

RESEARCH

Open Access



# Simulation studies, 3D QSAR and molecular docking on a point mutation of protein kinase B with flavonoids targeting ovarian Cancer

Suchitra Maheswari Ajjarapu<sup>1,2</sup>, Apoorv Tiwari<sup>1,3</sup>, Gohar Taj<sup>1</sup>, Dev Bukhsh Singh<sup>4</sup>, Sakshi Singh<sup>5</sup> and Sundip Kumar<sup>1\*</sup>

## Abstract

**Background:** Ovarian cancer is the world's dreaded disease and its prevalence is expanding globally. The study of integrated molecular networks is crucial for the basic mechanism of cancer cells and their progression. During the present investigation, we have examined different flavonoids that target protein kinases B (AKT1) protein which exerts their anticancer efficiency intriguing the role in cross-talk cell signalling, by metabolic processes through *in-silico* approaches.

**Method:** Molecular dynamics simulation (MDS) was performed to analyze and evaluate the stability of the complexes under physiological conditions and the results were congruent with molecular docking. This investigation revealed the effect of a point mutation (W80R), considered based on their frequency of occurrence, with AKT1 protein.

**Results:** The ligand with high docking scores and favourable behaviour on dynamic simulations are proposed as potential W80R inhibitors. A virtual screening analysis was performed with 12,000 flavonoids satisfying Lipinski's rule of five according to which drug-likeness is predicted based on its pharmacological and biological properties to be active and taken orally. The pharmacokinetic ADME (adsorption, digestion, metabolism, and excretion) studies featured drug-likeness. Subsequently, a statistically significant 3D-QSAR model of high correlation coefficient (R2) with 0.822 and cross-validation coefficient (Q2) with 0.6132 at 4 component PLS (partial least square) were used to verify the accuracy of the models. Taxifolin holds good interactions with the binding domain of W80R, highest Glide score of  $-9.63$  kcal/mol with OH of GLU<sup>234</sup> and H bond ASP<sup>274</sup> and LEU<sup>156</sup> amino acid residues and one pi-cation interaction and one hydrophobic bond with LYS<sup>276</sup>.

\* Correspondence: [malik.sundip@gmail.com](mailto:malik.sundip@gmail.com)

<sup>1</sup>Bioinformatics Sub-DIC, Department of Molecular Biology & Genetic Engineering, College of Basic Science and Humanities, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, Udham Singh Nagar 263145, Uttarakhand, India

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusion:** Natural compounds have always been a richest source of active compounds with a wide variety of structures, therefore, these compounds showed a special inspiration for medical chemists. The present study has aimed molecular docking and molecular dynamics simulation studies on taxifolin targeting W80R mutant protein of protein kinase B/serine- threonine kinase/AKT1 (EC:2.7.11.1) protein of ovarian cancer for designing therapeutic intervention. The expected result supported the molecular cause in a mutant form which resulted in a gain of ovarian cancer. Here we discussed validations computationally and yet experimental evaluation or in vivo studies are endorsed for further study. Several of these compounds should become the next marvels for early detection of ovarian cancer.

**Keywords:** AKT1, Point mutation, ADME, QSAR, Virtual docking, Dynamic simulations

## Background

Ovarian cancer marks the most lethal gynaecological malignancy which ranks the fifth marveling cause of cancer deaths in females [1]. It is estimated that there are 22,530 cases with a mortality rate of approximately 13,980 deaths in the United States in 2019 [1]. Ovarian cancers are categorized into 3 types based on cell origin: epithelial, stromal and germ cell [2]. The low survival rate and poor prognosis of ovarian cancer are due to a lack of screening methods at the early stages and ineffective treatments for advanced stages of disease [3]. Moreover it is very crucial to dissect the role of tumor-causing microenvironment during the early stage, proliferation, and metastasis. Thus, it becomes paramount to understand the root cause from different views of its molecular pathogenesis, histological subtypes, hereditary factors, epidemiology, methods of treatment, and diagnostic perspectives. The Cancer Genome Atlas (TCGA) revealed that the expression of AKT1, AKT2, and AKT3 was associated with poor patient survival [4]. The marveling cause of the disease is due to genetic and epigenetic changes of the cellular genome. Numerous small drug molecules of AKT gene targeting mutations such as FOXO, glucose metabolism (GSK3), and apoptotic proteins (BAD, NF- $\kappa$ B, FKHR) are available. Cell cycle arrest, apoptosis, DNA repair (MDM2) are critical in disease progression. Among various kinases, overexpression of AKT1 protein and associated mutations play a deciding role in cross-talk cell signalling in causing cancer. Recent studies have introduced assorted therapeutic agents as targets specific for cancer-driven factors involved in the inhibition of ovarian cancer development. One such factor of the kinase family is protein kinase B/serine-threonine (EC:2.7.11.1) (<https://www.brenda-enzymes.org/index.php>) serves as a decisive mediator of the P13K/AKT/mTOR cell signaling pathway that has distinct physiological functions such as cell growth, survival, proliferation, and metabolism [5]. Structurally AKT1 consists of three domains, including an N-terminal pleckstrin homology, a central catalytic kinases domain, and C-terminal domain [6].

AKT1 is the kinase that connects upstream signals from PI3K and mammalian targets of rapamycin complex2 (mTORC2) with downstream signals to mTORC1 and effectors such as mTOR, GSK3b along with phosphorylation cascade which acts as substrates that induce cell cycle progression, protein synthesis, lipid and protein phosphatases, glucose metabolism and cell growth [7]. AKT1 is mutated and AKT2 is amplified in about 40% AKT1 is inhibited by tumor suppressors including phosphatase and tensin homolog (PTEN) and inositol polyphosphate 4-phosphatase type 2 INPP4B [8, 9]. Therefore, targeting ATP binding cleft of AKT protein by inhibitors (natural/synthetic) has become an attractive strategy for treating patients in ovarian cancer. Interestingly, AKT1 protein inhibitors showed a strong binding affinity with mutant forms when compared to the native form. However, the emergence of acquired drug resistance in patients was found to limit its usage in the last phase of clinical trials. In ovarian cancer, overexpression of AKT is associated with advanced-stage platinum resistance [10, 11]. As an isoform of the AKT family, AKT1 is observed to be expressed unduly in a wide assortment of many human cancers including breast and ovarian cancers [12, 13]. The underlying molecular mechanism is assumed to cause conformational changes in native protein structure (AKT1) which modify covalent bond interaction by limiting their practical application. On that account, there is a need to search and develop novel as well as regimes that can counteract the drug resistance induced by the AKT1 gene. Even so, the molecular interactions and atomic stability for the W80R have also been determined for the present study.

W80R results in increased repression of FOXO 3 compared to wild type AKT1 in an invitro assay which is then predicted to result in a gain of AKT1 protein function. FOXO is a transcription factor in the nucleus that induces CGN2 transcription in epithelial ovarian cancer cells with enhanced catenin activity. The absence of Wnt ligand dissociates catenin from the destruction complex and translocates to the nucleus where it acts with the FOXO3 factor which is known to play a role in the W80R protein pathway. Abnormal activation of this

pathway marvels to hyper-activation of catenin, which has been reported in ovarian cancers. W80R is one of the reported mutants of AKT1 cancer which cause missense driver mutation with 238 T > C of the coding sequence, also CDS (change in the nucleotide sequence as a result of mutation, where the syntax here used is identical to the method used for the peptide sequence) mutation c.238 T > A with gene location 14q32.33 [14] in the uterus section causing endometrial cancer. It has been proved that W80R contains highly conserved residues damaged by polyphen2, targeting through PI3K/AKT1/mTOR pathway of substitution-missense variant type affecting exon of protein domain PH (the UniProt Consortium 2019) and SIFT prediction as 3 [12]. The mutant W80R-Q79K on combination found to be displayed a very strong membrane localization and hyperactivation in transfected HeLa cells in both presence and absence of serum under fluorescence microscopy [15]. The previous studies of AKT1 co-occurring mutations (like Q79K-W80R) found to be hyperactive equal to E17K mutant widely distributed in different tissues such as endometrium (homozygous and heterozygous), large intestine (caecum), prostate (with heterozygosity condition) breast cancers involving cross-talk signaling pathways [16]. The deleterious mutations of AKT1 (E17K and W80R) concluded to be of functional relevance exclusively in myxoid tumors [17]. The altering mutations promote growth factor independent cell proliferation as compared to wild type AKT [18]. AKT1 gene alterations account for most of the genetic drive contributing to the pulmonary sclerosing haemangioma which is a benign tumor development [19]. It was observed in the patients receiving genomically targeted therapy that W80R mutant found to be in clinical benefit of SD 4 mo + (stable disease), working efficiently with synthetic drugs temsirolimus and ixabepilone targeting ovary granulosa cell [20]. In line with, the inhibition of AKT1 or its mutant proteins has been recognized as a compelling strategy for the treatment of cancers with [21] induce ovarian tumor angiogenesis [22] and in immune evasion [23].

Existing chemotherapeutic drugs have developed resistance to the novel compounds along with side effects despite enormous progress in anticancer drug discovery. Hence more targeted strategies are required to develop with sensitivity and specificity. Most of the successful anticancer compounds were originated from natural sources or as their analogs. Natural products and their synthetic analogues are a rich source of biologically active compounds which have been recognised as cancer stem cells (CSCs). These anti-CSCs natural products include flavinoids, stilbenes, quinines, terpenoids, polyketide antibiotics, steroids and alkaloids [24]. Flavonoids are naturally occurring secondary metabolites consisting

of polyphenols having therapeutic benefits in multiple ways. These are low-molecular-weight compounds with non-nitrogenous properties consisting of C6-C3-C6 as a backbone with different classes [25] and their activities are structure-dependent. Chemically, flavonoids depend on their structural class, degree of hydroxylation, substitutions, and conjugations, and degree of polymerization [26]. Several mechanisms have been proposed for the effect of flavonoids at the initiation and promotion stages of the carcinogenicity including influences on development and hormonal activities [27]. Flavonoids fall under 6 different categories based on the functional group flavones (luteolin, apigenin), flavonols (quercetin, kaempferol), flavanones (naringenin), flavanoneol (taxifolin), isoflavones, and flavan-3-ols (genistein, epicatechin, catechin, wedelactone, ellagic acid, silibinin, folstein, parthenolids, oridonin, curcumin, resveratrol). The choice of this study has relied on the compounds of the family called flavonoids with a tremendous variety of pharmacological and biochemical consequences including hepatoprotective, antidiabetic, cardio protective, anti-tumor, neuroprotective, and anti-inflammatory and played a wonderful role in the preclusion of Alzheimer's disease. Equally studies on quercetin (QR) demonstrated as its effect on anti-inflammatory, anti-apoptotic, antioxidant, and anticancer agent. This also found to improve the quality of oocytes and embryos. It affects the proliferation and apoptosis and thereby decreases in oxidative stress in granulosa cells (GCs). Furthermore, it is also used as a complementary and alternative therapy in ovarian cancer with beneficial effects in treatment with PCOS (polycystic ovary syndrome) patient [28]. In an earlier investigation, this area has demanded series of chemical methods and animal models to synthesis marvel compounds with more time, investment, and level of exposure. To overcome this issue, computational approaches have opened doors for inquisition in predicting the mutation both in induced drug resistance and also to design resistance evading drugs. As a result of the above-mentioned shortfalls, the present study has aimed at the dynamic simulation at the molecular level and molecular docking studies on taxifolin targeting W80R mutant protein in protein kinase B/serine-threonine kinase/AKT1 protein of Ovarian cancer for designing therapeutic. This computational study relies on learning and pattern classification methods which can classify mutations create 3D protein structures.

## Materials and methods

### Sequence retrieval and structure analysis of selected protein

The amino acid sequence of AKT1 protein was retrieved from the Uniprot database with accession number P31749. The primary structure of the protein was

elucidated using the ProtParam tool [29] of the ExPASy server and the difference between physical and chemical properties of the AKT1 protein (wild) and mutant (W80R) were evaluated. Factors such as physicochemical properties, molecular weight, theoretical pI (isoelectric point), half-life, instability index (II), aliphatic index (AI), extinction coefficient (EI), grand average hydropathy (GRAVY), and site of origin were analyzed. The secondary structure properties prediction was carried out by the RAMPAGE server, which provides the configuration score like the total number of helices, turns, coils, predicted solvent accessibility, with the range, existed from 0 (highly buried) to 9 (exposed region) depending on the residue exposed. Normalized B-factor is measured for a selected protein as Z score which is a combination of template and profile-based prediction where residues are higher than zero are considered as less stable during experimental structures. The mutant protein W80R was edited manually at the amino acid position number and submitted to homology modelling [30].

### 3D modelling of W80R protein

Apart from successful experimental methods such as x-ray crystallography and nuclear magnetic resonance for 3D structures, there still exists the knowledge gap about the structural information about the protein. However, computational methodology fills the gap by an approach called Homology Modelling and makes it fit for the drug discovery purposes. The 480 amino acid residue length of W80R protein was retrieved to recognize the appropriate template sequence (PDB access code: 3O96) for structure modelling and functional prediction of the protein. This modelling depends mainly on a sequence alignment between the target and template sequence whose structure has been experimentally determined with [31, 32] the 3D structure of the target protein using its template was performed by MODELLAR (<https://salilab.org/modeller/>) and visualized by the PYMOL tool; based on template-target alignment. These theoretical structural models of the W80R protein were ranked based on the normalized discrete RMSD values. The model with the lowest RMSD score was considered as the best model [33].

### Evaluation of the structure model

The quality of AKT1 and mutant form W80R models were assessed by many tools to evaluate the stability and reliability of the model. PROCHECK suite [34] quantifies the residues in favorable zones of the Ramachandran plot, was used to evaluate the stereochemical quality of the model. ERRATA tool [35] finds the overall quality factor of the protein and was used to check the statistics of non-bonded interactions between different atom types. The compatibility of the atomic model (3D) with

its amino acid sequence was determined using the VERIFY 3D program. Swiss PDB viewer 4.1.07 was used for the energy minimization of the predicted AKT1 protein along with its mutant form. The W80R model was further subjected to structural analysis and verification server to evaluate its quality, before and after energy minimization. ProSA tool [36] was employed for the refinement and validation of the minimized structure to check the native protein folding energy. The superimposition of the proposed model of AKT1 protein along with mutant form with its closest-structural homolog was carried out using chimera 1.11 [37].

