoma: Carcinoma : Carcinogenesis, Prevention, and Treatment. *Updat Liver Cancer* 13
4. Reddy DCS (2019) Elimination of viral hepatitis: Evolution and India's response. *Indian J Public Health* 63:275–276
5. Wang Z, Li Z, Ye Y et al (2016) Oxidative stress and liver cancer: Etiology and therapeutic targets. *Oxid. Med. Cell. Longev.* 2016
6. Tarocchi M, Polvani S, Marroncini G, Galli A (2014) Molecular mechanism of hepatitis B virus-induced hepatocarcinogenesis. *World J Gastroenterol* 20:11630–11640
7. Ayub A, Ashfaq UA, Haque A (2013) HBV induced HCC: major risk factors from genetic to molecular level. *Biomed Res Int* 2013:810461. <https://doi.org/10.1155/2013/810461>
8. He G, Karin M (2010) NF- κ B and STAT3 – key players in liver inflammation and cancer. *Nat Publ Gr* 21:159–168. <https://doi.org/10.1038/cr.2010.183>
9. Fan Y, Mao R, Yang J (2013) NF- κ B and STAT3 signaling pathways collaboratively link inflammation to cancer. 4:176–185. <https://doi.org/10.1007/s13238-013-2084-3>
10. Ahn JY, Chung EY, Kwun HJ, Jang KL (2001) Transcriptional repression of p21 waf1 promoter by hepatitis B virus X protein via a p53-independent pathway. *Gene* 275:163–168
11. Bouchard M, Giannakopoulos S, Wang EH et al (2001) Hepatitis B Virus HBx Protein Activation of Cyclin A – Cyclin-Dependent Kinase 2 Complexes and G 1 Transit via a Src Kinase Pathway. *J Virol* 75:4247–4257. <https://doi.org/10.1128/JVI.75.9.4247>
12. Leach JK, Qiao L, Fang Y et al (2003) Regulation of p21 and p27 expression by the hepatitis B virus X protein and the alternate initiation site X proteins, AUG2 and AUG3. *J Gastroenterol Hepatol* 18:376–385
13. Mukherji A, Janbandhu VC, Kumar V (2007) HBx-dependent cell cycle deregulation involves interaction with cyclin E / A – cdk2 complex and destabilization of p27 Kip1. *Biochem J* 256:247–256. <https://doi.org/10.1042/BJ20061091>
14. Gearhart TL, Bouchard MJ (2011) The hepatitis B virus HBx protein modulates cell cycle regulatory proteins in cultured primary human hepatocytes. *Virus Res* 155:363–367. <https://doi.org/10.1016/j.virusres.2010.09.023>
15. Slagle BL, Bouchard MJ (2016) Hepatitis B Virus X and Regulation of Viral Gene Expression. *Cold Spring Harb Perspect Med* 6:a021402. <https://doi.org/10.1101/cshperspect.a021402>
16. Yeh C, Shen C, Tai D et al (2000) Identification and characterization of a prevalent hepatitis B virus X protein mutant in Taiwanese patients with hepatocellular carcinoma. *Oncogene* 19:5213–5220
17. Abdel-hamid NM, Nazmy MH, Mahmoud AW, Fawzy MA (2011) A survey on herbal management of hepatocellular carcinoma. *World J Hepatol* 3:175–183. <https://doi.org/10.4254/wjh.v3.i7.175>
18. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 4:682–690. <https://doi.org/10.1038/nchembio.118>

19. Ivanov SM, Lagunin AA, Rudik AV et al (2018) ADVERPred-Web Service for Prediction of Adverse Effects of Drugs. *J Chem Inf Model* 58:8–11. <https://doi.org/10.1021/acs.jcim.7b00568>
20. Liu T, Lin Y, Wen X et al (2007) BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* 35:. <https://doi.org/10.1093/nar/gkl999>
21. Otasek D, Morris JH, Bouças J et al (2019) Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol* 20:1–15
22. Patil VS, Khatib NA (2020) Triterpene saponins from *Barringtonia acutangula* (L.) Gaertn as a potent inhibitor of 11 β -HSD1 for type 2 diabetes mellitus, obesity, and metabolic syndrome. *Clin Phytoscience* 6:61. <https://doi.org/10.1186/s40816-020-00210-y>
23. Eswar N, Webb B, Marti-Renom MA et al (2007) Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Protein Sci* 50:. <https://doi.org/10.1002/0471140864.ps0209s50>
24. Hollingsworth SA, Dror RO (2018) Molecular Dynamics Simulation for All. *Neuron* 99:1129–1143
25. Bowers KJ, Chow DE, Xu H et al (2006) Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: *ACM/IEEE SC 2006 Conference (SC'06)*. IEEE, pp 43–43
26. Samdani A, Vetrivel U (2018) POAP: A GNU parallel based multithreaded pipeline of open babel and AutoDock suite for boosted high throughput virtual screening. *Comput Biol Chem* 74:39–48. <https://doi.org/10.1016/j.compbiolchem.2018.02.012>
27. Spicer JF, Rudman SM (2010) EGFR inhibitors in non-small cell lung cancer (NSCLC): The emerging role of the dual irreversible EGFR/HER2 inhibitor BIBW 2992. *Target Oncol* 5:245–255
28. Westphal M, Maire CL, Lamszus K (2017) EGFR as a Target for Glioblastoma Treatment: An Unfulfilled Promise. *CNS Drugs* 31:723–735. <https://doi.org/10.1007/s40263-017-0456-6>
29. Morris C (2002) The role of EGFR-directed therapy in the treatment of breast cancer. In: *Breast Cancer Research and Treatment*. Kluwer Academic Publishers, pp 51–55
30. Berasain C, Avila MA (2014) The EGFR signalling system in the liver: From hepatoprotection to hepatocarcinogenesis. *J Gastroenterol* 49:9–23
31. Komposch K, Sibilia M (2015) EGFR signaling in liver diseases. *Int. J. Mol. Sci.* 17
32. Gajiwala KS, Feng J, Ferre R et al (2013) Insights into the aberrant activity of mutant EGFR kinase domain and drug recognition. *Structure* 21:209–219. <https://doi.org/10.1016/j.str.2012.11.014>
33. Lu C, Fu W, Zhou R, Hu W (2020) Network pharmacology-based study on the mechanism of Yiganling capsule in hepatitis B treatment. *BMC Complement Med Ther* 20:1–17
34. Dandri M, Petersen J (2016) Mechanism of Hepatitis B Virus Persistence in Hepatocytes and Its Carcinogenic Potential. *Clin Infect Dis* 62:281–288. <https://doi.org/10.1093/cid/ciw023>
35. DeLeve LD (2003) Alterations in Hepatic Metabolism of Drugs
36. Li Y, li RCGM (2011) Herbal Medicine and Hepatocellular Carcinoma: Applications and Challenges. 2011:. <https://doi.org/10.1093/ecam/neq044>
37. Alshammari TM (2016) Drug safety: The concept, inception and its importance in patients ' health. *Saudi Pharm J* 24:405–412. <https://doi.org/10.1016/j.jsps.2014.04.008>

