

# SCIENTIFIC REPORTS



OPEN

## *In-Silico* Drug discovery approach targeting receptor tyrosine kinase-like orphan receptor 1 for cancer treatment

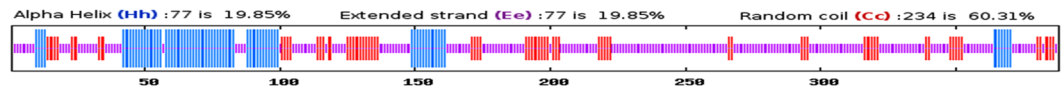
Onkar Nath<sup>1</sup>, Archana Singh<sup>2</sup> & Indrakant K. Singh<sup>3,4</sup>

Receptor tyrosine kinases (RTK) are important cell signaling molecules that influence many cellular processes. Receptor tyrosine kinase such as orphan receptor 1 (Ror1), a surface antigen, is a member of the RTK family of Ror, which plays a crucial role in cancers that have high-grade histology. As Ror1 has been implicated to be a potential target for cancer therapy, we selected this protein for further investigation. The secondary and tertiary structure of this protein was determined, which revealed that this protein contained three  $\beta$ -sheets, seven  $\alpha$ -helices, and coils. The prediction of the active site revealed its cage-like function that opens for ligand entry and then closes for interacting with the ligands. Optimized ligands from the database were virtually screened to obtain the most efficient and potent ones. The screened ligands were evaluated for their therapeutic usefulness. Furthermore, the ligands that passed the test were docked to the target protein resulting in a few ligands with high score, which were analyzed further. The highest scoring ligand, Beta-1, 2,3,4,6-Penta-O-Galloyl-D-Glucopyranose was reported to be a naturally occurring tannin. This *in silico* approach indicates the potential of this molecule for advancing a further step in cancer treatment.

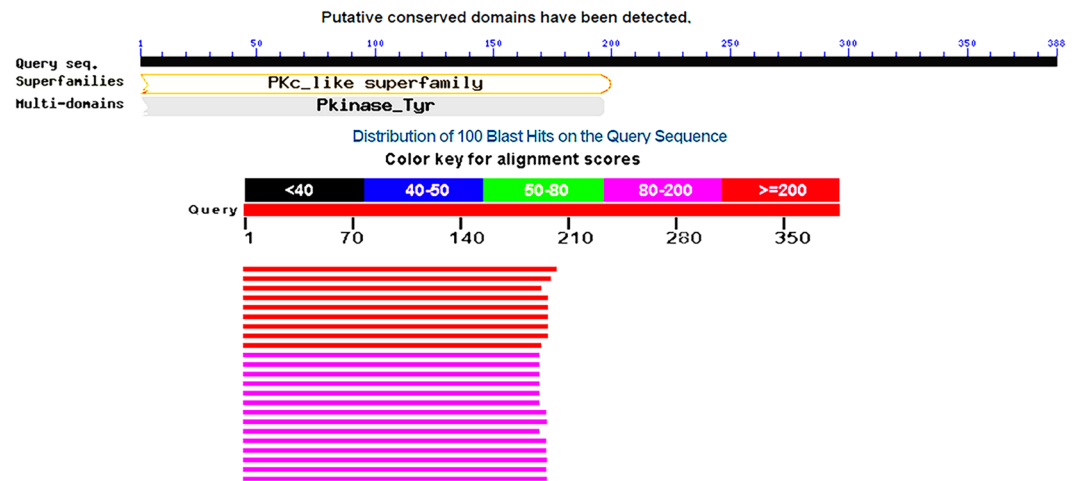
Ror1, a member of RTK family, is an orphan-receptor tyrosine-kinase-like surface antigen, which is primarily expressed during the early stages of embryogenesis. Ror1 is evolutionarily conserved among different species<sup>1–5</sup>. Mutations in human Ror2 have been implicated in certain congenital skeletal defects including shortened or missing digits and a form of short-limbed dwarfism<sup>6–8</sup> but Ror1 mutations have not been reported in any human disease. During mouse development, Ror1 protein is known to play an essential role<sup>9</sup>. Ror1 protein possess an extracellular immunoglobulin-like domain, a cysteine-rich Frizzled domain, and a membrane-proximal Kringle domain. In addition, Ror1 also possesses an intracellular portion with tyrosine kinase domain, two serine/threonine-rich domains and a proline-rich domain<sup>1,2,10</sup>. Notably, Ror1 lacks several key amino acids, shedding doubt on the actual enzymatic function of it<sup>11</sup>. Biochemical assays revealed that Ror1 is a pseudo-kinase that is devoid of catalytic activity<sup>11</sup>. Wnt5a acts as a potential ligand for Ror1 and Ror2<sup>2,12,13</sup> and interaction between Ror1 and Ror2 is required for Wnt5a signaling, which promotes leukemia chemotaxis and proliferation<sup>14</sup>.

Although, Ror1 does not express itself virtually in all normal adult tissues, it re-expresses in many tissues during some B-cell malignancies, and various cancer cell lines<sup>6,8,15,16</sup>. Ror1 was significantly more expressed in various tumours such as acute lymphocytic leukemia, renal carcinoma, breast cancer, lung cancer, adenocarcinoma and melanoma<sup>16–21</sup>. Ror1 has also been recognized as potential biomarker for lung adenocarcinoma<sup>22</sup>. Recent studies have reported the presence of natural humoral and cellular immunity against Ror1 in chronic lymphocytic leukemia (CLL) patients<sup>23</sup> and expression of high levels of Ror1 may promote cancer cell activation and survival enhancing disease progression in patients suffering from CLL<sup>24</sup>. Moreover, Ror1 has also been suggested to be associated with epithelial-to-mesenchymal transition (EMT) during embryogenesis and in cancer metastasis, maintaining the undifferentiated features of stem cells<sup>25,26</sup>. Some patients treated with vaccines of autologous

<sup>1</sup>Jawaharlal Nehru University, SCIS, New Delhi, 110067, India. <sup>2</sup>Department of Botany, Hans Raj College, University of Delhi, Delhi, 110007, India. <sup>3</sup>Molecular Biology Research Lab, Department of Zoology, Deshbandhu College, University of Delhi, Kalkaji, New Delhi, 110019, India. <sup>4</sup>Present address: Department of Entomology, University of Kentucky, S-225 AG. Science - North, Lexington, KY, 40546-0091, United States. Onkar Nath and Archana Singh contributed equally to this work. Correspondence and requests for materials should be addressed to I.K.S. (email: [iksingh@db.du.ac.in](mailto:iksingh@db.du.ac.in))



