Bioorganic Chemistry 105 (2020) 104391

Contents lists available at ScienceDirect

Bioorganic Chemistry

journal homepage: www.elsevier.com/locate/bioorg

Scaffold-based selective SHP2 inhibitors design using core hopping, molecular docking, biological evaluation and molecular simulation

Wei-Ya Li [a,1], Ying Ma [a,1,*], Hao-Xin Li [a], Xin-Hua Lu [b], Shan Du [a], Yang-Chun Ma [a], Liang Zhou [a], Run-Ling Wang [a,*]

[a] Tianjin Key Laboratory on Technologies Enabling Development of Clinical Therapeutics and Diagnostics (Theranostics), School of Pharmacy, Tianjin Medical University, Tianjin 300070, China

[b] New Drug Research & Development Center of North China Pharmaceutical Group Corporation, National Microbial Medicine Engineering & Research Center, Hebei Industry Microbial Metabolic Engineering & Technology Research Center, Shijiazhuang 050015, Hebei, China

ARTICLE INFO

Keywords:
SHP2
Inhibitor
Core hopping
MD simulation
Phosphatase activity assay

ABSTRACT

PTPN11 (coding the gene of SHP2), a classic non-receptor protein tyrosine phosphatase, is implicated in multiple cell signaling pathway. Abnormal activation of SHP2 has been shown to contribute to a variety of human diseases, including Juvenile myelomonocytic leukemia (JMML), Noonan syndrome and tumors. Thus, the SHP2 inhibitors have important therapeutic value. Here, based on the compound PubChem CID 8,478,960 (IC_{50} = 45.01 μM), a series of thiophene [2,3-d] pyrimidine derivatives (IC_{50} = 0.4–37.87 μM) were discovered as novel and efficient inhibitors of SHP2 through powerful "core hopping" and CDOCKER technology. Furthermore, the SHP2-PTP phosphatase activity assay indicated that Comp#5 (IC_{50} = 0.4 μM) was the most active SHP2 inhibitor. Subsequently, the effects of Comp#5 on the structure and function of SHP2 were investigated through molecular dynamics (MD) simulation and post-kinetic analysis. The result indicated that Comp#5 enhanced the interaction of residues THR357, ARG362, LYS366, PRO424, CYS459, SER460, ALA461, ILE463, ARG465, THR507 and GLN510 with the surrounding residues, improving the stability of the catalytic active region and the entrance of catalytic active region. In particular, the Comp#5 conjugated with residue ARG362, elevating the efficient and selectivity of SHP2 protein. The study here may pave the way for discovering the novel SHP2 inhibitors for suffering cancer patients.

1. Introduction

JMML is a fatal childhood disease clinically in which JMML cells hold the capabilities to infiltrate non-hematopoietic organs, impair their functions, and even develop into acute myeloid leukemia [1,2]. The main features of JMML are reflected in the over-production of myelomonocytic cells, hepatosplenomegaly, thrombocytopenia, progressive anemia and high fetal hemoglobin levels[3]. JMML has a poor prognosis, most of which have a survival period less than 2 years[4]. Noteworthy, due to the heterogeneity of the course of the disease, about 1/3 of the children died within a few months regardless of the treatment, which was manifested as rapid organ enlargement and cachexia and bone marrow failure.

The PTPN11, encoding tyrosine phosphatase SHP2 protein, was found to be ubiquitous in JMML[5,6]. Interestingly, the SHP2 phosphatase played a completely active role in receptor kinase-initiated signal transduction, especially in the Ras pathway. Although the pathogenesis of JMML was believed to be the Ras over-activation, Ras effect or similar pathways have also been implicated in the disease, like the PI3K/Akt signaling pathway[7]. Previous research showed that SHP2 was widely expressed in the cytoplasm downstream of multiple receptor tyrosine kinases and was implicated in many oncogenic cells signaling

Abbreviations: SHP2, protein tyrosine phosphatase-2; MD simulation, molecular dynamic; JMML, Juvenile myelomonocytic leukemia; PCA, principal component analysis; pNPP, p-nitrophenyl phosphate; RMSD, root mean square deviation; NVT/NPT, constant number of particles, volume/pressure, and temperature; ADMET, absorption, distribution, metabolism, excretion, toxicity; DCCM, dynamic cross correlation maps; RMSF, root mean square fluctuation; SPC, single-point charge; DS, Discovery Studio; MM-PBSA, molecular mechanics Poisson Boltzmann surface area; RING, residue interaction network generator.

