



Determination of promising inhibitors for N-SH2 domain of SHP2 tyrosine phosphatase: an in silico study

Emel Başak Gencer Akçok¹ · Hüseyin Güner^{1,3,4} · İsmail Akçok²

Received: 7 December 2023 / Accepted: 15 April 2024 / Published online: 13 May 2024
© The Author(s) 2024

Abstract

There are many genes that produce proteins related to diseases and these proteins can be targeted with drugs as a potential therapeutic approach. Recent advancement in drug discovery techniques have created new opportunities for treating variety of diseases by targeting disease-related proteins. Structure-based drug discovery is a faster and more cost-effective approach than traditional methods. SHP2 phosphatase, encoded by the *PTPN11* gene, has been the focus of much attention due to its involvement in many types of diseases. The biological function of SHP2 is enabled mostly by protein–protein interaction through its SH2 domains. In this study, we report the identification of a potential small molecule inhibitor for the N-SH2 domain of SHP2 by structure-based drug discovery approach. We utilized molecular docking studies, followed by molecular dynamics simulations and MM/PBSA calculations, to analyze compounds retrieved from the Broad's Drug Repurposing Hub and ZINC15 databases. We selected 10 hit compounds with the best docking scores from the libraries and examined their binding properties in the N-SH2 domain. We found that compound CID 60838 (Irinotecan) was the most suitable compound with a binding free energy value of -64.45 kcal/mol and significant interactions with the target residues in the domain.

✉ İsmail Akçok
ismail.akcok@agu.edu.tr

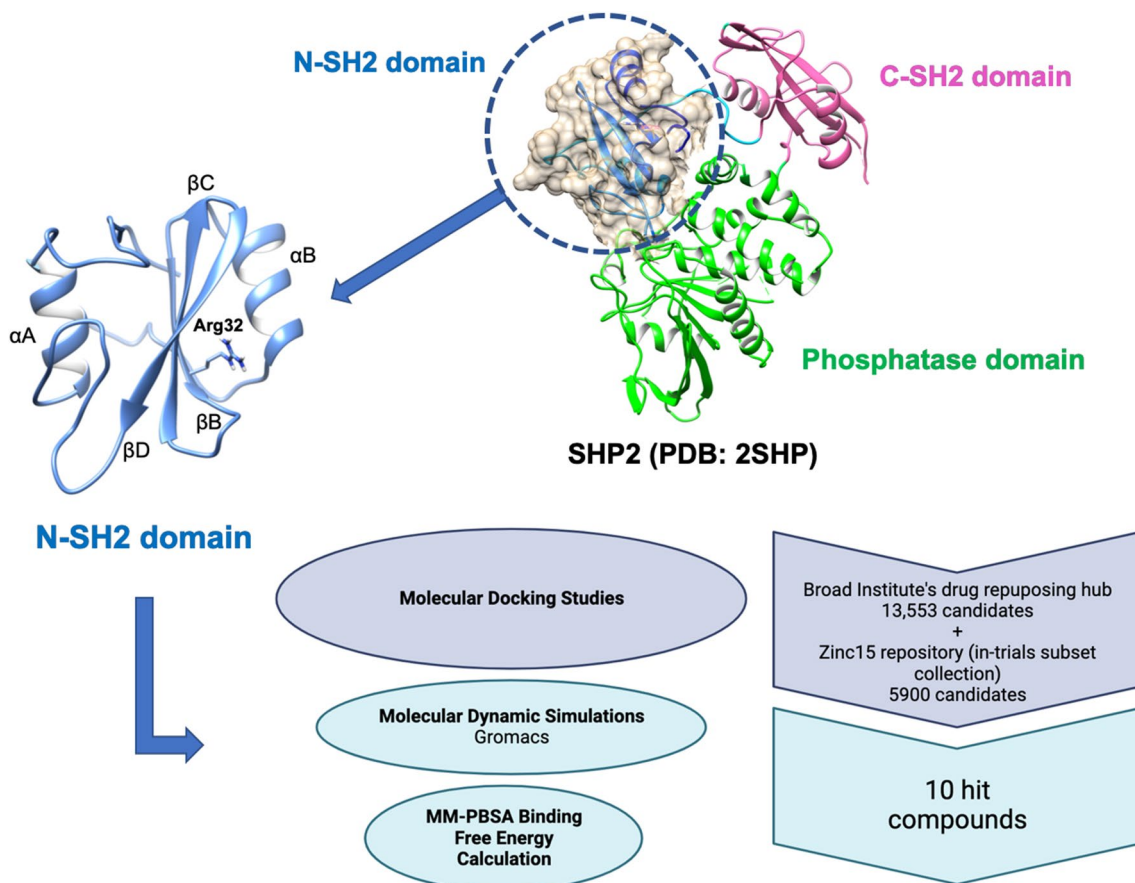
¹ Department of Molecular Biology and Genetics, Faculty of Life and Natural Sciences, Abdullah Gül University, 38080 Kayseri, Türkiye

² Department of Bioengineering, Faculty of Life and Natural Sciences, Abdullah Gül University, 38080 Kayseri, Türkiye

³ Izmir International Biomedicine and Genome Institute, Dokuz Eylül University, 35340 Balçova, İzmir, Türkiye

⁴ IzmirBiomedicine and Genome Center (IBG), 35340 Balçova, İzmir, Türkiye

Graphical abstract



Keywords Molecular docking · Molecular dynamics (MD) · In silico · SHP2 phosphatase · SH2 domain

Introduction

Specific protein–protein interaction is very important for the organization of intracellular signaling cascades and these interactions are mediated by protein domains [1]. The protein-mediated signaling processes sometimes require post-translational modifications and among these post-translational modifications, phosphorylation of specific residues such as tyrosine, serine, and threonine has great importance [2]. Tyrosine phosphorylation essentially has a critical importance for mammalian cells and studies demonstrated that it has particular importance in differentiation, proliferation, apoptosis and migration [3]. Tyrosine phosphorylation is mediated by proteins called protein tyrosine kinases and they are often called as “writers”. The phosphorylated tyrosine residues are recognized by the proteins via their SH2 (Src Homology 2) domains. There are also proteins that remove the phosphate

groups which are protein tyrosine phosphatases, and they are often referred as “erasers” [4].

SH2 domains are noncatalytic and highly conserved modules that contain approximately 100 amino acids and their structure is composed of a central antiparallel β -sheet flanked by two α -helices. The secondary structures, α -helices and β -sheets are connected by loop structures which are not folded and flexible regions of the target protein. Although the secondary structures are conserved in SH2 family of proteins, the loop sequences are not conserved. Protein phosphorylation is a fundamental post-translational modification that regulates many important signal transduction pathways, and the tyrosine phosphorylation has a particular importance in cell signaling, cell proliferation, growth, differentiation, cell survival, and even cell death. The protein interaction occurs through the interaction of phosphotyrosine and SH2 domains. Therefore, SH2 domains are very important regulators by maintaining specific phosphotyrosine signaling [4]. There are 121 SH2 domains in 111 different proteins [5]. SH2 domains recognize the

phosphorylated tyrosine residues and through this binding, they form multi-protein complexes [6]. By facilitating the protein–protein interactions, SH2 domains have a critical role in regulation of different essential cellular functions [7]. The specific recognition and binding to the phosphorylated tyrosine residues on target proteins are enabled by a specific amino acid sequence on SH2 domain. This conserved motif, G(S/T)FLVR(E/D)S, contains a positively charged arginine residue which is critical for binding to the negatively charged phosphate group on tyrosine residues [8, 9]. This conserved arginine residue on the strand β B plays a central role in forming bi-dentate hydrogen bonds with the phosphate moiety [10].