38. Raschi E, Ponti F, De (2015) Drug- and herb-induced liver injury: Progress, current challenges and emerging signals of post-marketing risk. *World J Hepatol* 7:1761–1771.
<https://doi.org/10.4254/wjh.v7.i13.1761>
39. Stickel F, Patsenker E, Schuppan D (2005) Herbal hepatotoxicity. *Hepatology* 43:901–910.
<https://doi.org/10.1016/j.jhep.2005.08.002>
40. Chen S, Jiang H, Cao Y et al (2016) Drug target identification using network analysis: Taking active components in Sini decoction as an example. *Nat Publ Gr* 1–14. <https://doi.org/10.1038/srep24245>
41. Todorov D, Hinkov A, Shishkova K, Shishkov S (2014) Antiviral potential of Bulgarian medicinal plants. *Phytochem Rev* 13:525–538
42. Schiffer E, Housset C, Cacheux W et al (2005) Gefitinib, an EGFR inhibitor, prevents hepatocellular carcinoma development in the rat liver with cirrhosis. *Hepatology* 41:307–314.
<https://doi.org/10.1002/hep.20538>
43. Zheng H, Yang Y, Hong Y, gang et al (2019) Tropomodulin 3 modulates EGFR-PI3K-AKT signaling to drive hepatocellular carcinoma metastasis. *Mol Carcinog* 58:1897–1907.
<https://doi.org/10.1002/mc.23083>
44. Zhangyuan G, Wang F, Zhang H et al (2020) VersicanV1 promotes proliferation and metastasis of hepatocellular carcinoma through the activation of EGFR-PI3K-AKT pathway. *Oncogene* 39:1213–1230. <https://doi.org/10.1038/s41388-019-1052-7>
45. Liu Z, Ning F, Cai Y et al (2021) The EGFR-P38 MAPK axis up-regulates PD-L1 through miR-675-5p and down-regulates HLA-ABC via hexokinase-2 in hepatocellular carcinoma cells. *Cancer Commun* 41:62–78. <https://doi.org/10.1002/cac2.12117>
46. Ghosh N, Ghosh R, Mandal V, Mandal SC (2011) Recent advances in herbal medicine for treatment of liver diseases. *Pharm Biol* 49:970–988. <https://doi.org/10.3109/13880209.2011.558515>
47. Jayakumar T, Hsieh C, Lee J, Sheu J (2013) Experimental and Clinical Pharmacology of *Andrographis paniculata* and Its Major Bioactive Phytoconstituent Andrographolide. *Evid Based Complement Altern Med* 2013
48. Jadhav AK, Karuppayil SM (2021) *Andrographis paniculata* (Burm. F) Wall ex Nees: Antiviral properties. *Phyther. Res*
49. Tan Y, Chiow K, Huang D, Wong S (2010) Andrographolide regulates epidermal growth factor receptor and transferrin receptor trafficking in epidermoid carcinoma (A-431) cells: Research paper. *Br J Pharmacol* 159:1497–1510. <https://doi.org/10.1111/j.1476-5381.2009.00627.x>
50. Kapil A, Koul IB, Banerjee SK, Gupta BD (1993) Antihepatotoxic effects of major diterpenoid constituents of *Andrographis paniculata*. *Biochem Pharmacol* 46:182–185.
[https://doi.org/10.1016/0006-2952\(93\)90364-3](https://doi.org/10.1016/0006-2952(93)90364-3)
51. Handa SS, Sharma A (1990) Hepatoprotective activity of andrographolide from *Andrographis paniculata* against carbontetrachloride. *Indian J Med Res* 92:276–283
52. Handa SS, Sharma A (1990) Hepatoprotective activity of andrographolide against galactosamine & paracetamol intoxication in rats. *Indian J Med Res* 92:284–292

53. Batkhuu J, Hattori K, Takano F et al (2002) Suppression of NO production in activated macrophages in vitro and ex vivo by neoandrographolide isolated from *Andrographis paniculata*. *Biol Pharm Bull* 25:1169–1174
54. Chen H, Ma Y-B, Huang X-Y et al (2014) Synthesis, structure–activity relationships and biological evaluation of dehydroandrographolide and andrographolide derivatives as novel anti-hepatitis B virus agents. *Bioorg Med Chem Lett* 24:2353–2359. <https://doi.org/10.1016/j.bmcl.2014.03.060>
55. Jun Liu Z-T, Wang, Li-Li Ji B-XG (2007) Inhibitory effects of neoandrographolide on nitric oxide and prostaglandin E 2 production in LPS-stimulated murine macrophage. *Mol Cell Biochem* 2:49–57. <https://doi.org/10.1007/s11010-006-9349-6>
56. Yuvaraj P, Subramoniam A (2009) Short communication hepatoprotective property of *thespesia populnea* against carbon tetrachloride induced liver damage in rats. *J Basic Clin Physiol Pharmacol* 20:169–178. <https://doi.org/10.1515/JBCPP.2009.20.2.169>
57. Jo S, Kim S, Shin DH et al (2020) Inhibition of SARS-CoV 3CL protease by flavonoids Inhibition of SARS-CoV 3CL protease by flavonoids. *J Enzyme Inhib Med Chem* 35:145–151. <https://doi.org/10.1080/14756366.2019.1690480>
58. Behbahani M, Sayedipour S, Pourazar A, Shanehsazzadeh M (2014) In vitro anti-HIV-1 activities of kaempferol and kaempferol-7-Oglucoside isolated from *Securigera securidaca*. *Res Pharm Sci* 9:463–469
59. Arthanari S, Renukadevi P, Vanitha J (2011) Evaluation of Antiviral and Cytotoxic Activities of Methanolic Extract of *Thespesia Populnea* (Malvaceae) Flowers. *J of Herbs. Spices Med Plants* 17:37–41. <https://doi.org/10.1080/10496475.2011.605514>

Table

Due to technical limitations, table 1 is only available as a download in the Supplemental Files section.

Figures

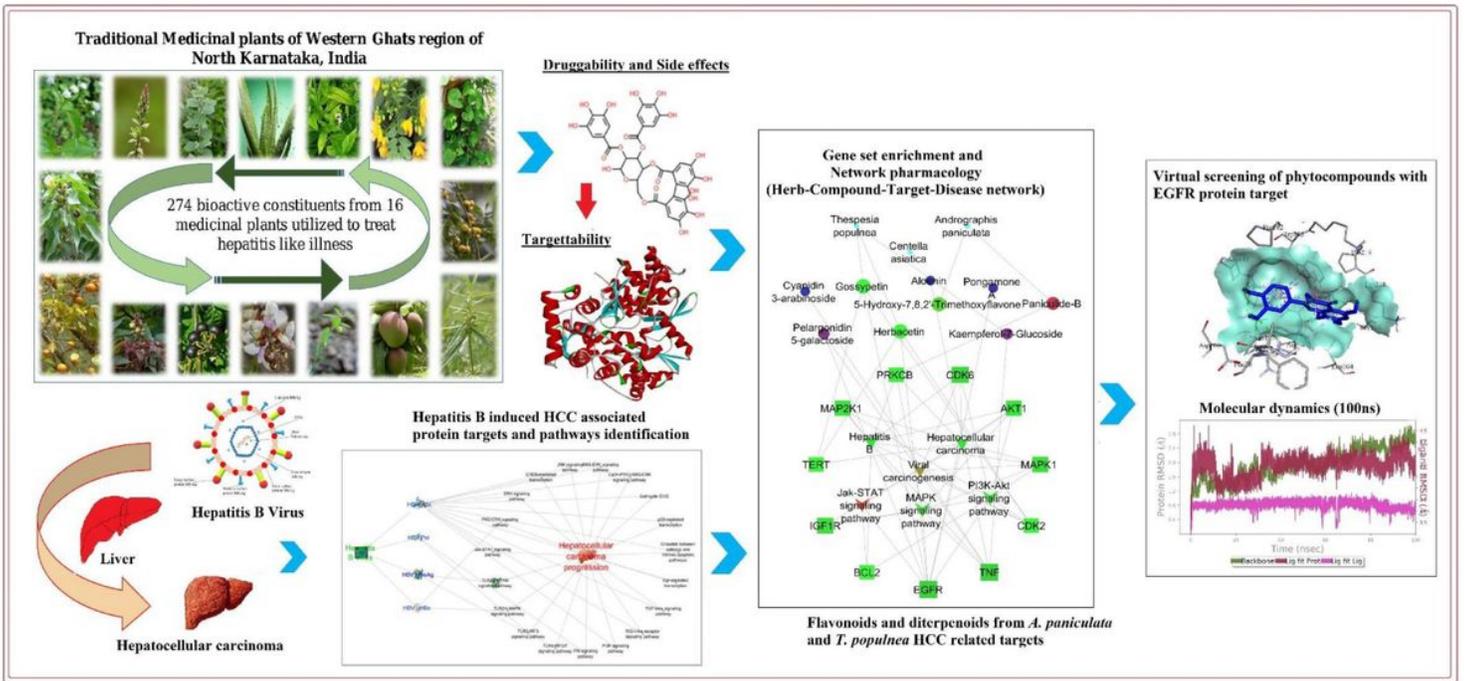


Figure 1

Workflow Implemented in the present study.

