

Network pharmacology of *Potentilla nepalensis* extract revealed p53, Nf- kB1, and HSP proteins as potential biomarkers

Mallari Praveen

Indira Gandhi National Tribal University

Muhammad Yaseen (✉ muhammadyaseen.my907@gmail.com)

University of Swat

Ricardo Buendia

Ricardo Enrique Buendia-Corona, Universidad de las Américas Puebla

Mian Gul Sayed

University of Swat

Mashooq A Bhat

King Saud University

Noha I Zeiden

University of Chester

Research Article

Keywords: TP53, HSPCB, Nf-kB1, *Potentilla nepalensis*, network pharmacology, computational studies, biomarkers

Posted Date: September 6th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3318745/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Potentilla nepalensis belongs to the Rosaceae family, and have numerous therapeutic applications as potent plant-based medicine. Forty phytoconstituents (PCs) from the root and stem through n-hexane (NR and NS) and methanolic (MR and MS) extracts were identified in our earlier studies. However, the PCs affecting human genes and their roles in the body are not disclosed till now. In this study, we employed network pharmacology, molecular docking, molecular dynamics simulations (MDS), and MMGBSA methodologies. SMILES format of PCs from the PubChem used as input to DIGEP-Pred, 764 identified as the inducing genes. Their enrichment studies have shown inducing genes gene ontology descriptions, involved pathways, associated diseases, and drugs. PPI networks constructed in String DB and network topological analysing parameters done in Cytoscape v3.10 revealed three biomarkers, TP53 from MS, NR, and NS induced genes; HSPCB and Nf-kB1 from MR induced genes. From 40 PCs, two PCs 1b (MR) and 2a (MS), showed better binding scores (*kcal/mol*) with p53 protein of -8.6, and - 8.0; three PCs 3a, (NR) 4a and 4c (NS) with HSP protein of -9.6, -8.7, and - 8.2. MDS and MMGBSA revealed these complexes are stable without higher deviations with better free energy values. Biomarkers identified in this study, have a prominent role in numerous cancers. Thus, further investigations such as in-vivo and in-vitro should be done on considering the PCs of *P.nepalensis*.

Introduction

Traditional medicine has significantly contributed to modern pharmaceuticals by providing valuable leads for the creation of effective drugs to eradicate diseases, indicating the potential of natural products in the realm of drug development (Harvey, 2008). Furthermore, phytoconstituents (PCs) derived from natural sources offer the greater advantage of accessibility and fewer or no side effects when compared to synthetic drugs. The integration of natural medicinal resources into drug discovery holds promise in addressing the eradication of antibiotic resistance (Mathur and Hoskins, 2017).

Potentilla nepalensis, belongs to the Rosaceae family, also known as Nepal cinquefoil that is native to the Himalayan regions of Nepal and Tibet, finding its habitat in alpine and subalpine environments. Despite being admired for its captivating flowers and often cultivated as an ornamental plant, *p.nepalensis* is esteemed for its medicinal value. The plant's therapeutic potential is closely linked to its abundance of secondary metabolite contents, including high amounts of phenols, flavonoids and terpenoid-derived compounds which were crucial for its therapeutic nature (Sharma et al. 2023).

P.nepalensis has been traditionally employed for its medicinal properties in various contexts. It has found use in wound healing and addressing skin-related diseased conditions, in addition to aiding digestion and promoting gastrointestinal well-being. Its potential as an anticarcinogenic agent has also been recognized (Tomczyk et al. 2010), and its antioxidant activity contributes to safeguarding cells from oxidative stress. Moreover, the plant possesses anti-inflammatory and analgesic properties, rendering it effective in alleviating pain and mitigating inflammation (Kumari et al. 2021). Within Tibetan traditional medicine, some species of the *Potentilla* genus, including *P.nepalensis*, have been utilized to treat ailments such as asthma, headache, dysentery, and common cold. Root extracts of *P.nepalensis* have exhibited promising anti-cancer (Tomczyk et al. 2013) as well as anti-microbial activity (Tomczyk et al. 2008).

In our present study, we delve into the exploration of n-hexane and methanolic extracts obtained from both the root and stem parts of the *p.nepalensis* using the Gas Chromatography–Mass Spectrophotometry (GC-MS) method determined in our previous research (Sharma et al. 2023). Our investigation aims to unveil the extracts' impact on

human genes through an integrated approach of network pharmacology, molecular docking, and dynamic simulation methodologies.

This study methodology encompasses several sequential steps. Firstly, we retrieved the pertinent information regarding the phytoconstituents (PCs) from the PubChem database. These PCs were scrutinized to identify respective inducing genes exhibiting significant pharmacological activity greater than 0.8. Subsequently, we generated protein-protein interaction networks for each extract using string DB. Our focus was then directed towards identifying the most important gene within these networks. Moreover, we performed molecular docking using Autodock Vina software. To facilitate this, we retrieved key proteins identified within the networks from the Protein Data Bank (PDB). Then Molecular dynamics simulations (MDS) and Molecular mechanics with generalized Born and surface area solvation (MMGBSA) free energy calculations of the complexes that exhibited better binding affinity values were calculated. The detailed information on the methodology employed in this research work was presented pictorially in Fig. 1.

Materials and methods

Retrieving the Compounds

PubChem IDs, 2D structures and Simplified Molecular Input Line Entry System (SMILES) formats of the PCs found in *P. nepalensis* were retrieved from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) (Kim et al. 2023). These compounds were earlier identified from the root and stem parts using methanolic and n-hexane extracts GC-MS method (Sharma et al. 2023). PubChem is an open-source chemical compounds library with a description.

Inducing genes of the PCs

The SMILES format of the PCs was utilized to identify induced genes, meeting the criteria of having pharmacological activity greater than 0.8 ($Pa > 0.8$). This analysis was performed in the DIGEP-Pred web server (<http://www.way2drug.com/GE>) (Lagunin et al. 2013), which is built based on the Prediction of Activity Spectra for Substances (PASS), calculated by using leave-one-out cross-validation and depended on both mRNA data and protein data.

Enrichment analysis

Enrichment analysis for the human genes induced by the PCs was conducted using Enrichr (<https://maayanlab.cloud/Enrichr/>), by applying a false discovery rate (FDR), and a significant (P) value, both set to be less than 0.05 ($FDR < 0.05$; $P\text{-value} < 0.05$) (Kuleshov et al. 2016). Duplicate genes were removed, resulting in a refined gene set that was used as input to ascertain the respective gene's biological process, molecular function, cellular components, pathways, diseases, and available drugs in the market.

Protein-protein interactions

The interactions among the predicted induced gene proteins were exploited using the STRING DB (<https://string-db.org/>) platform. A high confidence threshold of 0.7 was employed to construct the networks. These networks are formed from various sources, including text mining, experimental data, databases, co-expressions patterns, neighborhood analysis, gene fusion events, and co-occurrence patterns (Szkłarczyk et al. 2023). Subsequently, the resulting network was analysed by sending it into Cytoscape V3.10.0. (Shannon et al. 2003). With the use of the Analyzer plugin, a protein-protein interaction network was established based on the gene set. These networks

underwent further analysis to identify key gene's network topological parameters including Degree centrality, average shortest path length, clustering coefficient, closeness centrality, and betweenness centrality. These parameters offer valuable insights into the interactions and relationships among the genes within the network.

Molecular docking

We employed our previously performed molecular docking methods in defining the binding affinities between the protein-ligand complexes (Praveen and Morales-Bayuelo, 2023). The tertiary structures of the TP53-binding protein (PDB ID: 6MXY) (<https://www.rcsb.org/structure/6MXY>) with K6M ligand (N-[3-(tert-butylamino)propyl]-3-(trifluoromethyl)benzamide), Nuclear factor NF-kappa-BP (PDB ID: 3GUT) (Stroud et al. 2009) and Heat shock protein (PDB ID: 1UYM) with PU3 ligand (9-butyl-8-(3,4,5-trimethoxybenzyl)-9h-purin-6-amine) (Wright et al. 2004), were retrieved from the RCSB PDB (<https://www.rcsb.org/>). All these structures were determined experimentally through the X-ray diffraction method with resolutions of 1.62, 3.59 and 2.45.

The PCs' SMILES format obtained from PubChem and protein structures sourced from the PDB were not suitable for docking. To address this, the ligands were prepared by adding hydrogens at pH 7.4, generating the 3D geometries and the application of the MMFF94 forcefield using the Open Babel software (O'Boyle et al. 2011), a tool that converts the file to various chemical formats. For the proteins, the extra bound ligands and water molecules were removed in the Drug Discovery studio. Subsequently, the protein structures were repaired by adding missing atoms, polar hydrogens, and Gasteiger charges using the Autodock software (Morris et al. 2009).

Grid parameters play a crucial role in defining the precise interacting positions and binding affinity of the ligands. For instance, the bound chemical compounds, such as K6M ligand of TP53 protein (X=-10.80, Y = 26.77, and Z=-3.52), Heat shock protein with PU3 ligand (X = 3.60, Y = 11.13, and Z = 24.75) and the NFkB whole structure (X = 28.80, Y=-23.60, and Z = 58.23), were considered as grid centers with grid sizes considered as grid parameters in the present study to dock through the AutoDock Vina software (Trott and Olson, 2009). This strategic utilization facilitated the accurate protein-ligand interactions and binding affinities in the context of this study.

Molecular dynamics simulations and MMGBSA

Computational molecular simulations for MD were performed on an Ubuntu 22.04 LTS workstation equipped with an Intel core i7-13700k processor and an NVIDIA RTX 4080 graphics card. Amber 22 and AmberTools23 (Wang et al. 2004) packages were used to carry out all simulations and trajectory analyses. The ligands bound to p53 and HSPCB from docking complexes were separated, and any missing hydrogens were added using UCSF ChimeraX (Pettersen et al. 2021). The ligand parameters were assigned using General Amber Force Field (GAFF) (Wang et al. 2004), and these were calculated through Antechamber under the AM1-BCC charge method. The Amber force field FF19SB (Tian et al. 2020) was employed, along with the TIP3P water model. Neutralization was achieved by adding sodium and chlorine ions, and these parameters were assigned to the respective files for ligands, proteins, and solvents. To balance the system, a total of 30000 steps of minimization were performed, followed by an increase in temperature to 300 K and 1 atm pressure equilibrium for 200 ps of simulation. A production simulation of 300 ns was then conducted by pmemd software, accelerated by CUDA. Molecular Mechanics-Generalized Born Surface Area (MMGBSA) (Genheden and Ryde, 2015) analysis of trajectories was employed using module MMPBSA.py for binding energies calculation. Data visualization and figures were carried out by xmgrace (Turner, 2005) and R (Core Team, 2018).