SH2 domain-containing phosphatase 2 (SHP2), which is encoded by the protein tyrosine phosphatase non-receptor type 11 (*PTPN11*) gene, is a tyrosine phosphatase that is composed of two SH2 domains at the amino terminus of the protein tyrosine phosphatase (PTP) domain which is located at the carboxy terminus (Fig. 1) [11]. Under normal unstimulated conditions, SHP2 possesses a closed conformation in which the N-SH2 binds to the PTP domain and thus remains inactive. When the SH2 domains encounter a target protein that has phosphorylated tyrosine residues, the conformation of the protein is opened and the phosphatase becomes active [12].

SH2 domains of SHP2 have functional importance in several diseases [13]. SHP2 phosphatase is an important regulator of cytokine, growth factor, and integrin signaling and thus has crucial importance in cell survival, proliferation, and differentiation by regulating major signaling pathways [14]. Under normal unstimulated conditions, SHP2 possesses a closed conformation in which the N-SH2 binds to the PTP domain and thus remains inactive. However, when the SH2 domains

encounter a target protein that has phosphorylated tyrosine residues, the conformation of the SHP2 protein is opened and the phosphatase becomes constitutively active [12]. There are mutations that are identified on *PTPN11* gene which encodes the SHP2 phosphatase, identified in several diseases, as such leukemia, solid tumors, Noonan syndrome and LEOPARD syndrome [13]. SHP2 was the first phosphatase that was discovered to be oncogenic [15]. Gain-of-function *PTPN11* mutations are the cause of more than 30% of juvenile myelomonocytic leukemia (JMML) and can be also found in other childhood malignancies [16, 17]. Moreover, SHP2 has also a critical role in receptor tyrosine kinase-driven cancers [18]. Missense mutations of *PTPN11* are found in more than 50% of the Noonan syndrome cases, which is a genetic disorder resulting from hyperactivation of the Ras-MAPK pathway [19, 20]. These missense mutations are gain-of-function mutations that keep phosphatase active [21]. For all these involvements of SHP2 in various disorders, it is important to target SHP2 phosphatase. Although there are some compounds targeting SHP2, genomic screens demonstrate that resistance to SHP2 inhibition is conferred [22]. Considering the role of SHP2 in a variety of diseases, new candidates are required to be identified. Computational methods are useful tools to identify novel hit compounds for a target at the atomic level for drug repurposing.

In the current study, we aimed to find the potential inhibitor that targets the N-SH2 domain of the SHP2 protein. For this purpose, a virtual screening protocol that consists of molecular docking, protein–ligand interaction, molecular dynamics (MD) simulations, and molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) calculations were applied to Broad Repurposing Hub compounds, and in-trial subsets from ZINC15 library. The

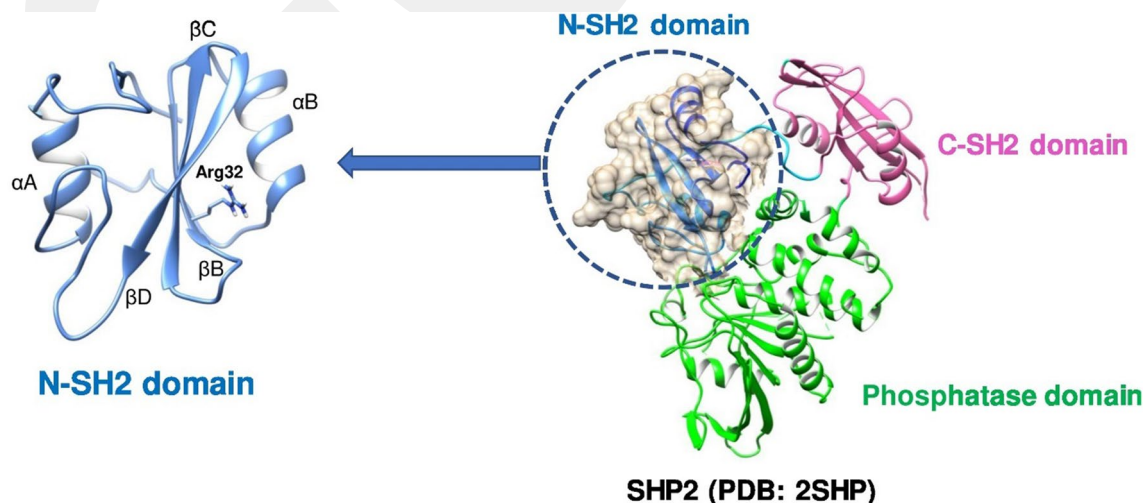


Fig. 1 Cartoon representation of SHP2 full length (PDB ID: 2SHP) with N-SH2 binding site

interactions of ligands with key residues were also shown and identified.

Methodology

Protein crystal structure preparation and ligand preparation

SHP-2 (PDB ID: 2SHP) [23] was selected as the target protein's X-ray crystallography model file and retrieved from the RCSB Protein Data Bank (www.rcsb.org). The protein was prepared for docking studies using PDBFixer (<https://github.com/openmm/pdbfixer>) [24]. It was cleaned off the existing water molecules and the missing residues and hydrogens were added to the protein. The Fpocket tool [25] was used to inspect and identify potentially druggable pockets. Pocket 15 was chosen as the potential protein–ligand binding site. PyMol software was used to visually inspect the druggable pockets and identify the binding area encompassing the biologically essential binding domain of N-SH2.

The Broad Institute's drug repurposing hub (<https://www.broadinstitute.org/drug-repurposing-hub>) provides a list of chemicals to screen potential drugs. Their repository contains a collection of 13553 FDA-approved drugs, clinical trials, and pre-clinical compounds. The structure files were downloaded of matching items of selected compounds from PubChem. A similar drug repurposing list of 5900 substances from the Zinc15 repository was also retrieved by browsing the in-trials subset collection [26]. All available structure files were downloaded accordingly. Because the three-dimensional structural files for some of the ligands were not available, they were processed using RDKit [27] and thus the hydrogens were added, and the energies of the 3D structures of each ligand in the screening library were minimized before docking.

Molecular docking studies

Smina [28], a variant of Autodock Vina [29], was used in molecular docking studies. The grid box of a rectangular parallelogram was selected with the coordinates (−2.5, −38.8, 39.3) and axis sizes (14.0, 14.6, 19.3) plus 4 Angstrom (Å) on each axis as the search space for the best binding poses. The exhaustiveness parameter was set as 16 and the seed was selected as 43.

MD simulation

MD simulations were carried out using Gromacs ver. 2021.03 [30]. The protein structure was cleaned and prepared for running simulations processing the script provided by the Chimera package. OPLS-AA/M as the force field and SPC216 as the explicit water models were selected to describe the ensemble. Ligand topologies and their force field parameter files for the force field selected were calculated by the LigParGen [31] command-line tool. (Please see the detailed MD methodology in Supporting Information).

MM/PBSA binding free energy calculation

The intensity of the binding free energy between ligands and receptors was estimated using MM/PBSA binding free energy calculations [32]. Sets of 200 configurations for each system were obtained as 1 nanosecond (ns) spaced snapshots, obtained directly from the molecular dynamics trajectories. The calculations were carried out using the *g_mmpbsa* GROMACS-compatible free energy program [33] with a grid space of 0.5 Å, salt concentration of 0.150 M, solute dielectric constant of 2 and employing the solvent accessible surface area (SASA) as estimate of the nonpolar solvation energy.