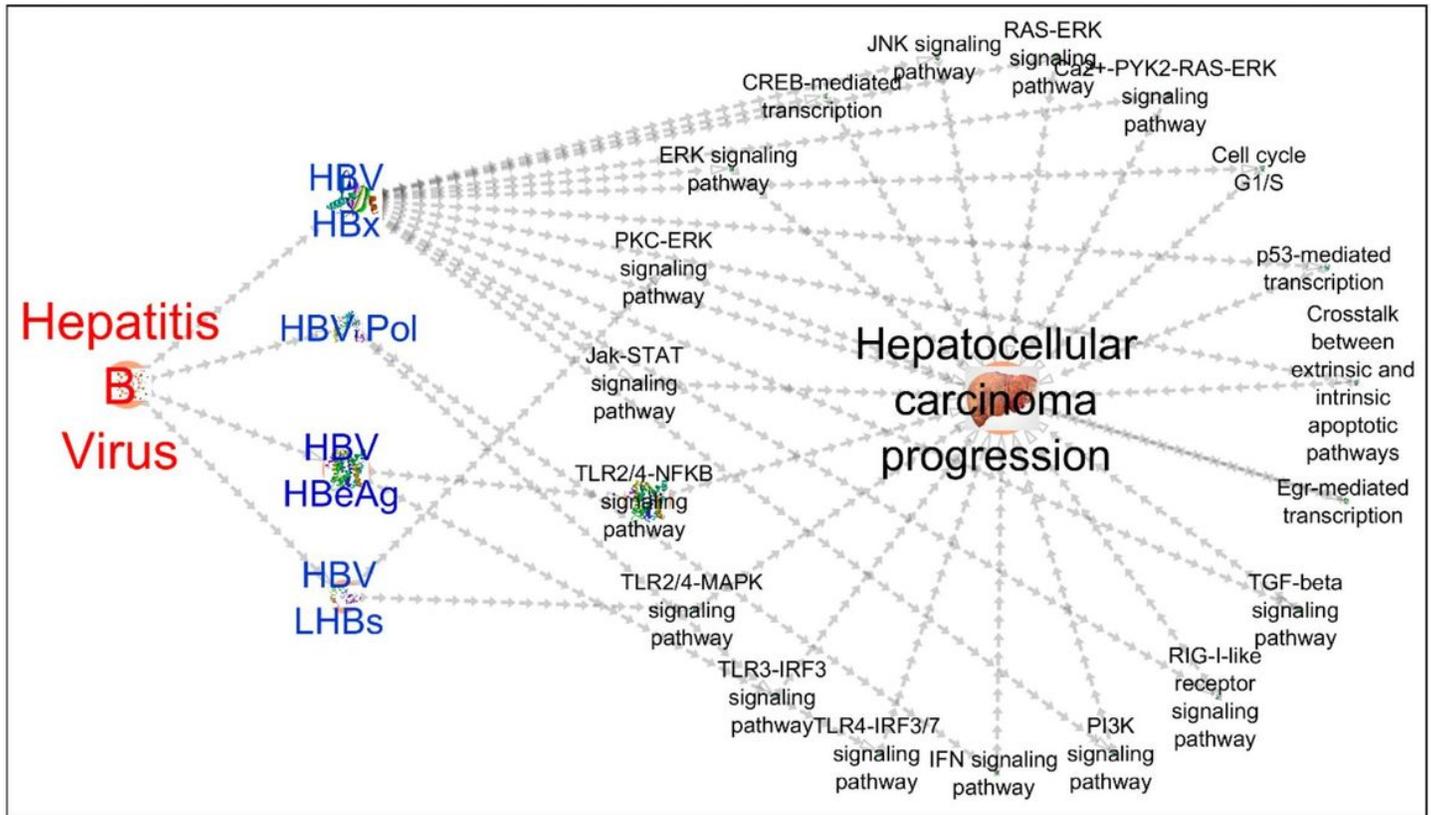


Figure 2

Network representation of pathways involved in HBV infection induced HCC

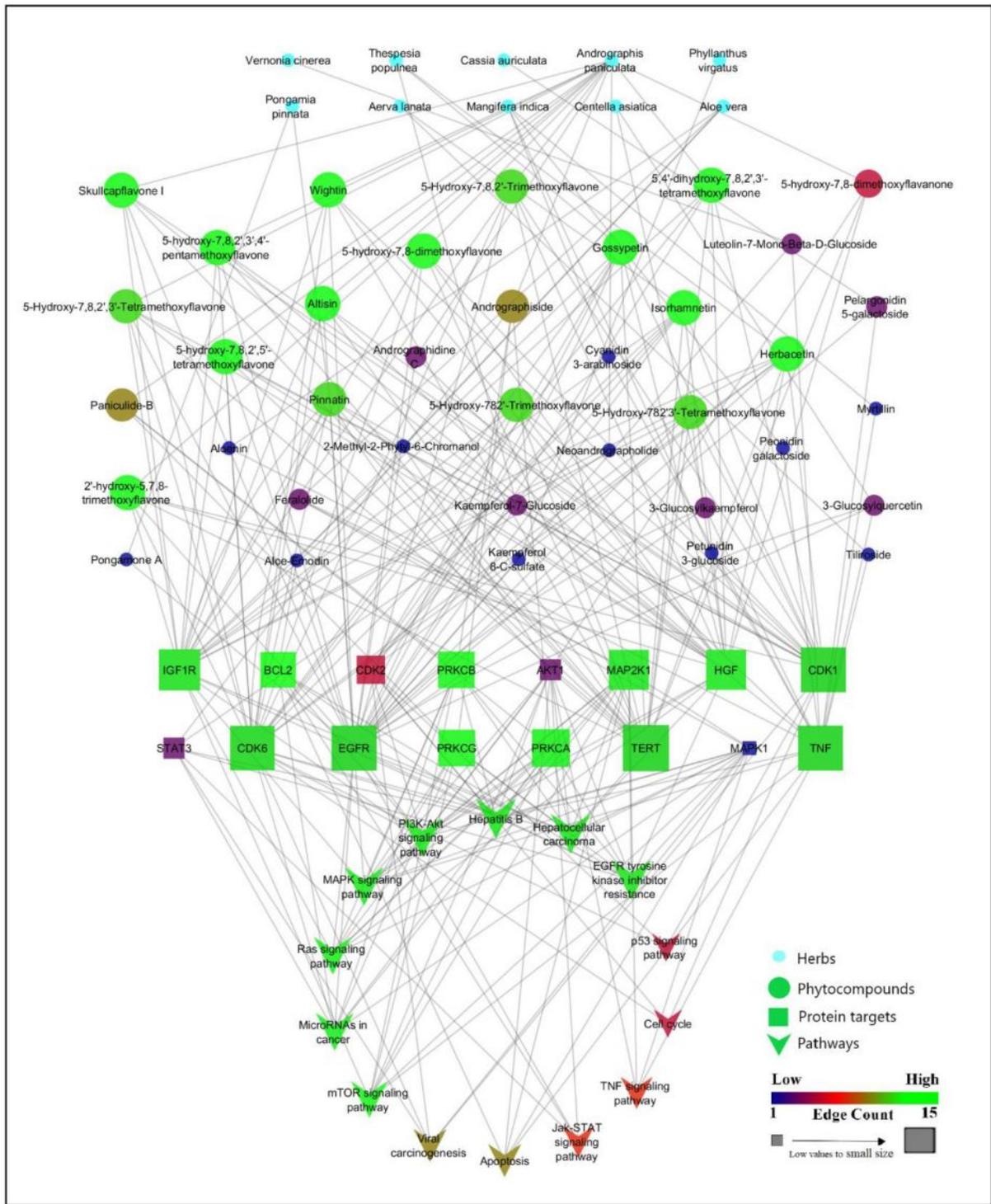


Figure 3

Network representation of herbs-phytocompounds-proteins-pathways of HBV infection and HCC.