Results

Retrieving the Compounds

The n-hexane and methanolic extracts from the root and stem parts of the *p.nepalensis* using Gas chromatography–mass spectrophotometry (GC-MS) method reported earlier in our earlier study outcomes a total of forty compounds, with ten compounds being identified in each extract from both parts of the plant. All these compounds' information such as their SMILES format, corresponding PubChem IDs, and 2D structures were retrieved from the PubChem database. For convenience, these compounds were classified into four categories, based on the extraction method applied to the plant part. Compounds designated as 1a-1j represent the n-hexane extract from the root. Similarly, compounds labeled as 2a-2j were methanolic extracts from the stem. Compounds 3a-3j were from the n-hexane root extract, while 4a-4j compounds denoted were part of the methanolic shoot extract of *p.nepalensis* in Supplementary Tables 1–4.

Inducing genes of the PCs

A comprehensive assessment of gene induction revealed a total of 764 genes that were induced. The derived root part methanolic (MR) extract led to the induction of 149 genes. Similarly, the shoot methanolic (MS) extract resulted in the induction of 217 genes. Furthermore, the n-hexane extract obtained from the root (NR) triggered the induction of 277 genes. Lastly, the n-hexane extract from the stem (NS) extract induced 121 genes. These findings were comprehensive understandings of gene induction associated with PCs of *p.nepalensis*.

Enrichment analysis

To ensure precision and relevance, the top ten descriptive terms were considered from the entire gene ontology study results. This curation was performed because of the high accuracy associated with these terms in predicting efficiently their gene attributes (Supplementary Excel 1). Among this maximum number of genes involved in enrichment, descriptors were described, providing insights into the genes involved in attribute description.

The biological process (BP) indicates which gene product is involved in biological work (Ashburner et al. 2000). From the extracts, a maximum number of genes' biological process were predominantly around two BPs; nucleic-acid templated transcription positive, and DNA-templated transcription regulation in Table 1; Supplementary Figs. 1A-4A. The MS extract exhibited a pronounced involvement in these two BPs.

Table 1

Enrichment analysis of methanolic and n-hexane extracts from root and stem plant parts of *p.nepalensis*

Descriptors	MR	MS	NR	NS
BP	nucleic acid-templated transcription positive regulation,	DNA-templated transcription regulation, Nucleic acid-templated transcription positive regulation	Nucleic acid-templated transcription positive regulation,	Nucleic acid-templated transcription positive regulation
MF	Protein homodimerization activity,	Protein Serine/Threonine Phosphatase activity	DNA binding, Protein Homodimerization Activity,	DNA-Binding transcription activator activity, Protein Serine/Threonine Phosphatase activity, Oxidoreductase activity
CC	Intracellular membrane-bounded organelle, Nucleus, secretory granule lumen	Nucleus, Azurophil granule lumen	Intracellular membrane-bounded organelle, Nucleus	Intracellular organelle lumen, Endoplasmic reticulum lumen, Secretory granule lumen
Pathways	Vegfa-Vegfr2 signalling pathway, Leptin signalling pathway, Micro-RNAs in cardiomyocyte hypertrophy, B-cell receptor signalling pathway	Nuclear receptors meta-pathway, Osteoblast differentiation, Vitamin-D receptor pathway	Nuclear receptors meta-pathway, Vitamin-D receptor pathway	Common pathways underlying drug addiction, Melanoma, pyrimidine metabolism
Diseases	Neoplasm metastasis, Liver carcinoma, Mammary neoplasms, Melanoma	Neoplasm metastasis, Breast carcinoma, Prostate malignant neoplasm	Neoplasm metastasis, Liver carcinoma	Breast carcinoma, Breast neoplasm malignant, Neoplasm metastasis
Drugs	Aprindine, Domperidone Trifluoperazine, Pitolisant, Cyproheptadine, Pimozide, Brompheniramine Buprenorphine, Lidoflazine, Chlorambucil	Mefenamic Acid, Diclofenac, Flufenamic Acid, Quercetin Mezlocillin,,Hydrochlorothiazide Hydroxycarbamide Bendroflumethiazide Benzthiazide, Chlorambucil	Bezafibrate, Rosiglitazone Stearic Acid, Dodecanoic Acid, Gamolenic Acid, Aprindine, Caffeine, Eicosapentaenoic Acid, Linolenic Acid, Mefenamic Acid	Stearic Acid, Epalrestat Dodecanoic Acid, Gamolenic Acid, Vemurafenib, Bezafibrate, Gemfibrozil, Linolenic Acid, Aprindine, Eicosapentaenoic Acid
BP – Biological process, MF – Molecular function, CC – Cellular components				

Molecular function (MF) refers to gene product activity (Ashburner et al. 2000). A converged MF emerged within the induced genes from the MR and NR extracts. Especially, in protein homodimerization and protein serine/threonine activity were common denominators, effectively linked between these extracts. Expanding on this, MS and NS showed similar attributes in their induced genes. In addition, MF has enriched with MR-NR and MS-NS extract-induced genes. DNA binding in NR, DNA binding transcription activator activity, phosphatase activity and oxidoreductase activity in NS-induced genes were additional MF results in Table 1; Supplementary Figs. 1B-4B. This interconnectedness suggests the potential areas of research focus within the study.

Cellular components (CC) represent the location of the gene in the cell (Ashburner et al. 2000). The commonality shared between the MR and NR extract-induced genes is reflected in their location within the intracellular membrane-bounded organelle and nucleus. Moreover, the MR extract demonstrates an additional secretory granule lumen. Moving to the MS extract-induced genes are mostly found in the nucleus and azurophil granule lumen. As for NS extract, the induced genes exhibit a more diverse localization in the intracellular organelle, endoplasmic reticulum, and secretory granule lumens in Table 1; Supplementary Figs. 1C-4C.

The genes targeted by the MR extract are involved in several significant pathways, such as Vegfa-Vegfr 2 signaling, leptin signaling, micro-RNAs in cardiomyocyte hypertrophy, and B-cell receptor signaling pathways. Nuclear receptors meta-pathway and Vitamin-D receptor pathways were common in the MS and NR targeted genes, and additionally the Osteoblast differentiation pathway from the MS extract-induced genes. Distinctly, the NS extract-induced genes contribute to the drug addiction pathways and melanoma pyrimidine metabolism pathway in Table 1; Supplementary Figs. 1D-4D.

Upon scrutinizing the diseases associated with the genes induced by all extracts, a notable involvement in cancer-related diseases, especially in the context of neoplasm metastasis. This underscores the significance of these genes in the context of cancer progression and dissemination. Effectively summarized in Table 1; Supplementary Figs. 1E-4E.

Protein-protein interactions

Protein-protein interaction networks were generated by interlinking the induced genes using String DB resulting in the network various containing called nodes (proteins) and edges called (interactions). The interactions have diverse sources from the experimental, text mining, gene fusion, co-expression, neighborhood, and databases. To achieve better insight into the constructed PPI networks, network topological parameters were employed, which include Degree centrality, average shortest path length, clustering coefficient, closeness centrality, and betweenness centrality on each gene's interaction proteins in the network in Table 2; Supplementary Excel 2.

In evaluating the four PPI network topological parameters descriptors, Degree centrality (DC) represents the number of interactions within the network. The average shortest path gauges the minimum number of edges required to connect between the nodes. The clustering coefficient (CC) measures the interconnections required to form a triangular sub-cluster. The shortest path numbers in a network can be measured between centrality (BC) values, which act as a transitional in switching the information. Closeness centrality (C. cen) signifies the closest node to the rest of the nodes in the network. Accounting for these descriptors, five actively interacting genes emerge as significant. These include HSPCB and NFKB1 from MR extract-induced genes in Fig. 2. TP53 is a central figure among MS, NR, and NS extracts-induced genes, illustrated in Figs. 3–5. These pivotal proteins are demarcated with centrally positioned rectangular boxes, highlighted with a yellow background and blue fonts were the proteins that played prominent roles in the network. Furthermore, Table 2 indicates the network topological parameter values of

these key proteins, further reaffirming their status as potent biomarkers sourced from the *p.nepalensis* PCs. This evaluation paves into the critical genes and their connectivity, enhancing our understanding of the molecular level within the network.

Table 2
Network topological parameters of the complete Protein-protein interaction network

Induced genes used	Proteins	Network topological parameters				
		DC	Avg short. path leng.	CC	C. cen	BC
MR	HSPCB	15	3.02	0.18	0.33	0.18
	NFKB1	15	2.85	0.25	0.34	0.33
MS	TP53	25	1.87	0.15	0.53	0.56
NR	TP53	20	2.32	0.16	0.42	0.58
NS	TP53	25	1.75	0.17	0.57	0.59
DC – Degree Centrality; Avg short. path leng. – Average shortest path length; CC - Clustering coefficient; C. cen - Closeness centrality; BC - Betweenness centrality						

Molecular docking

To explore the interactions between the PCs and potential biomarker proteins, molecular docking analyses were conducted on the methanolic and n-hexane extracts from the root and stem parts of *p.nepalensis*. The objective was to determine the binding affinity of the PCs concerning the proteins, namely p53 from the TP53 gene, heat shock protein from the HSPCB gene, and nuclear factor kappa light chain from the NFKB1 gene. In total 40 PCs were subjected, resulting in docking scores that indicate the strength of binding interactions in Supplementary Table 5. Among the array of PCs, five of them – 1b, 2a, 3a, 4a, and 4c stood out for resulting higher binding affinity with both p53 and heat shock proteins.