### Selection and preparation of ligands

Natural compounds database containing more than 12,000 ligands were aimed to the AKT1 protein family were downloaded from the Pubchem database [38] and subjected to ligand preparation by ligprep wizard application of the Maestro 9.3 [39]. Ligprep tool was used to prepare the high quality of ligands, such as the addition of hydrogen's, conversion of 2D to 3D structures, corrected bond angles and bond lengths, with lower energy structure, stereochemistry's, and ring conformation followed by minimization in the optimized potential of OPLS 2005 force field [40, 41]. Properties such as ionization did not change and tautomers were not generated, specifically retained chiralities. Compounds were selected based on the lowest energy.

### Preparation of protein molecule and active site prediction

The W80R protein was modelled by using the protein preparation wizard of Schrodinger Suite; by adding hydrogen atoms, optimizing hydrogen bonds, and verifying the protonation states of His, Gln, and Asn. Energy minimization was carried out using constraint 0.3 Å RMSD and OPLS 2005 force field with steepest descent algorithm. The sitemap tool was used to identify binding pockets of W80R protein [42].

### Receptor grid generation

Receptor grid generation was done by the Glide application [42]. The receptor grid for W80R was generated using active site residues which were identified by Sitemap tool. Once the grid has been generated, the ligands are docked to the protein (W80R) using Glide version 5.8 (Grid-based Ligand Docking with Energetics) docking protocol. The scaling factor (0.25) and partial charge (1 Å) represent cut-offs of Vander Waals radius scaling.

### Molecular docking

Docking is the popular method of molecular modelling to build ligands into the active site of receptor molecule by estimating energy for the ligand binding to the protein [43, 44]. The value of this energy determines the

biological activity of the molecules i.e. the higher energy, the more effective the drug based on the receptor will be considered. However, the term scoring (score) is used for calculation of binding energy by a ligand to a receptor molecule rather ranks assigned to position of ligand with their specific targets procedures were consistently carried out using a preparation of protein of Schrodinger and defining the grid on the active site of the protein. The reliability of the molecular docking is significantly affected by the accuracy of docking scores and the 3D structure of the receptor [45].

**GLIDE (Grid based ligand docking with energies)** molecular docking tool uses computational simulation methods for evaluating particular poses and ligand flexibility. GLIDE systematic method, a new approach for rapid, accurate molecular docking and its output G-score, is found to be an empirical scoring function, is a combination of diversified attributes. Glide uses the Emodel scoring function to select between protein-ligand complexes of a given ligand and the Glide Score function to rank-order compounds to separate compounds that bind strongly (actives) from those that don't (inactives). G-score is calculated in Kcal/mol, encompass ligand-protein interaction energies, hydrophobic interactions, hydrogen bonds, internal energy, pi-pi stacking interactions, root mean square deviation (RMSD), and desolvation. GLIDE modules of the XP visualizes analysis of the specific ligand-protein interaction [46]. The ligands were docked using Extra Precision mode (XP) and conformers were evaluated using the Glide (G) score. The G score is calculated using this formula as:

$$GScore = a * vdW + b * Coul + Lipo + Hbond + Metal + BuryP + RotB + Site$$

where vdW denotes van der Waals energy, Coul denotes Columb energy, Lipo denotes lipophilic contact, H-bond indicates hydrogen bonding, Metal indicates metal-binding, BuryP indicates penalty for buried polar groups, RotB indicates penalty for freezing rotatable bonds, site denotes polar interactions in the active site and  $a = 0.065$  while  $b = 0.130$  were the coefficients of vdW and Coul.

**ADME properties studies** Calculation of absorption, distribution, metabolism, excretion, and toxicity (ADME/T) properties was performed for best-docked ligand molecules by QikProp software. This software predicts various limiting factors such as QP log Po/w, QPlog BB, SASA, FOSA, FISA, PISA, WPSA, volume, donarHB, acceptorHB,  $dip^2/V$ ,  $AC * DN * 5$ , Caco, QlogS, rotors, rule of 5, rule of 3, the overall percentage of human oral absorption, etc. [47]. Lipinski's rule of five [48] measures the drug-likeness for the prediction of a chemical compound as an orally active drug based on biological compounds and pharmacological properties.

### Analysis of cancer-associated mutants

The deleterious W80R mutations that are specific for cancers were predicted using the FATHMM server (<http://fathmm.biocompute.org.uk/>) [49] which allows the distinct difference between cancer-promoting/driver mutations and other germline polymorphisms. The gene number identifiers (UniProt id) along with mutant form as a text were provided as the input for the prediction.

### Molecular alignment and 3D QSAR studies and validation

The key component of 3D QSAR analysis is the arrangement of the molecules based on the scaffold they share which generated using the training was set of 44 molecular poses with a grid spacing of 1 Å PLS (partial least square) algorithm to establish the relationship between biological activity and different structural features. The training set was adjusted to 50%. Three models were generated by Gaussian filed extension as Gaussian steric, electrostatic, hydrophobic, hydrogen bond donor, hydrogen bond acceptor, and aromatic ring fields. CoMFA and CoMSIA are the tools employed as independent variables in PLS regression analysis. The best model was chosen based on the criteria of statistical robustness and visualized using contour map modules. The predictive power and stable models were assessed using the leave one odd (LOO) cross-validation method. The crucial aspects for the test set statistics include RMSE, Q<sub>2</sub>, SD, R<sup>2</sup>, R<sup>2</sup>CV, R<sup>2</sup>scramble, stability, F, P, Q<sub>2</sub>, Pearson's r which indicates the predictive ability of the model. A Scatter plot was generated in correlation with predicted activity on the Y-axis and observed activity on the X-axis of the data set model [50].

### Contour maps visualisation

Representation of the fields as contours (surfaces) or as color intensities of the fields on the grid can be displayed in different styles. Based on the field type, the colors are designed and field intensities are shown for one field at a time. The fields with greater absolute values than the cut-off were presented at the maximum brightness.

### Molecular dynamics simulation

The simulation of protein-ligand complexes was implemented by GROMACS 4.5.5 (Groningen machine for Chemical Simulations) software [51]. The complex with the lowest binding energy was selected for molecular dynamics (MD) simulation. The ligand parameters were analyzed using PRODRG online server [52] in the framework of GROMACS force-field 43a. The ligand enzyme complex was solvated at a simple point charge as well as a water box under periodic boundary conditions using 1.0 nm distance protein to the box faces. The system was then neutralized by Cl<sup>-</sup> or Na<sup>+</sup> counter ions for the W80R complex with ligand respectively. To perform

energy minimization, the complex was equilibrated under volume, constant number of particles, and temperature condition for 100 ps at 300 k, followed by 100 ps. All the covalent bonds with hydrogen bonds were considered using a linear constraint solver algorithm. The electrostatic interactions were treated using the particle mesh Ewald method [53]. Further MD simulation studies were noted for 20 ns to check the accuracy and stability of the ligand-protein complexes. The potential of each trajectory produced after MD simulations were analyzed using *g\_rms*, *g\_rmsf*, and *g\_h* bond of GROMACS utilities [54] the root mean square deviation (RMSD), the root mean square fluctuation (RMSF), with hydrogen bonds formed between the ligand and protein complex.

## Results

### Mutant W80R sequence analysis

The development of anticancer compounds with varied pharmacological effects becomes a very paramount topic and hence main class of secondary metabolites, both dietary and synthetic flavonoids have been subjected to clinical trials [55]. Definite beneficial biological activities of dietary flavonoids including antioxidants [56] anticancer [57], and cardio-protective properties [58] have been identified in a series of previous studies. Flavonoids are known for their wide exposure to chemopreventive, chemotherapeutic activities, and the availability of the compound in plant sources for the human diet in routine consumption [59].

The analysis of the mutant W80R protein sequence of the AKT1 has 480 amino acid residue which plays a very crucial role in metabolism, cell proliferation, cell survival, growth, and angiogenesis, was downloaded from Uniprot with accession number (P31750). The amino acids in the protein sequence of W80R were composed of lysine, leucine, glutamic acid, and alanine. The ProtParam tool was used for the W80R protein sequence to compute physio-chemical parameters such as molecular weight of 5565.45 kD. The W80R had a pI (isoelectric point) of 5.99 indicating its acidic nature ( $pI < 7.0$ ) with

an aliphatic index (AI) (71.69). The protein volume is occupied by aliphatic side chains such as lysine, leucine, glutamic acid, and alanine. The instability index of W80R measured 35.76 of the unstable nature. The grand average of hydropathicity (GRAVY) of W80R protein was lower ( $-0.583$ ), which proves its high affinity with water. The comparison of statistical characteristics is showing the differences among wild AKT1 and mutant W80R using the ProtParam tool (Table 1). The comparison of sequence analysis of W80R mutant protein with AKT1(wild) at nucleotide and protein level was same with a slight difference, thus proving-T, C-G rich region, and properties such as molecular weight, amino acid composition, theoretical pI, aliphatic index, and grand average of hydropathicity (GRAVY) were found in an appropriate range of influencing the protein stability.

### 3D molecular modelling of W80R mutant protein

The 480 amino acid residue length of W80R protein was subjected to BLASTp analysis against RCSB PDB to identify the suitable template for comparative structural modelling and functional prediction. The result of the BLASTp search revealed a template (PDB id 3O96) of high-level identity with the target sequence of AKT1. The query coverage (100%) showed a high degree of identity between two proteins (AKT1 and W80R) of 480 sequence length, and E value ( $2e-60$ ) is the expected value obtained by hits, percentage identity defines the extent of two sequences, Modeller 9.13 has generated 5 models of W80R, among these the lowest score is considered as stable which is thermodynamically subjected to further refinement. The lowest RMSD as 0.18 score model was considered as the best one for further validation purposes [33]. Finally, three dimensional (3D) structure of the selected protein using its template was visualized by the PYMOL tool.

### Model assessment and validation

The stability of the protein was constructed based on the backbone of torsion angles psi and phi which were evaluated by the PROCHECK server that computes the

**Table 1** Comparison of primary sequence analysis using the ProtParam tool between AKT1 (wild) and W80R (mutant)

S.No	Parameters	AKT1	W80R
1	Molecular weight	5586.7kD	5565.4kD
2	pI	5.75	5.99
3	Aliphatic Index	71.69	71.69
4	Instability Index	35.47	35.76
5	GRAVY	-0.575	-0.583
6	Atoms	7772	7776
7	Total number of Asp+Glu residues in a protein content	77	76
8	Total number of Arg + Lysresidues in a protein content	66	68

amino acid residues in the existing zones of Ramachandran plot analysis of W80R mutant forms (Table 2). The information presented in Table 2 depicts the Ramachandran plot through RAMPAGE server where W80R mutant protein has 79.3% amino acids falls in the most favored region with located major active binding sites, while 13.8% in an allowed region and 6.9% residues in the outlier region of the plot with lesser significance. SAVES analysis was conducted to confirm the quality of the protein model followed by ProSA, RMSD assessment for a high-quality structural model for virtual screening. The quality of the predicted model of AKT1 protein and a W80R mutant was supported by a high ERRAT score of 81.99 in an acceptable protein environment. The VERIFY 3D results of W80R showed 81.88% of the residues with an average 3D-1D score  $> 0.2$ , indicating the stability of the model. 'WHAT IF' tool examines the coarse packing quality, the model protein structure, reflecting the acceptance of good quality. The reliability of the W80R form was confirmed by ProSA (Fig. 1) which achieved a Z score of  $-7.92$  kcal/mol compared to the wild form AKT1 having a Z score  $-7.2$  kcal/mol, wherein the energy is negative, reflects the best quality of the model. The quality of the model was evaluated through the comparison of predicted structure with experimentally determined structure followed by superimposition and atoms RMSD assessment using Chimera 1.11, which proved that the predicted model is good and quite similar to the wild protein.

#### Active site and score prediction

A proven algorithm for binding site identification and evaluation of the drug ability of those sites marvels to modify hit-compounds to enhance receptor complementarities. The active site was performed using a sitemap tool to assess each site by calculating attributes such as size, volume, amino acid exposure, hydrophobicity, hydrophilicity, donor/acceptor ratio. The most reliable score was obtained in the binding pockets of W80R. The predicted amino acids in the active region were LEU<sup>156</sup>, GLY<sup>157</sup>, GLU<sup>234</sup>, MET<sup>281</sup>, ASN<sup>279</sup>, GLU<sup>278</sup>, LYS<sup>276</sup>, ASP<sup>274</sup>, THR<sup>291</sup>, ASP<sup>292</sup>, PHE<sup>293</sup>, GLY<sup>294</sup>, LEU<sup>295</sup>, GLU<sup>298</sup> of site score for the selected model was 1.128, drug ability score  $-1.149$  with Volume 384.486 and size measured was 179 for further docking analysis.