Results and discussion

Selection of target protein and identification of drug candidates

In this work, we target the specific recognition and binding region of the N-SH2 domain of SHP2, which has a conserved motif of G(S/T)FLVR(E/D)S. The specific positively charged arginine residue (Arg32) which is critical for binding to the negatively charged phosphate group on tyrosine residues was targeted. Because of the importance of this residue in protein–protein interaction and SHP2 signaling, targeting Arg32 is the main motivation of this study. The most conserved residues reside on the β B strand, the FLVR motif is present with the conserved arginine residue that plays the central role in forming an interaction with the phosphate group of phosphotyrosine [34]. Moreover, thermodynamic studies have established that the ionic interaction between Arg β B5 and pTyr is crucial for the affinity, providing more than half of the total binding energy [35]. Thus, we targeted the phosphotyrosine binding site of SHP2 (N-SH2 domain) for *in silico* screening with the goal of finding molecules that inhibit the interaction between N-SH2 and pTyr.

Molecular docking analysis of drugs against N-SH2 domain of 2SHP

Molecular docking studies have become essential for the discovery of novel drug candidates with the advancement of computational tools. The experimental binding mechanism and affinity were predicted by this method and it provides invaluable information about the possible interactions.

The molecular docking study of the compounds from The Broad Institute's FDA-approved drug and ZINC15 in-trials subset was carried out SMINA docking program to get the lowest energy conformers of the compounds at the binding site of the N-SH2 domain of SHP2 protein. At this stage, all known kinase inhibitors were eliminated from the list as they are known to bind to binding site which is not targeted in the current study. The molecular docking data of the compounds were collected and the best ten candidates were chosen for further studies. All docking scores, main interacting residues of the top ten candidates and interaction types of these molecules are listed in Table 1 and demonstrated in Fig. 2. The best docking score against 2SHP was achieved by Irinotecan (CID 60838) with a value of -9.3 kcal/mol, followed by Rimegepant (CID 51049968) with -9.2 kcal/mol, CID 12940973 with -9.1 kcal/mol, Eltrombopag (CID 135449332) with -9.1 kcal/mol, Antrafenine (CID 68723) with -8.9 kcal/mol, Naldemedine (CID 54732242) with -8.8 kcal/mol, CID 73774610 with -8.7 kcal/mol, Estriol 3-sulfate 16-glucuronide (CID 151223) with -8.7 kcal/mol, Cholic acid glucuronide (CID 21252309) with -8.7 kcal/mol, Glp-0187 (CID 53340771) with -8.7 kcal/mol respectively. The results show that all these ten molecules make hydrogen bonds with Arg32 as well as other interacting residues such as Thr12, Gly13, Lys35, Lys55 and Ile11. (Table 1).

MD simulations of selected drugs against N-SH2 domain of 2SHP

MD simulation, which is accepted as the gold standard to test the stability of the complex formed between protein and the leading drug candidates [36], has been performed for the best docking poses of the top ten candidates. For a comprehensive evaluation of the ligand–protein complexes, 300 ns molecular dynamic simulation was performed.

There are a lot of data that can be retrieved from MD simulations. In our study, first, we evaluated the distance between the center of mass (COM) of the ligands and the COM of Arg32. As mentioned previously, with its positive charge Arg32 residue is essential for binding tyrosine phosphorylated substrates. Focusing on the distance between this important residue and the corresponding ligands during the simulation provided us an important information about the binding affinity and stability of the complex. By measuring

this distance, we could easily determine whether the ligands maintain their positions during the simulation. As shown in Fig. 3A, it is clearly seen that the COM of the ligands CID 12940973, CID 21252309, CID 54732242 move away from Arg32 residue at the beginning of the simulation. The distance between COM of CID 53340771 and Arg32 is maintained as 7 Å until 50 ns, then the distance is calculated as 15 Å approximately. The distance between COM of the ligands CID 68723, CID 60838, CID 51049968, CID 73774610, and CID 135449332 is very close with Arg32 (between 4 and 7 Å), and they retain their position during the simulation.

Next, the binding free energies of the ligands were calculated by MM/PBSA analysis (Fig. 3B). In Fig. 4, the bar plots clearly demonstrated that 6 ligands among 10 hits have presented good binding free energy. Ligand CID 60838 has shown the best binding free energy with -64.45 kcal/mol. Ligand CID 51049968 has shown the second-best binding free energy with -42.13 kcal/mol. The energies of other ligands are listed in Fig. 4 and Table S1. (see Supporting information Table S1). Surprisingly, ligand CID 21252309 has demonstrated huge positive binding energy (1974 kcal/mol). When MM/PBSA calculation energies were checked (see Supporting information Table S2) for this ligand, a very large polar solvation energy was observed. This could be the explanation for such a large positive energy. Although it had a remarkable docking score (-8.7 kcal/mol), it was observed as the least stable ligand during the MD simulation (see MD movie_supporting information). Besides, the energy calculations table (Supporting information Table S2) shows that ligands CID 12940973, CID 53340771 and CID 54732242 have relatively small van der Waal energy, and high electrostatic energies compared to other ligands. Therefore, these three ligands have exhibited weaker binding energies with -2.21 kcal/mol, 0.39 kcal/mol, and 0.70 kcal/mol, respectively. These results also proved why a long-time MD simulation analysis and MM/PBSA calculations are necessary to decide the stability of the ligand–protein complexes.

The root mean square deviation (RMSD) plots of ten hit ligands were created to measure the difference between the protein's backbone atoms from its initial conformation to its final conformation. The stability of the protein can be determined by the fluctuations observed during the simulations. The smaller the fluctuations, the more stable the protein structure. In Fig. 5, RMSD plots show that the fluctuations of all proteins do not pass beyond 6 Å for the proteins, and 4 Å for the ligands. In the figure, it is clearly seen that the ligands which have closer contact with Arg32 have smaller RMSD values, and show fewer fluctuations during the 300 ns simulations, except for CID 73774610. Analysis of the dynamic behavior of the compounds helped us to understand the stabilization of the ligands in the binding site. Superpositions of the frames of the selected ligands were collected during

Table 1 Molecular docking results and interaction details of top 10 hit molecules against 2SHP protein