Illustrations, Fig. 6 and Table 3 indicate the binding affinity, and amino acid involved in interaction types. The binding affinity of 1b with the p53 protein was – 8.6 kcal/mol. This interaction involved a single conventional hydrogen bond with the Aser1503 residue. Furthermore, two alkyl bonds formed, connecting the alkyl ends of 1a with the alkyl groups of BMet1584 in the p53 protein. Seventeen π -alkyl bonds emerged, linking the π -alkyl groups of ⁴ATrp1495, ATyr1502, ²APhe1519, ATyr1523, ³BTrp1495, ³BTyr1502, ²BPhe1519, BTyr1523 with the π -orbitals of 1b. PC 2a demonstrated – 8.0 kcal/mol of binding affinity with the p53 protein. AMet1584 participated in an alkyl bond formation with alkyl ends. Additionally, a set of twelve π -alkyl bonds connecting the π -alkyl groups of ²ATrp1495, ²ATyr1502, ²APhe1519, ²BTrp1495, BTyr1502, ²BPhe1519, BTyr1523 with the π -orbitals of 2a.

PCs 3a, 4a, and 4c exhibited impressive docking scores of -9.6, -8.7, and – 8.2 kcal/mol with HSP protein. 3a engaged in nine interactions with the HSP residues. AMet98 is involved in three different bond types, including a sulfur bond, a π -Sigma, and an alkyl bond with 3a. Three additional alkyl bonds were formed between the alkyl ends of HSP residues (²ALeu107, AAla111) with 3a. The π -orbitals between APhe138 with 3a are involved to form a π - π Stacked bond in enchaining their stability. Two π -alkyl bonds merged between the π -alkyl groups of APhe138 and AVal150 and the π -orbitals of 3a. For 4a, three π -alkyl bonds and a π - π Stacked bond were formed with the APhe138 residue. Additionally, two π -alkyl and two π - π Shaped bonds were made by 4a with ATrp162. One π -alkyl and two alkyl bonds with AMet98, followed by an alkyl bond with AVal186. Another π -alkyl and one alkyl bond connected 4a and Aleu107.

Lastly, 4c shows two CHB interactions with the ATrp162 residue. Six π -alkyl bonds emerged, connecting the π -alkyl groups of APhe22, APhe170, ²ALeu107, AMet98, and AVal150 with the π -orbitals of 4c. Furthermore, an alkyl between the alkyl groups of 4c and Alle26 formed. Two π - π Stacked (²APhe138) and three π - π T Shaped (ATyr139, ²ATrp162) by the 4c. These results provide binding modes and interactions of PCs 3a, 4a, and 4a with the HSP protein, contributing potential roles in therapeutic applications.

Table 3

Binding affinity (B.A.) and interactions of the Heat shock protein (HSP), Nuclear factor NF-kappa-BP (NFKB1), and TP53-binding protein (p53) with the extracted PCs of *p.nepalensis*

Complex		B.A. (<i>kcal</i> <i>/mol</i>)	Hydrogen Bonds	Hydrophobic bonds					Other bonds
Prot- eins	PCs		CHB	π -Alkyl	Alkyl	π - π Stacked	π - π T Shaped	π - Sigma	
p53	1b	-8.6	ASer1503	⁴ ATrp1495, ATyr1502, ² APhe1519, ATyr1523, ³ BTrp1495, ³ BTyr1502, ² BPhe1519, BTyr1523	² BMet1584	-	-	-	-
	2a	-8.0	-	² ATrp1495, ² ATyr1502, ² APhe1519, ² BTrp1495, BTyr1502, ² BPhe1519, BTyr1523	AMet1584	-	-	-	-
HSP	3a	-9.6	-	APhe138, Aval150	AMet98, ² ALeu107, AAla111	APhe138	-	AMet98	AMet98
	4a	-8.7	-	³ APhe138, ² ATrp162, AMet98, ALeu107	² AVal186, ² AMet98, AVal150, ALeu107	APhe138	-	² Trp162	-
	4c	-8.2	² ATrp162	APhe22, APhe170, ² ALeu107, AMet98, AVal150	Alle26	² APhe138	ATyr139, ² ATrp162	-	-
CHB – conventional hydrogen bond									

Molecular dynamics simulations and MMGBSA

To study the dynamic nature and protein-ligand stability of the complex in the water molecular dynamics simulations were employed. It provides insights into interaction informatic at an atomic level. The current study, from the results of molecular docking best pose docking score complexes were taken for MD simulations and MMGBSA calculations. A total of five complexes were employed for MDS at 300ns, they are i. p53 with 1b, ii. p53 with 2a, iii. HSPCS with 3a, iv. HSPCB with 4a, and v. HSPCB with 4c. These complexes were pre-processed into three stages along with energy minimization, NPT, and NVT equilibrium. Root-Mean square deviations (RMSD), Root-Mean square fluctuations (RSMF), and MMGBSA values were analyzed by the trajectories.

The Computed trajectories RMSDs of the complexes are represented in Fig. 7. For p53 protein bound with 1b (p53 + 2b) and 2a (p53 + 2a) in Fig. 7A, initially to the complex p53 + 2a, deviations were higher compared to the p53 + 1b. p53 + 1b complex maintained constant deviation not exceeding greater than 5Å. Both complexes reached equilibrium at 6000 frames (~ 130 ns) at 2–3 Å. Both ligands exhibit stable complexes and did not leave the binding site during the whole simulation.

For the complexes, HSPCB protein with 3a (HSPCB + 3a), with 4a (HSPCB + 4a), and with 4c (HSPCB + 4c) in Fig. 7B. Their MDS analysis represents the convergence achieved at the end of 100ns. At the beginning of the simulation, the deviations were increasing in three complexes but not exceeding greater than 3.5 Å in HSPCB + 4c. All the three complexes protein-ligand (s) simulations were stable between 2–3 Å. This indicates the stability of the complexes and without leaving of ligands are exhibited during simulation from their binding regions.

RMSF by bfactor was employed to determine residue displacement in Fig. 8. The RMSF of complexes, p53 + 1b, and p53 + 2a exhibited similar RMSF without major fluctuations in Fig. 8A. Suggesting that the binding site was not affected differently by any of both ligands. The three complexes, HSPCB + 3a, HSPCB + 4a, and HSPCB + 4c. Their RMSF in Fig. 8B, shows that in the HSPCB + 4c, the residue 97 suffered considerable fluctuation, nonetheless not involved in the binding site cavity.

Free energy calculations like MMGBSA runs performed by cpptraj module are employed in analyzing the trajectories of the five complexes. p53 + 1b and p53 + 2a, frame 6000 was declared as starting with 5 frames intervals. For HSPBC + 3a, HSPBC + 4a, and HSPBC + 4c, frame 4000 was also started with 5 frames intervals. Frames taken for analysis relied on autocorrelation for every complex. Energy fluctuation calculated by MMGBSA demonstrated the difference between p53 + 1b and p53 + 2a complexes since p53 + 1b revealed better binding energies. Nonetheless, at the end of the simulation, these energies began to fall somehow equal to the p53 + 2a ligand in Fig. 9A. Both ligands are considered sufficiently thermodynamically favorable to the target receptor, the t-test for both complexes demonstrates a significant difference between energies in Supplementary Table 6. MMGBSA calculations demonstrated the ability to target HSPCB protein while maintaining thermodynamically favorable energies during the whole simulation. One-way ANOVA analysis demonstrates that ligand 3a is significantly different from 4a and 4c in Fig. 9B, nevertheless, 4a and 4c are not different from each other in Supplementary Table 7. The summary of the MMGBSA free energies results of the complexes were presented as boxer-whisper plots in Fig. 10.

Discussion

Considering the root and stem PCs by the n-hexane (NR and NS) and methanolic (MR and MS) extracts through Gas chromatography mass spectrometry (GC-MS) on *P. nepalensis*, the present study has proceeded to identify the inducing genes/protein as biomarkers by enrichment analysis, protein-protein interaction networks, and molecular docking followed by MDS and MMGBSA studies.

Of the total 764 inducing genes in TextS1-S4, prediction from the DIGEP-Pred at $P_a > 0.8$. Enrichment analysis was performed to find the functional annotations of the genes identified by descriptors of gene ontology, pathways, diseases, and drugs. Subjecting each extract to StringDB to build PPI networks.

Through the protein-protein interactions (PPI) network the gene-encoded proteins and their functions and interactions can be defined. In the network of PPI, proteins are represented by nodes and connecting lines referred to as edges. The current study intended to comprehend the interplay among the genes. The network topological analysis of the more interacting genes among the network in Table 2). DC of HSPCB and NFKB1 is 15 of MR extract-induced genes, lesser than the TP53 from the rest of the extracted-induced genes. Three proteins have average shortest path values in the range of 1.75 to 3.02 means the connected nodes are interconnected nearer compared to the other genes. The CC of TP53 is relatively lesser than HSPCB and NFKB1. The C. cen values of TP53 (0.53, 0.42 and 0.57) are little far with GSPCB and NFKB1 (0.33 and 0.34). TP53 has a greater BC value of 0.56, 0.58, and 0.59, compared to HSPCB and NFKB1 of 0.18 and 0.33. Targeting these three genes will have a greater significance in the network. Hence, three genes have been identified as potential biomarkers for the PCs of *P. nepalensis*.

HSPCB belongs to the heat shock protein 90 family, considered a pseudogene like a heat shock 90kD protein 1 beta, that has a prominent role in signaling, gastric apoptosis, protein folding, and inflammation. Previous studies have reported its involvement in various cancer cell lines as potential targets such as breast cancer (Liu et al. 2015) and Ovarian cancer tissues (Fu et al. 2010).

Nuclear factor kappa B subunit 1 (Nf-kB1) known as the transcription regulator which activates in the nucleus by translocation into it by stimuli substances (cytokines, and oxidant free radicals) followed by the transcription (Cartwright et al. 2016). Problems in Nf-kB1 activation have been linked to numerous inflammatory diseases. While continuous inhibitory affects immune cell development or delayed cell growth (Li et al. 2021). Widely used as a biomarker in diabetic cardiomyopathy (Guo et al. 2022), and cancer (Gaptulbarova et al. 2020). Thus, inhibition of the Nf-kB1 helps to reduce the anti-inflammatory activity of targeting Nf-kB1 in the associated pathways (Weinmann et al. 2018). Designing the antagonist to the Nf-kB1 changes the central gene expression in the leukemic process (Reikvam, 2020).