**Figure 1.** Secondary structure analysis of the Ror1 protein.



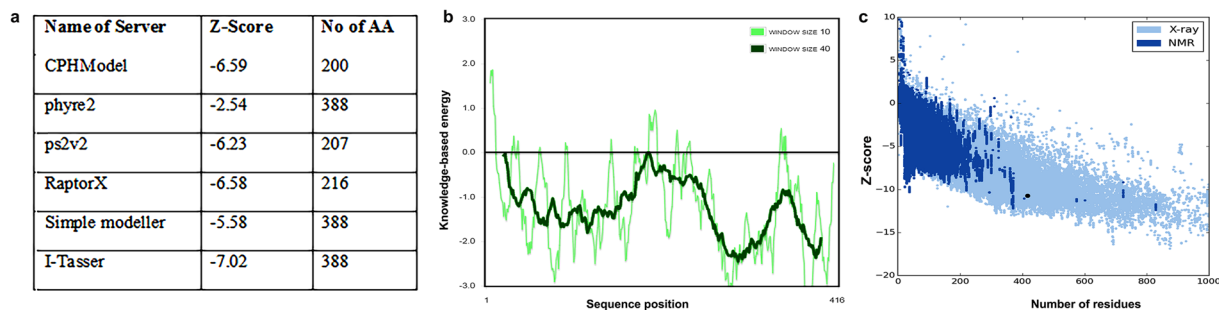
**Figure 2.** Template search using BLASTp tool.

Name	Volume [Å <sup>3</sup> ]	Surface [Å <sup>2</sup> ]	Lipo surface [Å <sup>2</sup> ]	Depth [Å]	Drug Score
P0	870.66	1010.30	701.62	23.83	0.84
P1	834.24	1092.71	801.51	16.42	0.80
P2	620.35	600.22	427.95	17.60	0.80
P3	584.77	570.70	329.67	14.97	0.74
P4	399.36	630.28	374.36	17.24	0.71
P5	225.86	381.32	263.09	8.65	0.37
P6	175.81	338.94	188.84	10.70	0.37
P7	140.54	371.06	253.04	6.81	0.19
P10	105.15	248.34	204.86	7.61	0.19
P8	127.81	192.05	125.80	6.07	0.15
P9	107.52	289.59	191.98	6.45	0.14

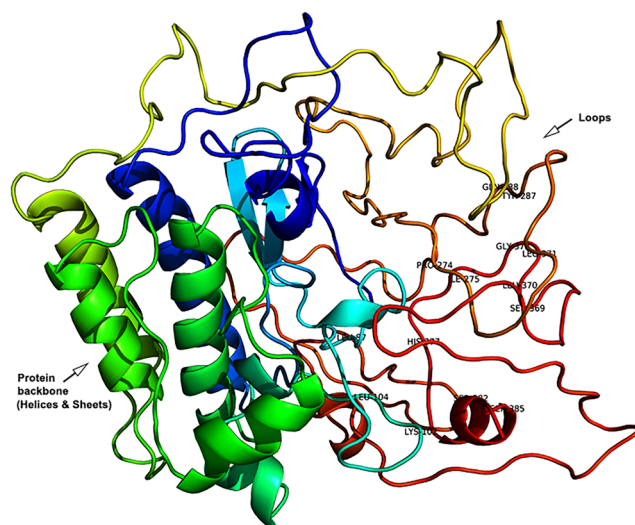
**Table 1.** Active Site analysis of the Protein Ror1 using DogSiteScorer.

leukemia cells genetically engineered to promote anti-leukemia immune responses generated auto-antibodies specific for Ror1 that did not react with non-tumor tissues showing that this receptor is specific to cancer cells<sup>6</sup>. These receptors may enhance chemoresistance, and its knockdown may sensitize these cells to cisplatin<sup>27</sup>. Primary cancers in which Ror1 is up-regulated, high levels of phosphorylated AKT/PKB (a serine/threonine-specific protein kinase B) (p-AKT) and phosphorylated cAMP response element binding-factor (p-CREB) are also expressed<sup>28</sup>. The association between Ror1 with activated AKT has also been reported<sup>21</sup>, which suggests that Ror1 could associate with epidermal growth factor receptor, thereby, enhancing signaling in response to relevant ligands. Ror1 could enhance the survival of tumor cells by either kinase-dependent or kinase-independent pathways.

Many studies support the notion that Ror1 plays a functional role in promoting tumor cell growth and suggest that it may be a potential target for diagnosis and development of therapies against a variety of different human cancers<sup>10,28</sup>. In this study, we have investigated a truncated Ror1 ('t-Ror1'), as not much information regarding this isoform of Ror1 (a 2373 bp transcript encoding 388 aa) is available. 't-Ror1' is identical with the cytosolic, C-terminal region of Ror1 but lacks the transmembrane and the entire extracellular domain. It has been demonstrated that its mRNA level is up-regulated in fetal and adult human CNS, in human leukemia, lymphoma cell lines, and in a variety of human cancers derived from neuroectoderm<sup>29</sup>. Our study provides clues to a potential ligand, Beta-1,2,3,4,6-Penta-O-Galloyl-D-Glucopyranose, which is a naturally occurring tannin, and can inhibit the activity of Ror1. We also studied the dynamics of its structure, which helps in understanding the nature of the molecule. The major role of this protein is in cancer development and proliferation, and its role as a gateway for cancer indicates that the ligands can prove to be a remedy for the disease.



**Figure 3.** (a) Z-score and number of amino acids indicating accuracy and completeness of predicted structure (b) Z-Score plot of model predicted by I-Tasser with structures available in database (c), Energy plot for amino acid residues for the 3D-structure of I-Tasser.



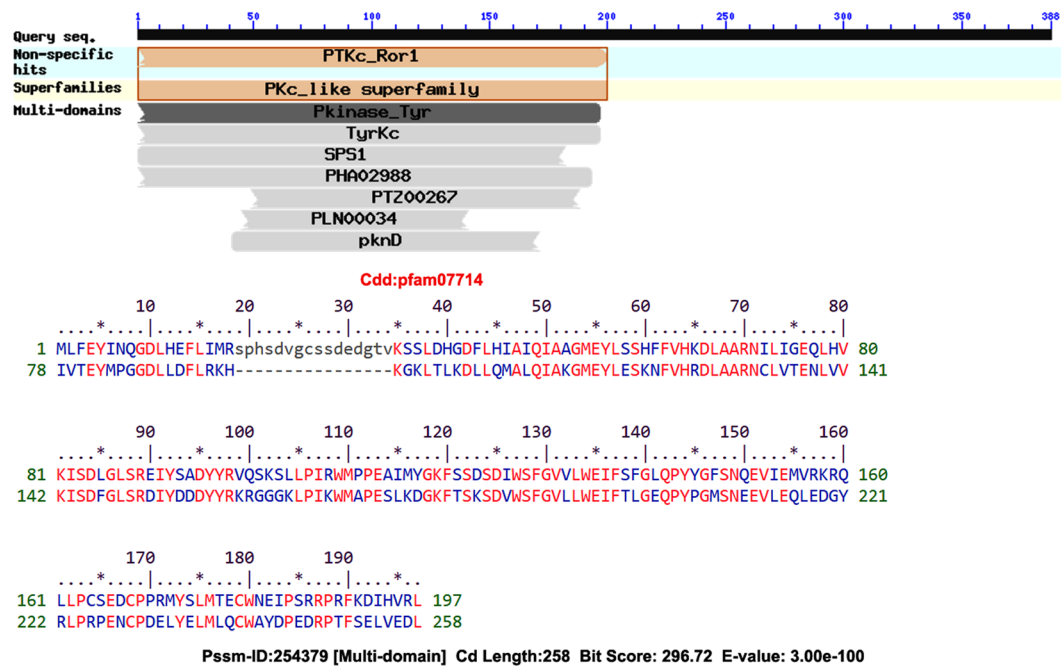
**Figure 4.** 3D-Structure of Ror1 protein and its active sites.