#### Analysis of cancer-associated mutants

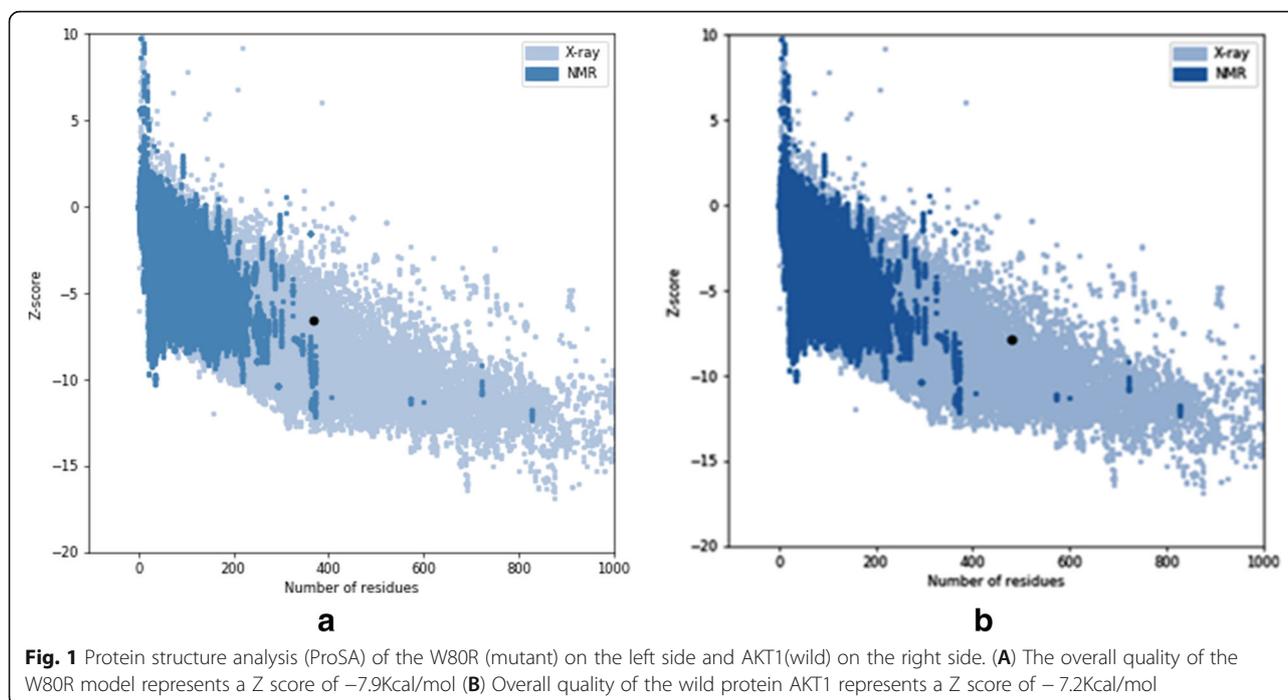
The mutation impact for the protein W80R was classified using the FATHMM server derived from the new FATHMM-MKL algorithm. It distinguishes between cancer-promoting/driver mutations and other germline polymorphisms. This algorithm predicts the functional, molecular, and phenotypic consequences of the missense mutation of a functional protein using hidden Markov models (HMMs), representing the alignment of homologous sequences and conserved protein domains with "pathogenicity weights", representing overall tolerance of protein/domain to mutations [49]. The gene number identifier (UniProt id) along with mutant form as a text was provided as the input for the prediction based on the FATHMM server predictions with a score  $-1.12$  responsible for benign cancer. The functional scores for individual mutations were obtained from the FATHMM-MKL server which falls in the range of 0–1 known as single p-values fall in the range of (0–1) where the values below 0.5 are predicted as benign and above 0.5 are deleterious.

#### Determination of ADME profile

Molecular properties of the selected compounds were studied using Qikprop and chosen based on the Lipinski rule of five which marks the most important activity in drug discovery and development. Multifarious Insilco techniques have been employed to measure the drug-likeness for a compound based on numerous descriptors. Calculation of absorption, distribution, metabolism, excretion, and toxicity (ADME/T) properties was predicted for best-docked ligand molecules using Qikprop software. Qikprop computes almost 20 physical descriptors over a wide range of predicted properties unlike a fragment-based approach, by screening compound libraries for hits and play a marvel optimization that can be used to improve predictions by fitting to experimental data and also to generate QSAR models. The detailed analyses of chemical and molecular descriptors and also solubility properties were tabulated in Table 3, 4, and 5. The results of ADME properties are an important index to check the clinical candidates have reached the required standard. It is revealed that compounds in the table were ranked based on the potential drug properties. According to a previous study,  $\sim 40\%$  of failures to develop medicine in the development phase are due to

**Table 2** Comparison of secondary structure using RAMPAGE server between AKT1 and W80R mutants

S.No	Protein Properties	AKT1(wild)	W80R(mutant)
1	Total amino acids	480	480
2	Number of residues in favoured region ( $-98.0\%$ expected)	388 (81.2%)	379 (79.3%)
	Number of residues in allowed region ( $-2.0\%$ expected)	61 (12.8%)	66 (13.8%)
3	Number of residues in outlier region	29 (6.1%)	33 (6.9%)



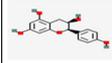
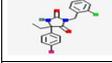
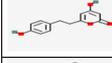
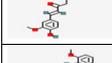
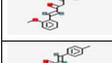
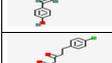
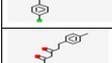
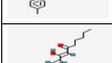
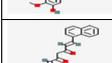
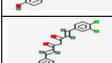
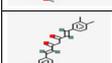
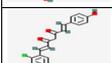
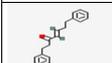
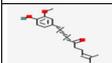
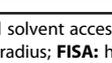
poor biopharmaceutical properties (pKa-dissociation constant and bioavailability) [60]. The ADME as a deal medicine has the following characteristics, hydrogen bond donor  $< 5$ ; hydrogen bond acceptor  $< 10$ ; molecular weight  $< 500$  Da; lipid water partition coefficient  $< 5$ ; water solubility partition coefficient  $-6.5 < \log s < 0.5$ ; and polar surface area 7.0–20.

#### QSAR studies and validation

A dataset of 44 ligand compounds was chosen for statistical studies and classified as the training set and test set into 50% for suitable 3D QSAR model development. The graphical interface allowed building dataset into training and testing equally for 50% by generating a correlation coefficient. The graph obtained for all/training models/test models was observed in Fig. 2A. Molecular descriptors (ligands) were divided into a training set and test set (Table 6) with parameters such as phase QSAR, phase activity, % extrapolation, predicted error and predicted activity. QSAR built model was generated based on docking poses and substructure alignment was represented with standard deviation for the regression distributed over  $n-m-1$  degrees of freedom ( $n$  ligands,  $m$  PLS factors) as 10.7913,  $R^2$  (the coefficient of determination) gives 0.8226 means the model accounts for 82% of the variance in the observed activity, which falls between 0 and 1,  $R^2C$  yields 0.2055 for cross-validated where  $R^2$  is obtained by leaving an  $N$ -out approach,  $R^2$ scramble ( $R^2$  is regression or coefficient of determination) obtained as 0.4889 which computes the average value obtained using scrambled activities of Fig. 2B. This parameter is

calculated from a series of models built using scrambled activities. The lower values indicate that the model cannot fit for random data while higher the value means that the variable set can be fairly complete and allows fitting anything.  $R^2$  value indicates the data set is over fit. Larger the value, the larger the data occupancy. The stability provides the degree for molecular fields that can fit random data, however, statistics observed to be 0.379 for the model prediction of the changes obtained in the training set composition. The descriptor F gives the ratio of the model variance to the observed activity variance. The model variance is calculated for the distributed data over  $m$  degrees of freedom and the activity variance is distributed over  $n-m-1$  ( $n$  ligands,  $m$  PLS factors) hence obtained value of 92.7 means indicates more significant regression. Pearson descriptor measured  $5.95e-09$ , where the significance level treated as a ratio of Chi-squared distributions. The smaller values indicate a greater degree of confidence while a  $P$  value of 0.05 means F is significant at a 95.6% level. RMSE- root mean square error predictions for the test set were to be 22.02,  $Q^2$  for predicted activities with 0.2915. If the value becomes negative, then the variance in the errors is larger than the variance from the observed activity. Pearson-r correlated with predicted activity, and observed activity observed for test set with 0.7508. The test set was determined within the maximum range of the training set. The field values for the ligand were estimated outside the range found in the training set in percentages was calculated under Extrapolation of the complete data set.

**Table 3** Detailed analysis of ADME properties of ligands using QIKPROP software

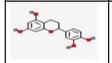
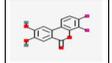
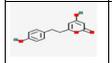
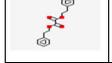
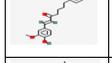
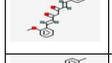
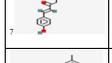
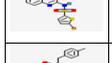
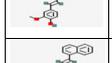
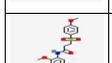
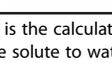
Chemical Structure	CID id	Mol.wt	Dipole	SASA	FOSA	FISA
	443637	310.392	7.125	515.633	442.178	73.455
	44264122	276.334	2.395	457.915	415.498	42.417
	71424203	336.387	5.934	489.544	447.123	42.421
	5907705	336.387	0.667	489.55	447.131	42.419
	6421929	298.338	0.002	600.799	533.867	66.932
	6440365	310.392	6.34	532.097	458.644	73.453
	10688261	336.387	3.009	553.24	511.992	41.249
	10852057	304.388	8.606	478.22	435.803	42.417
	14407192	280.366	4.742	505.615	428.846	76.769
	19792482	349.256	21.006	553.576	388.72	42.774
	19792563	308.419	2.685	554.716	511.942	42.774
	22321203	290.358	7.377	519.093	427.419	91.673
	24952793	342.393	4.9	577.054	369.818	89.296
	24952797	391.25	24.947	516.006	322.823	88.262
	24966389	350.413	7.928	576.902	488.639	88.262
	44538447	361.224	9.432	540.471	339.66	90.009
	44560954	280.366	2.771	484.13	407.362	76.769
	44610342	318.412	7.987	624.49	536.048	88.441
	53693682	286.37	5.469	532.126	460.735	71.391

\***SASA**: total solvent accessible surface area in square angstroms using a probe with a 1.4Å radius; **FISA**: hydrophilic component of the SASA (SASA on N, O, and H on heteroatom); **FOSA**: hydrophobic component of the SASA (saturated carbon and attached hydrogen); **CID ID**: compound Id from PubChem database.

### Contour visualisation

The contour maps (Fig. 3) were used to illustrate the fields required for biological activity. Field-based QSAR interface creates electrostatic, hydrophobic, and steric

**Table 4** Detailed analysis of ADME properties of ligands using QIKPROP software

Chemical Structure	CID ID	VOLUME	DonorHB	AcceptorHB	dip2v	AC*DN
	443637	846.643	1	2.75	0.059593	0.005333
	44264122	721.579	0	4	0.007948	0
	71424203	799.111	2	6	0.044058	0.017333
	5907705	799.119	2	6	0.000557	0.017333
	6421929	914.122	0	3.5	0	0
	6440365	855.378	1	2.75	0.046994	0.005168
	10688261	914.124	2	6	0.009905	0.015337
	10852057	769.386	0	4	0.096272	0
	14407192	805.364	1	2.75	0.02792	0.005439
	19792482	891.173	0	4	0.495123	0
	19792563	896.185	0	4	0.008045	0
	22321203	812.517	1	2.5	0.066971	0.004816
	24952793	899.071	1	4.75	0.026703	0.008232
	24952797	833.649	1	4.75	0.746536	0.009205
	24966389	906.417	1	4.75	0.069341	0.008234
	44538447	846.213	1	4.75	0.105131	0.008789
	44560954	791.867	1	2.75	0.000697	0.00568
	44610342	980.133	1	4.75	0.06508	0.007606
	53693682	821.324	1	2.75	0.056413	0.005168

\***Donor HB**: it is the calculated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution, values are averages take over many configurations, so they can be non-integer; **Acceptor HB**: it is estimated as the number of hydrogen bonds that would be accepted by the solute from water molecules in aqueous solution; **dip2/v**: square of the dipole moment divided by the molecular volume. This is the key term given in Kirkwood-Onsager equation for the free energy described of solvation of a dipole moment with volume V; **AC\*DN**: index of cohesive interaction in solids; **Volume**: total solvent-accessible volume in the cubic angstroms using a probe with 1.4 Å radius.

**Table 5** Solubility prediction parameters for molecular descriptors

CID	QPlogpC1	QPlogPW	Qplogpoc	QPlogPw	QPlogPo	QPlogS	CIQlogS	QPlogHER	QPPCaco	QPPMDCK	QPlogKp
5,318,278.0	27.1	6.7	11.8	5.4	2.9	-4.5	-3.0	-4.1	1992.1	1042.1	-2.8
5,469,422.0	22.7	5.0	8.7	5.1	1.8	-2.8	-1.9	-4.0	3923.3	2167.8	-2.3
5,469,423.0	25.8	6.3	14.3	10.0	1.5	-3.3	-2.2	-3.9	3923.0	2167.6	-2.3
5,907,705.0	25.8	6.3	13.7	10.0	1.5	-3.3	-2.2	-3.9	3923.1	2167.7	-2.3
443,637.0	30.4	6.9	10.9	5.1	3.1	-5.8	-2.6	-5.4	2297.1	1215.5	-2.8
44,264,122.0	27.4	6.7	11.7	5.5	3.0	-4.8	-3.0	-4.5	1992.2	1042.1	-2.8
71,424,203.0	30.4	7.3	15.4	10.2	2.1	-4.3	-2.2	-4.4	4024.7	2228.4	-2.3
10,852,057.0	24.6	5.5	10.6	5.1	2.1	-3.2	-2.3	-4.0	3923.3	2167.8	-2.3
14,407,192.0	25.4	6.2	10.8	5.4	2.6	-4.3	-2.6	-4.3	1853.1	963.6	-2.8
19,792,482.0	29.5	7.5	18.3	5.3	3.3	-5.1	-3.4	-4.6	3892.8	10,000.0	-2.3
19,792,563.0	29.7	6.6	11.1	5.3	2.9	-4.6	-2.4	-4.6	3892.8	2149.6	-2.3
22,321,203.0	25.1	6.3	11.1	5.1	2.7	-4.5	-2.9	-4.6	1338.3	677.9	-3.0
24,952,793.0	30.3	-7.7	13.7	8.4	2.4	-4.9	-3.3	-18.3	1409.6	717.0	-2.7
24,952,797.0	26.6	-9.3	22.7	7.5	2.3	-4.3	-3.9	-17.7	1441.8	2759.8	-3.1
24,966,389.0	29.5	-8.5	13.8	7.6	2.4	-4.9	-2.9	-17.7	1441.8	734.7	-3.1
44,538,447.0	27.1	-8.6	14.0	7.6	2.4	-4.8	-3.4	-17.7	1387.8	2852.3	-3.1
44,560,954.0	24.9	-10.9	10.4	5.4	2.5	-4.0	-2.6	-17.7	1853.1	963.7	-2.8
44,610,342.0	32.4	-7.3	14.7	7.7	2.8	-5.8	-2.5	-17.7	1436.1	731.6	-3.1
53,693,682.0	26.1	-9.8	11.0	5.4	2.7	-4.8	-2.7	-17.7	2084.0	1094.1	-2.739

\***QPlogPoc**: predicted octanol/gas partition coefficient; **QPlogPw**: predicted water/gas partition coefficient; **QPlogPo/w**: predicted octanol/water partition coefficient; **CIQlogS**: conformation-independent predicted aqueous solubility,  $\log S$  in  $\text{mol dm}^{-3}$  is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid; **QPlogHERG**: predicted IC<sub>50</sub> value for blockage of HERG K<sup>+</sup> channels; **QPPCaCo**: predicted apparent CaCo-2 cell permeability in nm/sec; CaCo-2 cells are a model for the gut blood barrier; **QPlogKp**: predicted skin permeability,  $\log Kp$ ; **QPlogS**: Predicted aqueous solubility,  $\log S$ ,  $S$  in  $\text{mol dm}^{-3}$  is the concentration of solute in the saturated solution that is in equilibrium with the crystalline solid; **QPPMDCK**: Predicted apparent MDCK cell permeability in nm/sec, MDCK cells are considered to be a good mimic for the blood-barrier; **QPlogpCI**: Predicted hexadecane/gas partition coefficient

fields for optimization and marvels discovery. The represented green contour indicates the bulky group in a favourable region. The contour map depicts hydrophobicity in the solvent-accessible hydrophobic pocket steric fields are considered as the most favourable regions with a high Glide score. The obtained results have shown the steric and Gaussian field fractions are much larger than other fields suggesting most of the binding energy has been contributed from hydrophobic interactions.