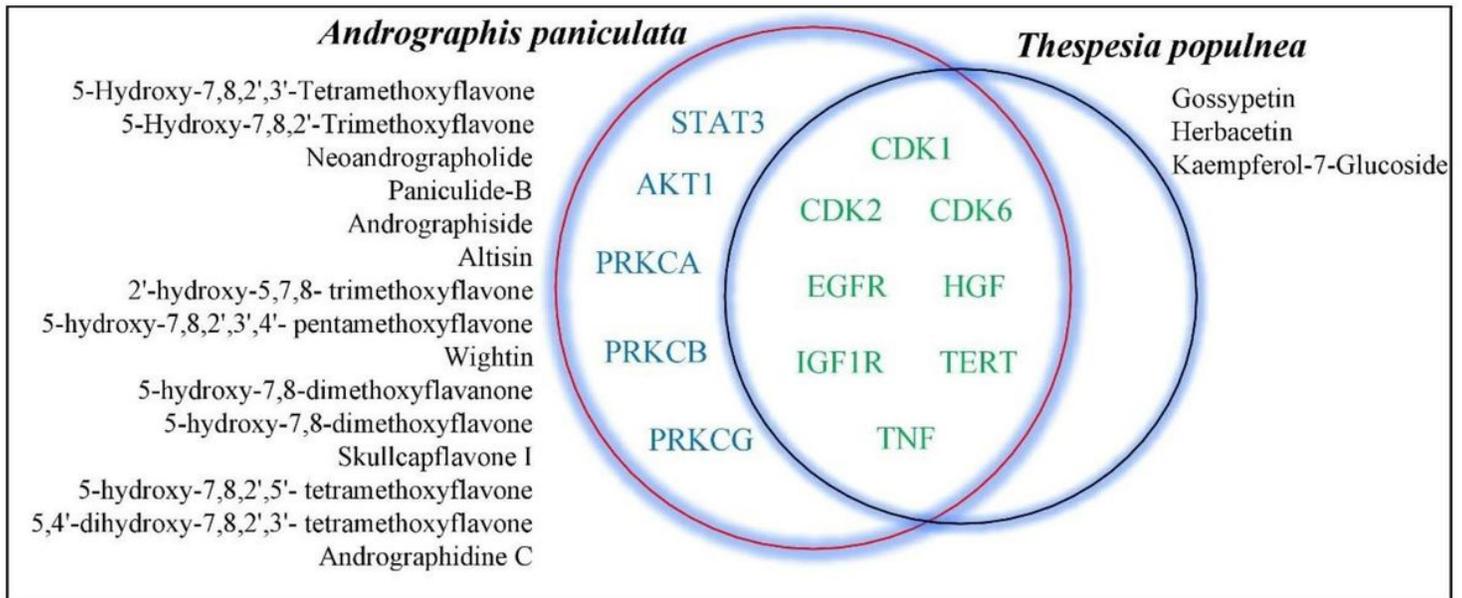


Figure 4

Venn diagram represents the modulated protein targets involved in HBV-induced HCC by the phytochemicals from *Andrographis paniculata* and *Thespesia populnea*.

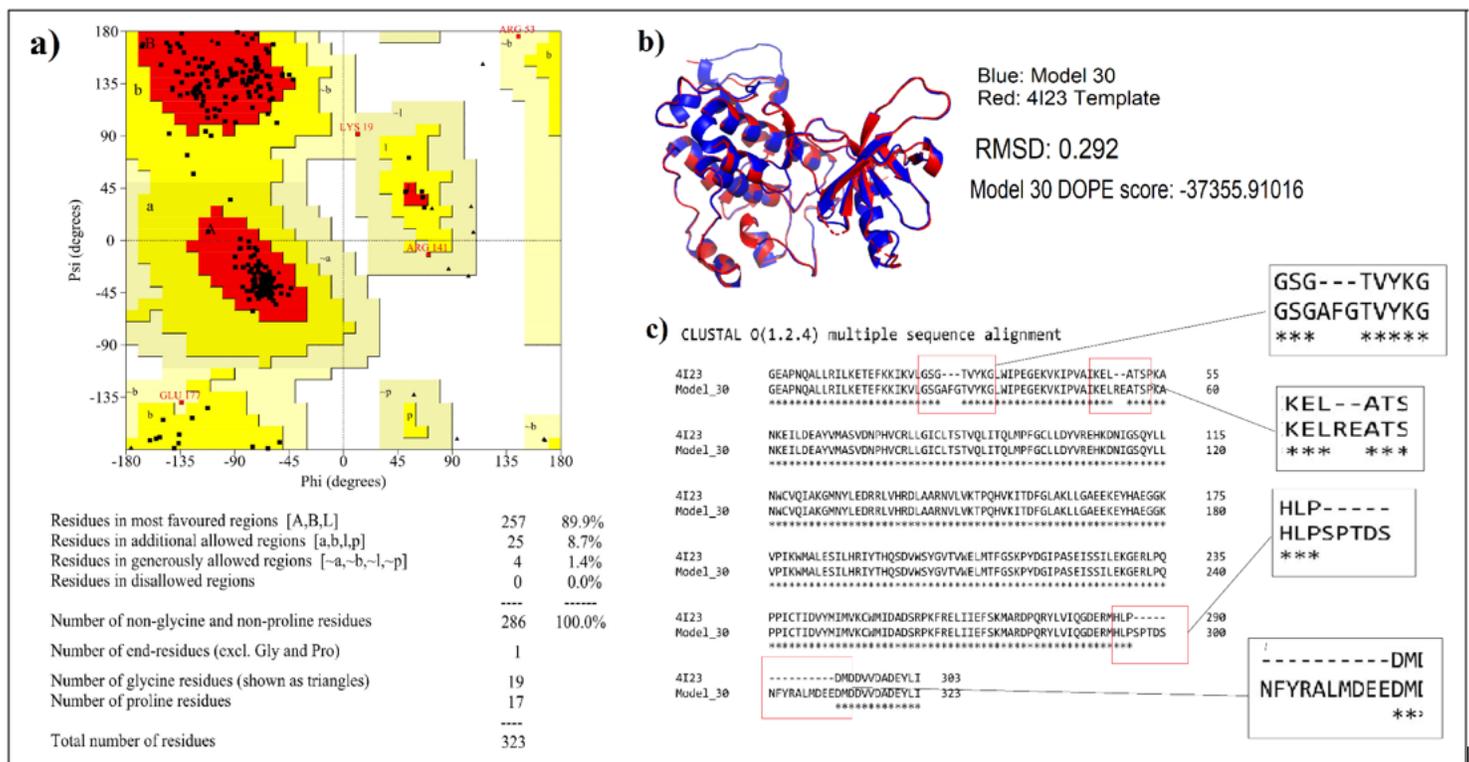


Figure 5

WT- EGFR model 30. a) Ramachandran plot for amino acid distribution b) PyMOL superimpose (Red: Template (PDB file 4123) and Blue: Model 30), c) Clustal W template and model 30 alignment.