| PubChem ID/name | Main interacting residues | Distance (Å) | Interaction types | Binding affinity (kcal/mol) |
|-----------------------|---------------------------|--------------|----------------------------|-----------------------------|
| 60838/Irinotecan | Arg32 | 2.20 | Conventional hydrogen bond | −9.3 |
| | Thr12 | 2.92 | Conventional hydrogen bond | |
| | Trp248 | 3.84 | Π—sigma | |
| | Trp248 | 4.19 | Π—alkyl | |
| | Lys35 | 4.40 | Π—alkyl | |
| | Lys35 | 4.30 | Π—alkyl | |
| | Lys35 | 5.32 | Π—alkyl | |
| | Pro33 | 5.22 | Alkyl | |
| | Gly13 | 3.27 | Π donor—hydrogen bond | |
| | Gly13 | 3.25 | Π donor—hydrogen bond | |
| 51049968/Rimegepant | Arg32 | 2.45 | Conventional hydrogen bond | −9.2 |
| | Lys35 | 4.54 | Alkyl | |
| | Lys35 | 5.08 | Π—alkyl | |
| | Pro33 | 4.36 | Π—alkyl | |
| | Lys55 | 4.24 | Π—alkyl | |
| | His53 | 2.97 | Halogen (fluorine) | |
| 12940973 | Pro33 | 2.20 | Conventional hydrogen bond | −9.1 |
| | Pro33 | 2.80 | Conventional hydrogen bond | |
| | Ile11 | 2.26 | Conventional hydrogen bond | |
| | Lys35 | 3.82 | Alkyl | |
| | Lys35 | 3.52 | Alkyl | |
| | Lys55 | 4.18 | Alkyl | |
| | Lys55 | 4.45 | Alkyl | |
| | Trp248 | 4.35 | Π—alkyl | |
| 135449332/Eltrombopag | Arg32 | 2.75 | Conventional hydrogen bond | −9.1 |
| | Gly13 | 2.53 | Conventional hydrogen bond | |
| | Ser34 | 2.24 | Conventional hydrogen bond | |
| | Lys35 | 2.38 | Conventional hydrogen bond | |
| | Lys35 | 4.36 | Π—alkyl | |
| | Ly35 | 5.42 | Π—alkyl | |
| | Ser36 | 2.53 | Conventional hydrogen bond | |
| | Thr12 | 2.78 | Π—sigma | |
| | Pro33 | 4.78 | Π—alkyl | |
| | Lys55 | 4.19 | Alkyl | |
| 68723/Antrafenine | Arg32 | 2.37 | Conventional hydrogen bond | −8.9 |
| | Gly13 | 2.00 | Conventional hydrogen bond | |
| | Ile11 | 2.32 | Conventional hydrogen bond | |
| | Lys35 | 2.25 | Conventional hydrogen bond | |
| | Lys35 | 4.43 | Π—alkyl | |
| | Lys35 | 4.70 | Π—alkyl | |
| | Lys55 | 4.21 | Alky | |
| | Lys55 | 4.26 | Π—alkyl | |
| | Asn10 | 4.43 | Amide—Π stacking | |
| | Try248 | 5.59 | Π—alkyl | |
| | Pro33 | 5.19 | Π—alkyl | |
| Thr12 | 2.53 | Π—sigma | | |
| 54732242/Naldemedine | Arg32 | 1.84 | Conventional hydrogen bond | −8.8 |
| | Arg32 | 1.88 | Conventional hydrogen bond | |
| | Gly13 | 2.51 | Π donor—hydrogen bond | |
| | Gly13 | 2.88 | Carbon hydrogen bond | |
| | Lys35 | 2.67 | Conventional hydrogen bond | |
| | Lys35 | 3.37 | Π donor—hydrogen bond | |
| | Lys35 | 4.51 | Π—alkyl | |
| | Ser36 | 2.66 | Conventional hydrogen bond | |
| | Lys55 | 4.45 | Alkyl | |
| | Tyr66 | 4.71 | Alkyl | |

Table 1 (continued)

| PubChem ID/name | Main interacting residues | Distance (Å) | Interaction types | Binding affinity (kcal/mol) |
|---|---------------------------|----------------------------|----------------------------|-----------------------------|
| 73774610 | Arg32 | 2.96 | Conventional hydrogen bond | − 8.7 |
| | Lys35 | 2.44 | Conventional hydrogen bond | |
| | Lys35 | 4.98 | Π—alkyl | |
| | Lys35 | 5.01 | Π—alkyl | |
| | Gly13 | 2.61 | Π donor–hydrogen bond | |
| | Pro33 | 3.00 | Halogen (fluorine) | |
| | Lys55 | 4.10 | Π—alkyl | |
| | Lys55 | 5.24 | Π—alkyl | |
| 151223/Estriol 3-sulfate 16-glucuronide | Arg32 | 2.30 | Conventional hydrogen bond | − 8.7 |
| | Gly13 | 2.23 | Conventional hydrogen bond | |
| | Lys35 | 2.15 | Conventional hydrogen bond | |
| | Ser34 | 2.23 | Carbon hydrogen bond | |
| | His53 | 2.31 | Conventional hydrogen bond | |
| | Lys55 | 2.35 | Conventional hydrogen bond | |
| | Lys55 | 5.36 | Π—alkyl | |
| | Lys55 | 5.36 | Alkyl | |
| | Ile54 | 2.88 | Carbon hydrogen bond | |
| 21252309/Cholic acid glucuronide | Arg32 | 2.15 | Conventional hydrogen bond | − 8.7 |
| | Thr12 | 2.30 | Conventional hydrogen bond | |
| | Thr12 | 2.60 | Carbon hydrogen bond | |
| | Gly13 | 2.03 | Conventional hydrogen bond | |
| | Lys35 | 2.29 | Conventional hydrogen bond | |
| | Lys55 | 1.99 | Conventional hydrogen bond | |
| | Lys55 | 3.65 | Alkyl | |
| | Ile54 | 2.80 | Carbon hydrogen bond | |
| 53340771/Glpg-0187 | Arg32 | 3.89 | Π—cation | − 8.7 |
| | Arg32 | 1.92 | Conventional hydrogen bond | |
| | Ser34 | 2.17 | Carbon hydrogen bond | |
| | Ile11 | 2.20 | Conventional hydrogen bond | |
| | Lys35 | 1.95 | Conventional hydrogen bond | |
| | Lys35 | 4.60 | Π—alkyl | |
| | Thr12 | 2.32 | Conventional hydrogen bond | |
| | Trp248 | 5.34 | Π—alkyl | |
| | Lys55 | 5.35 | Alkyl | |
| | Lys55 | 3.91 | Π—alkyl | |
| | Tyr66 | 4.12 | Π—alkyl | |
| His53 | 2.48 | Conventional hydrogen bond | | |

the MD simulations shown in Fig. S4. These results also reflect the Fig. 5 which shows small RMSD fluctuations for compounds such CID 60838, CID 68723 (see the movies supplied as Supporting Information).

The root mean square fluctuation (RMSF) method predicts the protein's stability by estimating the average atomic flux of individual amino acids across the simulation period compared to the initial structure. When we examine the top three ligands, it can be easily observed that the fluctuation of amino acids for each protein–ligand complex mostly under 4 Å. These results indicate that the binding of ligands does not affect the stability of the protein. The RMSF plot of top three ligands is given in the supporting information (Fig. S5).

The radius of gyration (R_g) is also calculated from the MD simulation trajectories, which calculates the RMSD

values between the protein center of gravity and the protein residues. When a ligand binds to a protein, there is a conformational change that affects the radius of gyration. Investigation of R_g determines the compactness of the protein–ligand complex, and gives an idea about the stability of the protein. A small fluctuations in R_g indicates that the ligand does not change the protein conformation. R_g analysis of top 10 ligands showed similar R_g values in the range 2.5 to 2.6 nm. The radii of gyration showed a stable trend in all complexes during the 300 ns simulation (Fig. S6).

The plots of contact heatmaps of the ligands and 2SHP show the close interactions between the ligands and all amino acids in the phosphotyrosine binding site (Fig. 6). CID 60838 shows very close interaction with all amino acids in the binding site. The contact heatmap of CID 60838 shows that this ligand has demonstrated a tight interaction