TP53 gene on transcription gives p53 which is a tumor suppressor protein that is involved in the regulation of cell division in an uncontrolled way (Damineni et al. 2014). p53 protein is widely known for its action in cancer regulation, well renowned as a target for various cancer types because of its involvement in earlier causes of cancer (Hassin and Oren, 2023). At present few drugs like piperdinone analogs, spirooxindole, nutlin, and isoquinoline are actively using inhibitors against the p53-Mdm2 complex (Nayak et al. 2018). Wide research studies are being conducted in designing the small molecules (ligands) targeting mutated p53 protein to restore tumor-suppressing activity (Zhao et al. 2017).

Considering the above-mentioned research study statement and our study purpose correlated in PCs of *P. nepalensis* found biomarkers can effectively exhibit their activity upon binding. Thus, finding the interaction between the PCs and identified biomarker proteins subjected to the docking studies resulted in PCs 1b and 2a, exhibiting better binding scores with p53 protein and PCs 3a, 4a, and 4c with HSPCB in Table 3. Most of the interactions made by the PCs with the p53 and HSPCB are hydrophobic interactions. Thus, designing these PCs with suitable chemical functional groups that could make hydrogen bonds might increase the binding stronger. MDS of these complexes, inferring RMSD in Fig. 7 and RSMF in Fig. 8 have better stability between them without higher deviations that are between 2–3 Å. MMGBSA calculations in Fig. 9, demonstrated the ability to target HSPCB protein while maintaining thermodynamically favorable energies during the whole simulation.

Conclusion

P. nepalensis renowned for its therapeutic applications, with its composition of PCs identified in our earlier studies. In this study, integrative computational methodologies were employed, leading to the identification of 764 inducing genes. Through the PPI networks analysis, three biomarkers were revealed: TP53 from MS, NR, and NS-induced genes; HSPCB and Nf-kB1 from MR-induced genes. Among the 40 PCs studied, two PCs 1b (MR) and 2a (MS), exhibited superior binding scores with the p53 protein demonstrating values of -8.6, and - 8.0 respectively. Additionally, three PCs, 3a (NR) 4a, and 4c (NS) notable binding scores with HSP protein of -9.6, -8.7, and - 8.2. MDS and MMGBSA revealed these complexes remain stable without significant deviations, featuring favorable free energy values. The biomarkers identified in this study play significant roles in numerous cancer types.

Declarations

Acknowledgments

We are thankful to the Research supporting project number (RSPD2023R740), King Saud University, Riyadh, Saudi Arabia.

Author Contributions

Conceptualization, M.P. and M.Y.; methodology, M.P.; software, M.P. and R.D.; validation, M.P., R.R. and Z.Z.; formal analysis, M.P. and R.D.; investigation, M.P.; data curation, M.P. and R.D.; writing—original draft preparation, M.P. and R.D.; writing—review and editing, M.Y. and M.G.S.; visualization, M.P. and R.D.; supervision, M.Y.; project administration, M.Y., M.A.B, M.G.S.; N.I.Z.; funding acquisition, M.A.B All authors have read and agreed to the published version of the manuscript

Funding

Research supporting project number (RSPD2023R740), King Saud University, Riyadh, Saudi Arabia.

Data Availability

All the data used in this study are included in the manuscript and freely available in the online web servers.

Declarations

Conflict of interest - All the authors declare no conflict of interest.

Ethical approval - Not applicable

References

1. Ashburner M et al (2000) Gene Ontology: Tool for the unification of biology. Nat Genet 25(1):25–29. <https://doi.org/10.1038/75556>
2. Augustynowicz D et al (2023) Phytochemical Profiling of Extracts from Rare Potentilla Species and Evaluation of Their Anticancer Potential. Int J Mol Sci 24(5):4836. <https://doi.org/10.3390/ijms24054836>
3. Cartwright T, Perkins ND, Wilson C (2016) NFKB1: A suppressor of inflammation, ageing and cancer. FEBS J 283(10):1812–1822. <https://doi.org/10.1111/febs.13627>

4. Damineni S et al (2014) Germline mutations of TP53 gene in breast cancer. *Tumor Biology* 35(9):9219–9227. <https://doi.org/10.1007/s13277-014-2176-6>
5. Fu J, Bian L, Zhao L, Dong Z, Gao X, Luan H, Sun Y, Song H (2010) Identification of genes for normalization of quantitative real-time PCR data in ovarian tissues. *Acta Biochim Biophys Sin* 42(8):568–574. <https://doi.org/10.1093/abbs/gmq062>
6. Gaptulbarova KA, Tsyganov MM, Pevzner AM, Ibragimova MK, Litviakov, and N.V (2020) NF-kB as a potential prognostic marker and a candidate for targeted therapy of cancer. *Exp Oncol* 42:263–269. <https://doi.org/10.32471/exp-oncology.2312-8852.vol-42-no-4.15414>
7. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 10(5):449–461. <https://doi.org/10.1517/17460441.2015.1032936>
8. Guo Q, Zhu Q, Zhang T, Qu Q, Cheang I, Liao S, Chen M, Zhu X, Shi M, Li X (2022) Integrated bioinformatic analysis reveals immune molecular markers and potential drugs for diabetic cardiomyopathy. *Front Endocrinol* 13:933635. <https://doi.org/10.3389/fendo.2022.933635>
9. Harvey AL (2008) Natural products in drug discovery. *Drug Discovery Today* 13(19–20):894–901. <https://doi.org/10.1016/j.drudis.2008.07.004>
10. Hassin O, Oren M (2023) Drugging p53 in cancer: One protein, many targets. *Nat Rev Drug Discovery* 22(2):127–144. <https://doi.org/10.1038/s41573-022-00571-8>
11. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE (2023) PubChem 2023 update. *Nucleic Acids Res* 51(D1):D1373–D1380. <https://doi.org/10.1093/nar/gkac956>
12. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A (2016) Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 44(W1) W90-97. <https://doi.org/10.1093/nar/gkw377>
13. Kumari S, Seth A, Sharma S, Attri C (2021) A holistic overview of different species of *Potentilla* a medicinally important plant along with their pharmaceutical significance: A review. *J Herb Med* 29:100460. <https://doi.org/10.1016/j.hermed.2021.100460>
14. Lagunin A, Ivanov S, Rudik A, Filimonov D, Poroikov V (2013) DIGEP-Pred: Web service for in silico prediction of drug-induced gene expression profiles based on structural formula. *Bioinformatics* 29(16):2062–2063. <https://doi.org/10.1093/bioinformatics/btt322>
15. Li J, Lei W-T, Zhang P, Rapaport F, Seeleuthner Y, Lyu B, Asano T, Rosain J, Hammadi B, Zhang Y, Pelham SJ, Spaan AN, Migaud M, Hum D, Bigio B, Chrabieh M, Béziat V, Bustamante J, Zhang S-Y, ..., Boisson B (2021) Biochemically deleterious human NFKB1 variants underlie an autosomal dominant form of common variable immunodeficiency. *J Exp Med* 218(11):e20210566. <https://doi.org/10.1084/jem.20210566>
16. Liu LL, Zhao H, Ma TF, Ge F, Chen CS, Zhang YP (2015) Identification of valid reference genes for the normalization of RT-qPCR expression studies in human breast cancer cell lines treated with and without transient transfection. *PLoS ONE* 10(1):e0117058. <https://doi.org/10.1371/journal.pone.0117058>
17. Mathur S, Hoskins C (2017) Drug development: Lessons from nature. *Biomedical Rep* 6(6):612–614. <https://doi.org/10.3892/br.2017.909>
18. Morris GM et al (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 30(16):2785–2791. <https://doi.org/10.1002/jcc.21256>

19. Nayak SK et al (2018) P53-Mdm2 Interaction Inhibitors as Novel Nongenotoxic Anticancer Agents. *Curr Cancer Drug Targets* 18(8):749–772. <https://doi.org/10.2174/1568009617666170623111953>
20. O'Boyle NM et al (2011) Open Babel: An open chemical toolbox. *J Cheminform* 3(1):33. <https://doi.org/10.1186/1758-2946-3-33>
21. Pettersen EF et al (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science: A Publication of the Protein Society* 30(1):70–82. <https://doi.org/10.1002/pro.3943>
22. Praveen M, Morales-Bayuelo A (2023) Drug Designing against VP4, VP7 and NSP4 of Rotavirus Proteins – Insilico studies. *Moroccan Journal of Chemistry*, Vol. 11, Mor. J. Chem. 11 (3): 729–741 Pages. <https://doi.org/10.48317/IMIST.PRSM/MORJCHEM-V11I3.40088>
23. Reikvam H (2020) Inhibition of NF- κ B Signaling Alters Acute Myelogenous Leukemia Cell Transcriptomics. *Cells* 9(7):1677. <https://doi.org/10.3390/cells9071677>
24. Shannon P et al (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* 13(11):2498–2504. <https://doi.org/10.1101/gr.1239303>
25. Sharma S et al (2023) Phytochemical Analysis, In Vitro Biological Activities, and Computer-Aided Analysis of *Potentilla nepalensis* Hook Compounds as Potential Melanoma Inhibitors Based on Molecular Docking, MD Simulations, and ADMET. *Molecules* 28(13):5108. <https://doi.org/10.3390/molecules28135108>
26. Sharma S et al (2023) Phytochemical Analysis, In Vitro Biological Activities, and Computer-Aided Analysis of *Potentilla nepalensis* Hook Compounds as Potential Melanoma Inhibitors Based on Molecular Docking, MD Simulations, and ADMET. *Molecules* 28(13):5108. <https://doi.org/10.3390/molecules28135108>
27. Stroud JC et al (2009) Structural Basis of HIV-1 Activation by NF- κ B—A Higher-Order Complex of p50:RelA Bound to the HIV-1 LTR. *J Mol Biol* 393(1):98–112. <https://doi.org/10.1016/j.jmb.2009.08.023>
28. Szklarczyk D et al (2023) The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 51(D1):D638–D646. <https://doi.org/10.1093/nar/gkac1000>
29. Tian C et al (2020) ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* 16(1):528–552. <https://doi.org/10.1021/acs.jctc.9b00591>
30. Tomczyk M, Leszczyńska K, Jakoniuk P (2008) Antimicrobial activity of *Potentilla* species. *Fitoterapia* 79(7–8):592–594. <https://doi.org/10.1016/j.fitote.2008.06.006>
31. Tomczyk M et al (2013) The influence of aqueous extracts of selected *Potentilla* species on normal human colon cells. *Acta Pol Pharm* 70(3):523–531
32. Tomczyk M, Pleszczyńska M, Wiater A (2010) Variation in total polyphenolics contents of aerial parts of *Potentilla* species and their anticariogenic activity. *Molecules* 15(7):4639–4651. <https://doi.org/10.3390/molecules15074639>
33. Trott O, Olson AJ (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* NA-NA. <https://doi.org/10.1002/jcc.21334>
34. Turner PJ (2005) XMGRACE. Center for Coastal and Land-Margin Research. Oregon Graduate Institute of Science and Technology, Beaverton
35. R Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
36. Wang J et al (2004) Development and testing of a general amber force field. *J Comput Chem* 25(9):1157–1174. <https://doi.org/10.1002/jcc.20035>