WT EGFR model 30 RMSD

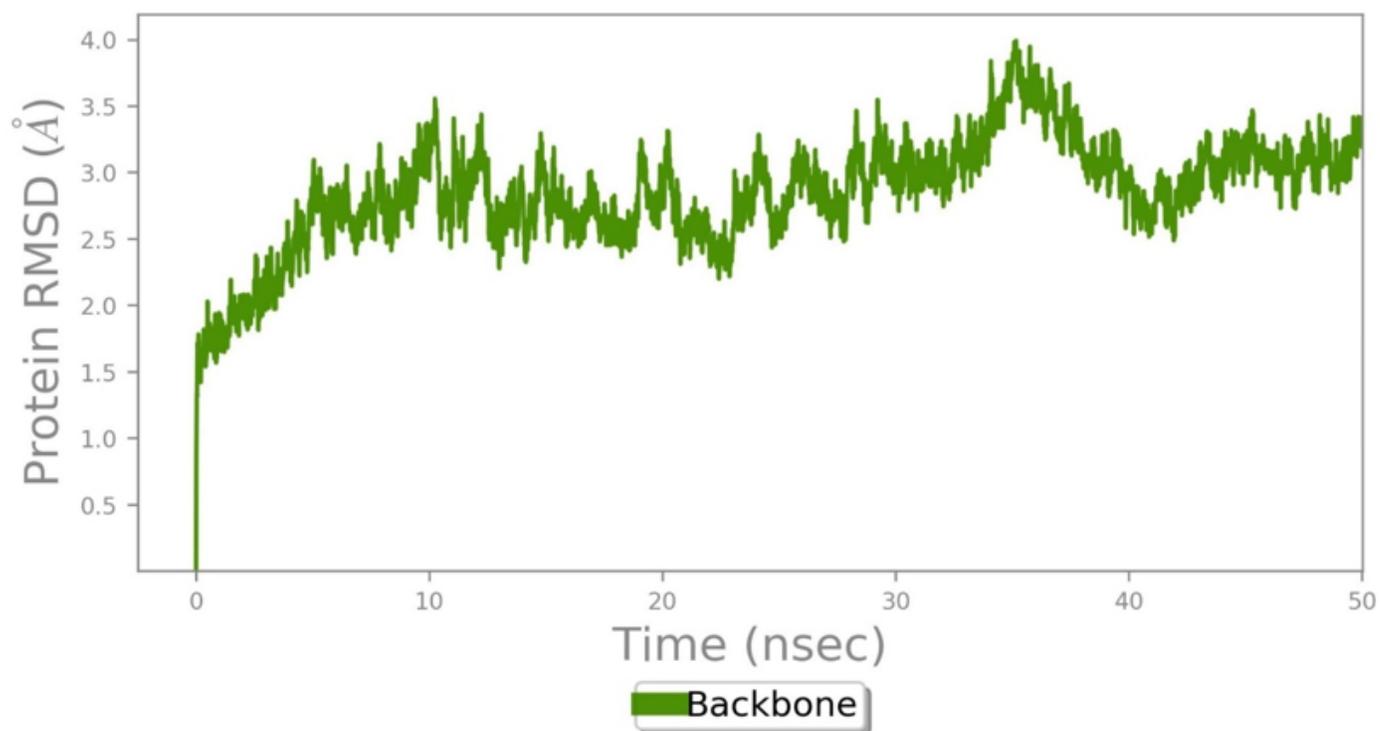


Figure 6

WT EGFR model-30 RMSD plot showing the deviation and minimum energy conformation observed at 45252.0thps.

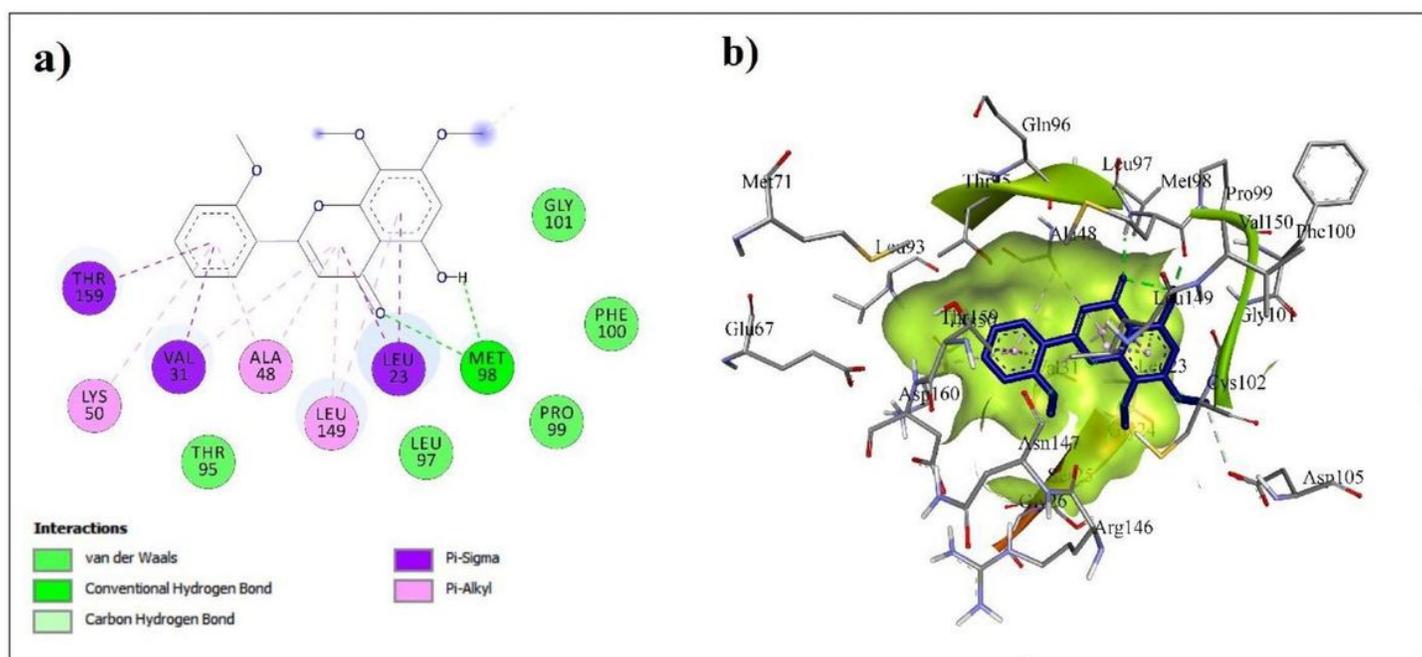


Figure 7

Intermolecular interactions of Andrographin with WT EGFR. a) 2D representation b) 3D representation of Andrographin within the binding pocket.

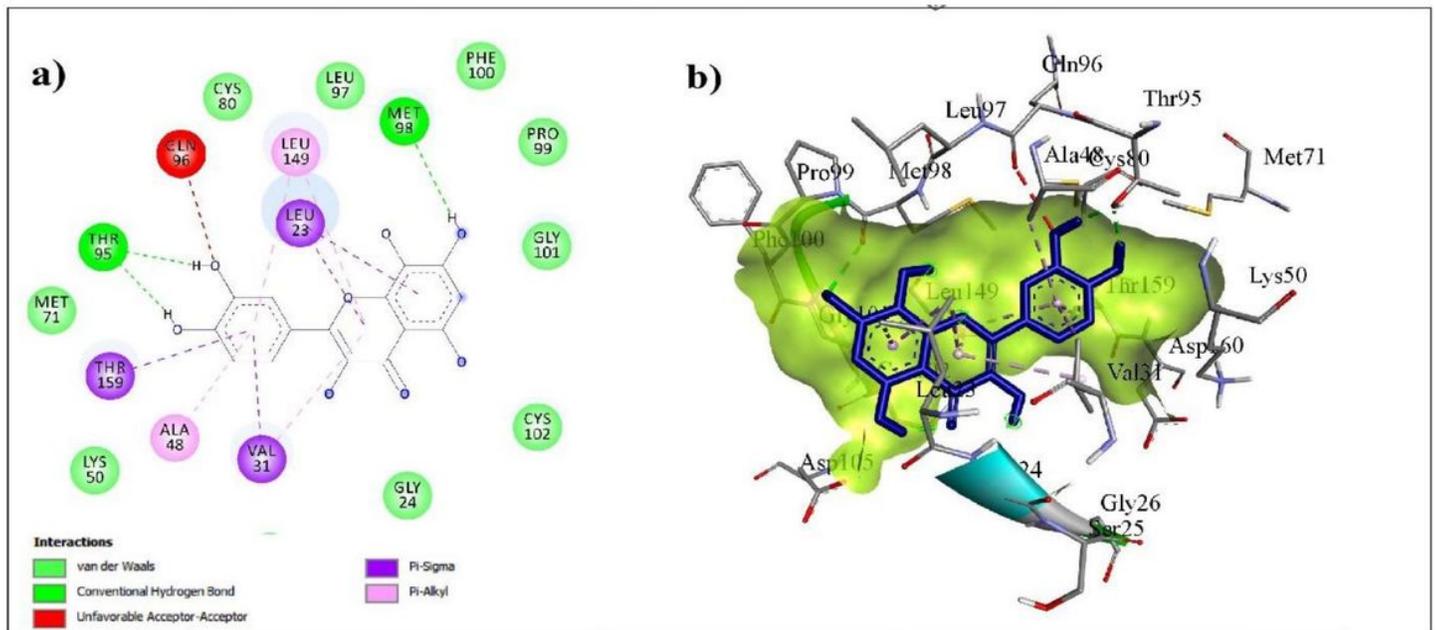


Figure 8

Intermolecular interactions of Gossypetin with WT EGFR. a) 2D representation b) 3D representation of Andrographin within the binding pocket.