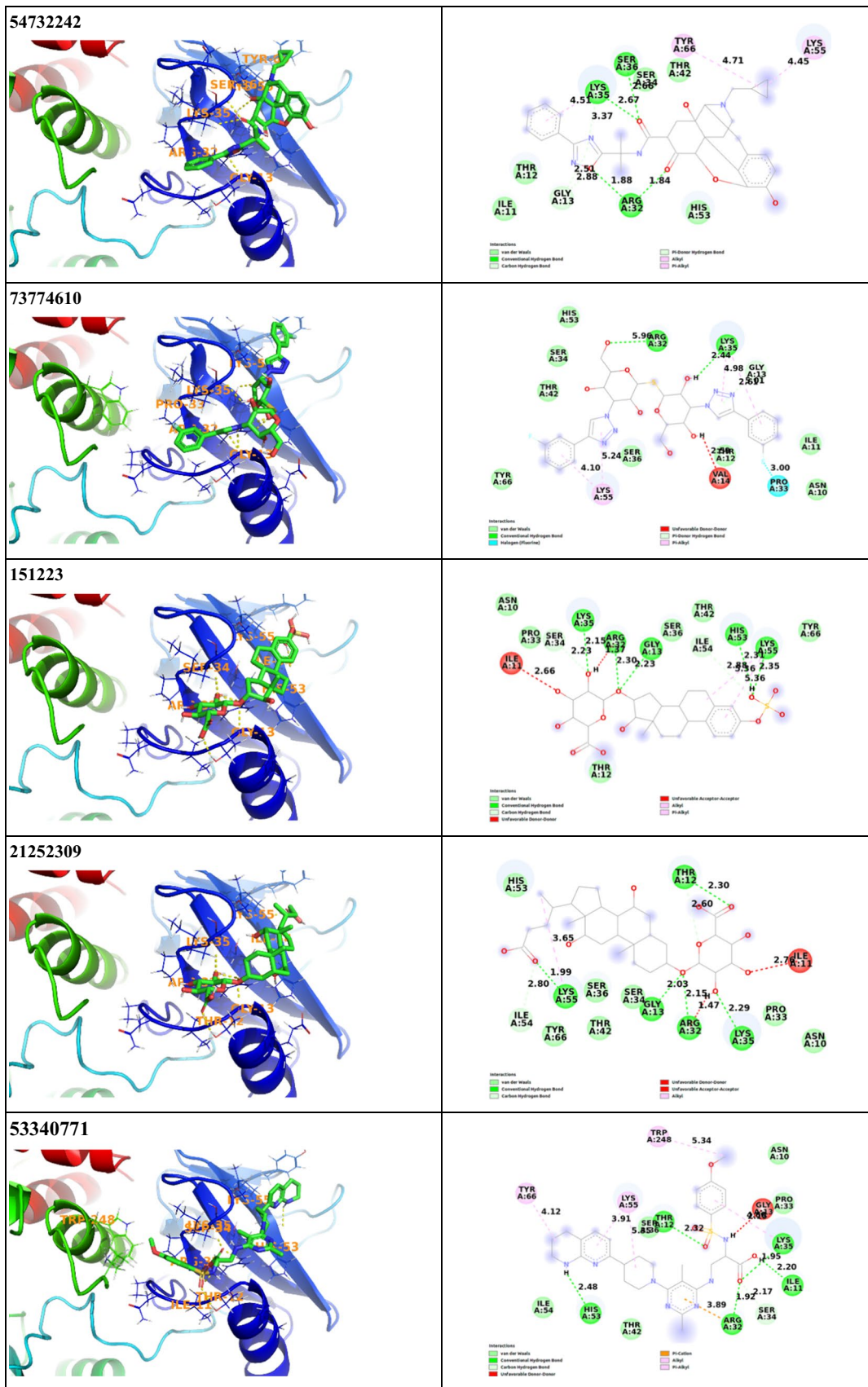


Fig. 2 (continued)

Fig. 3 **A** Average distance between COM of Arg32 and COM of the ligands and **B** MM/PBSA binding free energy plots of top ten ligands

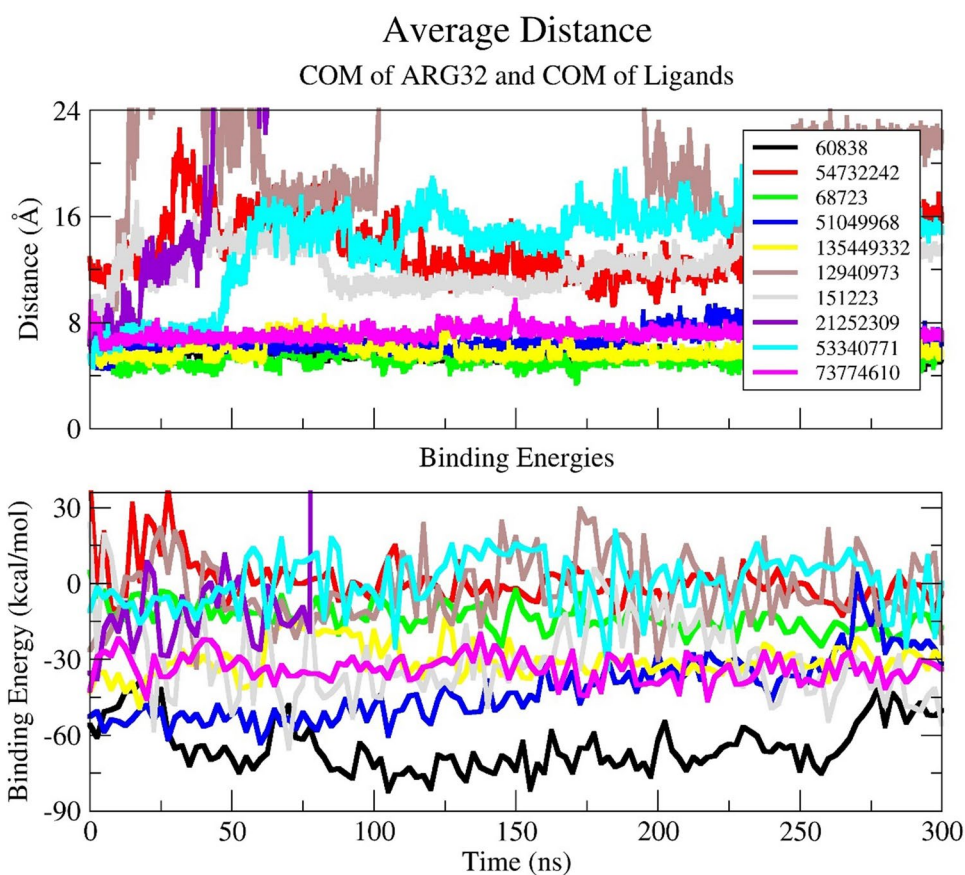
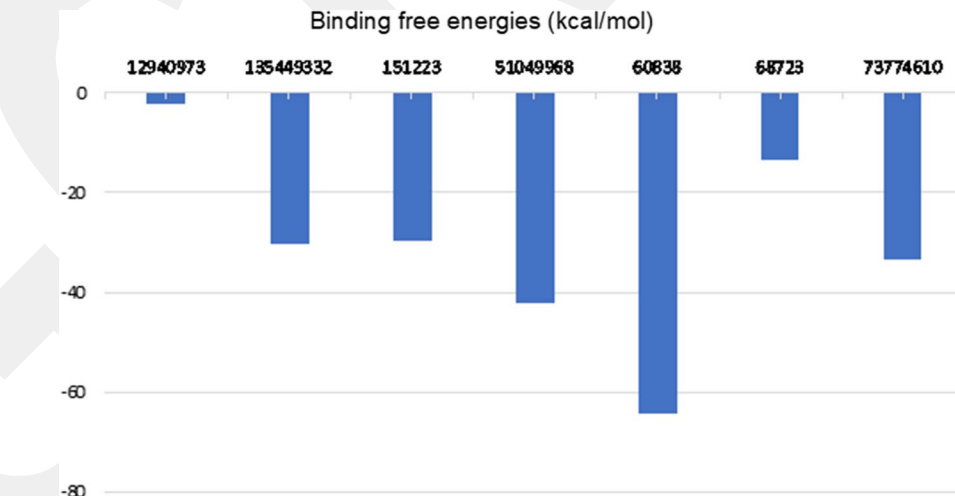


Fig. 4 MM/PBSA results for the 2SHP complexed with hit compounds. Only negative binding free energies are listed



(1–1.5 Å) with Arg32 from 100 to 300 ns of the simulation. Other close interactions were observed with CID 51049968, CID 135449332, CID 68723 and CID 73774610. The interaction distances of other ligands were observed longer than the interactions of CID 60838. These results are in line with the findings in Fig. 3A in which the distance between COM of ligands and COM of Arg32. All ligands which have a closer distance with COM of Arg32, have closer contacts

with all amino acids in the target phosphotyrosine binding site.