## Results

**Secondary and tertiary structure analysis.** The Ror1 protein of *Homo sapiens* is a 388 amino acid long sequence, which was retrieved from the NCBI protein database (<https://www.ncbi.nlm.nih.gov/protein>). The GenBank ID for Ror1 is AAC50714.1<sup>29</sup> listed as tyrosine kinase t-Ror1. The secondary structure of the protein was analyzed by PROMOTIF tool<sup>30</sup>. The sequence was predicted to contain four strands, seven alpha helices and two  $3_{10}$  helices. Disulphide bridges were found between cysteine 164 and cysteine 168. The secondary structure predicted by the GOR4<sup>31</sup> method contained 77 amino acids in helix and 77 in sheet regions (as shown in Fig. 1); most of the regions are predicted to form coils. Presence of a higher percentage of coils in a protein indicates a flexible structure, which is key to molecular interactions with substrates and ligands. Due to unavailability of the 3D structure of the target protein on RCSB PDB<sup>32</sup> and SCOP<sup>33</sup> databases, the 3D structure of this protein was predicted from the primary sequence. In order to acquire a more accurate structure, different tools including CPHModel<sup>34</sup>, phyre2<sup>35</sup>, ps2v2<sup>36</sup>, RaptorX<sup>37</sup>, Modeller<sup>38</sup> and I-Tasser<sup>39</sup> were used for structure prediction<sup>40</sup>. A search for template (homologous protein sequences with known structures) resulted in sequences that covered only, approximately, 50% of initial region of the sequence (Fig. 2).

Thus, it was necessary to check the accuracy of the predicted structures by comparing them on the basis of different parameters. Z-Score is considered as a standard parameter for quality assessment. To check the quality of the predicted structures, Prosa Server<sup>41,42</sup> was used (see Fig. 3a–c). As the structure predicted by I-Tasser<sup>39,43,44</sup> had the lowest Z-score and it also modelled the most complete structure this was considered as the best predicted 3D-structure. The SAVES server<sup>45</sup> was used to analyze other parameters of the protein 3D-Structure and predicted that 97.16% of the residues had an averaged 3D-1D score greater than or equal to 0.2 (For detail overview, see Supplementary Fig. 1a and b). As indicated in Supplementary Fig. S1b, Procheck<sup>46</sup> analysis predicted 67.4% of the residues to be in the most favorable region whereas 27.3% were in additionally allowed, 3.8% were in generously allowed and 1.6% (i.e. 5 amino acid residues) were in disallowed regions. The RMSD value was predicted to be 1.5 angstroms from the native structure.

Before using this structure for further analysis, it was prepared and optimized using Schrodinger's Protein Preparation wizard<sup>47</sup> (shown in Fig. 4). The structure was found to be stable, and no unstable or sterically



**Figure 5.** Domain Analysis by CD Search indicating similarity to tyrosine kinase domain.

200 ≤ Molecular Weight ≤ 500	0 < Rings < 6
Rotatable bonds ≤ 12	MW > 100.0
35 > Carbon Atoms > 3	0 < Charges < 3
-2 < Formal Charge < 2	Max Size Ring < 18
15 > Hetero Atoms > 1	30 > Rigid Bonds > 0
180.0 > tPSA > 0.0	1.1 > Ratio H/C > 0.0
-3.0 < LogP < 6.0	11 > Rotational Bonds > 0
0 < HBA < 12	No Lipinski Violations < None
HBD > 0	HBD > 0

**Figure 6.** Parameters used to screen ligands based on drug likeliness.

disallowed regions were noted except the steric clash between OD2 (ASP 66) and OH (TYR 96), which is due to their location in an intersecting field i.e. at a distance less than the allowed.

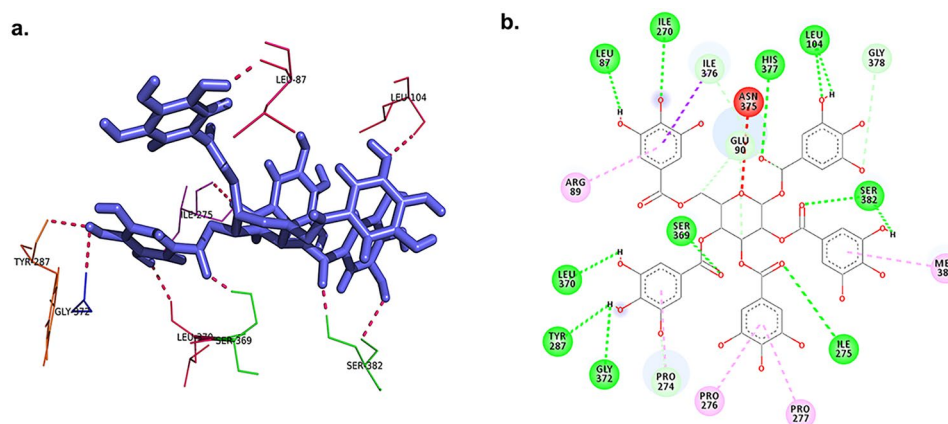
**Domain and Active Site Prediction.** Other researchers have also analyzed this protein and some of the domains of this protein have been deciphered. However, in order to find additional domains CDSearch was performed<sup>48</sup>. Domain analysis showed that the target protein belongs to a Pkc\_like superfamily as shown in Fig. 5 It can be seen that only the N-terminal part of the sequence shows similarity with other conserved domains. The E-value of the domains indicated that only partial or modified regions of the domains are present. These results suggest that the domains present are modified and thus, may not play the expected functional role. The active site of the protein is not known thus it is necessary to predict the active sites of the protein for ligand docking.

The Active Sites of the protein were predicted using DogSiteScorer<sup>49</sup> and Schrodinger's siteprep<sup>50, 51</sup>. DogSiteScorer predicts the druggability of the active sites thus predicting confidence on sites. In this calculation, the largest site contained the first three active site residues/regions predicted by DogSiteScorer. Due to higher druggability scores of this site (shown in Table 1), it was definitely considered as the only active site. Thus, the total volume of the active site was predicted to be 2325 cubic angstroms and the surface area to be 2703 cubic angstroms.