* Corresponding authors.
E-mail addresses: maying@tmu.edu.cn (Y. Ma), wangrunling@tmu.edu.cn (R.-L. Wang).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bioorg.2020.104391
Received 28 April 2020; Received in revised form 7 October 2020; Accepted 15 October 2020
Available online 21 October 2020

cascades[8,9]. It maintained the activation of Ras and regulates PD-1 and BTLA to inhibit the immune receptor pathway, as well as played a vital role in the PI3K-AKT, JAK-STAT and NF-κB signaling pathways [10,11]. In summary, accumulated biochemical studies demonstrated that SHP2 regulated the growth factor receptor signaling, and was considered as a potential target, and thus, SHP2 has key implications for the development and treatment of JMML and other SHP2-related cancers.

In recent years, PTP phosphatase inhibitors have been widely applied in research. Compound II-B08, as a reversible inhibitor of SHP2 (IC50 = 5.5 μmol/L), could inhibit the activation of ERK1/2 stimulated by growth factors (such as EGF), and hinder the growth of NSCLC cells in vitro and in vivo[12]. At the same time, the compound has an in vivo inhibitory effect on mast cell leukemia model. Indole salicylic acid SHP2 inhibitor 11a-1 blocked growth factor-mediated activation of ERK1/2 and AKT, and exhibited antiproliferative activity in lung cancer, breast cancer, and leukemia cell lines[13]. Furthermore, the compound PHPS1 effectively inhibits the activation of ERK1/2 by the leukemia-related SHP2 mutant SHP2-E76K, and blocks the non-adherent growth of various human tumor cell lines[14,15]. The results indicated that the application of PTP inhibitors will play an important role in future cancer research.

In this study, the lead compound PubChem CID 8,478,960 (IC_{50} = 45.01 μM) was obtained by searching PubChem database. The Core Hopping method and CDOCKER technology were employed to obtained a variety of thiophene [2,3-d] pyrimidine derivatives by modifying the polar head and hydrophobic tail of the PubChem CID 8478960. The PTP phosphatase activity assay showed that the compound 2-(7,7-difluoro-4-oxo-5,6,7,8-tetrahydrobenzo[4,5]thieno[2,3-d]pyrimidin-3(4H)-yl)-N-(2,4-difluorophenyl)acetamide (Comp#5) was identified as the most effective inhibitor of SHP2 with an IC_{50} value of 0.4 μM. Furthermore, the ADMET prediction indicated that Comp#5 was consistent with the rule of drug synthesis. Afterward, MD simulation and post-kinetic analysis were conducted to study the conformational changes caused by the interaction between Comp#5 and SHP2, and to clarify the molecular mechanism of Comp#5 inhibiting SHP2 activity. With Comp#9 (IC_{50} value > 100 μM) as a negative control, the SHP2-Comp#9 system was also subjected to MD simulation and a series of post-analysis. Fortunately, Comp#5 properly attached to the active pocket of SHP2 protein. The interactions between Comp#5 and the residue ARG362 located in the entrance of catalytic active region improved the efficiency and even the selectivity of the SHP2 inhibitors. While, Comp#9 was far from the active pocket of SHP2 and formed weak interactions with key amino acids. In the RMSF analysis, the fluctuations in the β5-β6 loop, P-loop and Q-loop region of the SHP2 protein were larger than those in the SHP2-Comp#5 protein, indicating that these regions were more stable in the SHP2-Comp#5 protein. Nevertheless, the fluctuations of these regions in SHP2-Comp#9 system were not significantly different from those in SHP2-without ligands system. In addition, DCCM analysis showed that the β5-β6 loop region and the WPD-loop region, the P-loop region and the β5-β6 loop region, the Q-loop region and the WPD-loop region were weakly correlated with each other in the SHP2-without ligands system and SHP2-Comp#9 system, respectively. However, the correlations between amino acids were heightened in SHP2-Comp#5 protein. RIN analysis showed that there were slight interactions between amino acids in the SHP2-without ligands system and SHP2-Comp#9 system, whereas the residues THR357, ARG362, LYS366, PRO424, CYS459, SER460, ALA461, ILE463, ARG465, THR507 and GLN510 formed hydrogen bonds or van der Waals interactions with surrounding amino acids in SHP2-Comp#5. The results suggested that the Comp#5 enhanced the interactions of amino acids within the catalytic activity region (P-loop and Q-loop) and the entrance of catalytic active region (β5-β6 loop) in SHP2-without ligands protein, improving the stability of the system structure. Therefore, it could be concluded that Comp#5 inhibited the activity of SHP2 protease.