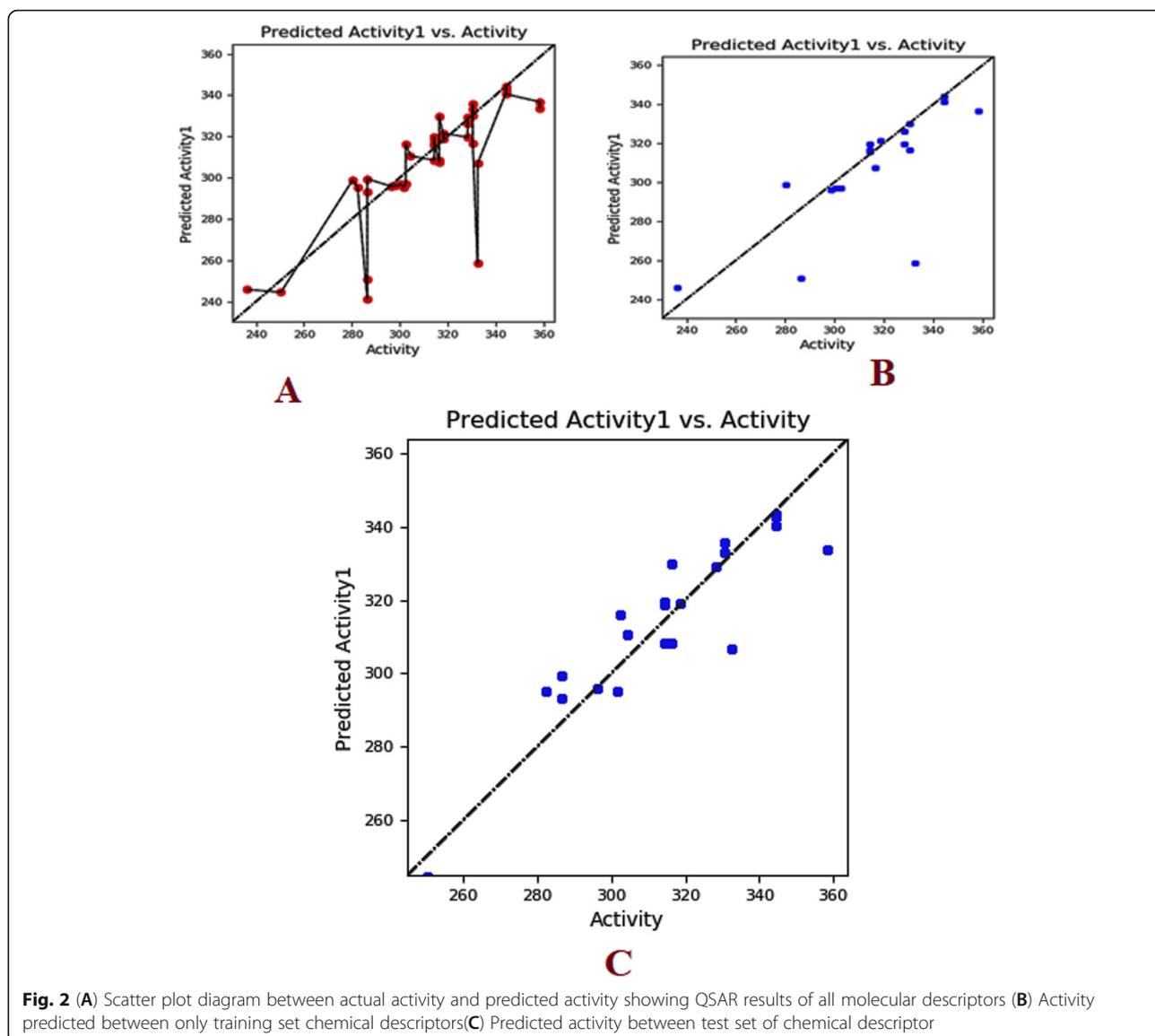
### Molecular docking studies

Molecular docking is the paramount computational tool to configure (Fig. 4) all the possible active conformations of binding at the active site for the receptor molecule. Before performing the docking protocol, the co-crystallized ligand was re-docked into the crystal structure of the 4GV1 (AKT1) receptor molecule to evaluate the reliability of the standard precision algorithm of the Glide. A dataset of flavonoids family along with its structural analogs comprising 7000 ligands was selected. Upon generation of Epik for suitable tautomeric states per 16 for each ligand, 12,000 ligands were chosen entirely as a whole set for virtual screening with W80R mutant protein. The top three ligands with the best

binding energy were considered for further analysis (Fig. 4).

The major contribution is incorporation of interaction energies of Coulomb and vdW between ligand and the receptor. The wide disparities from the original interaction energies seem to be reduced greatly, although charge-charge interactions were found to be favoured to an extent. The coulomb-vdW energies used in Glide score 2.5 employ these reductions in net charge except in anionic ligand-metal interactions, for which glide uses the full interaction energy.

Several hydrogen bond interactions were found in the docking result. The top-scoring compound belongs to CID-443637 was having lower binding energy with a Glide score of -9.63 Kcal/mol. The hydroxyl group of LEU<sup>156</sup>, GLU<sup>234</sup> and ASP<sup>274</sup> forms hydrogen bond revealing the strongest stability with the receptor molecule. The three hydrogen bond interactions provide the guarantee for stable conformation of a binding ligand molecule to protein structure which influences the activity of the ligand. The interaction with 1 pi ~ cation LYS<sup>268</sup> recognized as energetically significant [61] and noncovalent binding interaction proves to exist in a quite strong platform both in the gas phase and liquid



media [62] which is a special hydrophobic interaction having a cationic side chain amino acid, indicating that the geometry is biased towards aromatic amino acid, one that experiences a favorable  $\pi$ -cation interaction [63] having IUPAC name 2-(3,4-dihydroxy phenyl)-3,4-dihydro-2H-chromene-5,7-diol of C<sub>15</sub>H<sub>14</sub>O<sub>5</sub> (Fig. 5). The second highest molecule of 44264122 has a binding energy of -9.43 kcal/mol with hydrophobic contacts with residues such as THR<sup>291</sup>, ILE<sup>290</sup>, THR<sup>211</sup> and oxy bond with LYS<sup>268</sup> having IUPAC name 3,4 Difluoro-8,9-dihydrobenzo[c] chromen-6-one of C<sub>13</sub>H<sub>6</sub>F<sub>2</sub>O<sub>4</sub> (Fig. 6). The third compound CID 71424203 has the binding energy of -9.36 kcal/mol forming hydrogen bond interactions with amino acid residues THR<sup>211</sup>, SER<sup>215</sup> aromatic amino acid residue, and TYR<sup>474</sup> of 1  $\pi$ - $\pi$  stacking interaction. The residue TRP<sup>80</sup> between two aromatic

amino acids has a separation of -3.35Å (vDw) having IUPAC name 2,5,7-trihydroxy-3-(4-hydroxyphenyl)-2,3-dihydrochromen-4-one of C<sub>13</sub>H<sub>12</sub>O<sub>4</sub> (Fig. 7). After the comparison of all three models, the compound CID with 443,637 with the lowest energy is chosen for further molecular dynamics simulation studies.

The lower Glide score represents the most and highest favorable binding affinity. Hydrogen bond interactions,  $\pi$ -interactions,  $\pi$  staking of the best poses were visualized and interpreted using XP visualizer with descriptors (Table 7) in ascending order. It rewards the topmost ligands for hydrogen bond with lengths and angles deviating significantly from "ideal" hydrogen-bond interaction (1.65Å H-A distance, 180 D-H A angle) [47]. The PhobEn measures hydrophobic enclosure reward on the protein. The lipophilic EvdW is the term for the

**Table 6** List of statistical analysis generated using field-based QSAR

CID ID	Chemical Structure	Phase QSAR	Phase Activity	% Extrapolated	Prediction Error	Prediction Activity
368569		Test	314.424	0.653	5.16645	319.59
443637		Trainin g	358.477	0.458	-24.7856	333.692
461867		Trainin g	296.408	0	-0.612247	295.796
2751794		Trainin g	314.424	0	-5.94993	308.478
44264122		Test	280.366	0.049	18.5276	298.894
6324890		Trainin g	330.423	0	5.28888	335.712
71424203		Test	332.482	0.227	-74.0301	258.452
11566133		Test	330.466	0.055	-13.8518	5.51
14379076		Trainin g	344.45	0	-3.96902	340.481
14633296		Test	286.456	0.246	-35.6831	250.773
15690458		Test	298.381	0.044	-2.43817	295.943
20975206		Test	236.353	0.22	9.44186	245.795
22973545		Test	300.483	0.037	-3.7.753	296.775
2430736		Trainin g	250.38	0.039	-5.88.58	296.775
25303072		Trainin g	286.456	0	12.9111	299.3672
44254444		Test	328.407	0.044	-2.11377	326.293
44254874		Test	344.407	0.045	-0.262634	344.144

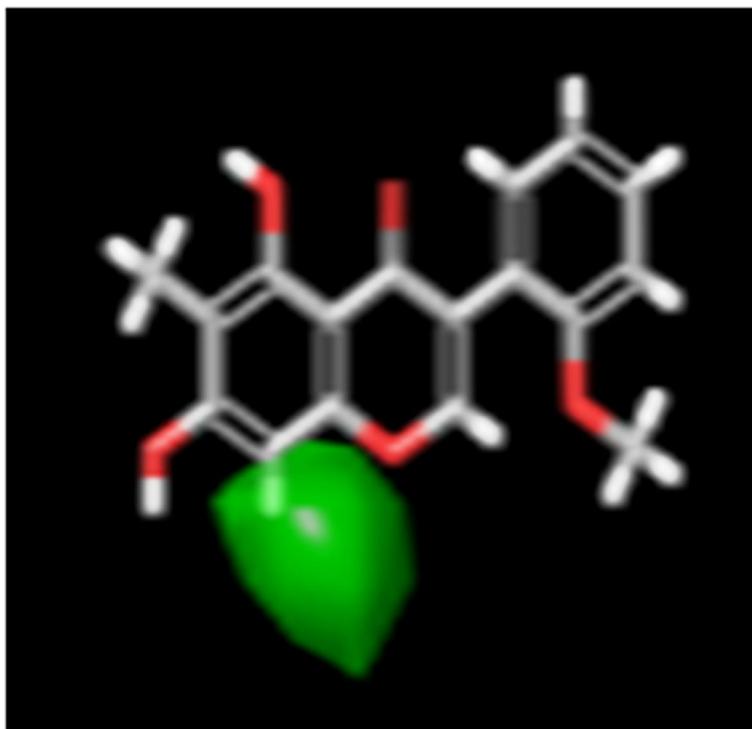
hydrophobic region that lies within receptor and Ligand proximity. For the obtained data, PiCat, ClBr, PhobEnPa, penalties, HB penal, exposed penal, zprot remained at zero, whereas other properties of descriptors were exhibited accordingly.

### Molecular dynamics simulation

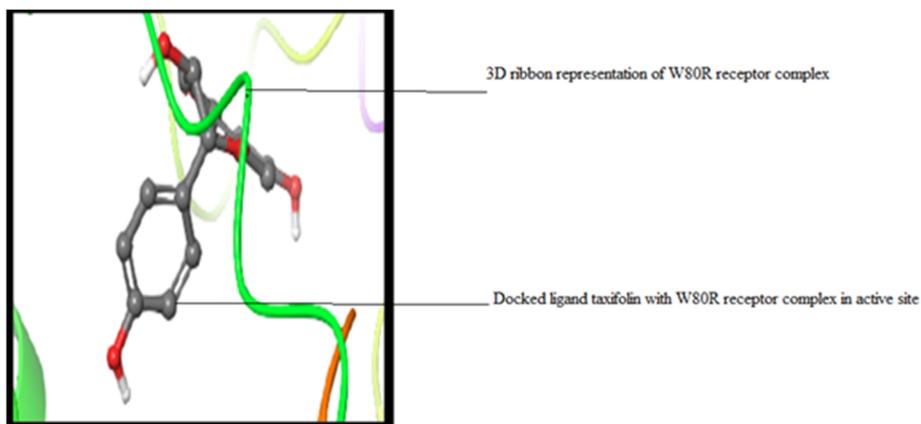
MD simulations were performed to W80R protein-ligand complex with the least binding energy (Fig. 8A). The results of MD trajectories were evaluated by the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) plot which could provide significant insights into understanding structural changes in atomic details. The RMSD is a significant parameter to analyze the equilibrium in MD trajectories, which is estimated for backbone atoms of W80R protein and taxifolin ligand complex. For the W80R protein complex, the fluctuations were raised about 0.3 to 0.4 nm during the initial stage (Fig. 8A). Clear and noticeable deviations were observed in the residues of RMSD values with an increase in time from 200 ps to 600 ps. The majority of residues resulted to attain a stable state at 600 ps between 0.45 nm to 0.5 nm. At the same time, W80R protein-ligand complex fluctuated from 700 ps to 900 ps at 0.4 nm and remained stable between 0.4 nm to 0.45 nm until the end of the simulation [65].

RMSF results were obtained by considering the average of all backbone residues of atoms to inspect the local variations of protein flexibility (Fig. 8B). The fluctuations observed above have an important role in protein complex flexibility and thus affect protein-ligand activity and stability. The high RMSF value shows more flexibility with a maximum level of fluctuation in the residue positions of 355 and 405 at 6 Å of the backbone structure, while the minimum RMSF shows very limited movements. The RMSF graph for the W80R-ligand complex was shown in Fig. 8B. The W80R-ligand complex has attained the amino acid residues at 455 and 500 also shows a fluctuation at 5 Å of RMSF. While at positions 305 and 355 at 4 Å indicate a similar steep-up graph at 5 Å. The amino acid residues between 15, 55, and 105, 155, have shown medial deviation at 3 Å.