**Ligand Preparation.** Prior to docking, a search for a molecule previously reported against Ror1 was carried out in the drug databases. No chemical molecules are known as its possible inhibitor, however, a natural molecule Wnt5a has been reported to be a possible inhibitor of Ror1 and Ror2. For docking, ligands were downloaded from the ZINC database<sup>52</sup>. To reduce the computational intensity and to maintain the accuracy of the prediction, the ligands were screened to obtain a single conformation for all ligands with the same scaffold. Then these ligands were prepared and optimized using LigPrep tool<sup>53</sup>. The optimized 3D structure of ligand molecules was obtained from LigPrep and 45,000 unique ligands were obtained.

Source File	Glide gscore (kcal/mol)	Glide energy (kcal/mol)	XP HBond	XP PoseRank
ligand_15	-20.6641	-84.4288	-8.16	6
ligand_302	-19.8955	-83.0535	-7.54205	8
ligand_1725	-19.5035	-74.0717	-7.23445	5
ligand_276	-18.8408	-91.4571	-7.96953	2
ligand_1259	-16.953	-65.4968	-1.57924	6
ligand_307	-16.2679	-69.7247	-1.88054	11
ligand_256	-15.8306	-66.6586	-2.22251	11
ligand_2126	-15.8234	-64.2842	-1.39871	9
ligand_682	-15.8049	-63.5301	-1.66	11
ligand_146	-15.7883	-85.6785	-1.66	11

**Table 2.** Glide scores of the best 10 docked ligands.



**Figure 7.** Hydrogen bond interactions within the Protein and ligand 15 complex (a) 3D-image (b) 2D-image

**Protein-Ligand Docking.** It is computationally very intensive to dock a ligand library as large as ours and thus, before going for intensive docking the ligands were screened using high throughput technique to obtain the ligands that are best to carry forward. These ligands were initially screened using the High Throughput Virtual Screening (HTVS) module of Schrodinger<sup>54</sup>. The screening was conducted in flexible docking mode in order to analyze all the possible conformations of the ligands. This will remedy the initial cut down performed to reduce the database size. The results of this screening were obtained by selecting a cut off docking score value of  $-7$  kcal/mol. These ligands were analyzed for druggability parameters (shown in Fig. 6) using MobyL Server (FAF-Drugs3@rpbs a Free ADME/toxicity Filtering tool 3)<sup>55,56</sup>.

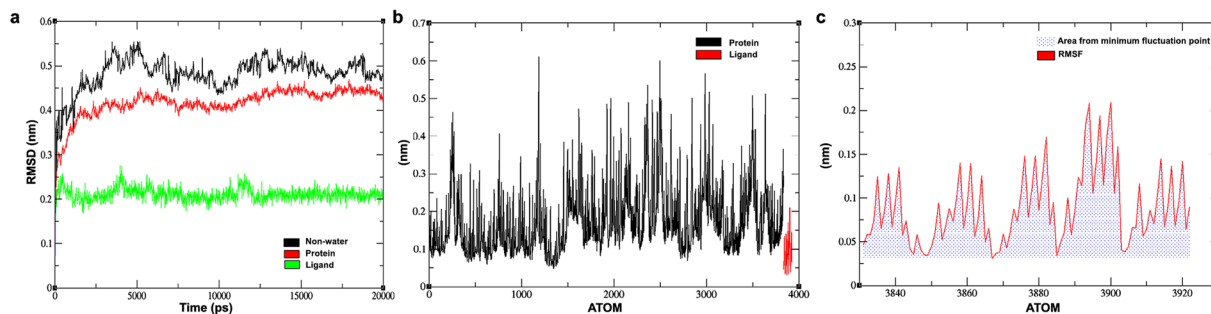
A total of 7779 ligands were obtained after HTVS and absorption, distribution, metabolism and excretion/toxicity (ADMET) screening. These ligands were docked again using AutoDockVina<sup>57</sup>. This program keeps the ligand flexible while docking. Docking results obtained were consolidated and sorted in accordance with affinity score and root-mean-square deviations (RMSDs). Only ligands with an affinity value of 10.9 and above were considered. Duplicate ligand occurrences fulfilling the cutoff were removed i.e. only first occurrence was considered. All unique molecules fulfilling the criterion were considered for further analysis. These 300 ligands were then scored and screened using SP docking<sup>58</sup>. The top 50 ligands with highest glide score were taken for flexible XP (extra precision) docking<sup>59</sup> to ensure a high accuracy of docking. In XP docking, flexibility of the interacting regions were considered, the ligand is kept flexible while some atoms of the protein that directly interact are also kept flexible. XP docking produced a high docking score and glide score (Shown in Table 2).

The best hit was Ligand 15 with a glide score of  $-20.66$  kcal/mol (Fig. 7a and b). Ligand 15 is the ligand with DrugBank accession number DB03208. It is named as Beta-1,2,3,4,6-Penta-O-Galloyl-D-Glucopyranose. This compound belongs to the class of organic compounds known as tannins<sup>60</sup>. It is a naturally occurring polyphenol, aromatic heteromonocyclic compound<sup>61</sup>. It is a non-carcinogenic ligand with absorption in human intestine probability as  $-0.8347$ , Blood Brain Barrier value as 0.5216. Its water solubility is 0.679 mg/ml, logP 3.43 and potential energy 396.15 kcal/mol.

**Protein Simulation.** The protein was simulated using Gromacs<sup>62</sup> to check its conformational changes. The predicted structure of the protein was used as the initial structure for MD simulation analysis. Extended Simple Point Charge (SPC/E) water model and “AMBER99SB-ILDN protein” force-field was used to prepare the system.

The system consisted of a cubic water box containing a protein molecule at its center and Cl<sup>-</sup> and Na<sup>+</sup> for charge stability. To keep the whole system neutral, the 3 Cl atoms were added. The protein was energetically