2. Materials and methods

2.1. Virtual screening and molecular docking based on the core hopping method

The drug-like database (subset of ZINC database) was used to screen the lead compound based on the conformation of the active site in SHP2 [16]. The "Core Hopping" approach was then preformed to modify the lead compound.

The primary characteristic of the Core Hopping was to utilize the principle of bioisosteres to replace the inefficient or ineffective groups in the template ligands, aiming at reducing the toxic and side effects and improving the curative effect [17,18]. In the research, the Core Hopping method was applied during the molecular docking procedure to receive molecules with good activity, including the improvement of binding affinity, the increase/decrease of water solubility, the improvement of Pharmacokinetics and pharmacokinetics [19]. In general, the core hopping process could be divided into four major steps. In the first step, the possible points where the core attached to the scaffold were defined. It was performed in the module of "Template combinatorial definition file". In the second step, the Receptor Preparation panel was employed to define "the receptor grid file". In detail, the original state of SHP2 complexed with JZG (PDB ID: 3O5X) interaction was kept to determine the active site, which would interfere with the search space of docking algorithm. The protocol "from current selection" was used to determine the binding position (spherical box). In the third step, the cores connected to the scaffold was prepared using a fragment database from ZINC [19,20]. In the final step, the entire molecular structures constructed from the core and scaffold were submitted to the "Protocore Docking" module to align and dock with the SHP2 receptor. The final molecules with better docking scores than the original molecule was kept [21].

All compounds obtained by core hopping process were pretreated by LigPre module to generate possible states. Afterwards, the pose of each compound was responsible to dock into the active site of the SHP2 protein [22]. Ultimately, discovery studio v3.5 (DS v3.5) visualizer was applied to analyze the interactions between the inhibitor and the SHP2 protein.

2.2. Chemistry

Unless otherwise specified, all the reagents were purchased from commercial suppliers and could be used without further purification. The TLC were used to monitored reaction courses on silica gel precoated F254 Merck plates. Developed plates were examined with UV lamps (254 nm). The physical, analytical and spectral data (^{1}H NMR, ^{13}C NMR and ESI-MS) were employed to characterized the intermediates and all the new compounds. The solvents were DMSO-d_6 and $CDCl_3$.

2.3. SHP1/SHP2 phosphatases activity assay

Purified SHP1 full-length protein and PTP domain protein by GST-tag Protein Purification Kit and Ni-NTA-Sefinosen Column were used as enzymes, respectively. It was determined that p-nitrophenyl phosphate (pNPP) was tested as substrates for phosphatase activity assay [23]. In this assay, protease dephosphorylate with pNPP to generate PNP, which had specific fluorescence at 405 nm. Briefly, Purification of recombinant SHP1 full-length protein and SHP2-PTP domain protein (0.4 μg) in 60 μL buffer containing 50 mM NaCl, 25 mM Tris-HCl (pH 7.0), 1 mM dithiothreitol (DTT), 2.0 mM EDTA, and 0.05% Tween-20 and test compounds were mixed, and then incubated under the condition of 37 °C. Detect A_{405} value after 2 min. Afterwards, the pNPP was maintained at a concentration of 2 mM and incubated at 37 °C for 30 min. The reaction was stopped by Add 50 μL 3 M NaOH to the reactant and chilled it quickly on ice to stop the reaction. The OD_{405} was then measured.