There are a lot of interactions to stabilize the ligand–protein complex including hydrogen bonds (H-bond), hydrophobic interactions, and electrostatic interactions. H-bond formation is especially crucial for the formation of more stable complexes. H-bond formation was investigated during the MD simulation and hydrogen bond plots

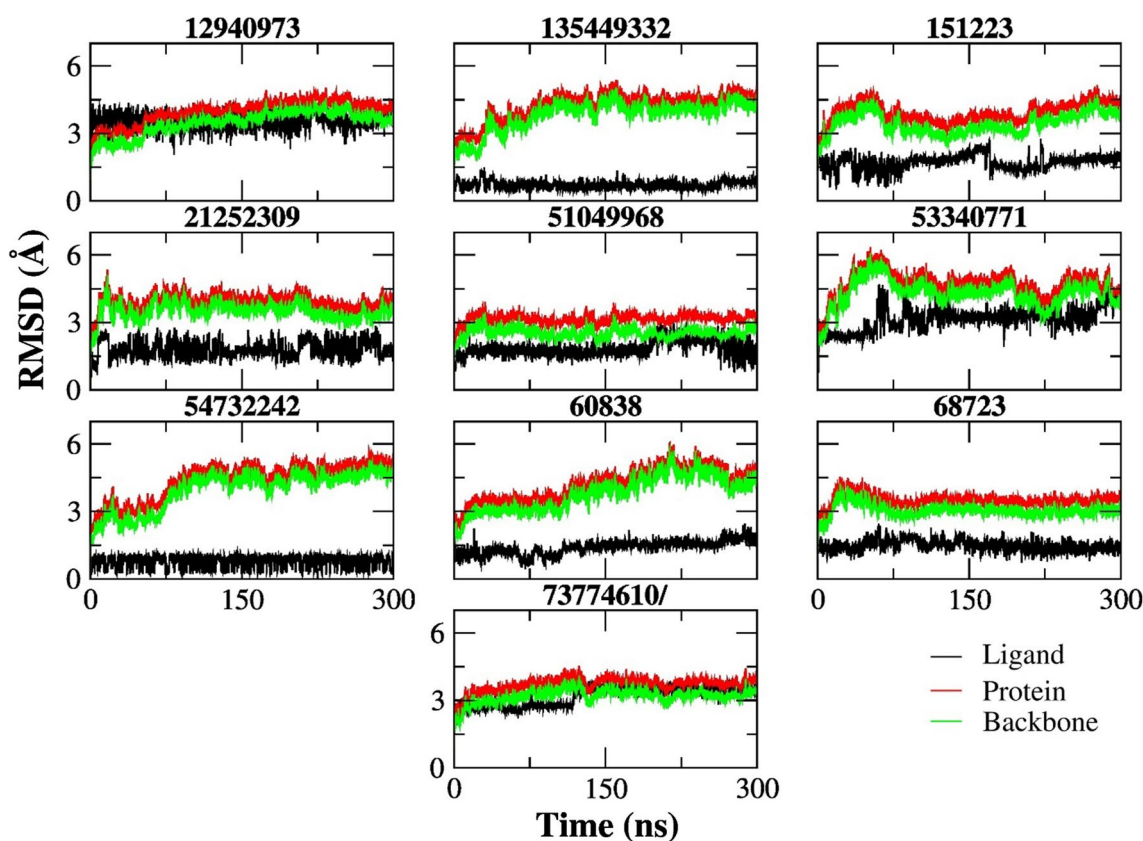


Fig. 5 RMSD plots of 10 hit compounds during 300 ns simulations

[number of the H-bonds vs. time (ns)] were demonstrated (Figs. S1–S3). Hydrogen bond plots show that all ligands have formed H-bonds with the target amino acids. These results were correlated with the molecular docking data (see Table 1 and Fig. 1). As mentioned before, all 10 hit ligands were able to make H-bonds. However, it was observed that some of the ligands lost their H-bond-making ability, or they could not make stable H-bonds during the simulation. According to H-bond analysis, the number of the H-bonds between the following ligands and protein decreased; CID 53340771, CID 21252309, and CID 12940973. After 100 ns of the simulation, the number of the H-bonds increased for the ligands CID 60838, CID 135449332, and CID 151223. In particular, CID 151223 formed 6 stable H bonds for the last 75 ns of the simulations.

Principal component analysis (PCA) was performed on the backbone atoms of the protein–ligand complexes to explore their conformational flexibility and the variety of conformations sampled during the 300 ns molecular dynamics simulations. The analysis focused on first eigenvectors of the top three ligands. The results are visualized in Fig. 7, which shows a 2D projection of the first two principal components.

To understand the protein's conformational changes, we performed a detailed analysis of its backbone motion throughout the simulations. We analyzed backbone motion for a deeper understanding of conformational dynamics in the simulations. This involved extracting the reference structure of the backbone from the first frame of each trajectory. This allowed us to track the relative motion of the backbone throughout the simulation for each complex. We then employed the complete trajectory data, encompassing atomic positions at each simulated time step, to calculate a covariance matrix. This matrix captures the coordinated motions of the backbone atoms, revealing the statistical relationships between individual atomic fluctuations. In simpler terms, the covariance matrix provides insights into how different backbone segments move together during the simulation. Building upon this information, we calculated the principal components (PCs) of the backbone motion. These PCs effectively identify the directions of greatest mobility. Finally, the trajectory was projected onto the top two PCs, resulting in a simplified 2D representation of the backbone's movement over time, as depicted in Fig. 7.

As evident in Fig. 7, the CID 60838/protein complex occupies a larger area in the PCA plot compared to the other complexes. This observation aligns well with the