**Figure 8.** (a) RMSD plot of protein, ligand and complex (b), RMSF plot of protein and ligand (c), RMSF of ligand.

minimized for 0.1 nanoseconds (ns) and for a maximum force greater than or equal to 10.0 kJ/mol. The production MD simulations were then carried out for 200 ns. The RMSD plot of the complete simulation indicates that the deviation from the initial structure continues to increase for the full simulation length. The deviation halted for short time intervals during the simulation but the plot direction faced upwards. The deviation was almost halted during 70 to 118 ns and then towards the end after 174 ns. The RMSF plot indicates that the initial region of the sequence has very less fluctuations whereas the later regions show higher fluctuations. Lower movement of helix and sheet region shows that loop regions provide flexibility to the protein structure. The residues that show maximum fluctuation are 210, 209, 212–215, 345, 344 and 256 respectively. The less-frequently fluctuated residues in the second half of the sequence indicate that these residues act as hinge points between the moving loop regions. The major hinge points in the sequence are residues 195, 221, 274, 310, 322, 332, and 361. The first half of the sequence showed lower fluctuations of around 0.1 nm to the minimum fluctuation. The area per atom plot indicates that the standard deviation of atoms in the second half of the protein was higher, in the first half of the sequence only a few atoms fluctuated more. The average accessible area plot showed that some residues were never accessible during the simulation at the same time some were less accessible and others were highly accessible. The high number of peaks in the second half of the sequence shows that the loop regions interacted the most with other molecules whereas there were some residues in the first that also showed high accessibility. The average displacement that protein has faced during the simulation is plotted smoothly thus there is no strong structural activity detected during the simulation. The protein slowly and gradually changes its structural conformation in response to its surrounding. The secondary structure plot indicates that the number of residues in coils and bend regions decreases whereas the number of residues in turn and 3-helix regions increased. This change was not very prominent because only around 10–15 residues changed their state. Other structures did not show any significant change during simulation. The lack of major changes indicates that they are very stable and conserved. Plotting Hydrogen bonds between protein with itself and with other residues indicated a very stable hydrogen bond count during the simulation. There were higher numbers of hydrogen bonds between the protein residues than with non-protein residues. This indicated that the protein is more internally stabilized than from outside forces.

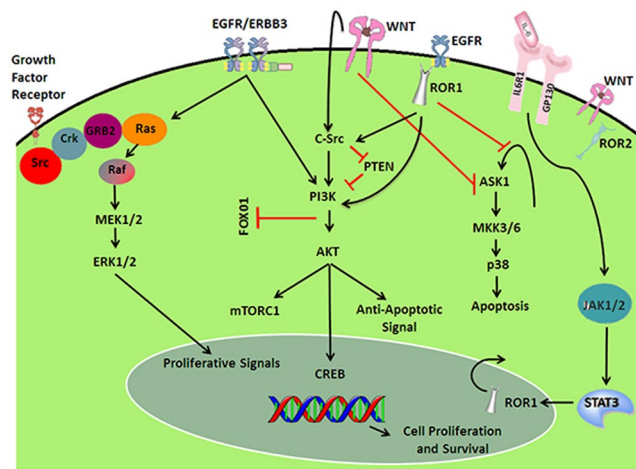
**Protein-Ligand Complex Simulation.** The protein-ligand complex simulation analysis shows that the complex is very stable (Fig. 8a). The protein became stable around 5 ns and the ligand around 2 ns. The protein was almost stable during the simulation, whereas, the ligand showed higher fluctuations at some points during the simulation. At the same time non-water molecules became stable around 6 ns. Nonetheless, the complex continued to show some fluctuations during the remaining 15 ns of simulations. The Root Mean Square Fluctuation (RMSF) analysis showed that the first half of the protein had lower fluctuations than the second half (Fig. 8b), which was comprised of loop regions. The ligand showed lower fluctuations. Most atoms had some fluctuations when they were analyzed together indicating that the protein show acceptability to other molecules and the loop region of the protein is flexible to accept ligands and adopt conformational changes based on interactions.

The hydrogen bond analysis confirmed that the high score of the complex is explainable and based on interactions. The ligand atoms that are towards the sheet region and forming a hydrogen bond with the protein showed higher fluctuations than other atoms. The atoms forming the central ring showed the least fluctuations or none at all (Fig. 8c) whereas the '-OH' group attached to the outermost rings showed higher fluctuations than other atoms. The plot shows that during the simulation the movements of the outer rings are similar to one another. The amino acid residues interacting with the ligand were observed to have a bond length range of 1.5–2.8 Å.

**High-Throughput MM-PBSA Calculations by GMPBSA.** The top three hundred ligands were analyzed using the gmpbsa method<sup>63,64</sup>. This method produces a better score with respect to negating false positives i.e. top scoring ligands have higher chances to pass the experimental test. It aims to integrate high-throughput molecular dynamics (MD) simulations with binding energy calculations. The screening for 300 ligands selected after docking generated the top hundred ligands based on binding energy with energy in the range of 200–484 kcal/mol (for a detailed overview, see supplementary information: Tables S1 and S2).

## Discussion

In this work we have examined the target protein Ror1, whose expression changes during certain types of cancers. In a previous study, it was determined that when over-expression of Ror1 without HER2/neu, and hormone



**Figure 9.** A pathway showing role of Ror1 in promoting tumor-cell growth<sup>16,71</sup>. During tumorous growth interleukin 6 (IL-6) binds with IL6R1 and GP130 and induces signal transducer activator of transcription 3 (STAT3) phosphorylation through the involvement of JAK proteins and upregulates Ror1 protein levels in a time- and dose-dependent manner by activating Ror1 promoter activity. STAT3 also induces expression of Wnt5a that interacts with Ror1 to trigger PI3K/AKT through C-Src leading to activation of CREB that promotes tumor cell growth. Ror1 is synthesized in the nucleus and then it is transported to the cytoplasm where it inhibits the activity of ASK1, which follows a cascade that leads to apoptosis. At the same time, Ror1 activates S-src and PI3K, which in turn influences the CREB cycle to enhance expression of genes that enhance resistance to tumor cell apoptosis and/or promote tumor cell growth. Therefore, Ror1 functions to keep a balance between pro-survival PI3K-AKT and pro-apoptotic p38 signaling. [DNA Helix used in Fig. 9 is downloaded from <https://pixabay.com/en/photos/?q=dna%20helix>: All images on Pixabay website are released free of copyrights under Creative Commons CC0.]

receptors on the cell surface were introduced in the breast cancer this protein served as an appropriate candidate for designing a cancer vaccine and it was concluded that Ror1 with an enterotoxin B could be a potent vaccine<sup>65</sup>. Recently, it was also reported that knockdown of Ror1 significantly inhibited cell migration and invasion and when both Ror1 and Ror2 were knocked down, the cells got significantly sensitized to cisplatin. However, Ror1 over expression in the parental cell line increased cell invasion, indicating that Ror1 and Ror2 have potential as novel drug targets in metastatic and recurrent ovarian cancer patients<sup>29</sup>. In a different report, it was shown that knockdown of Ror1 resulted in reduction of stemness and sphere formation capacity. Moreover, it was shown that down-regulation of Ror1 suppressed the expression of EM related genes and the migratory and invasive abilities of the tumour. The results of this study indicated that targeting Ror1 could induce differentiation of cancer stem cells (CSCs) and inhibit metastasis in glioblastoma<sup>15</sup>. In a recent report, it was also shown that siRNAs targeting Ror1 in CLL induced apoptosis and a small molecule inhibitor, Ror1 tyrosine kinase inhibitor dephosphorylated Ror1, down-regulating the activated PI3K/AKT/mTOR signaling pathway and inducing specific apoptosis of CLL cells<sup>66</sup>. Therefore, we can conclude that Ror1 may be developed as a potential marker for several types of cancer and could be a potential target for stem cell therapy and drug discovery.