2.4. Fluorescence titrations assay

Fluorescence quenching experiment referred to the physical or chemical reaction process between fluorescent substance and solvent molecule that lead to the change of fluorescence intensity or related change of excitation peak or fluorescence peak. To confirm the binding degree of the inhibitor to PTP domain protein, fluorescence titration was employed to determine whether the binding of the compound changed the fluorescence of PTP domain protein. In this experiment, the purified PTP protein was diluted into the reaction solution and maintained at a protein concentration of 0.2 mg/mL. Titration was then performed by changing the concentration of the inhibitor by a gradient. Fluorescence spectra were recorded with Synergy 2 purchased by BioTek. The fluorescence contribution of the inhibitor as background fluorescence will be subtracted.

2.5. Molecular dynamic simulation

Using the prepared SHP2 model and SHP2-Comp#5 complex model as the initial structures, the GROMACS v4.5.5 software was applied for the 100 ns MD simulation [24]. First, the GROMOS 43a1 force field was used to obtain the documents required by Gromacs (molecular topology file, molecular structure file, kinetic parameter file, running input file and trajectory file). Afterwards, the model was confined to a dodecahedral box solvated with SPC216 water model (the box boundary was at least 1 nm from the model surface)[25] and the system was neutralized with appropriate Na^+ or CL^- ions. To ensure the rationality of protein conformations and solvent, the equilibrium was carried out before MD simulation. The steepest descent algorithm was applied on the initial structures to escape the steric clashes and the unsuitable geometry, thereby relaxing the initial structures with 10,000 steps with a maximum force below 1000 kJ/mol or no drastic energy changes. Subsequently, NVT and NPT simulation were conducted to equilibrate the system, respectively [26]. The V-rescale and the Parrinello–Rahman was used to equilibrate the system was under constant temperature (310 K) and pressure (1 bar) for 100 ps, respectively. Meanwhile, the LINCS algorithm constrained all bonds and angles of the system [27]. The long-range electrostatic potential was calculated using the PME method [28,29]. The neighbors list updated every 10 steps. Finally, the models performed 100 ns MD simulation with the trajectory recorded at intervals of 20 ps.

2.6. MM-PBSA free energy calculations

Based on the molecular mechanics Poisson-Boltzmann surface area algorithm, the binding free energy of receptor-ligand interaction was estimated by the g_mmpbsa tool[30]. A total of 950 frames was extracted from the last 95 ns MD trajectory with an interval of 100 ps [31]. The major residues playing an important role in binding free energy provided clear insight into the molecular mechanisms by which proteins interacted with their ligands [32]. For the calculation of MMPBSA free energy, brief instructions were provided as followed [33]:

$$\Delta G_{bind} = G_{complex} - \left(G_{receptor} + G_{ligand}\right) \quad (1)$$

$$\Delta G_{bind} = \Delta E_{gas} + \Delta G_{sol} - T\Delta S \quad (2)$$

$$\Delta E_{gas} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdw} \quad (3)$$

$$\Delta G_{sol} = \Delta G_{ele,sol} + \Delta G_{nonpl,sol} \quad (4)$$

$$\Delta G_{nonpl,sol} = \gamma \Delta SASA \quad (5)$$

The binding free energy (ΔGbind) was calculated by the gas-phase free energy, the solvation free energy and the entropy term (−TΔS). The gas-phase free energy (ΔEgas) was composed of internal energy, the electrostatic and VDW interactions. The solvation free energy was divided into polar contribution and non-polar contribution.

2.7. ADMET analysis

ADMET stands for the five major characteristics of pharmacokinetics: absorption, distribution, metabolism, excretion, and drug toxicity. The pharmacokinetic properties of drugs in the human body were assessed through the module of "ADMET Descriptions" in the DS v3.5, including human intestinal absorption (HIA), blood–brain barrier (BBB), water solubility, plasma protein binding (PPB), cytochrome P450 CYP2D6 binding and liver toxicity. Through ADMET screening, compounds conforming to drug potency would be obtained, which reduced the loss in drug discovery and development.