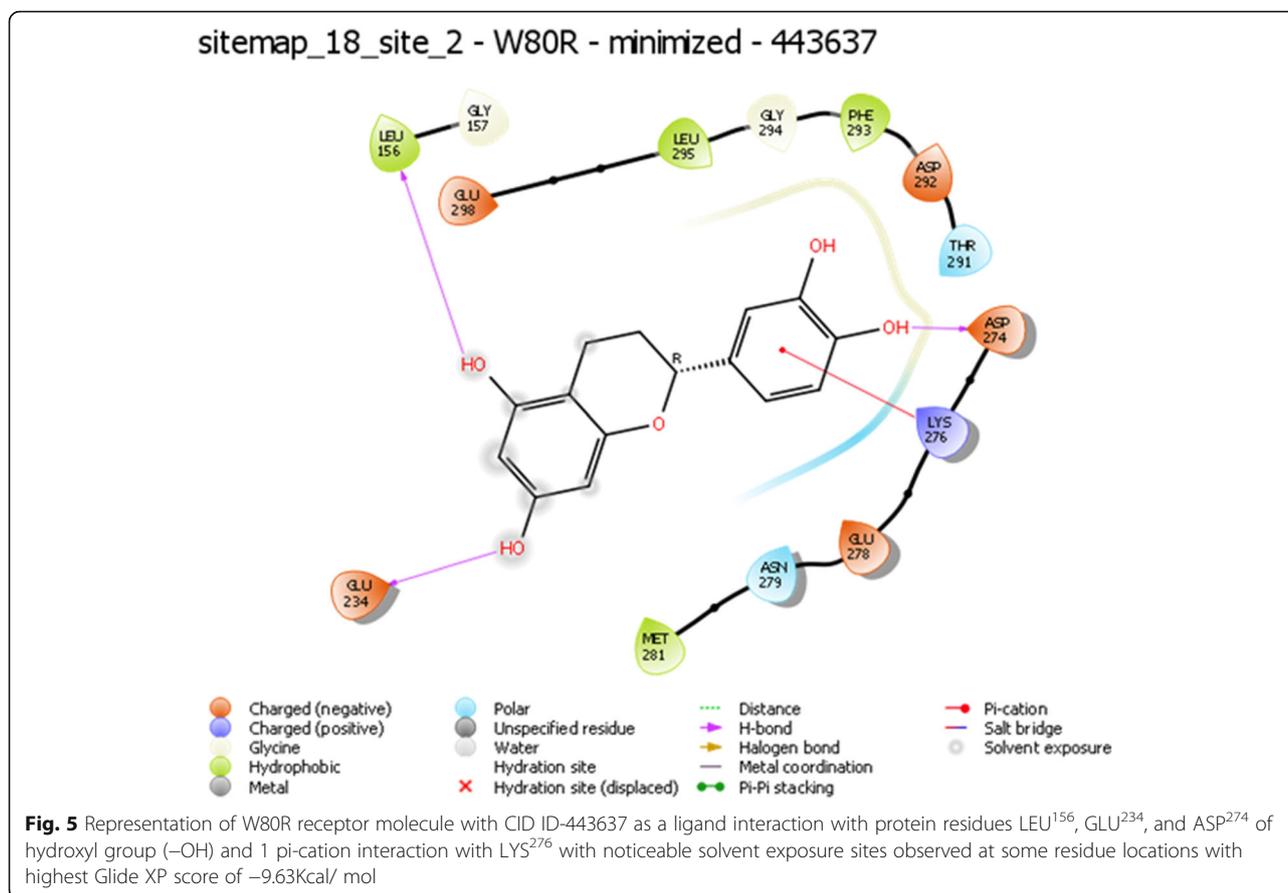
To determine the residue interaction network, RING2.0 software identifies all types of non-covalent interactions in atomic levels which have wide different energies and lengths. The output has been visualized in two different ways (i) interaction network which has been visualized using different labels and (ii) structural contacts using RING\_viz-script for pymol (Fig. 9). The applications of RING 2.0 have a growth in protein folding patterns, domain-domain communication and catalytic activity, inter-intrachain interactions that combine both solvent and ligand atoms. Residue interaction network (RIN) describes the single amino acid as nodes and



**Fig. 3** The CoMFA steric field with 2 Å grid spacing is displayed for the compound CID: 5482167, which depicts hydrophobicity via contour mapping and demonstrates the C7 substituted group was in the solvent accessible hydrophobic pocket. Steric fields are considered as most favorable regions with high glide score



**Fig. 4** A docked complex of W80R protein in ribbon model with inhibitor CID ID 443637 (Taxifolin) at an active site of binding pocket with XP score – 9.63 kcal/mol

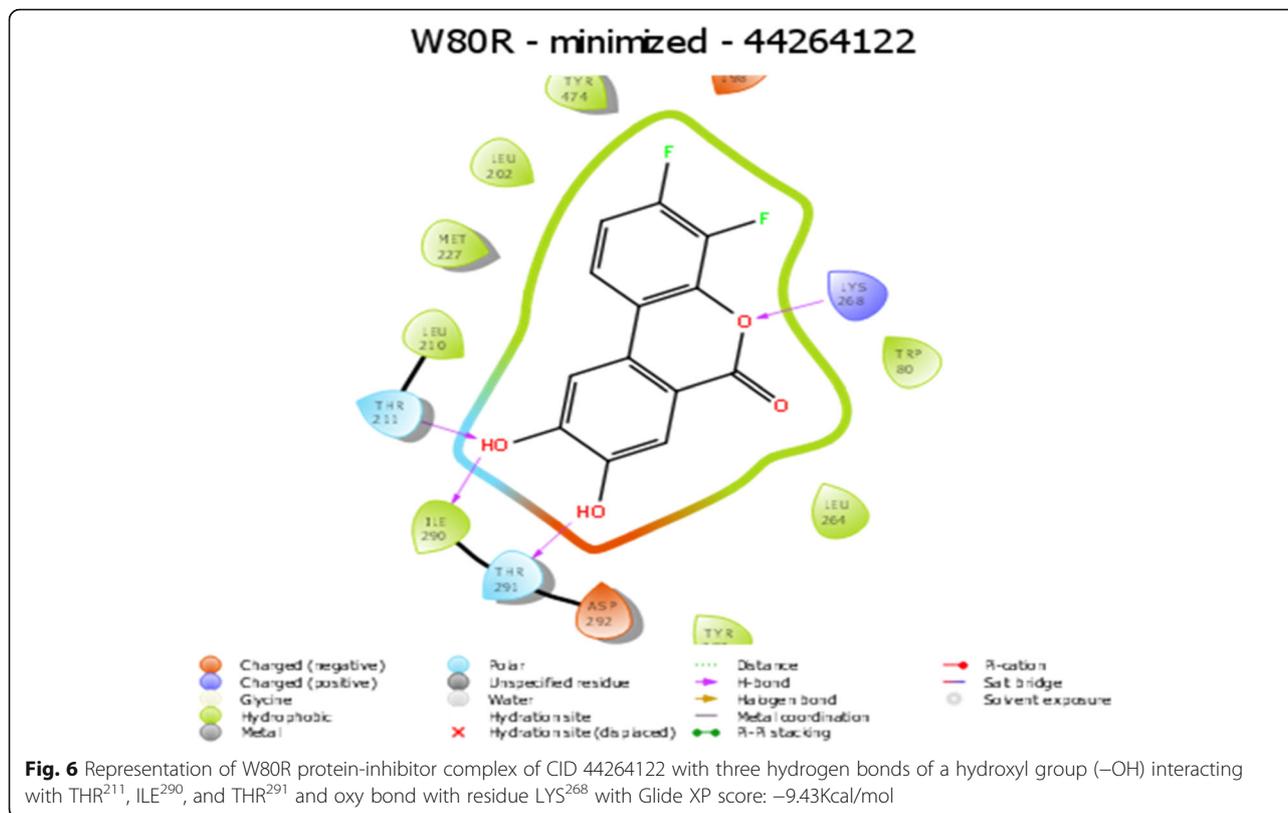


physicochemical properties as edges including covalent and non-covalent bonds. RIN has become common practice to explore the complexity inherent in macromolecular systems [66].

#### Validation of marvel compound with AKT1-inhibitors

This study has compared the marvel compound (taxifolin) with co-crystallised protein-ligand (inhibitors) complex. Among the proposed models for AKT1 protein, the best and high 3D resolution of protein 4GV1 pdb format is considered for better understanding the specificity and potency of the inhibitor when compared other authentic protein ligands further evaluation. This co-crystallised pdb molecule is subjected to auto dock programming software for evaluating binding energy, RMSD values and  $K_i$  (nm). The ligands were chosen and downloaded from RSCB database. The Auto dock program 4.2 [67] has been employed for this purpose which utilizes a semi-empirical free energy. The force field is based on thermodynamic model which takes into account of intermolecular energies. The difference between bound and unbound states of ligands associated with their calculations of binding energies. However, the force field employs the simple method of utilizing

atomic charges. The force field assesses the binding energy of two or more water molecules by pair-wise atom interaction energy along with empirical approach to evaluate the surrounding water molecule. The results depicted that obtained marvel compound shown higher binding energy than co-crystallised ligands. Therefore, it is redocked to check for the validation of model. The docking position accuracy is indicated by the shift in the position of the accuracy of the native ligand from its position in the respective complex of the co-crystallised protein and ligand. The results depicted that the high resolution protein structure 4GV1 of AKT1 is used for redocking purpose with co-crystallised ligands. 4GV1 protein structure has also used to check the obtained marvel compound (taxifolin). The validation studies confirmed the highest binding energy with taxifolin about -13.94 kcal/mol and reference RMSD as 27.43 and  $K_i$  (nm) as 60.57. All the other inhibitors were redocked and checked for RMSD and binding energy and results obtained as shown in the table. This provides the strong evidence that natural flavonoid has high binding energy while other synthetic inhibitors such as CQU, SMH, SM9 fall between -10.54 kcal/mol to -9.33 kcal/mol range of energies. The detailed obtained values are given in the Table 8.



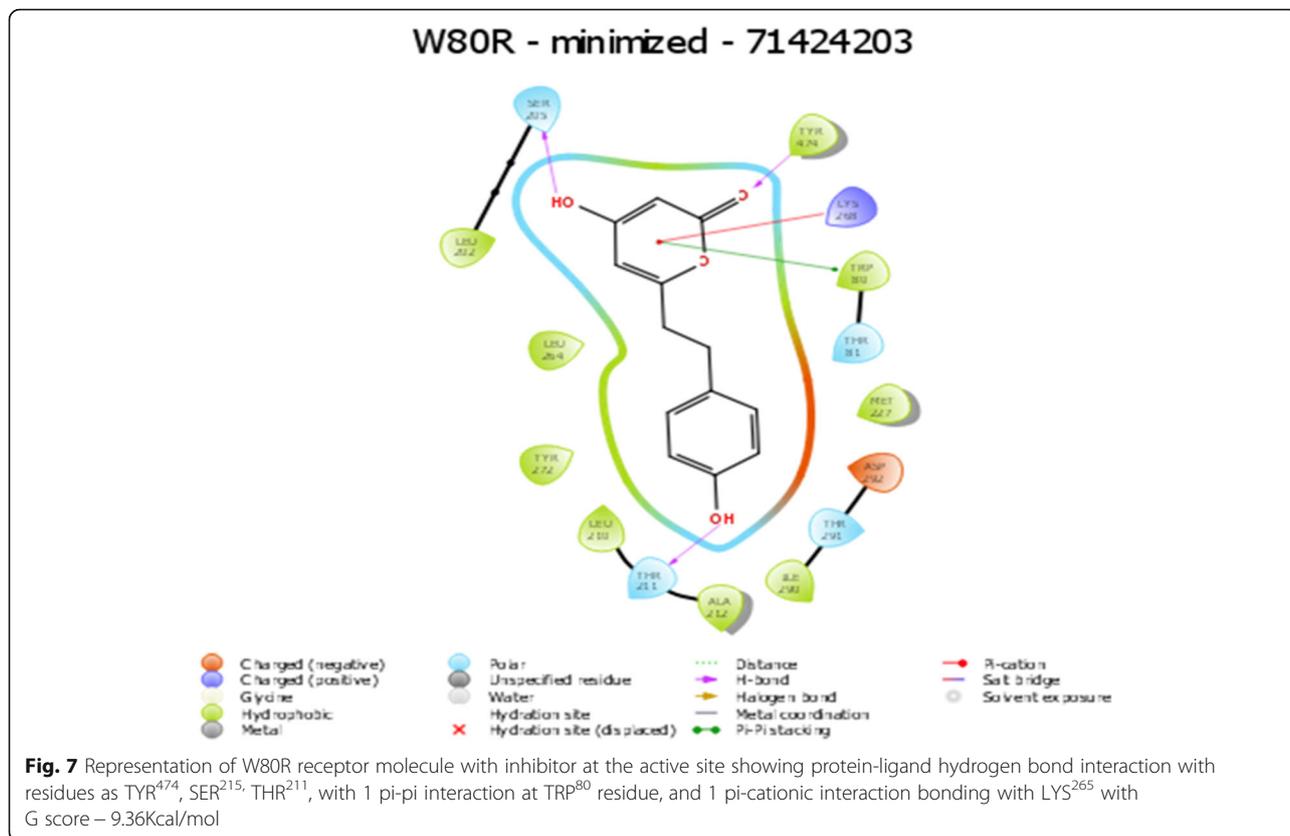
Reference RMSD means the difference rms between the taken structure and the input structure. The most proper way of evaluate geometry is to measure the root means square deviation (RMSD) of a ligand from its reference position after the superimposition of the receptor molecule. It is usually considered best < 1.5–2.0 a for good accuracy

## Discussion

The enigmatic in ovarian cancer is that in nearly 75% of patients, cancer does recurse during the first two years and fail to respond to available therapeutic drugs due to acquired resistance [68, 69] in addition to late diagnosis in advanced clinical stages and metastasis within the peritoneal cavity [70]. Therefore, there is an immediate need to design novel drugs to deal with the existing problem. Numerous studies since a decade have reported that Flavonoids as candidates are meant to block, retard, or reverse the progression of carcinogenesis [71]. Although various studies have been carried out using flavonoids the anticancer mechanisms have not been defined clearly. However, it was found that the flavonoids such as quercetin and silymarin induce anti-cancer mechanisms in the ovarian cancer cell [72, 73]. Consequently, the effects of apigenin, luteolin, and myricetin on ovarian cancer have to uncover the link between potential mechanisms underlying their anticancer effects.