Here, a detailed study of the nature of the receptor protein and scrutinizing of probable inhibitory small molecules has been performed. The target protein (Ror1) was selected on the basis of previous reports concluding that Ror1 can act as a novel target for designing an efficient drug for cancer treatment (Fig. 9). Under normal conditions this protein expresses itself during childhood to support fast growth in children and it is not expressed in normal adult cells. Thus, it is expected that targeting this molecule will control cells from becoming cancerous as well as remedying a cancerous cell conditions. Therefore, it can be used in early stage cancers as well as in late stage cancers. As this target is not expressed in normal cells and has a small role to play under normal conditions, targeting it will have minimal side effects.

In this study, the protein structure was modeled and found to be stable as analyzed by energy plot and other analyses (e.g. Ramachandran plot and Z-score). Considering an active site with high volume and high accessible surface area ensured the freedom of the ligand to find the best sites and positions to interact. The structure of this protein is very interesting as it can be divided into two halves of which the first half contains the helix and sheet regions whereas the other half contains the coil regions. Our predictions show that the active sites lie in the coil region. The helix and sheet region from its back and a coil lock from its front support the active site that we have predicted. The protein-ligand complex showed stability and minimal fluctuations during simulation; therefore, the complex is expected to be stable. Stepwise screening of a considerable number of ligands ensured that the best ligands from the database were obtained. Glide score for top hits reached a high score of  $-20$  kcal/mol which indicates that the ligands interact with very high affinity towards the protein. Energy of the docked complexes was lower than the unbound protein and ligand, which increases the feasibility of their interaction. The best hit thus obtained is ligand 15 (DrugBank accession number DB03208), which is designated as Beta-1, 2,3,4,6-Pent a-O-Galloyl-D-Glucopyranose, a naturally occurring organic compound (polyphenol) belonging to the class of

tannins. Tannins are potent collagen cross-linking agent and found in many plants and plant products that are used as food. In recent reports, it was concluded that tannins are potential anticancer agents and apoptotic activity in breast cancer and prostate cancer cells is enhanced when they are exposed to tannin extracts<sup>67–70</sup> indicating the role of tannins as prospective anti-cancer therapy. This discovery also supports our finding that since Beta-1,2,3,4,6-Penta-O-Galloyl-D-Glucopyranose interacts with Ror1 and inhibits its action, this compound could potentially serve as a potent drug for cancer treatment. This work can further be extended for wet lab experimentation and clinical trials. We also expect these molecules as the potent drug molecules in the market for both cancer inhibition and for decreasing chances of cancer.

## Methods

The sequence for Ror1 (ID AAC50714.1) was retrieved from NCBI protein database (<https://www.ncbi.nlm.nih.gov/protein>). Secondary structure predictions were performed by PROMOTIF<sup>30</sup> and GOR4 tool<sup>31</sup>. Then the 3D-structure of the protein was predicted using following servers and tools CPHModel<sup>34</sup> phyre2<sup>35</sup>, ps2v2<sup>36</sup>, RaptorX<sup>37</sup>, Modeller<sup>38</sup> and I-Tasser<sup>39</sup>. The structures were analyzed and validated by Prosa<sup>41,42</sup> and SAVES<sup>45</sup> server. The most correctly modelled structure was then optimized using Schrodinger's Protein Preparation wizard<sup>47</sup>. Domains were searched using CDSearch<sup>48</sup>. Then the active site of the protein was predicted with help of DogSiteScorer<sup>49</sup> and Schrodinger's siteprep<sup>50,51</sup>. The ligands were obtained from Zinc database<sup>52</sup>. Ligprep<sup>53</sup> was used to prepare the ligands after having the ligands screened for multiple scaffolds. Then the ligands were screened using Schrodinger's HTVS<sup>54</sup>. The ligands having docking score below  $-7$  kcal/mol were analyzed on druggability parameters using Mobyl Server<sup>55,56</sup>.

Then the ligands passing the druggability parameters were docked using AutoDockVina<sup>57</sup>. Again the screening was carried out at an affinity value 10.9 and above. The screened ligands were then screened for duplication and carried forward for SP docking<sup>58</sup>. Top 50 ligands obtained through screening with SP docking were docked again using XP (Extra precision) docking<sup>59</sup>. Then, the protein and protein-ligand15 complex were simulated using Gromacs<sup>62</sup>. Extended Simple Point Charge (SPC/E) water model and "AMBER99SB-ILDN protein" force-field was used to prepare the system. The system contained a cubic water box with protein, Cl<sup>-</sup> and Na<sup>+</sup>. The system was kept neutral by adding ions. The protein was energetically minimized for 0.1 ns before the production MD simulations of 200 ns. In similar fashion the protein-ligand complex was simulated for 20 ns. The top 300 ligands were analysed using g\_mmpbsa<sup>63,64</sup> to obtain any probable false positives and to score better docked complexes with low energy scores.