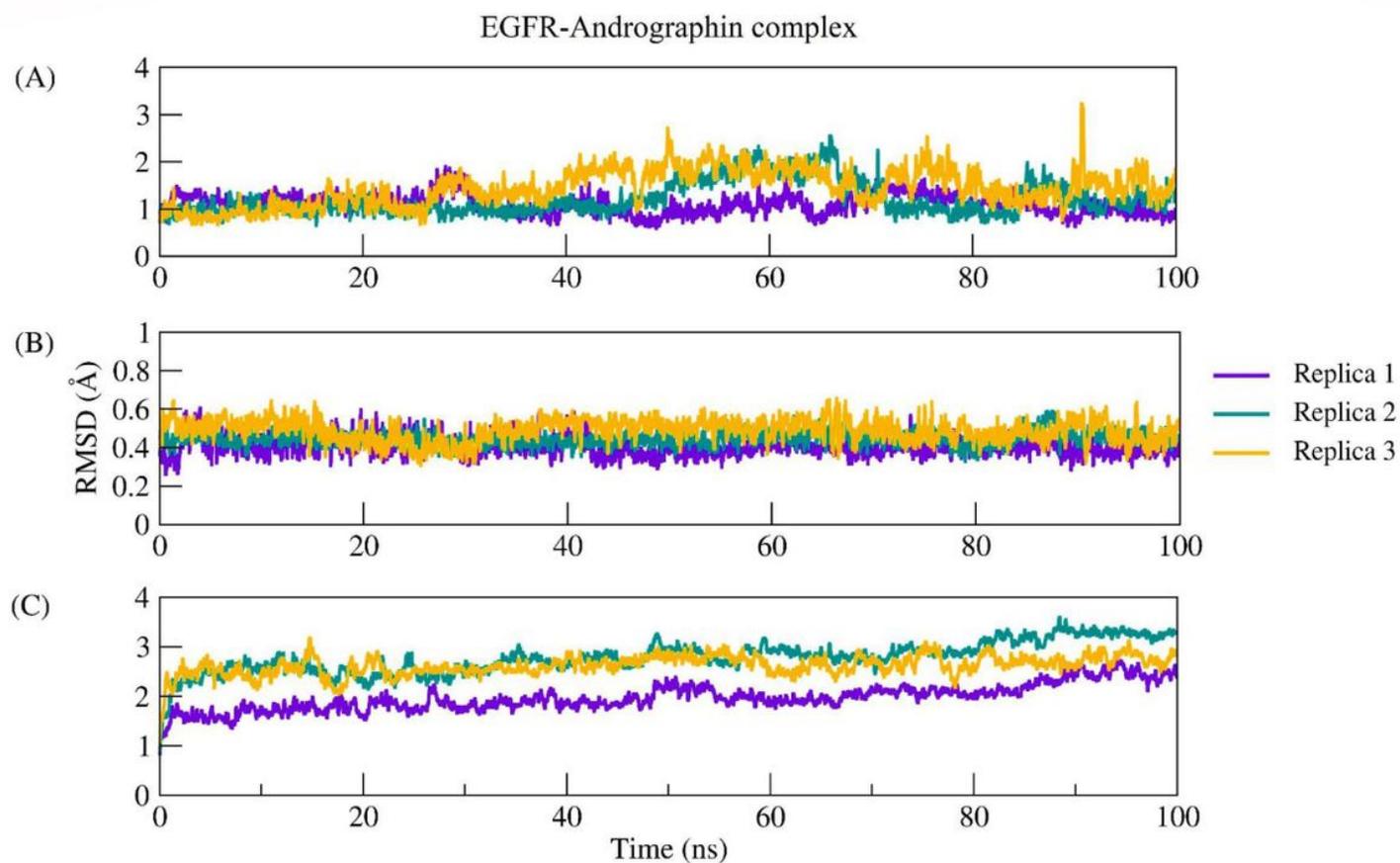


Figure 9

Andrographin with WT EGFR 100ns MD production run. (A) Plot depicting the docked complex stability "Ligand RMSD w.r.t. Protein", (B) Plot depicting the stability of "Ligand w.r.t. Ligand", and (C) RMSD plot depicting the "Protein Backbone".

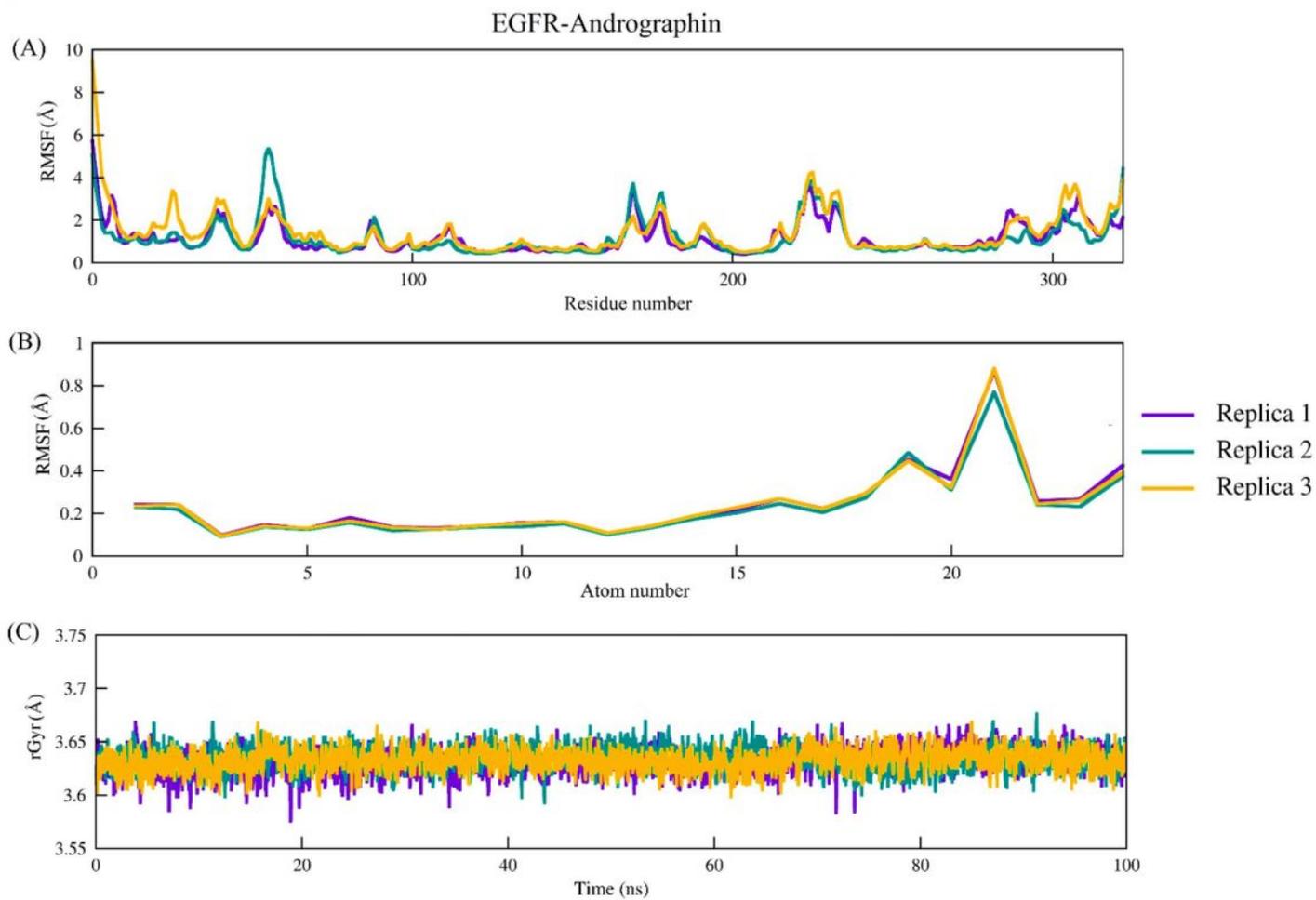


Figure 10

EGFR-Andrographin complex 100ns MD production run. (A) Protein RMSF plot, (B) Ligand RMSF plot, and (C) A gyration plot depicting the extendedness' of a ligand during a 100ns production run.