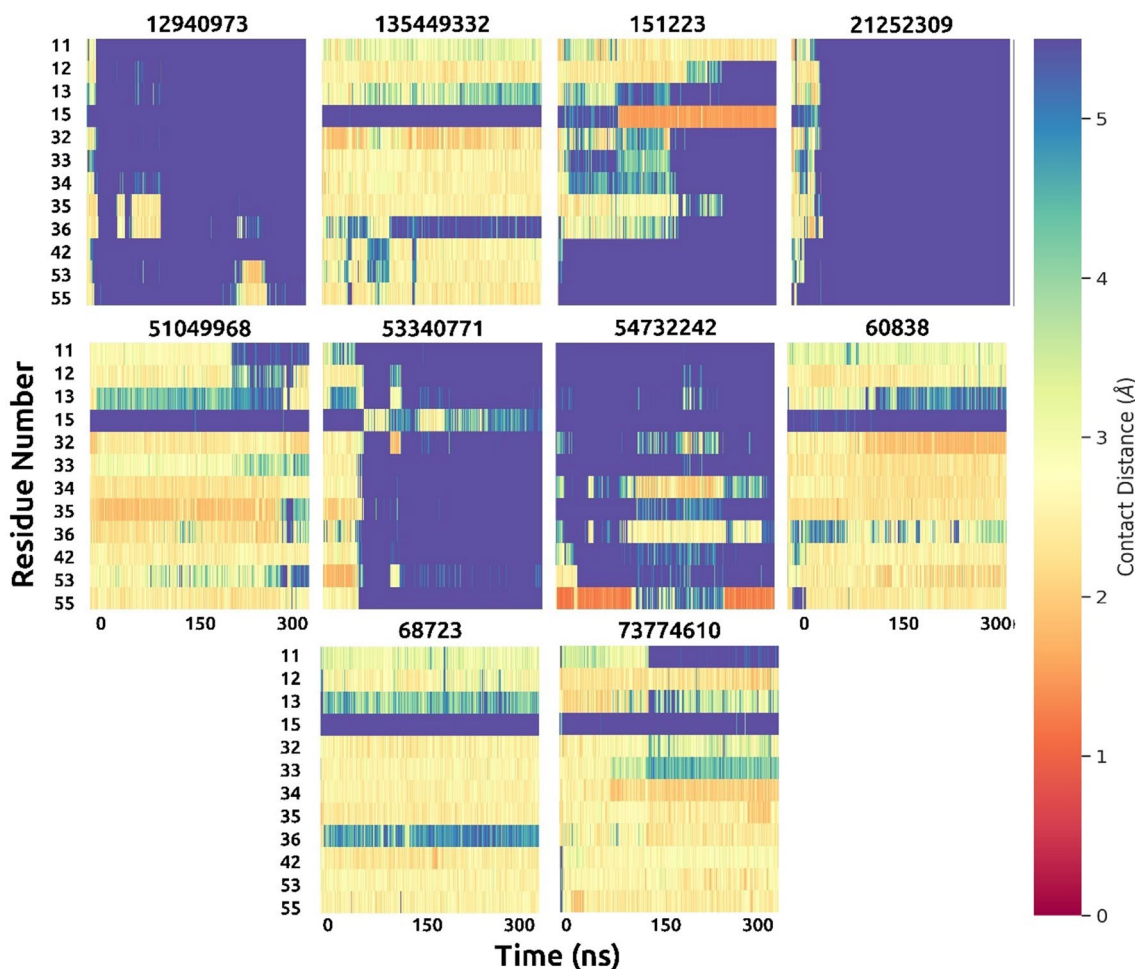


Fig. 6 Contact heatmaps of 10 hit compounds during 300 ns simulations. The contact distance is scaled between 0 and 5.5 Å. The warm colors (red to yellow) indicate closer contacts, whereas cold colors (green to violet) indicate farther distance, relatively

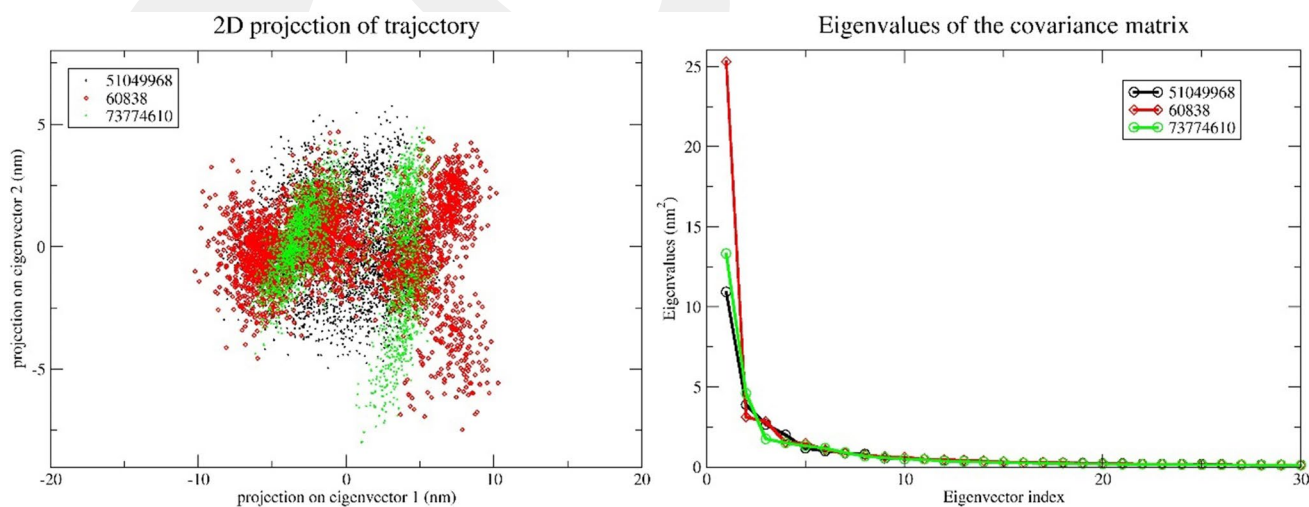


Fig. 7 Principal component analysis of top three protein–ligand complexes. The collective motion of CID 60838/protein (red), 7CID 3774610/protein (green), and CID 51049968 (black) using projec-

tions of MD trajectories on two eigenvectors corresponding to the first two principal components

detailed analysis of the backbone motion. The larger area signifies a higher diversity of conformations explored by the CID 60838 ligand during the simulation. Conversely, the tighter clustering of the other complexes suggests a more restricted conformational space. This implies that the CID 60838 ligand exhibits greater flexibility within the binding pocket, while the other ligands adopt more well-defined conformations. Interestingly, this increased flexibility of the CID 60838 ligand might be a contributing factor to its well-stabilized interaction with the protein throughout the simulation. The ability to sample a wider range of conformations could allow the ligand to form favorable interactions with different residues within the binding pocket, distributing the energetic burden and enhancing the overall stability of the complex.

It is worth mentioning that the compound CID 68723 shows why H-bond formation is very important for the stability of the molecules. When we examine the whole molecular docking and MD simulations results of CID 68723; it demonstrates 4 H-bonds with its best docking pose (Table 1), it is the closest ligand to the COM of Arg32 (Fig. 3A), it has very low RMSD values (Fig. 5), and it has remarkable close distance interactions with important residues in the binding site (Fig. 6). However, its binding free energy is higher (−13.43 kcal/mol) compared to other ligands since it makes a smaller number of H-bonds (Fig. S3) during 300 ns simulations.

Conclusions

SH2 domains are capable of recognizing phosphorylated tyrosine residues, and thus play important roles in cellular functions. This critical recognition occurs through the very well-conserved universal Arg residue which resides at the β B5 chain of SH2 domains at position 32 [37]. SH2 domains attract attention for targeting in a wide variety of disorders such as cancer, autoimmune diseases, and are potential targets for drug development [38]. The motivation of this study is to target Arg32 specifically to potentially inhibit the protein–protein interactions that occur via this specific residue. For this purpose, we have investigated the potential small molecule inhibitors of the N-SH2 domain of SHP2. After the molecular docking studies of a set of compounds retrieved from Broad Repurposing Hub and ZINC15, we focused on 10 compounds that have the best docking scores among the tested compounds for further MD analysis and binding free energy calculations by MM/PBSA. After all investigations and interpretations of *in silico* studies, CID 60838 (Irinotecan) has shown the best inhibitory potential against the N-SH2 domain of SHP2. Having a remarkable binding free energy (−64.45 kcal/mol), especially

interacting directly with Arg32 and other residues in the N-SH2 domain during 300 ns MD simulation makes this compound special. In conclusion, although further *in vitro* and *in vivo* studies are required, the gathered information in this study might provide a new direction for the design and development of the new inhibitors targeting the N-SH2 domain of SHP2.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11030-024-10880-2>.

Author contributions Conceptualization and methodology: İsmail Akçok; formal analysis and investigation: İsmail Akçok and Hüseyin Güner; writing—original draft preparation: Emel Başak Gencer Akçok and İsmail Akçok; writing—review and editing: Emel Başak Gencer Akçok and İsmail Akçok; supervision: İsmail Akçok.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). The authors did not receive support from any organization for the submitted work.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Consent to participate Not applicable.

Consent to publish Not applicable.

Ethical approval Not applicable.

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/>.