## References

- Forrester, W. C. The Ror receptor tyrosine kinase family. *Cell Mol Life Sci* **59**, 83–96, doi:10.1007/s00018-002-8407-9 (2002).
- Katoh, M. & Katoh, M. Identification and characterization of rat Ror1 and Ror2 genes *in silico*. *Int J Mol Med* **15**, 533–538 (2005).
- Masiakowski, P. & Carroll, R. D. A novel family of cell surface receptors with tyrosine kinase-like domain. *J Biol Chem* **267**, 26181–26190 (1992).
- Oishi, I. *et al.* Spatio-temporally regulated expression of receptor tyrosine kinases, mRor1, mRor2, during mouse development, implications in development and function of the nervous system. *Genes Cells* **4**, 41–56, doi:10.1046/j.1365-2443.1999.00234.x (1999).
- Wilson, C., Goberdhan, D. C. & Steller, H. Dror, a potential neurotrophic receptor gene, encodes a Drosophila homolog of the vertebrate Ror family of Trk-related receptor tyrosine kinases. *Proc Natl Acad Sci USA* **90**, 7109–7113, doi:10.1073/pnas.90.15.7109 (1993).
- Fukuda, T. *et al.* Antisera induced by infusions of autologous Ad-CD154-leukemia B cells identify ROR1 as an oncofetal antigen and receptor for Wnt5a. *Proc Natl Acad Sci USA* **105**, 3047–3052, doi:10.1073/pnas.0712148105 (2008).
- Afzal, A. R. *et al.* Recessive Robinow syndrome, allelic to dominant brachydactyly type B, is caused by mutation of ROR2. *Nat Genet* **25**, 419–422, doi:10.1038/78107 (2000).
- Oldridge, M. *et al.* Dominant mutations in ROR2, encoding an orphan receptor tyrosine kinase, cause brachydactyly type B. *Nat Genet* **24**, 275–278, doi:10.1038/73495 (2000).
- Nomi, M. *et al.* Loss of mRor1 enhances the heart and skeletal abnormalities in mRor2-deficient mice, redundant and pleiotropic functions of mRor1 and mRor2 receptor tyrosine kinases. *Mol Cell Biol* **21**, 8329–8335, doi:10.1128/MCB.21.24.8329-8335.2001 (2001).
- Aghebati-Maleki, L. *et al.* Receptor tyrosine kinase-like orphan receptor 1 (ROR-1): An emerging target for diagnosis and therapy of chronic lymphocytic leukemia. *Biomed. Pharmacother.* **88**, 814–822, doi:10.1016/j.biopha.2017.01.070 (2017).
- Gentile, A., Lazzari, L., Benvenuti, S., Trusolino, L. & Comoglio, P. M. Ror1 is a pseudokinase that is crucial for Met-driven tumorigenesis. *Cancer research* **71**, 3132–3141, doi:10.1158/0008-5472.CAN-10-2662 (2011).
- Mikels, A. J. & Nusse, R. Purified Wnt5a protein activates or inhibits beta-catenin-TCF signaling depending on receptor context. *PLoS Biol* **4**, e115, doi:10.1371/journal.pbio.0040115 (2006).
- Nomachi, A. *et al.* Receptor tyrosine kinase Ror2 mediates Wnt5a-induced polarized cell migration by activating c-Jun N-terminal kinase via actin-binding protein filamin A. *J. Biol Chem* **283**, 27973–81, doi:10.1074/jbc.M802325200 (2008).
- Yu, J. *et al.* Wnt5a induces ROR1/ROR2 heterooligomerization to enhance leukemia chemotaxis and proliferation. *J. Clin. Invest.* **126**, 585–598, doi:10.1172/JCI83535 (2016).
- Jung, E. H., Lee, H. N., Han, G. Y., Kim, M. J. & Kim, C. W. Targeting ROR1 inhibits the self-renewal and invasive ability of glioblastoma stem cells. *Cell Biochem Funct.* **34**, 149–57, doi:10.1002/cbf.3172 (2016).
- Zhang, S. P. *et al.* ROR1 is expressed in human breast cancer and associated with enhanced tumor-cell growth. *PLoS One* **7**, e31127, doi:10.1371/journal.pone.0031127 (2012).
- Daneshmanesh, A. H. *et al.* Ror1, a cell surface receptor tyrosine kinase is expressed in chronic lymphocytic leukemia and may serve as a putative target for therapy. *Int. J. Cancer* **123**, 1190–1195, doi:10.1002/ijc.v123:5 (2008).
- Hojjat-Farsangi, M. *et al.* The Tyrosine Kinase Receptor ROR1 Is Constitutively Phosphorylated in Chronic Lymphocytic Leukemia (CLL) Cells. *PLoS ONE* **8**, e78339, doi:10.1371/journal.pone.0078339 (2013).
- Rabbani, H. *et al.* Expression of ROR1 in patients with renal cancer—a potential diagnostic marker. *Iran Biomed. J.* **14**, 77–82 (2010).
- Shabani, M. *et al.* Overexpression of orphan receptor tyrosine kinase Ror1 as a putative tumor-associated antigen in Iranian patients with acute lymphoblastic leukemia. *Tumour Biol.* **28**, 318–326, doi:10.1159/000121405 (2007).
- Yamaguchi, T. *et al.* NKX2-1/TITF1/TTF-1-induced ROR1 is required to sustain EGFR survival signaling in lung adenocarcinoma. *Cancer Cell* **21**, 348–361, doi:10.1016/j.ccr.2012.02.008 (2012).