37. Weinmann D et al (2018) Brazilin blocks catabolic processes in human osteoarthritic chondrocytes via inhibition of NFKB1/p50: BRAZILIN BLOCKS CATABOLIC PROCESSES VIA INHIBITION OF NFKB1/p50. *J Orthop Research®* 36(9):2431–2438. <https://doi.org/10.1002/jor.24013>
38. Wright L et al (2004) Structure-Activity Relationships in Purine-Based Inhibitor Binding to HSP90 Isoforms. *Chem Biology* 11(6):775–785. <https://doi.org/10.1016/j.chembiol.2004.03.033>
39. Zhao D et al (2017) Molecularly targeted therapies for p53-mutant cancers. *Cell Mol Life Sci* 74(22):4171–4187. <https://doi.org/10.1007/s00018-017-2575-0>
40. Zhou Y, Xiao D, Jiang X (2022) LncRNA RP3-525N10.2-NFKB1-PROS1 triplet-mediated low PROS1 expression is an onco-immunological biomarker in low-grade gliomas: A pan-cancer analysis with experimental verification. *J Translational Med* 20(1):335. <https://doi.org/10.1186/s12967-022-03536-y>

Figures

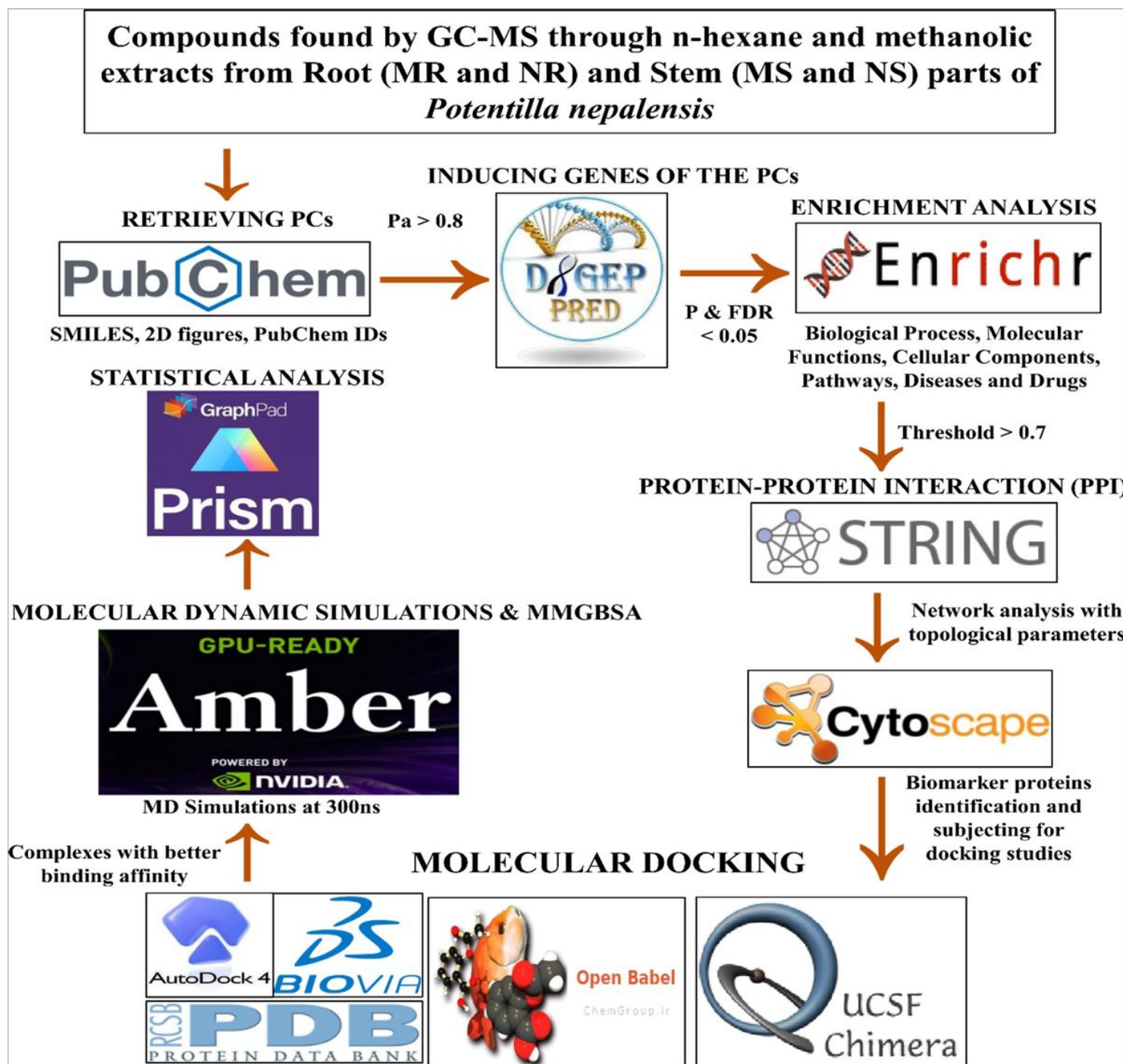


Figure 1

Flow chart representing the sequential step methodologies followed in current study