EGFR-Andrographin complex

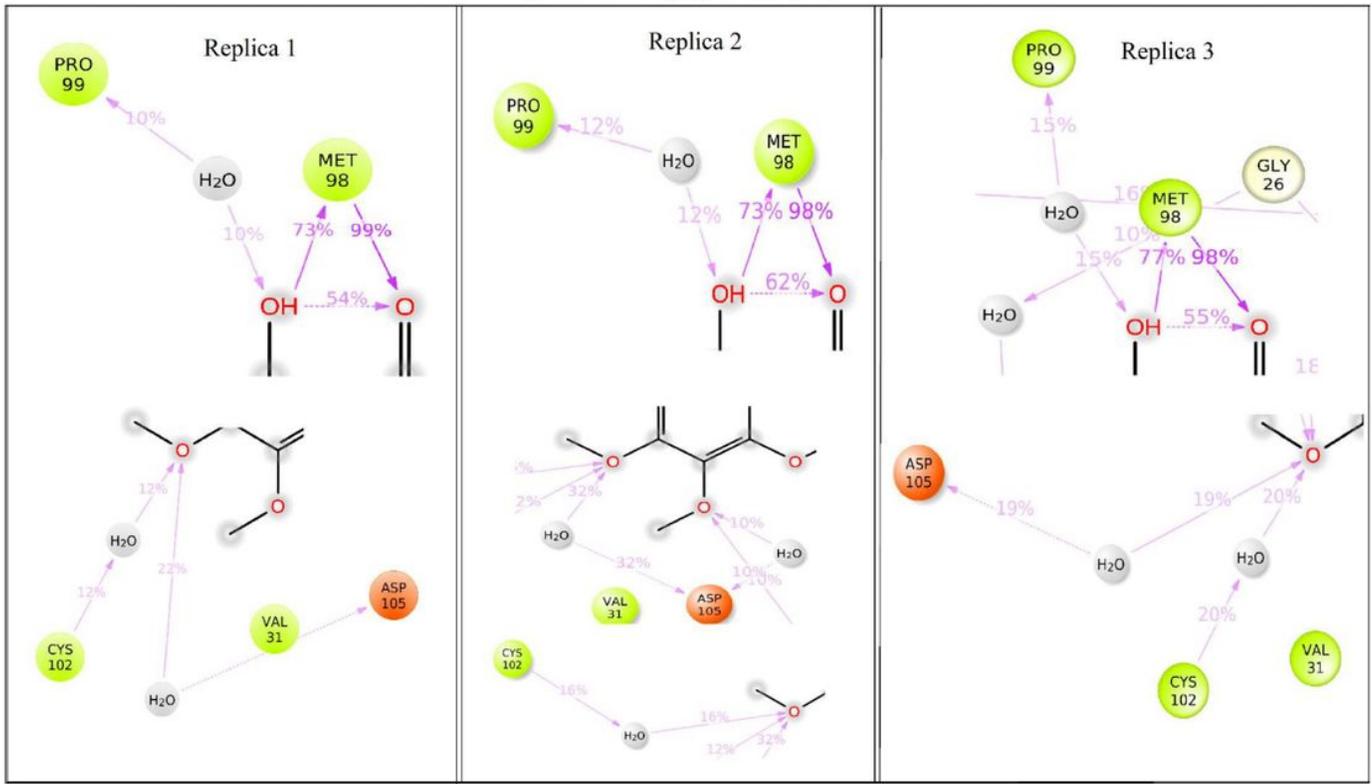


Figure 11

EGFR-Gossypetin complex sustained intermolecular interactions and percentage interaction fraction throughout 100ns production run

EGFR-Gossypetin complex

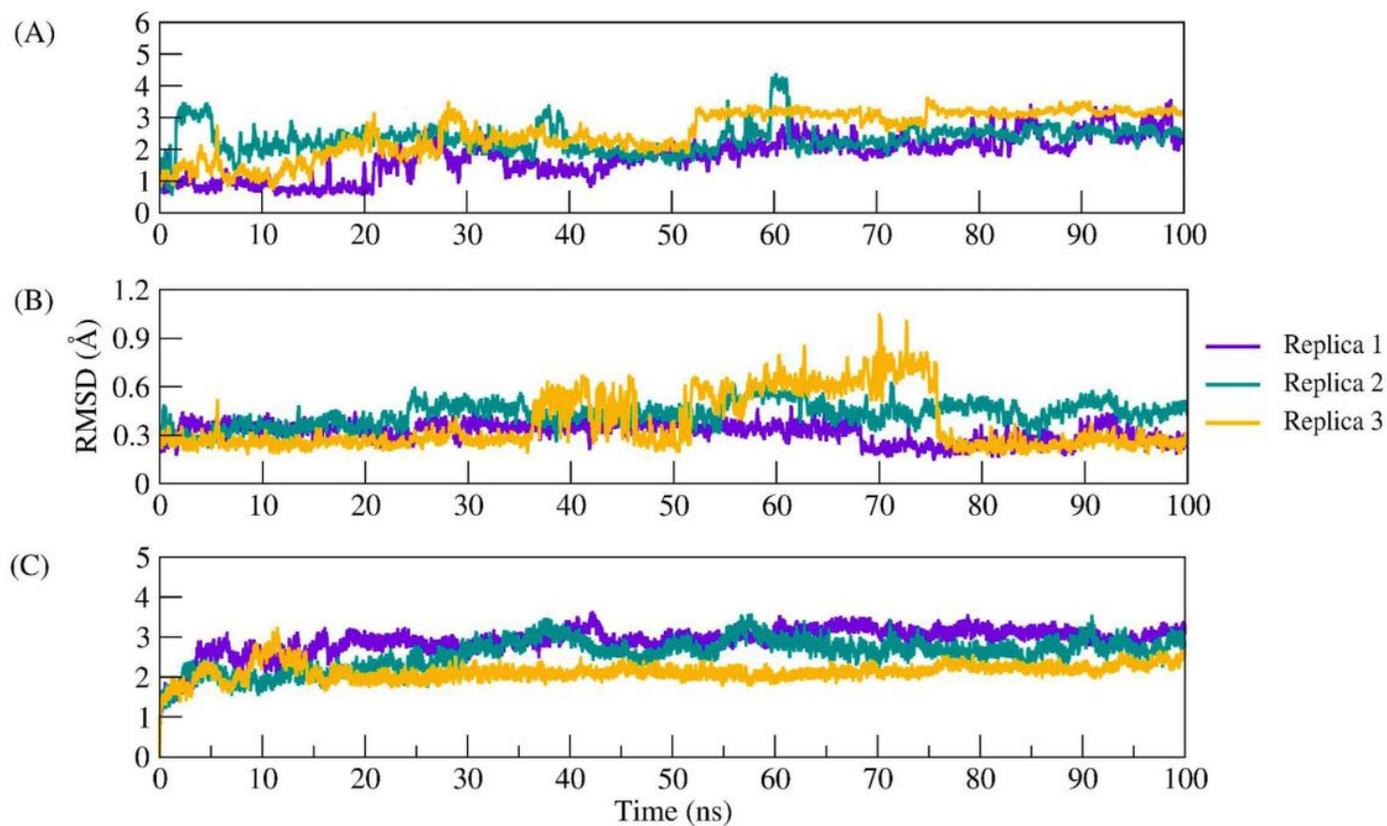


Figure 12

Gossypetin with WT EGFR 100ns MD production run. (A) Plot depicting the docked complex stability “Ligand RMSD w.r.t. Protein”, (B) Plot depicting the stability of “Ligand w.r.t Ligand”, and (C) RMSD plot depicting the “Protein Backbone”.