22. Zheng, Y.-Z. *et al.* ROR1 is a novel prognostic biomarker in patients with lung adenocarcinoma. *Sci. Rep.* **6**, 36447, doi:10.1038/srep36447 (2016).
23. Hojjat-Farsangi, M. *et al.* Spontaneous Immunity Against the Receptor Tyrosine Kinase ROR1 in Patients with Chronic Lymphocytic Leukemia. *PLoS One* **10**, e0142310, doi:10.1371/journal.pone.0142310 (2015).
24. Cui, B. *et al.* High-level ROR1 associated with accelerated disease-progression in chronic lymphocytic leukemia. *Blood* **128**, 2931–2940, doi:10.1182/blood-2016-04-712562 (2016).
25. Cui, B. *et al.* Targeting ROR1 inhibits epithelialmesenchymal transition and metastasis. *Cancer Res* **73**, 3649–3660, doi:10.1158/0008-5472.CAN-12-3832 (2013).
26. Endo, M., Doi, R., Nishita, M. & Minami, Y. Ror family receptor tyrosine kinases regulate the maintenance of neural progenitor cells in the developing neocortex. *J. Cell Sci.* **125**, 2017–2029, doi:10.1242/jcs.097782 (2012).
27. Henry, C. E., Llamas, E., Djordjevic, A., Hacker, N. F. & Ford, C. E. Migration and invasion is inhibited by silencing ROR1 and ROR2 in chemoresistant ovarian cancer. *Oncogenesis* **5**, e226, doi:10.1038/oncsis.2016.32 (2016).
28. Zhang, W., Kluz, P. & Kolb, R. ROR1 is an Intriguing Target for Cancer Therapy. *Molecular enzymology and drug targets* **2**, 11 (2016).
29. Reddy, U. R., Phatak, S. & Pleasure, D. Human neural tissues express a truncated Ror1 receptor tyrosine kinase, lacking both extracellular and transmembrane domains. *Oncogene* **13**, 1555–1559 (1996).
30. Hutchinson, E. G. & Thornton, J. M. PROMOTIF A program to identify and analyze structural motifs in proteins. *Protein Sci* **5**, 212–220, doi:10.1002/pro.v5:2 (1996).
31. Garnier, J., Gibrat, J. F. & Robson, B. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* **266**, 540–53, doi:10.1016/S0076-6879(96)66034-0 (1996).
32. Berman, H. M. *et al.* The Protein Data Bank. *Nucl Acids Res* **28**, 235–242, doi:10.1093/nar/28.1.235 (2000).
33. Andreeva, A., Howorth, D., Chothia, C., Kulesha, E. & Murzin, A. G. SCOP2 prototype, a new approach to protein structure mining. *Nucl Acids Res* **42**, D310–D314, doi:10.1093/nar/gkt1242 (2014).
34. Nielsen, M., Lundegaard, C., Lund, O. & Petersen, T. N. CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles. *Nucl Acids Res.* **38**, W576–81, doi:10.1093/nar/gkq535 (2010).
35. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858, doi:10.1038/nprot.2015.053 (2015).
36. Chen, C. C., Hwang, J.-K. & Yang, J.-M. (PS)<sup>2</sup>: protein structure prediction server. *Nucl Acids Res.* **34**, W152–W157, doi:10.1093/nar/gkl187 (2006).
37. Yang, J. *et al.* The I-TASSER Suite, Protein structure and function prediction. *Nature Methods* **12**, 7–8, doi:10.1038/nmeth.3213 (2015).
38. Chen, C. C., Hwang, J.-K. & Yang, J.-M. (PS) 2-v2: template-based protein structure prediction server. *BMC Bioinformatics* **10**, 366, doi:10.1186/1471-2105-10-366 (2009).
39. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nature protocols* **7**, 1511–1522, doi:10.1038/nprot.2012.085 (2012).
40. Benjamin, W. & Sali, A. Comparative protein structure modeling using. *Modeller. Curr. Protoc. bioinfo.* **47**, 5–6 (2014).
41. Sippl, M. J. Recognition of Errors in Three-Dimensional Structures of Proteins. *Proteins* **17**, 355–362, doi:10.1002/prot.340170404 (1993).
42. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucl Acids Res* **35**, W407–W410, doi:10.1093/nar/gkm290 (2007).
43. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40, doi:10.1186/1471-2105-9-40 (2008).
44. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER, a unified platform for automated protein structure and function prediction. *Nature Protocols* **5**, 725–738, doi:10.1038/nprot.2010.5 (2010).
45. Luthy, R., Bowie, J. U. & Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83–85, doi:10.1038/356083a0 (1992).
46. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK, a program to check the stereochemical quality of protein structures. *J Appl Cryst.* **26**, 283–291, doi:10.1107/S0021889892009944 (1993).
47. Schrödinger Suite 2012 Protein Preparation Wizard; Epik version 2.3. Schrödinger, LLC, New York, NY, 2012; Impact version 5.8, Schrödinger, LLC, New York, NY, 2012. Prime version 3.1, Schrödinger, LLC, New York, NY (2012).
48. Marchler-Bauer, A. *et al.* CDD, NCBI's conserved domain database. *Nucl Acids Res.* **43**, 222–226, doi:10.1093/nar/gku1221 (2015).
49. Volkamer, A., Kuhn, D., Thomas Grombacher, T., Rippmann, F. & Rarey, M. Combining global and local measures for structure-based druggability predictions. *J Chem Inf Model.* **52**, 360–372, doi:10.1021/ci200454v (2012).
50. Halgren, T. New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des.* **69**, 146–148, doi:10.1111/jpp.2007.69.issue-2 (2007).
51. Halgren, T. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model.* **49**, 377–389, doi:10.1021/ci800324m (2009).
52. Irwin, J. J. & Shoichet, B. K. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *J Chem Inf Model.* **45**(1), 177–182, doi:10.1021/ci049714+ (2005).
53. Schrödinger Release LigPrep, version 2.3, Schrödinger, LLC, New York, NY (2009).
54. Friesner, R. A. *et al.* Glide, A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem.* **47**, 1739–1749, doi:10.1021/jm0306430 (2004).
55. Lagorce, D., Sperandio, O., Galons, H., Miteva, M. A. & Villoutreix, B. O. FAF-Drugs2, free ADME/tox filtering tool to assist drug discovery and chemical biology projects. *BMC Bioinformatics* **9**, 396, doi:10.1186/1471-2105-9-396 (2008).
56. Lagorce, D. *et al.* The FAF-Drugs2 server, a multistep engine to prepare electronic chemical compound collections. *Bioinformatics* **27**, 2018–20, doi:10.1093/bioinformatics/btr333 (2011).
57. Trott, O. & Olson, A. J. AutoDockVina, improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* **31**, 455–461, doi:10.1002/jcc.21334 (2010).
58. Halgren, T. A. *et al.* Glide, A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J Med Chem.* **47**, 1750–1759, doi:10.1021/jm030644s (2004).
59. Friesner, R. A. *et al.* Extra Precision Glide, Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J Med Chem.* **49**, 6177–6196, doi:10.1021/jm051256o (2006).
60. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucl Acids Res.* **34**, D668–72, doi:10.1093/nar/gkj067 (2006).
61. Kim, S. *et al.* PubChem Substance and Compound databases. *Nucl Acids Res.* **44**, D1202–13, doi:10.1093/nar/gkv951 (2016).
62. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854, doi:10.1093/bioinformatics/btt055 (2013).
63. Kumari, R. *et al.* g\_mmpbsa A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J Chem Inf Model.* **54**, 1951–1962, doi:10.1021/ci500020m (2014).
64. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041, doi:10.1073/pnas.181342398 (2001).
65. Fooladi, A. A. I., Hosseini, H. M. & Amani, J. An *In silico* Chimeric Vaccine Targeting Breast Cancer Containing Inherent Adjuvant. *Iran J Cancer Preven* **8**, e2326 (2015).

66. Hojjat-Farsangi, M. *et al.* First-in-Class ROR1 Small Molecule Inhibitor (KAN0439834) Down regulated Wnt-Canonical and Non-Canonical Signaling Pathways and Induced Apoptosis of CLL Cells. In: ASH 57<sup>th</sup> Annual Meeting & Exposition, Orlando, FL, December 5-8 (2015).
67. Huang, W.-Y., Cai, Y.-Z. & Zhang, Y. Natural phenolic compounds from medicinal herbs and dietary plants: potential use for cancer prevention. *Nutrition and Cancer* **62**, 1–20, doi:10.1080/01635580903191585 (2010).
68. Weng, C. J. & Yen, G. C. Chemopreventive effects of dietary phytochemicals against cancer invasion and metastasis: phenolic acids, monophenol, polyphenol, and their derivatives. *Cancer Treatment Reviews* **38**, 76–87, doi:10.1016/j.ctrv.2011.03.001 (2012).
69. Fresco, P., Borges, F., Diniz, C. & Marques, M. P. M. New insights on the anticancer properties of dietary polyphenols. *Medicinal Research Reviews* **26**, 747–766, doi:10.1002/med.20060 (2006).
70. Bawadi, H. A., Bansode, R. R., Trappey, A. II, Truax, R. E. & Losso, J. N. Inhibition of Caco-2 colon, MCF-7 and Hs578T breast, and DU 145 prostatic cancer cell proliferation by water-soluble black bean condensed tannins. *Cancer Letters* **218**, 153–162, doi:10.1016/j.canlet.2004.06.021 (2005).
71. Karachaliou, N. *et al.* ROR1 as a novel therapeutic target for EGFR-mutant non-small-cell lung cancer patients with the EGFR T790M mutation. *Transl. Lung Cancer Res.* **3**, 122–130, doi:10.3978/j.issn.2218-6751.2014.03.02 (2014).

## Acknowledgements

We thank Dr. Ajay K. Arora, Deshbandhu College, University of Delhi; SCFBIO, IIT Delhi; and JNU New Delhi for permission to use the computational facility. The authors acknowledge Dr. Paula Mitchell (Winthrop University, USA) for language corrections. The authors also acknowledge Dr. Rajesh Sinha (University of Delhi, India) and Ms Chetna Tyagi (University of Szeged, Hungary) for their critical inputs to the manuscript and Dr Amit Roy, University of Kentucky, USA for formatting and improvement of figures.

## Author Contributions

O.N., A.S. and I.K.S. planned experiments, analyzed data and prepared manuscript. O.N. and A.S. performed experiments, and prepared the figures. O.N., A.S. and I.K.S. reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-01254-w

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017