Quercetin inhibits cell proliferation of ovarian cancer cell line of SKOV-3 which correlated with findings of (Yi, 2014) caused on concentration and time-dependent manner [74] showed to inhibit UVB induced skin cancer cell proliferation and induce apoptosis in vivo models upon apigenin treatment. Taxifolin in-vitro studies have been efficient especially in anticancer, antimicrobial activities but leave a strong gap in the *in vivo* studies at the root level.

The marvel compound in the current study was recognized as taxifolin which has potent to exhibit anti-cancer effects on U2OS and Saos-2 in osteosarcoma cell lines by inhibiting the proliferation and disrupting colony formation. In vivo studies exhibit intraperitoneal administration in nude mice bearing U2OS xenograft that resists tumor growth. This potency is known to arrest the G1 phase of the cell cycle in U2OS and Saos-2 cell lines. Taxifolin has known to function by inhibiting colon carcinogenesis by NF- $\kappa$ B mediated Wnt/b catenin signalling through upregulation of Nrf2 pathway while downregulation in genes such as TNF- $\alpha$ , COX-2,  $\beta$ -catenin, and cyclin-D1 was inhibited by NF- $\kappa$ B and Wnt signalling pathway [75]. It is also reported that injection of taxifolin has reduced the proliferative activity on Wistar rats with benign prostatic hyperplasia [76]. Taxifolin also has an excellent report on antiangiogenic effect by new blood vessels and its branches per area of chick

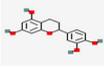
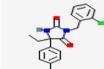
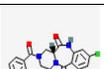
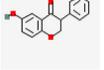
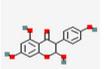
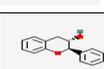
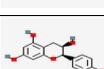
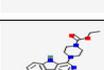
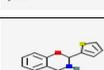
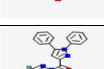
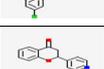
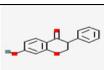
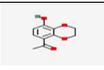


chorioallantoic membrane assay which is inhibited by tube formation on matrigel matrix in the human umbilical vein of endothelial cells which were evaluated against tachyzoites in vitro with IC<sub>50</sub> of 1.39 μg/mL ( $p \leq 0.05$ ) along with pyrimethamine. Taxifolin has known to express an anti-proliferative effect on cancer cell types by inhibiting cell lipogenesis and inhibits the fatty acid synthesis in cancer cell lines which is able to prevent the growth of cancer cells [77].

An extensive animal (rat) study of antioxidant activity on taxifolin acid has shown the decreased lipid peroxidation in the serum and liver levels. The presence of OH groups at position 5th and 7th together with 4-OXO function in the A and C rings were meant for scavenging effect while the O-dihydroxy group in the B ring provided stability [78]. Consequently, in vivo studies on taxifolin induced in apoptosis of HCT116 and HT 29 cells revealed PARP1 overexpression is responsible for ovarian cancer. AKT and catenin proved that down-regulated expression by taxifolin on HCT 116 and HT 29 cells demonstrates a decline in p-AKT and catenin in a dose of 40 μM against DMSO altering in the G2 cell cycle and its regulators [79]. The expression levels of AKT, SKP-2, v-mc avian myelocytomatosis viral oncogene homolog (c-myc) and P-SER<sup>473</sup>, have reduced activity on AKT gene by taxifolin [80]. Although the above-

mentioned experimental outcomes have contributed to diversified pharmacological activities with AKT1 protein, we still lack the detailed and molecular changes with respect to the W80R mutant protein of the AKT1 family. Consequently, the marginal overview of the molecular mechanism and atomic level with W80R mutation has aimed to identify hits for optimization from large data set of compounds from the PubChem database screening of flavonoids in parallel to W80R mutant protein of AKT1 targeting ovarian cancer. Table 1 for the receptor molecule W80R of 480 amino acid sequence provides detailed knowledge about the stability of protein using Protparam tools of ExPASy server. The extensive evaluation of the W80R sequence at the nucleotide level reveals its density, while other parameters such as A-T, C-G rich region, molecular weight, amino acid composition, theoretical pI, aliphatic index, instability index, and GRAVY significantly stand up for stability factor. The most favored region by RAMPAGE server was assessed to be 79.3% (Table 2) with active site binding. Furthermore, the reliability of the protein model has been assessed by 3D or homology modelling. Therefore, the Generation of 3D protein structure from sequence information, in the absence of experimentally determined structures in protein data bank through computational approaches has become the topmost priority in the

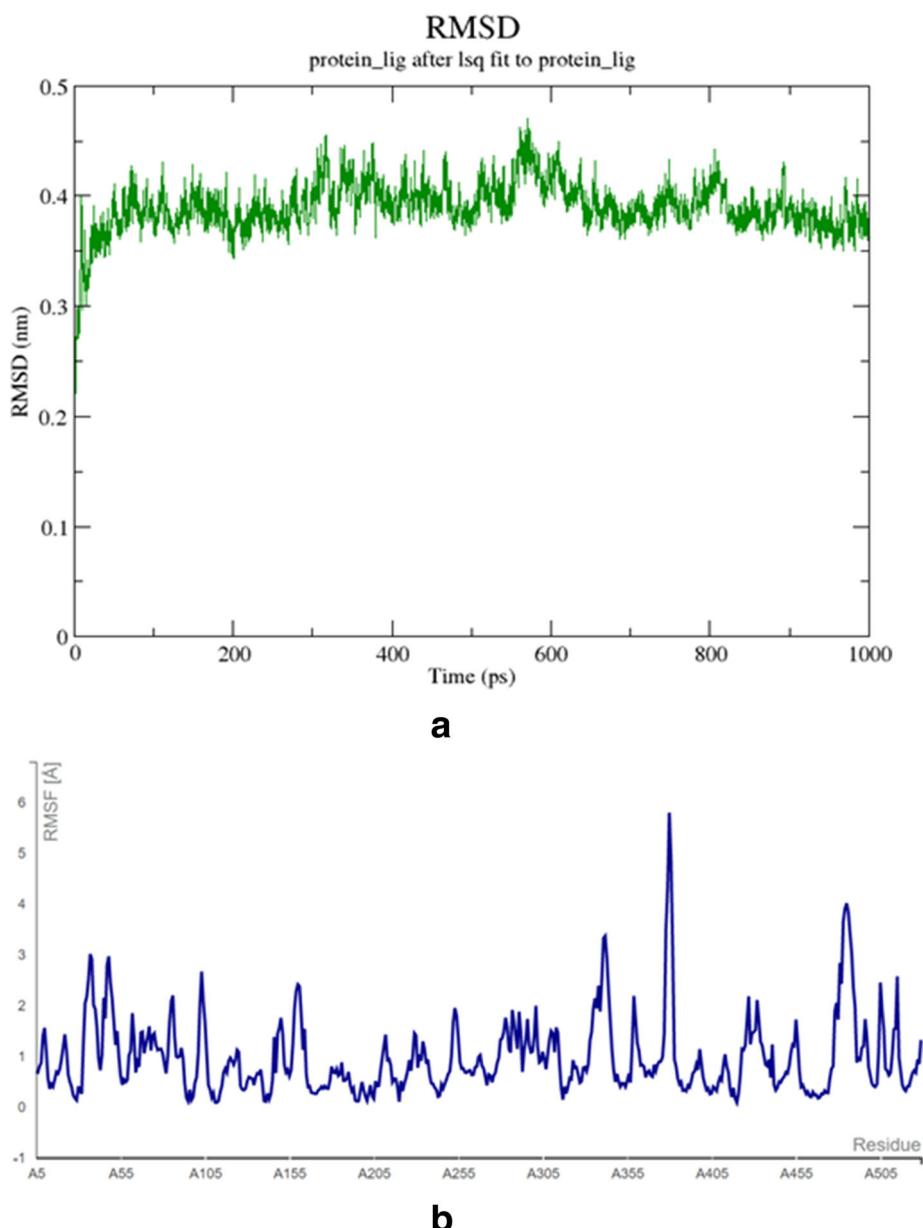
**Table 7** Top-ranked hit compounds of docking with protein W80R obtained using XP visualize

Chemical Structure	CID ID	Dock Score	G Score	Lipophilic EvdW	Phob En	H Bond	Elec tro	Rot Penl	Acti vity
	443637	-9.63	-9.63	-3.22	-2.17	-3.02	0.72	0	0.08
	442641 2	-9.43	-9.43	-3.11	-2.11	-2.11	0.57	0	0.06
	7142420 3	-9.36	-9.36	-3.98	-0.86	-3.2	0.64	0.12	0
	586089	-9.34	-9.34	-3.82	-2.55	-1.04	0.15	0.13	-0.1
	443902	-9.32	-9.32	-3.93	-1.7	-2.05	0.45	0.09	0.31
	847733	-9.26	-9.26	-5.25	-2.7	-0.7	0.25	0.14	0.32
	443639	-9.2	-9.2	-4.95	-2.09	-1.08	-0.7	0.1	0.16
	690180 2	-9.18	-9.18	-4.5	-1.65	-2.4	0.32	0.52	0
	323987 0	-9.15	-9.15	-3.27	-1.97	-0.7	0.18	0.14	0.32
	588032 6	-9.14	-9.14	-4.7	-1.84	-1.87	0.64	0.41	0.85
	401388 8	-8.97	-8.97	-4.24	-1.87	-0.7	0.23	0.14	0.32
	594226 6	-8.9	-8.9	-4.73	-2.65	-1.27	0.32	0.57	0
	586088	-8.71	-8.71	-4.18	-1.98	-0.85	0.43	0.13	0.11
	80337	-8.71	-8.71	-3.45	-1.85	-1.7	0.35	0	0
	229016	-8.66	-8.66	-4.66	-2.31	-0.67	0.65	0.13	0.24

\*G Score-total G score along with the sum of XP terms(G score = a\*vdW + b\*Coul + Lipo + Hbond + Metal + BuryP + RotB + Site, where vdW is van der Waals energy, Coloumb energy, Lipo is lipophilic contact, Hbond is hydrogen bonding, Metal, is metal-binding, BuryPis penalty for buried polar groups, RotBis penalty for freezing rotatable bonds, the site is polar interactions in the active site and a = 0.065 while b = 0.130 were the coefficients of vdW and Coul Dock score: Vanderwaals + Coulombic + HBonds represent the potentiality of bonding. In simple rigid systems, the ligand is searched in a 6 dimensional rotational or translational space to fit in the binding site, which can serve as a marvel compound for drug design [64]. The lipophilic term is derived from the hydrophobic grid potential and the fraction of the total protein-ligand vdWenergy, PhobEn- can be as hydrophobic enclosure reward for the penalty for ligands with large hydrophobic contacts and low hydrogenbond scores phobic penal for the penalty for exposed hydrophobic ligand groups, Rot Penal for the rotatable bond penalty.

scientific community based on structural biology research for several decades [81, 82]. The protein was henceforth evaluated with SAVES server (structural

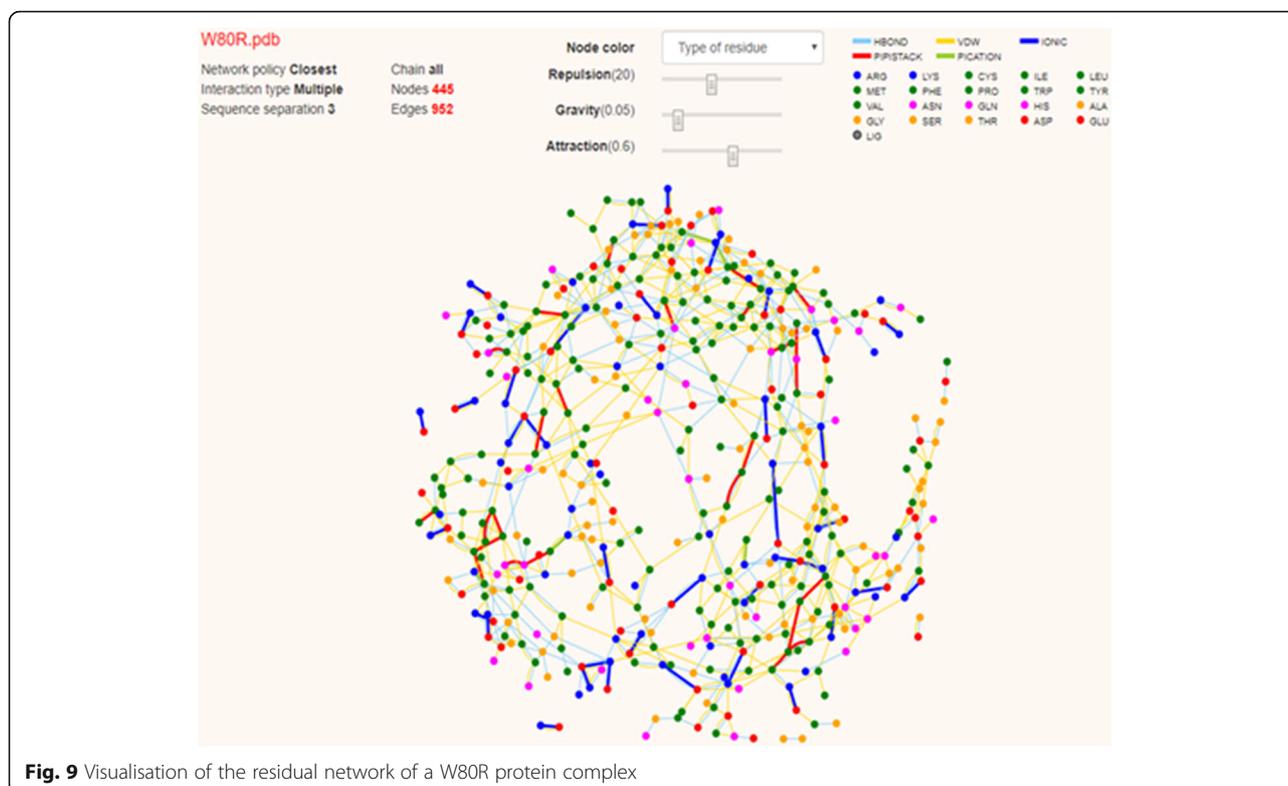
analysis and verification) for quality check, structural refinement through energy minimization in lowest energy state in its stable conformation, followed by ProSA (Fig.