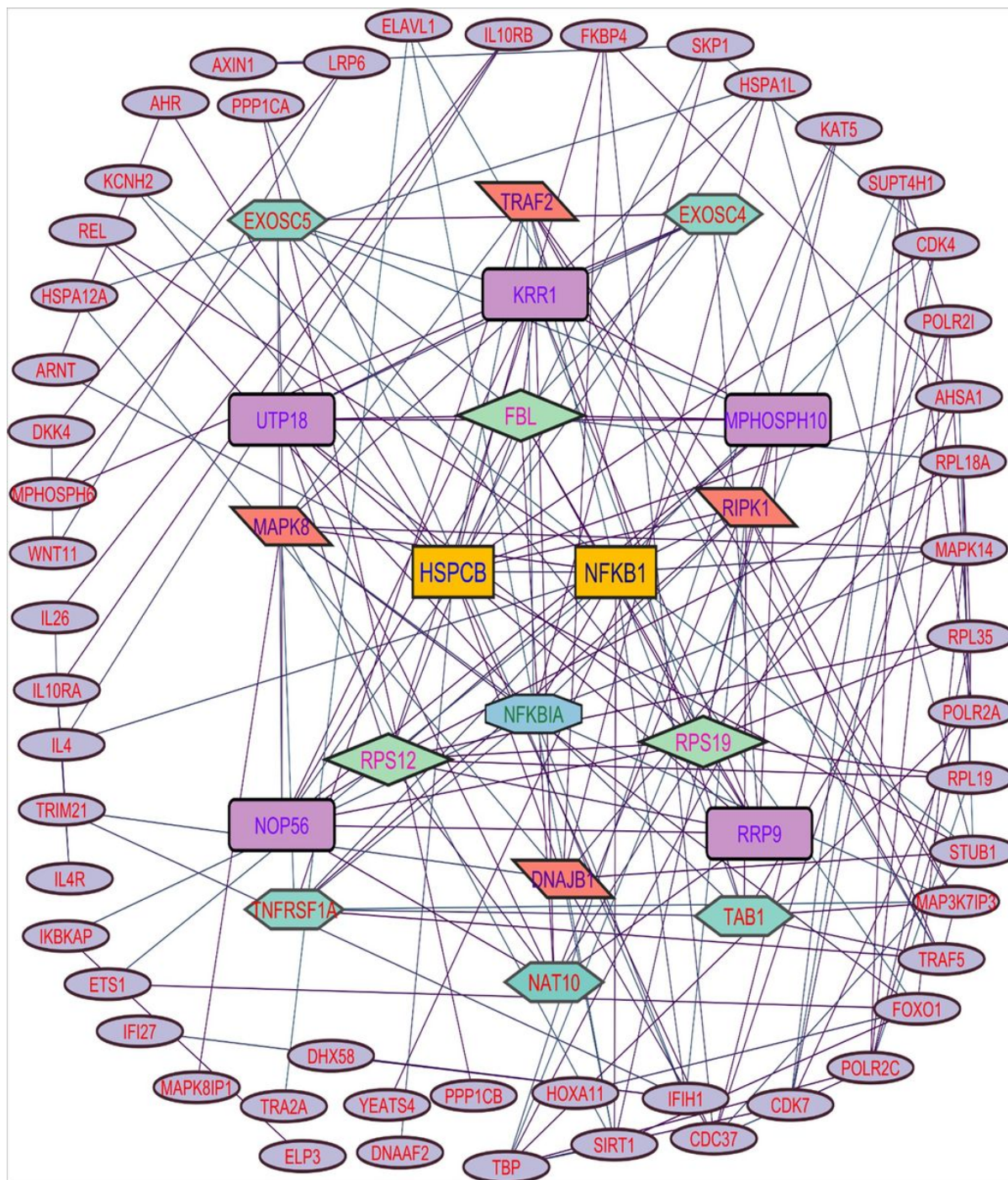


Figure 2

MR extract induced genes protein-protein interaction network

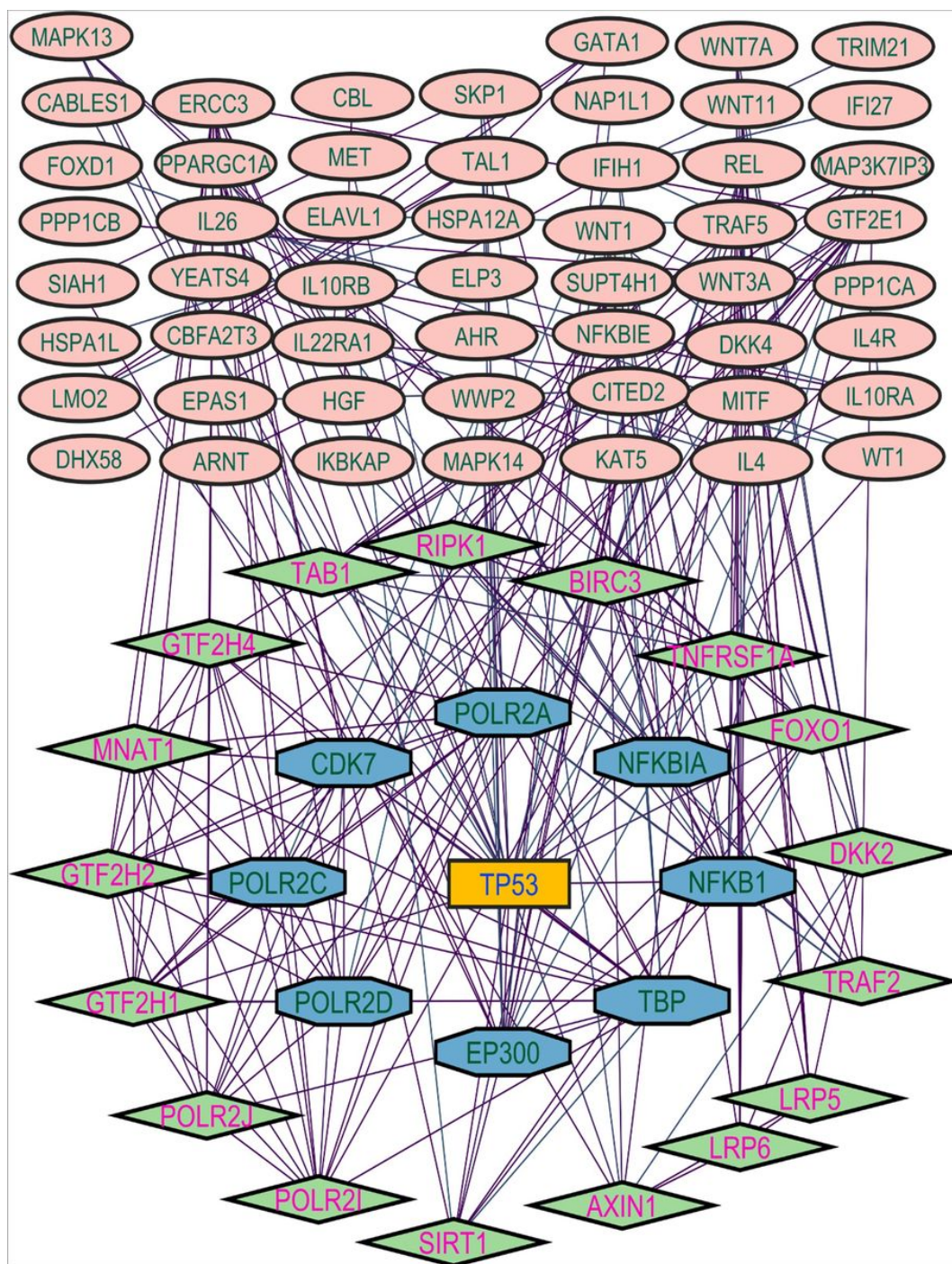


Figure 3

MS extract induced genes protein-protein interaction network

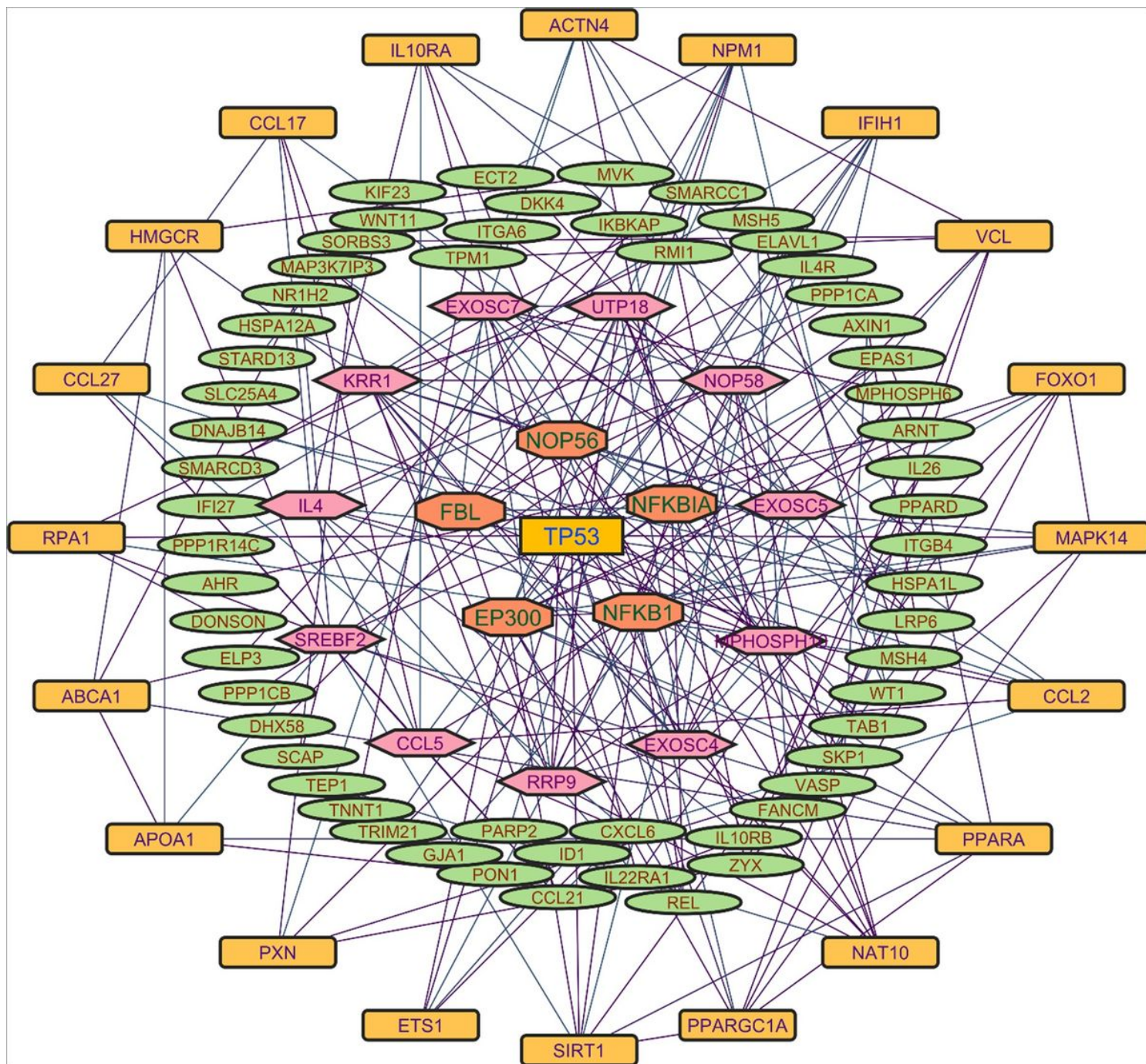


Figure 4

NR extract induced genes protein-protein interaction network

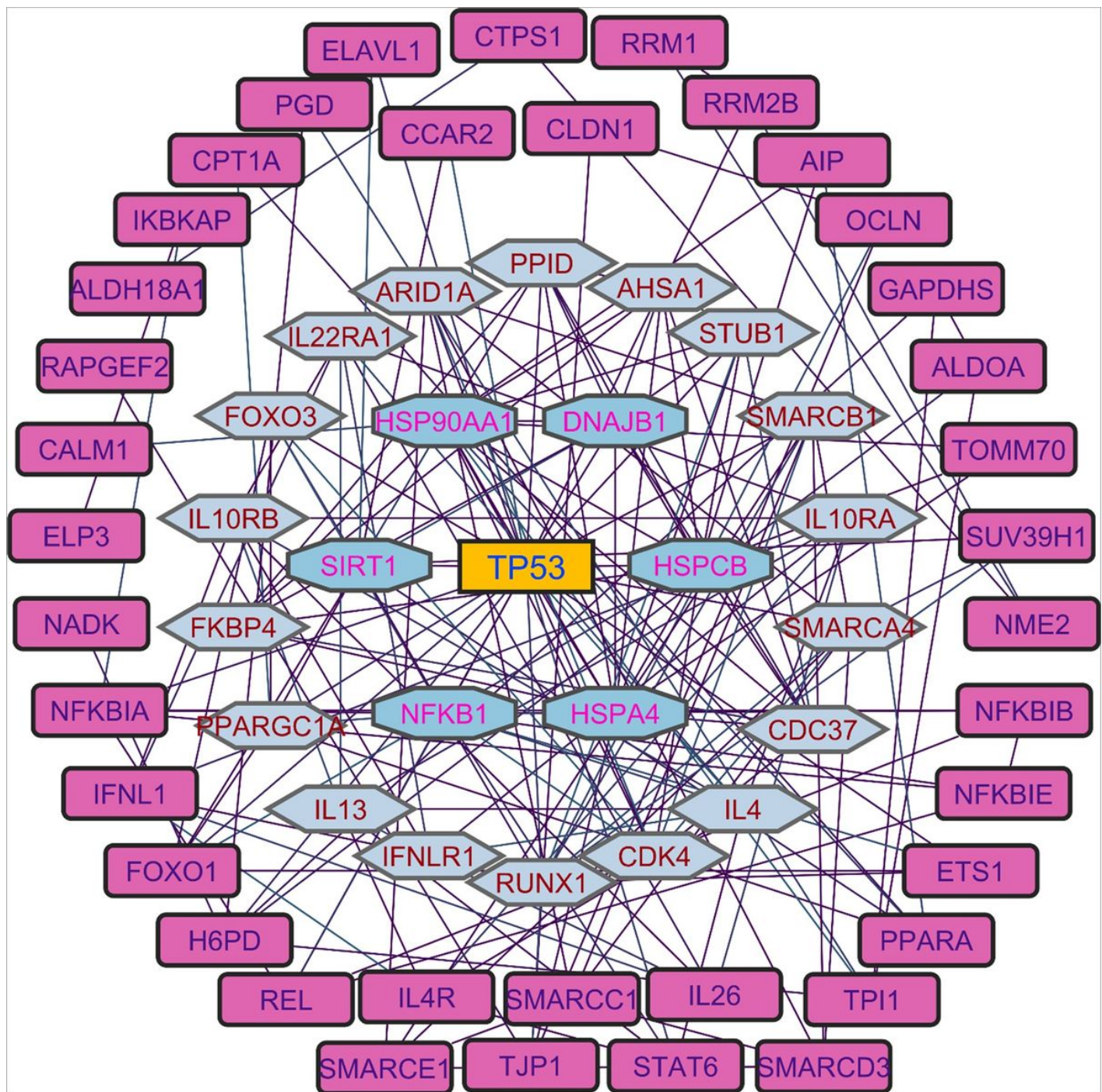


Figure 5