References

1. Pawson T (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* 116(2):191–203. [https://doi.org/10.1016/s0092-8674\(03\)01077-8](https://doi.org/10.1016/s0092-8674(03)01077-8)
2. Seet BT, Dikic I, Zhou MM, Pawson T (2006) Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* 7(7):473–483. <https://doi.org/10.1038/nrm1960>
3. Hunter T (2009) Tyrosine phosphorylation: thirty years and counting. *Curr Opin Cell Biol* 21(2):140–146. <https://doi.org/10.1016/j.ceb.2009.01.028>
4. Lim WA, Pawson T (2010) Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* 142(5):661–667. <https://doi.org/10.1016/j.cell.2010.08.023>

5. Liu BA, Shah E, Jablonowski K et al (2011) The SH2 domain-containing proteins in 21 species establish the provenance and scope of phosphotyrosine signaling in eukaryotes. *Sci Signal* 4(202):ra83. <https://doi.org/10.1126/scisignal.2002105>
6. Sha F, Gencer EB, Georgeon S et al (2013) Dissection of the BCR-ABL signaling network using highly specific monoclonal inhibitors to the SHP2 SH2 domains. *Proc Natl Acad Sci USA* 110(37):14924–14929. <https://doi.org/10.1073/pnas.1303640110>
7. Kraskouskaya D, Duodu E, Arpin CC, Gunning PT (2013) Progress towards the development of SH2 domain inhibitors. *Chem Soc Rev* 42(8):3337–3370. <https://doi.org/10.1039/c3cs35449k>
8. Jaber Chehayeb R, Boggon TJ (2020) SH2 domain binding: diverse FLVRs of partnership. *Front Endocrinol (Lausanne)* 11:575220. <https://doi.org/10.3389/fendo.2020.575220>
9. Waksman G, Kominos D, Robertson SC et al (1992) Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* 358(6388):646–653. <https://doi.org/10.1038/358646a0>
10. Bradshaw JM, Waksman G (2002) Molecular recognition by SH2 domains. *Adv Protein Chem* 61:161–210. [https://doi.org/10.1016/s0065-3233\(02\)61005-8](https://doi.org/10.1016/s0065-3233(02)61005-8)
11. Neel BG, Gu H, Pao L (2003) The ‘Shp’ing news: phosphatases in cell signaling. *Trends Biochem Sci* 28(6):284–293. [https://doi.org/10.1016/S0968-0004\(03\)00091-4](https://doi.org/10.1016/S0968-0004(03)00091-4)
12. Barford D, Neel BG (1998) Revealing mechanisms for SH2 domain mediated regulation of the protein tyrosine phosphatase SHP-2. *Structure* 6(3):249–254. [https://doi.org/10.1016/s0969-2126\(98\)00027-6](https://doi.org/10.1016/s0969-2126(98)00027-6)
13. Tartaglia M, Martinelli S, Stella L et al (2006) Diversity and functional consequences of germline and somatic PTPN11 mutations in human disease. *Am J Hum Genet* 78(2):279–290. <https://doi.org/10.1086/499925>
14. Frankson R, Yu ZH, Bai Y et al (2017) Therapeutic targeting of oncogenic tyrosine phosphatases. *Cancer Res* 77(21):5701–5705. <https://doi.org/10.1158/0008-5472.CAN-17-1510>
15. Chan RJ, Feng GS (2007) PTPN11 is the first identified proto-oncogene that encodes a tyrosine phosphatase. *Blood* 109(3):862–867. <https://doi.org/10.1182/blood-2006-07-028829>
16. Loh ML, Vattikuti S, Schubert S et al (2004) Mutations in PTPN11 implicate the SHP-2 phosphatase in leukemogenesis. *Blood* 103(6):2325–2331. <https://doi.org/10.1182/blood-2003-09-3287>
17. Tartaglia M, Niemeyer CM, Fragale A et al (2003) Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat Genet* 34(2):148–150. <https://doi.org/10.1038/ng1156>
18. Zhang J, Zhang F, Niu R (2015) Functions of Shp2 in cancer. *J Cell Mol Med* 19(9):2075–2083. <https://doi.org/10.1111/jcmm.12618>
19. Tartaglia M, Mehler EL, Goldberg R et al (2001) Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nat Genet* 29(4):465–468. <https://doi.org/10.1038/ng772>
20. Roberts AE, Allanson JE, Tartaglia M, Gelb BD (2013) Noonan syndrome. *The Lancet* 381:333–342. [https://doi.org/10.1016/S0140-6736\(12\)61023-X](https://doi.org/10.1016/S0140-6736(12)61023-X)
21. Gelb BD, Tartaglia M (2006) Noonan syndrome and related disorders: dysregulated RAS-mitogen activated protein kinase signal transduction. *Hum Mol Genet* 15:R220–R226. <https://doi.org/10.1093/hmg/ddl1197>
22. Wei W, Geer MJ, Guo X et al (2023) Genome-wide CRISPR/Cas9 screens reveal shared and cell-specific mechanisms of resistance to SHP2 inhibition. *J Exp Med* 220(5):e20221563. <https://doi.org/10.1084/jem.20221563>
23. Hof P, Pluskey S, Dhe-Paganon S et al (1998) Crystal structure of the tyrosine phosphatase SHP-2. *Cell* 92(4):441–450. [https://doi.org/10.1016/S0092-8674\(00\)80938-1](https://doi.org/10.1016/S0092-8674(00)80938-1)
24. Eastman P, Swails J, Chodera JD et al (2017) OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* 13(7):e1005659. <https://doi.org/10.1371/journal.pcbi.1005659>
25. Schmidtke P, Le Guilloux V, Maupetit J, Tufféry P (2010) fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res* 38(Web Server):W582–W589. <https://doi.org/10.1093/nar/gkq383>
26. Sterling T, Irwin JJ (2015) ZINC 15—ligand discovery for everyone. *J Chem Inf Model* 55(11):2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>
27. Bento AP, Hersey A, Félix E et al (2020) An open source chemical structure curation pipeline using RDKit. *J Cheminform* 12(1):51. <https://doi.org/10.1186/s13321-020-00456-1>
28. Koes DR, Baumgartner MP, Camacho CJ (2013) Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J Chem Inf Model* 53(8):1893–1904. <https://doi.org/10.1021/ci300604z>
29. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461. <https://doi.org/10.1002/jcc.21334>
30. Abraham MJ, Murtola T, Schulz R et al (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. <https://doi.org/10.1016/j.softx.2015.06.001>
31. Dodda LS, De Vaca IC, Tirado-Rives J, Jorgensen WL (2017) LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res* 45:331–336. <https://doi.org/10.1093/nar/gkx312>
32. Homeyer N, Gohlke H (2012) Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. *Mol Inform* 31(2):114–122. <https://doi.org/10.1002/minf.20110135>
33. Kumari R, Kumar R, Consortium OSDD, Lynn A (2014) g_mmpbsa—a GROMACS tool for MM-PBSA and its optimization for high-throughput binding energy calculations. *J Chem Inf Model* 54(7):1951–1962. <https://doi.org/10.1021/ci500020m>
34. Songyang Z, Shoelson SE, McGlade J et al (1994) Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Mol Cell Biol* 14(4):2777–2785. <https://doi.org/10.1128/mcb.14.4.2777>
35. Grucza RA, Bradshaw JM, Fütterer K, Waksman G (1999) SH2 domains: from structure to energetics, a dual approach to the study of structure–function relationships. *Med Res Rev* 19(4):273–293. [https://doi.org/10.1002/\(sici\)1098-1128\(199907\)19:4%3c273::aid-med2%3e3.3.co;2-7](https://doi.org/10.1002/(sici)1098-1128(199907)19:4%3c273::aid-med2%3e3.3.co;2-7)
36. Salo-Ahen OMH, Alanko I, Bhadane R et al (2021) Molecular dynamics simulations in drug discovery and pharmaceutical development. *Processes* 9(1):71. <https://doi.org/10.3390/pr9010071>
37. Marasco M, Carlomagno T (2020) Specificity and regulation of phosphotyrosine signaling through SH2 domains. *J Struct Biol X* 4:100026. <https://doi.org/10.1016/j.yjsbx.2020.100026>
38. Diop A, Santorelli D, Malagrino F et al (2022) SH2 domains: folding, binding and therapeutical approaches. *Int J Mol Sci* 23(24):15944. <https://doi.org/10.3390/ijms232415944>

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

GCPRIS