**Fig. 8 (A)** Root mean square deviation (RMSD) of the C-alpha backbone of the W80R protein complex and ligand taxifolin (X-axis time scale in ps and Y-axis in RMSD in nm). **(B)** Root mean square fluctuation (RMSF) for C-alpha backbone atom of a W80R protein complex with ligand taxifolin (X-axis shows amino acid residue number and Y-axis shows RMSF in nm)

1) and superimposition analysis with experimentally determined template structure as well as atoms and RMSD assessment to obtain a high-quality structural model for virtual screening [83, 84]. The predicted score for the 3D homology model of RMSD for the W80R protein was 0.18, the model was considered as the best one for further validation purposes. 3D QSAR studies have been performed with structural similarity to predict the unknown/untested ligands for better potency by correlating mathematical and statistical values. QSAR models can prioritize ideas in virtual screening as well in the

optimization of marvel compounds. Thus it has gained acceptance in *in-silico* drug discovery. The scatter plot QSAR tool (Fig. 2) assessed the molecular fields for the compounds which estimate the stability and establish statistical value to be 0.379 predicting the changes obtained in the training set composition with 92.7 measured higher F indicates more statistical significant regression. The dataset of 44 ligands was classified into test and training models randomly with combined mathematical and statistical approaches for the drug candidate represents phase activity of 358.477% extrapolated



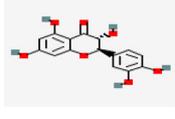
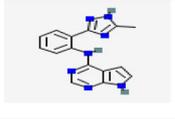
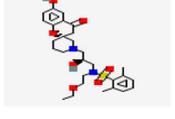
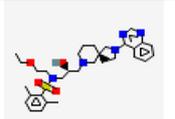
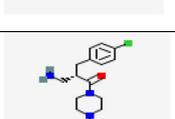
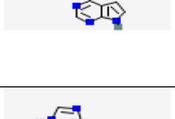
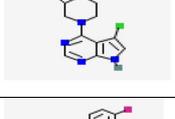
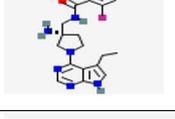
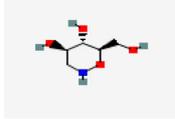
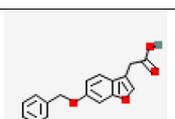
**Fig. 9** Visualisation of the residual network of a W80R protein complex

for 0.458 with the predicted activity of 333.692 and predicted error of  $-24.7856$  which was a good combination as a marvel compound.  $R^2$  (Coefficient of determination) value measured to be 0.822 which means the obtained model accounts for 82.2% of the variance for the observed activity, while standard  $R^2$  falls between 0 and 1 always (Table 6). The increase in a correlation coefficient ( $R^2$ ) as an increase in the number of PLS factors, with the decrease in values of standard deviation (SD) and the increase in the number of variables, allows the user to reduce the error of the fit model. Consequently, the stability tends to increase to 3 factors, and then decrease. Here, the 3–3-model predictions were the least sensitive to the training set composition based on leave 1- out tests. But for the 4 or above factors, the  $R^2$  larger values than stability indicates the over-fit begins at a4 factor. Of the complete training set data, the descriptor Stability indicates the model sensitivity while larger  $R^2$  indicates the over-fit data set. Therefore, it is concluded that obtained data holds quite good for marvel compounds statistically. As per Lipinski's rule of five, a drug is a good molecule if it possesses ADME (absorption, distribution, metabolism, and excretion) properties [48]. All the physicochemical properties and drug-likeness were listed in Tables 3, 4, and 5 consequently; it becomes easy for the marvel compound to enter the mammalian cell to interact with proteins and regulating gene expression in metabolic pathways. The top 10 hits

obtained by molecular docking were further docked into the active binding sites of protein using a sitemap tool of above score 1 and grid generation followed by XP protocol (Table 7). Nevertheless, a contour map is one such tool used in the present study to determine favorable regions based on field-based QSAR which depends on steric, electrostatic, hydrophobicity in solvent-accessible pockets based on least binding energy. This application plays a vital role in combination therapies of multi-drug-resistant conditions as well as in drug discovery.

The evaluated hydrophobicity gives an accurate check for the drug-ability of a compound (Fig. 3). Sitemap tool treats entire protein to locate binding sites whose size, the extent of solvent exposure is assessed based on scoring function by ranks. Active sites are ranked based on ligand propensity of binding measured by their ability to bind tightly for passively absorbed small molecules. Among the predicted combinations, active site amino acid residues of site score 1.128, drug-ability score  $-1.149$ , volume 384.486, and size 179 (Fig. 4) were taken for further analysis. Taxifolin holds good interactions with the binding domain of W80R, highest Glide score of  $-9.63$  kcal/mol with O-H of GLU<sup>234</sup> and H bond LEU<sup>156</sup> and ASP<sup>274</sup> amino acid residues and one pication interaction and one hydrophobic bond with LYS<sup>276</sup> (Fig. 5). The marvel molecule satisfied all the surface area calculations using QIKPROP tool of SASA, FISA, FOSA, PSA and partition coefficient of Qplogpact,

**Table 8** Validation of 4GV1 receptor with co crystallised inhibitors with the obtained marvel compound

Ligand/Compound ID	Chemical structure	Binding energy (Kcal/mol)	Reference RMSD	Ki (nM)
TAXIFOLIN 443637		-13.94	27.43	60.57
CQU 24798741		-10.54	25.88 A	18.72
SMH 50990925		-9.88	34.65	57.11
SM9 51000425		-9.33	26.34	144.54
XM1 46870040		-8.94	37.90	278.63
CQW 24798742		-8.68	39.81	433.71
WFE 25061501		-8.50	37.67	587.37
OXZ 449101		-8.20	40.48	974.09
OR4 139592521		-7.88	28.42	1.67
SMV 737237		-6.29	27.52	24.42

QPlogPw, QPlogPo, QPlogS, ClQPlog, QPlogHER, QPPCaco, QPPMDCK, QPlogKp, wherefore, this inhibitor of the PI3K/AKT pathway has shown diverse aptitudes for anticancer activity in both preclinical and clinical experimental values and also supported through *in-silico* analysis.

The reports by the administration of taxifolin in colorectal cancer cell lines and in the HCT 116 xenograft mouse model had shown excellent antitumor activity. The studies proved that the taxifolin hindered the mRNA expression of  $\beta$ -catenin thus compiling anti-proliferative activity which was arbitrated by PI3K/AKT signal by jamming Wnt/ $\beta$ -catenin signaling transduction through hampering the  $\beta$  expression [79]. The elucidation of suppression by taxifolin on nuclear factor- $\kappa$ B, C-Fos, and mitogen-activated protein kinase also decreased osteoclast specific gene expression including Trap, Mmp-9, Cathepsin K, C-Fos, Nfatc1, and Rank; taxifolin osteoclastogenesis via regulation of many RANKL signaling pathways was also confirmed [85]. Taken together, these studies demonstrated that Wnt/catenin pathway plays a crucial role in ovarian cancer development and this idea also laid a strong platform for the development of targeted curatives.

CID- 44264122 with three hydrogen bonds of a hydroxyl group ( $-\text{OH}$ ) interacting with THR<sup>291</sup>, ILE<sup>290</sup>, and THR<sup>211</sup> where ILE<sup>290</sup> forms oxy bond with residue LYS<sup>268</sup> (Fig. 6) with Glide XP score  $-9.43$  Kcal/mol. CID-71424203 forms the hydrogen bond interactions with residues of TYR<sup>474</sup>, SER<sup>215</sup>, THR<sup>211</sup>, with 1 pi-pi interactions at TRP<sup>80</sup> residue, and 1 pi-cationic interaction bonding with LYS<sup>265</sup> with G score  $-9.36$  Kcal/mol showed good hydrophobic interactions (Fig. 7). The molecular dynamics simulation was performed to obtain the lowest error and data loss. The fluctuations in relative positions of atoms in protein-ligand complex explain the structural stability (RMSD) at 0.45 nm to 0.50 nm between 600 to 800 ps (Fig. 8A). The RMSF has shown a steep up graph at 5Å with a slight medial deviation and not much structural change in protein cavity was observed [86, 87] (Fig. 8B). Finally, the high resolution receptor molecule 4GV1 is used for validation purpose and estimated binding energy for the marvel compound as  $-13.94$  Kcal/mol after redocking. This study supported strong evidence against other synthetic inhibitors found in the database for AKT1 molecule.

Residue interaction networks (RINs) consider single amino acids as nodes and physio-chemical interactions as edges (Fig. 9) representing the protein structure as RINs have become common practice to explore the complexity inherent in macromolecular systems. Henceforth, the taxifolin has been suggested as a drug for human use in clinical trials.

Above all, assessing protein-ligand binding affinity has become the main challenge during early stages of drug

discovery. Machine learning approach has been introduced in contributing to this type of prediction by two approaches. By exploring the experimental structures with binding affinity and thermodynamic data accessed using BindingDb, Binding MOAD, and PDBbind with open sources like SANdRes and Taba, while the second method protein ligand docking simulations. Therefore this combination has outruled classical scoring functions with high predictive performance.

## Conclusions

The mutant forms of the amino acid were found to induce pathological outcomes disrupting the native conformation of a protein. The W80R mutation in the PH domain of AKT1 had been reported to cause ovarian cancer by *in-vitro* studies and recorded in the Cancer genome database. The clinical significance of W80R in ovarian cancer and synthetic drugs used has laid as the platform for this study. Therefore, the modelling of W80R protein has been made for the first time. To examine the detailed molecular mechanism of W80R, we conducted molecular docking along with dynamic simulation studies to understand the stability of the mutant structure, which is known to cause a damaging effect of the mutation.

Furthermore, a rise in RMSD values for stability in trajectory and conformational drifts were observed in W80R protein. The expected result supported the molecular cause in a mutant form which resulted in a gain of ovarian cancer. However, experimental evaluation or *in vivo* studies is recommended for further validation.

## Acknowledgements

The authors acknowledge the financial support provided by the BTISNet, Department of Biotechnology, Government of India, to the Bioinformatics Sub-DIC G. B. Pant University of Agriculture & Technology, Pantnagar.

## Authors' contributions

SK and SMA conceived the idea, SMA and AT designed the study, SMA performed the docking study, SS and DBS done the molecular dynamics. AT, GT, DBS, and SK helped in the critical revision of the manuscript. All authors read and approved the final manuscript.

## Funding

Not applicable.

## Availability of data and materials

Provided with the manuscript.

## Declaration

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

All authors on the research paper have approved the manuscript for submission.

### Competing interests

Authors declare no conflict of interest.

**Author details**

<sup>1</sup>Bioinformatics Sub-DIC, Department of Molecular Biology & Genetic Engineering, College of Basic Science and Humanities, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, Udham Singh Nagar 263145, Uttarakhand, India. <sup>2</sup>Department of Biotechnology, Andhra University, Vishakhapatnam 530003, Andhra Pradesh, India. <sup>3</sup>Department of Computational Biology and Bioinformatics, Jacob Institute of Biotechnology and Bio-Engineering, Sam Higginbottom University of Agriculture, Technology and Sciences, Prayagraj, Uttar Pradesh 211007, India. <sup>4</sup>Department of Biotechnology, Siddharth University, Kapilvastu, Siddharth Nagar 272202, Uttar Pradesh, India. <sup>5</sup>Department of Molecular and Human Genetics, Banaras Hindu University, Varanasi 221005, India.