NR extract induced genes protein-protein interaction network

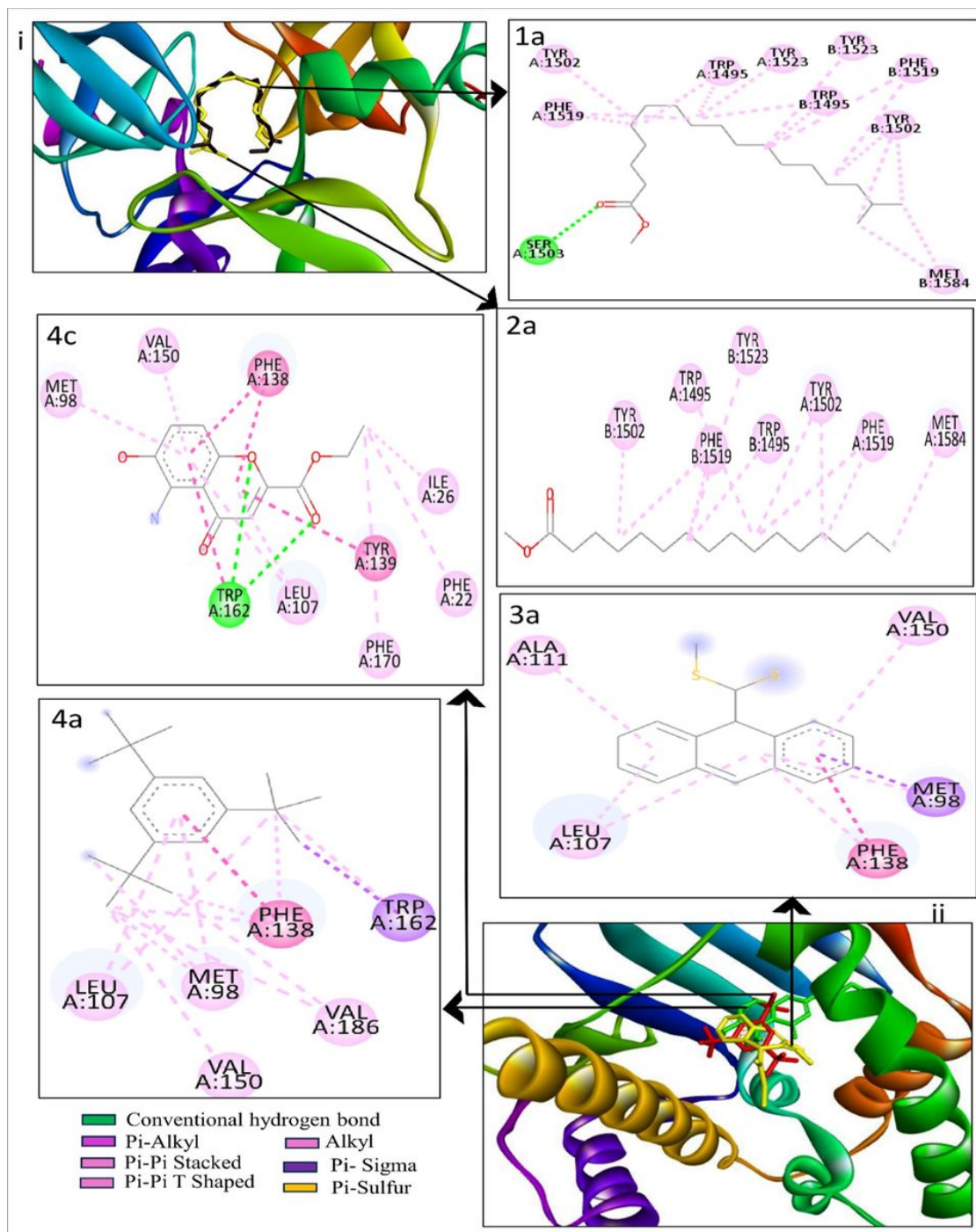


Figure 6

Interactions between i. p53 with 1a (black) and 2a (yellow) PCs and ii. HSP with 3a (yellow), 4a (green), and 4c (red) PCs.

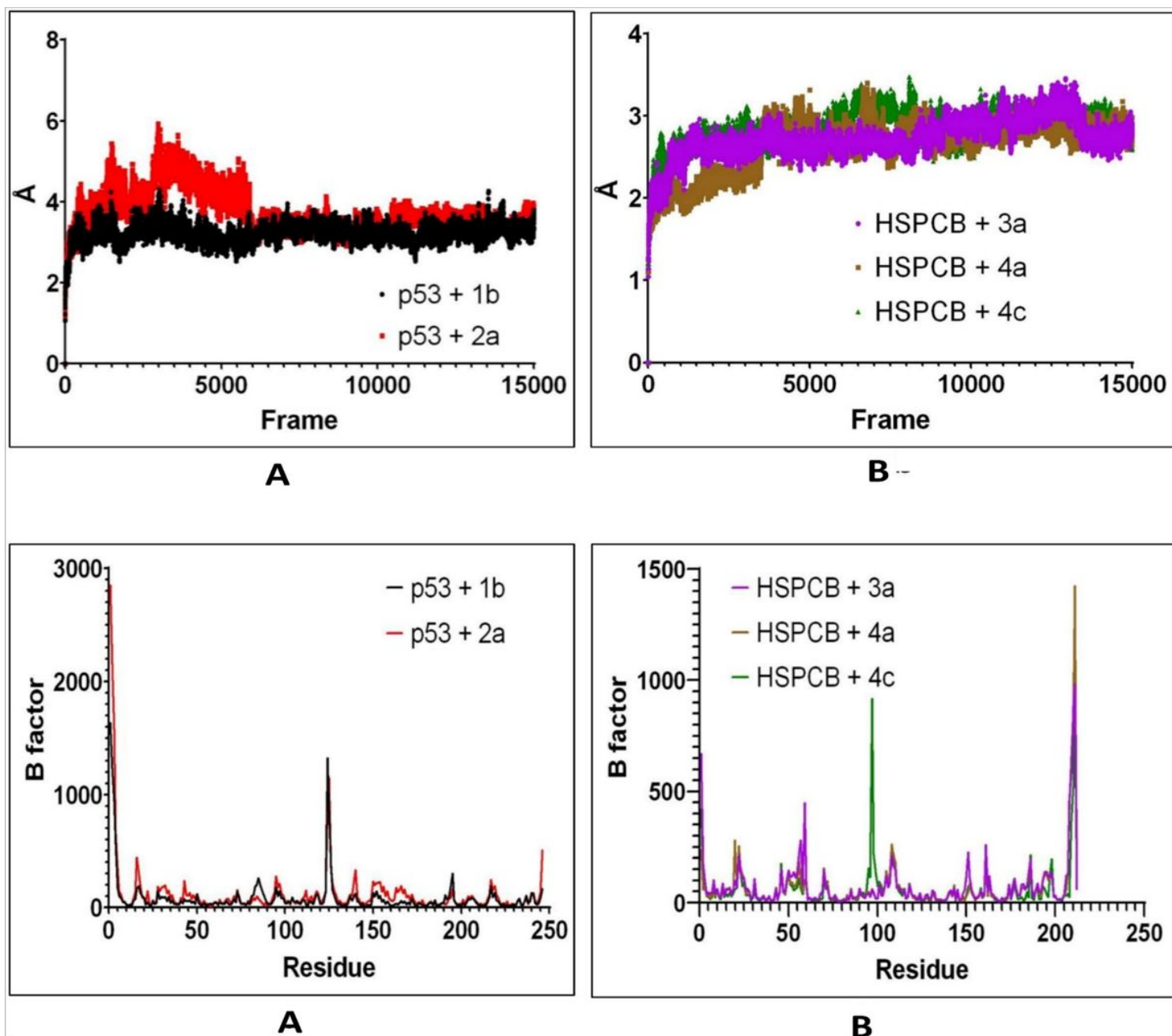


Figure 7

RMSD values from simulated p53 (A) and HSPCB (B) complexes throughout 300 ns production runs divided in 15000 frames.