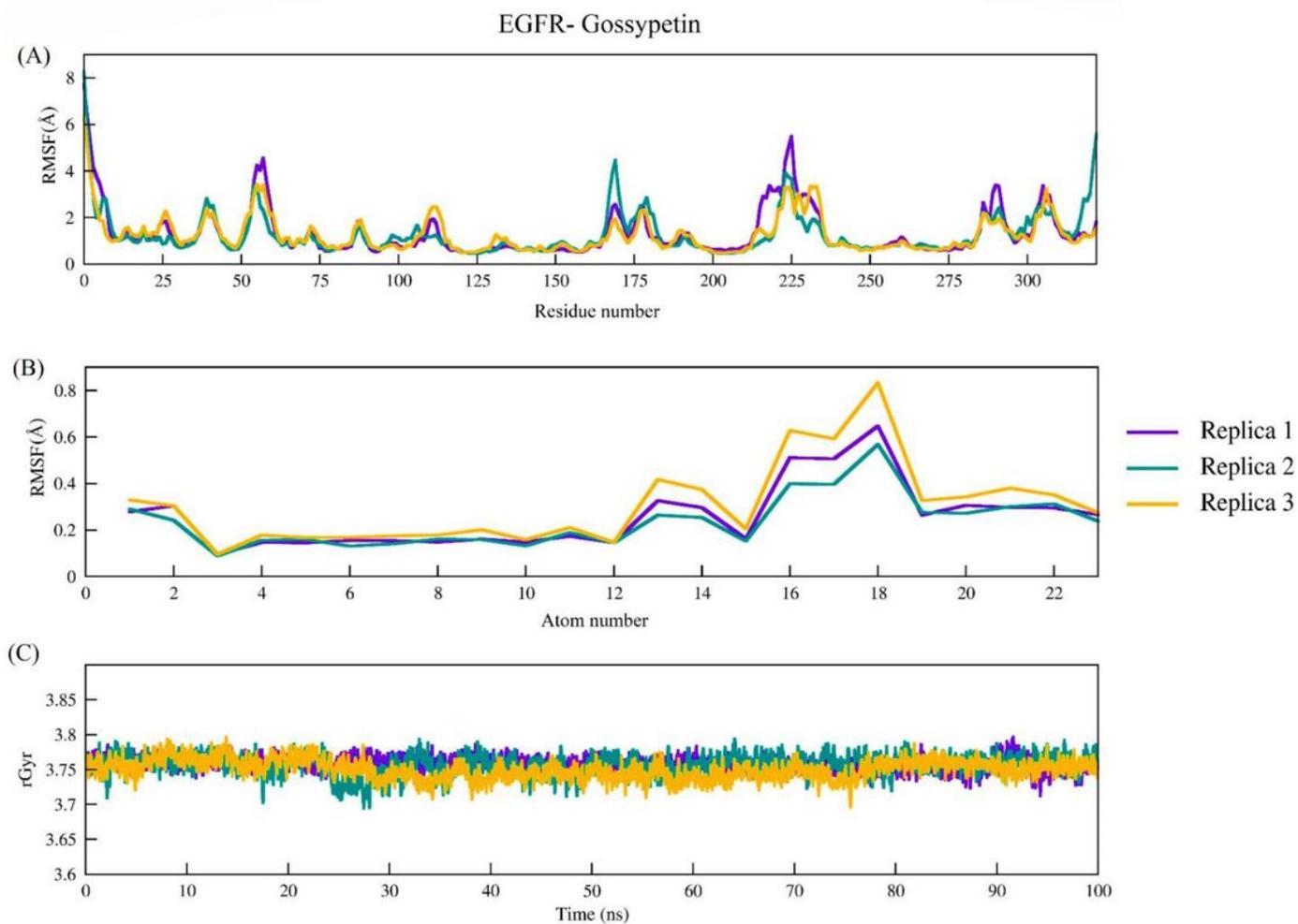


Figure 13

EGFR- Gossypetin complex 100ns MD production run. (A) Protein RMSF plot, (B) Ligand RMSF plot, and (C) A gyration plot depicting the extendedness' of a ligand during a 100ns production run.

EGFR-Gossypetin complex

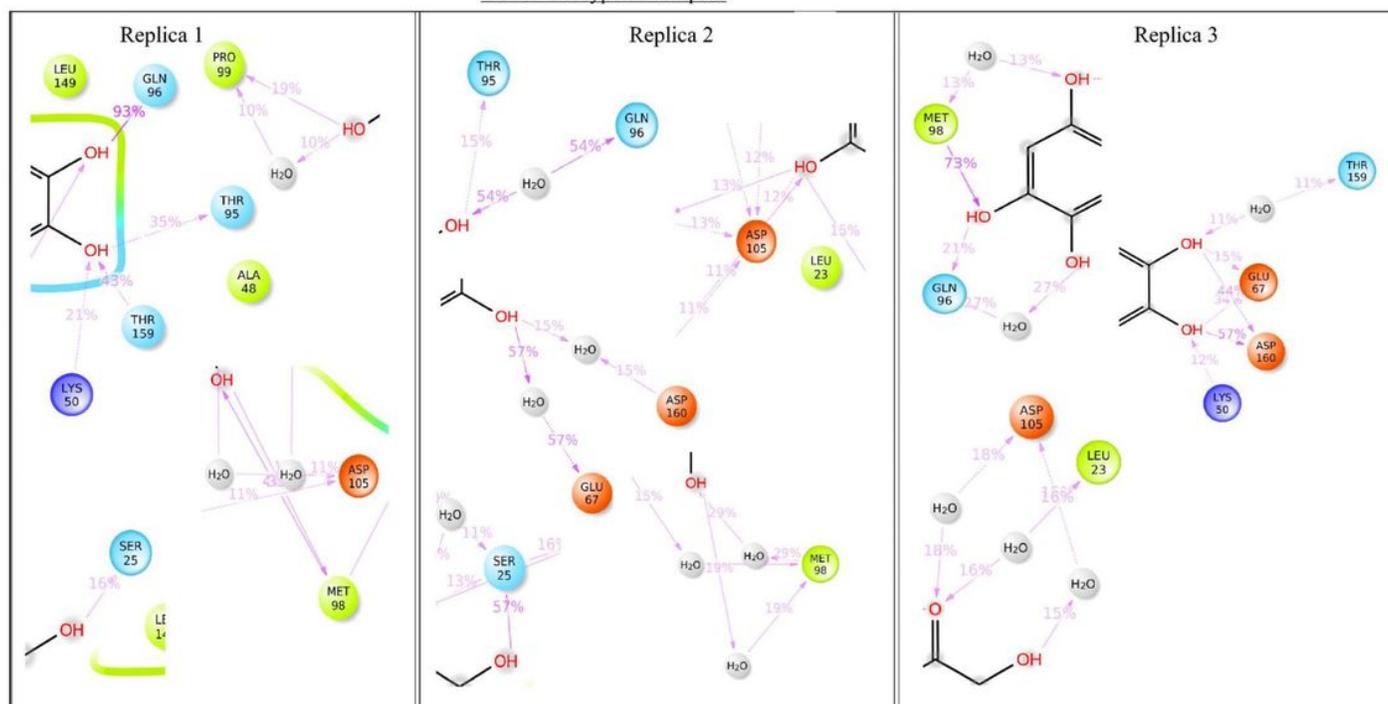


Figure 14

EGFR-Gossypetin complex sustained intermolecular interactions and percentage interaction fraction throughout 100ns production run

Supplementary Files

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

- [Table1Phytocompoundspositivedrug.docx](#)
- [SupplementaryTables.pdf](#)