Received: 2 April 2021 Accepted: 9 July 2021

Published online: 02 November 2021

**References**

- Siegel RL, Miller KD, Jemal A. Cancer statistics. *CA Cancer J Clin*. 2019;69(1):7–34. <https://doi.org/10.3322/caac.21551>.
- Gilks CB, Prat J. Ovarian carcinoma pathology and genetics: recent advances. *Hum Pathol*. 2009;40(9):1213–23. <https://doi.org/10.1016/j.humpath.2009.04.017>.
- Howlader N, et al. SEER Cancer Statistics Review, 1975–2010, National Cancer Institute. Bethesda: National Cancer Institute; 2013.
- Ghoneum A, Said N. PI3K-AKT-mTOR and NFκB Pathways in Ovarian Cancer: Implications for Targeted Therapeutics. *Cancers (Basel)*. 2019;11(7):949.
- Dancey JE. Therapeutic targets - mTOR and related pathways. *Cancer Biol Ther*. 2006;5(9):1065–73. <https://doi.org/10.4161/cbt.5.9.3175>.
- Du-Cuny L. Computational modelling of novel inhibitors targeting the Aktpleckstrin homology domain. *Bioorg Med Chem*. 2009;17(19):6983–92. <https://doi.org/10.1016/j.bmc.2009.08.022>.
- Mundi PS. AKT in cancer: new molecular insights and advances in drug development. *Br J Clin Pharmacol*. 2016;82(4):943–56.
- Martini M. PI3K/AKT signaling pathway and cancer: an updated review. *Ann Med*. 2014;46(6):372–83. <https://doi.org/10.3109/07853890.2014.912836>.
- Hanrahan AJ, Schultz N, Westfal ML, et al. Genomic complexity and AKT dependence in serous ovarian cancer [published correction appears in *Cancer Discov*. 2012 Apr;2(4):376. Janikariman, Manickam [corrected to Janikariman, Manickam]]. *Cancer Discov*. 2012;2(1):56–67. <https://doi.org/10.1158/2159-8290.CD-11-0170>.
- Rodgers SJ. Regulation of PI3K effector signalling in cancer by the phosphoinositide phosphatases. *Biosci Rep*. 2017;37(1):BSR20160432.
- Brasseur K. Chemoresistance and targeted therapies in ovarian and endometrial cancers. *Oncotarget*. 2017;8(3):4008.
- Liu J. The role of mTOR in ovarian neoplasms, polycystic ovary syndrome, and ovarian aging. *Clin Ana*. 2018;31(6):891–8. <https://doi.org/10.1002/ca.23211>.
- Hsu JH, Shi Y, Hu L, Fisher M, Franke TF, Lichtenstein A. Role of the AKT kinase in expansion of multiple myeloma clones: effects on cytokine-dependent proliferative and survival responses. *Oncogene*. 2002;21(9):1391–400. <https://doi.org/10.1038/sj.onc.1205194>.
- Hart R, Prlic A. Universal Transcript Archive Repository. 2015 Version uta\_20180821. San Francisco CA: Github; <https://github.com/biocommons/uta>.
- Laurianne B, et al. A hot-spot of in-frame duplications activates the Oncoprotein AKT1 in juvenile granulosa cell tumors. *EBioMedicine*, Elsevier. 2015;2(5):421–31.
- Tate JG, et al. COSMIC: the catalogue of somatic mutations in Cancer. *Nucleic Acids Res*. 2018;47(D1):D941–7.
- Trautmann M, Cyra M, Isfort I, Jeiler B, Krüger A, Grünwald I, et al. Phosphatidylinositol-3-kinase (PI3K)/Akt signaling is functionally essential in MyxoidLiposarcoma. *Mol Cancer Ther*. 2019;18(4):834–44. <https://doi.org/10.1158/1535-7163.MCT-18-0763>.
- Parikh C, Janakiraman V, Wu W, Foo CK, KJavin NM, Chaudhuri S, et al. Disruption of PH-kinase domain interactions marvels to oncogenic activation of AKT in human cancers. *Proc Natl AcadSci U S A*. 2012;109(47):19368–73. <https://doi.org/10.1073/pnas.1204384109>.
- Jung SH, Kim MS, Lee SH, Park HC, Choi HJ, Maeng L, et al. Whole-exome sequencing identifies recurrent AKT1 mutations in sclerosinghemangioma of lung. *Proc Natl AcadSci U S A*. 2016;113(38):10672–7. <https://doi.org/10.1073/pnas.1606946113>.
- Bryce AH, Egan JB, Borad MJ, Stewart AK, Nowakowski GS, Chanan-Khan A, et al. Experience with precision genomics and tumor board, indicates frequent target identification, but barriers to delivery. *Oncotarget*. 2017; 8(16):27145–54. <https://doi.org/10.18632/oncotarget.16057>.
- Craig WL. The Akt/PKB family of protein kinases: a Review of small molecule inhibitors and Progress towards target validation: a update. *Curr Top Med Chem*. 2010;10(4):458–77.
- Tang MKS, et al. Soluble E-cadherin promotes tumor angiogenesis and localizes to exosome surface. *Nat Commun*. 2018;9(1):2270.
- Cannon MJ, et al. Signaling Circuits and Regulation of Immune Suppression by Ovarian Tumor-Associated Macrophages. *Vaccines (Basel)*. 2015;3(2):448–66.
- Cui J, Qian J, Chow LM, Jia J. "Natural Products Targeting Cancer Stem Cells: A Revisit". *Curr Med Chem*. 2021;28:1–32. <https://doi.org/10.2174/0929867328666210405111913>.
- Yang J, et al. The I-TASSER suite: protein structure and function prediction. *Nat Methods*. 2015;1:7–8.
- Kelly MG, Mor G, Husband A, O'Malley DM, Baker L, Azodi M, et al. Phase II evaluation of phenoxodiol in combination with cisplatin or paclitaxel in women with platinum/taxane-refractory/resistant epithelial ovarian, fallopian tube, or primary peritoneal cancers. *Int J Gynecol Cancer*. 2011;21(4):633–9. <https://doi.org/10.1097/IGC.0b013e3182126f05>.
- Kumar S, et al. Chemistry and biological activities of flavonoids: an overview. *Sci World J*. 2013;162750:1–16.
- Rashidi Z, Khosravizadeh Z, Talebi A, Khodamoradi K, Ebrahimi R, Amidi F. Overview of biological effects of quercetin on ovary. *Phytother Res*. 2021; 35(1):33–49. <https://doi.org/10.1002/ptr.6750>.
- Khurana R, Hate AT, Nath U, Udgaonkar JB. pH dependence of the stability of barstar to chemical and thermal denaturation. *Protein Sci*. 1995;4(6): 1133–44. <https://doi.org/10.1002/pro.5560040612>.
- Lovell SC, Davis IW, Arendall WB III, de Bakker PIW, Word JM, Prisant MJ, et al. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins*. 2003;50:437–50.
- Bitencourt-Ferreira G, de Azevedo WF Jr. Homology modeling of protein targets with MODELLER. *Methods Mol Biol*. 2019;2053:231–49. [https://doi.org/10.1007/978-1-4939-9752-7\\_15](https://doi.org/10.1007/978-1-4939-9752-7_15).
- Bitencourt-Ferreira G, de Azevedo Junior WF. Electrostatic Potential Energy in Protein-Drug Complexes. *Curr Med Chem*. 2021;28(24):4954–71. <https://doi.org/10.2174/0929867328666210201150842>.
- Webb B, Sali A. Comparative Protein Structure Modelling Using MODELLER. *Curr Protoc Bioinformatics*. 2014;47(56):1–32.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK a programme to check the stereochemical properties of protein. *J Appl Crystallogr*. 1993;26(2):283–91. <https://doi.org/10.1107/S002188982009944>.
- Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 1993;9:1511–9.
- Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*. 2007;35(Web Server issue):W407–10.
- Petterson EF, et al. UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;13:1605–12.
- Sayers EW, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2020;48(D1):D9–D16.
- Friesner RA, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem*. 2004;47:1739–49.
- Jorgensen WL, Tirado-Rives J. The OPLS Potential function for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem Soc*. 1988;110(6):1657–66.
- Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J Chem Theory Comput*. 2010;6(5): 1509–19. <https://doi.org/10.1021/ct900587b>.
- Schrödinger Release SiteMap, Schrödinger, LLC, New York, NY, 2020.
- Chen YC. Beware of docking! *Trends Pharmacol Sci*. 2015;36(2):78–95.
- Yuriev E, Holien J, Ramsland PA. Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. *J Mol Recognit*. 2015;28(10):581–604. <https://doi.org/10.1002/jmr.2471>.
- Zhan W, Li D, Che J, Zhang L, Yang B, Hu Y, et al. Integrating docking scores, interaction profiles and molecular descriptors to improve the accuracy of molecular docking: toward the discovery of novel Akt1

- inhibitors. *Eur J Med Chem.* 2014;75:11–20. <https://doi.org/10.1016/j.ejmech.2014.01.019>.
46. Schrödinger Release Glide, Schrödinger, LLC, New York, NY, 2020.
  47. Schrödinger Release:Maestro, Schrödinger, LLC, New York, NY, 2020.
  48. Lipinski CA. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 2001;46(1–3):3–26. [https://doi.org/10.1016/S0169-409X\(00\)00129-0](https://doi.org/10.1016/S0169-409X(00)00129-0).
  49. Shihab HA. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Human Genomics.* 2014;8(1):1–11.
  50. Schuler, et al. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J Comput Chem.* 2001;22:1205–18.
  51. Hess B. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput.* 2008;4(3):435–47. <https://doi.org/10.1021/ct700301q>.
  52. Schuttelkopf AW, Van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr.* 2004;60(8):1355–63.
  53. Perera L, et al. A smooth particle mesh Ewald method. *J Chem Phys.* 1995; 103(19):8577–93.
  54. Van Der Spoel D, et al. GROMACS: fast, flexible, and free. *J Comput Chem.* 2005;16:1701–18.
  55. Mamounas E, Wieand S, Wolmark N, Bear HD, Atkins JN, Song K, et al. Comparative efficacy of adjuvant chemotherapy in patients with Dukes' B versus Dukes' C colon cancer: results from four National Surgical Adjuvant Breast and bowel project adjuvant studies (C-01, C-02, C-03, and C-04). *J Clin Oncol.* 1999;17(5):1349–55. <https://doi.org/10.1200/JCO.1999.17.5.1349>.
  56. Rigano D, et al. Antioxidant Flavonoids and Isoflavonoids from Rhizomes of *Iris pseudopumila*. *Planta Medica.* 2007;73:93–6.
  57. Boseetti C, Gallus S, la Vecchia C. Aspirin and cancer risk: an updated quantitative review to 2005. *Cancer Causes Control.* 2006;17(7):871–88. <https://doi.org/10.1007/s10552-006-0033-7>.
  58. Gonzalez FD, et al. Urinary 6-sulfatoxymelatonin and total antioxidant capacity increase after the intake of a grape juice cv. Tempranillo stabilized with HHP. *Food Funct.* 2012;3(1):34–9. <https://doi.org/10.1039/C1FO10146C>.
  59. Karthikeyan S, et al. Chemopreventive potential of chrysin in dimethyl benzanthracene induced hamster buccal pouch carcinogenesis. *Int J Nutr Pharmacol Neurol Dis.* 2013;3(4):46–53.
  60. Gomeni R, Goyal N, Bressolle F, Fava M. A novel methodology to estimate the treatment effect in presence of highly variable placebo response. *Neuropsychopharmacol.* 2015;40(11):2588–95. <https://doi.org/10.1038/npp.2015.105>.
  61. Zhong W, Gallivan JP, Zhang Y, Li L, Lester HA, Dougherty DA. From ab initio quantum mechanics to molecular neurobiology: a cation- $\pi$  binding site in the nicotinic receptor. *Proc Natl Acad Sci USA.* 1998;95(21):12088–93. <https://doi.org/10.1073/pnas.95.21.12088>.
  62. Sussman JL, et al. Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science.* 1991; 253(5022):872–9.
  63. Dougherty DA. Cation- $\pi$  interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. *Science.* 1996;271(5246):163–8. <https://doi.org/10.1126/science.271.5246.163>.
  64. Alberg DG, Schreiber SL. Structure-based design of a cyclophilin–calcineurin bridging ligand. *Science.* 1993;262(5131):248–50. <https://doi.org/10.1126/science.8211144>.
  65. Cheng S, Niv MY. Molecular dynamics simulations and elastic network analysis of protein kinase B (Akt/PKB) inactivation. *J Chem Inf Model.* 2010; 50(9):1602–10. <https://doi.org/10.1021/ci100076j>.
  66. Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex networks. *Nature.* 2000;406(6794):378–82. <https://doi.org/10.1038/35019019>.
  67. Morris, et al. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem.* 2009;16:2785–91.
  68. Banno K, Yanokura M, Iida M, et al. Application of microRNA in diagnosis and treatment of ovarian cancer. *Biomed Res Int.* 2014;2014:232817. <https://doi.org/10.1155/2014/232817>.
  69. Stronach EA, Cunnea P, Turner C, et al. The role of interleukin-8 (IL-8) and IL-8 receptors in platinum response in high grade serous ovarian carcinoma. *Oncotarget.* 2015;6(31):31593–603. <https://doi.org/10.18632/oncotarget.3415>.
  70. Kellenberger, L D et al. The role of dysregulated glucose metabolism in epithelial ovarian cancer. *Journal of oncology.* 2010;2010:514310. <https://doi.org/10.1155/2010/514310>.
  71. Atawodi SE, et al. Evaluation of the polyphenol composition and antioxidant activity of African variety of *Dacryodes edulis* (G.Don) H. J Lam fruit. *J. Med. Food.* 2009;12:1321–5.
  72. Cho HJ, Suh DS, Moon SH, Song YJ, Yoon MS, Park DY, et al. Silibinin inhibits tumor growth through downregulation of extracellular signal-regulated kinase and Akt in vitro and in vivo in human ovarian cancer cells. *J Agric Food Chem.* 2013;61(17):4089–96. <https://doi.org/10.1021/jf400192v>.
  73. Fan L, Ma Y, Liu Y, Zheng D, Huang G. Induces cell cycle arrest and apoptosis in ovarian cancer cells. *Eur J Pharmacol.* 2014;743:79–88. <https://doi.org/10.1016/j.ejphar.2014.09.019>.
  74. Czyz JZ, et al. Flavonoid apigenin inhibits motility and invasiveness of carcinoma cells in vitro. *Int. J. Cancer.* 2005;114:12–8.
  75. Manigandan KD. Kaphle Taxifolin curbs NF- $\kappa$ B-mediated Wnt/ $\beta$ -catenin signaling via up-regulating Nrf2 pathway in experimental colon carcinogenesis. *Biochimie.* 2015;119:103–12.
  76. Borovskaya TG, Krivova NA, Zaeva OB, Fomina TI, Kamalova SI, Poluektova ME, et al. Dihydroquercetin effects on the morphology and antioxidant/prooxidant balance of the prostate in rats with sulphide-induced benign hyperplasia. *bull. Exp Biol Med.* 2015;158(4):513–6. <https://doi.org/10.1007/s10517-015-2797-9>.
  77. Brusselmans, et al. Induction of Cancer Cell Apoptosis by Flavonoids is Associated with Their Ability to Inhibit Fatty Acid Synthase Activity. *J Biol Chem.* 2005;280(7):5636–45.
  78. Salah N, Miller NJ, Paganga G, Tijburg L, Bolwell GP, Riceevans C. Rice-Evans polyphenolic flavanols as scavengers of aqueous phase radicals and as chain-breaking antioxidants. *Arch. Biochem Biophys.* 1995;322(2):339–46. <https://doi.org/10.1006/abbi.1995.1473>.
  79. Razak S, et al. Taxifolin, a natural flavonoid interacts with cell cycle regulators causes cell cycle arrest and causes tumor regression by activating Wnt/ $\beta$ -catenin signaling pathway. *BMC Cancer.* 2018;18:1043.
  80. Chen X, et al. Plant flavonoid taxifolin inhibits the growth, migration and invasion of human osteosarcoma cells. *Mol. Med Rep.* 2018;17:3239–45.
  81. McGuffin, et al. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000;16(4):404–5.
  82. Hekkelman ML, Te Beek TA, Pettifer SR, Thorne D, Attwood TK, Vriend G. WWIWS: a protein structure bioinformatics Web service collection. *Nucleic Acids Res.* 2010;38(Web Server issue):W719–W723. <https://doi.org/10.1093/nar/gkq453>.
  83. Bagaria A, et al. Protein structure validation by generalized linear model root-mean-square deviation prediction. *Protein Sci.* 2012;2:229–38.
  84. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 2007;35(Web Server issue):W407–W410. <https://doi.org/10.1093/nar/gkm290>.
  85. Cai C, et al. Effects of Taxifolin on Osteoclastogenesis in vitro and in vivo. *Front Pharmacol.* 2018;12(9):1286.
  86. Pantavos A, et al. Total dietary antioxidant capacity, individual antioxidant intake and breast cancer risk: the Rotterdam study. *Int. J. Cancer.* 2015;136: 2178–86.
  87. Mourouti N, et al. Breast cancer: a systematic review. *Int. J. Food Sci Nutr.* 2015;66:1–42.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