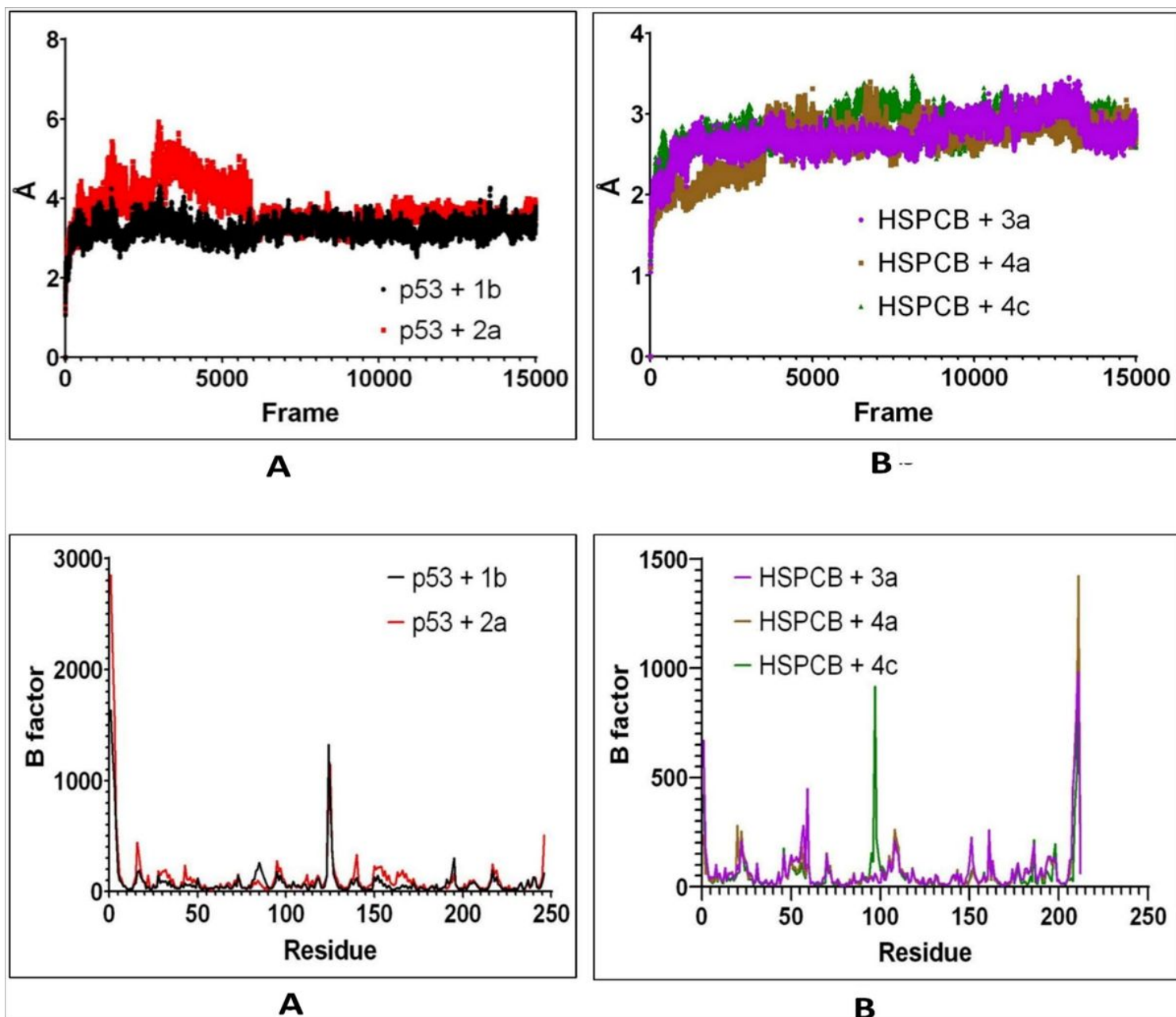


Figure 8

RMSF values by bfactor from every residue in p53 (A) and HSPCB (A) complexes

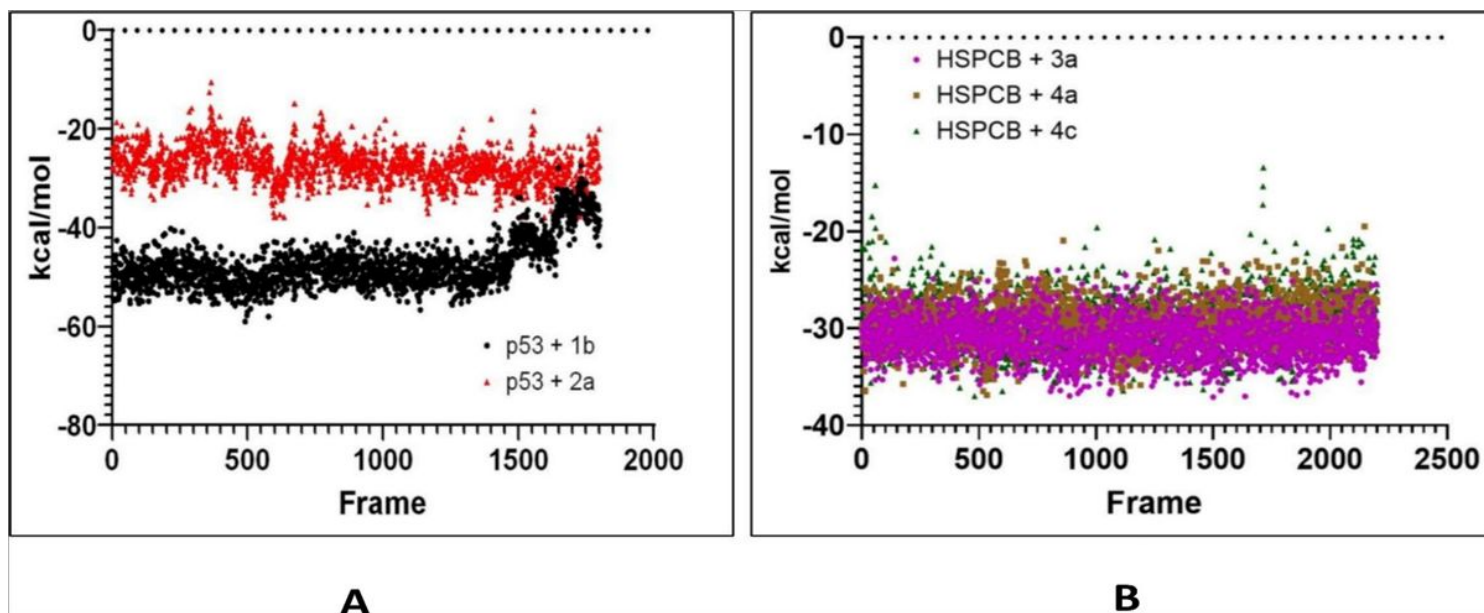


Figure 9

MMGBSA energy values from p53 (A) and HSPCB (B) complexes during simulation time.

MMGBSA

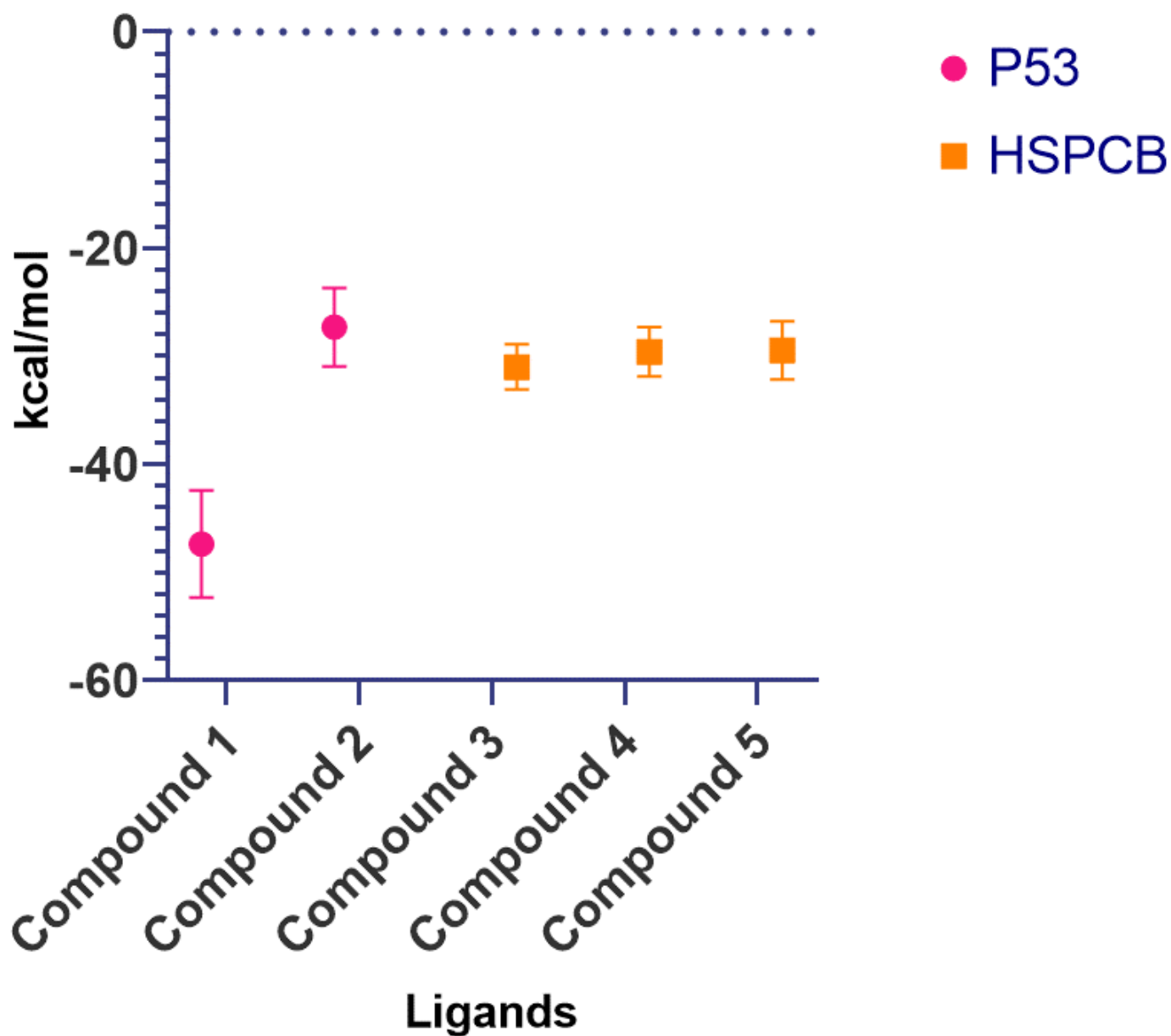


Figure 10

Summary of MMGBSA results

Supplementary Files

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

- [Supplementaryfile.docx](#)
- [SupplementaryExcel1.Enrichmentanalysis.xlsx](#)
- [SupplementaryExcel2.ProteinProteininteractionnetworkanalysis.xlsx](#)