2.8. Principal component analysis

The Principal component analysis (PCA analysis), a technique for reducing data complexity and extracting collaborative motion, was used to reveal the movements of structures[34]. In general, the first few principal components (PCs) were used to describe the most important slow-mode biological functional movements of the system[35]. Furthermore, PCA analysis could identify the most significant fluctuation pattern of protein in terms of planar motion (eigenvector) and amplitude (eigenvalue). Specifically, the PCA analysis first removed the global rotation and motion of the system through the least square fitting, and then constructed the covariance matrix according to the three-dimensional position fluctuation of the given set of atoms at its overall average position[36]. The internal motion of the correlation between the molecule and N atoms is obtained by diagonalizing the covariance matrix[37]:

$$\sigma_{ij} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle$$

where the ri/rj represents the Cartesian coordinate of the i-th/j-th atom of the system, and "<>" calculates the average of ensemble. The PCA scatter plots were obtained by the Bio3D library of R base package software.

2.9. Correlation motion analysis within the system

Dynamic cross correlation matrix (DCCM) illustrated the time-correlated information between system atoms i and j, and it could identify concerted and correlated motion. DCCM analysis had three typical characteristics, including strong cross-correlation along the diagonal, cross-correlation emanating from the diagonal, and non-diagonal cross-correlation. The atoms along the diagonal always showed strong correlated fluctuations with the Cij values are equal to 1. These interrelationships provide dynamical information about an average and/or the tertiary structure derived from the experiment. The normalized correlation information between the protein atoms i and j were calculated by[38]:

$$C_{i,j} = \frac{\langle \Delta r_i * \Delta r_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}}$$

where the $\Delta r_i/\Delta r_j$ expressed the mean position of i-th/j-th atom and the angle brackets "<>" indicated the global average of MD trajectory. Ci, j = 1 means completely coupled, while Ci, j = − 1 means anti-correlated motion. The DCCMs were constructed by the Bio3D library of R base package software[39].

2.10. The interaction of amino acids within a protein

The average MD trajectory of the 5 ns-100 ns simulation was extracted to submit to the Residue Interaction Network Generator v2.0.1 (RING v2.0.1) server for post dynamic analysis[40]. The interactions

between residues constituted a complex network matrix, which was visualized by the plugin "NetworkAnalyzer"[41] integrated in Cytoscape. Nodes were used to represent residues or ligands, and lines were employed to describe the residue-residue interactions, including hydrogen bonding, van der Waals interactions, ionic interaction and salt bridges. In addition, the "RINspector" and "RINalyzer" plug-ins[42] in Cytoscape were applied to find required network parameters for each system, such as closeness centrality and shortest path betweenness. Subsequently, to highlight and investigate the similarity and difference of internal relations between SHP2-without ligands and SHP2-ligand systems, a combined comparison network of the two systems was generated based on the superposition and alignment of corresponding network matrix. In undirected networks, the betweenness centrality reflected the degree to which a node controled the interaction of other nodes, while the closeness centrality measured the speed at which information spreaded from a given node to other reachable nodes[43].

3. Result and discussion

3.1. Structural design and virtual screening of SHP2 inhibitors

The researches have reported that the active site of SHP2 protein contained four major loop regions, including pTyr recognition loop (ARG277-PRO284), WPD-loop (ARG421-ASP431), P-loop (HIS458-ARG465) and Q-loop (ARG501-THR507) (Fig. S8). Specifically, pTyr-loop region identified Tyr specifically. The P-loop region contains active nucleophiles CYS459 and ARG465, which can be used to identify the phosphorylated base in the substrate. The general acid-base catalyst ASP425 located in the WPD-loop region. Besides, the GLN506 located in the Q-loop region could hydrolyze phosphatase intermediate. The second binding site (ARG362/LYS364) was responsible for the enhanced selectivity[44]. Therefore, the development of inhibitors that could occupy these active sites simultaneously and the second binding site might yield high affinity and selectivity. In this paper, thiophene [2,3-d] pyrimidine as lead compounds were obtained by virtual screening. Divided the structure of the lead compound PubChem CID 8,478,960 was into three parts, Core A, Core B, and Core C, as marked by red line, blue line and purple line, respectively (Fig. 1). The core A and core B were placed in the polar region and formed hydrogen bond interaction with residues ARG465, GLN506 and GLN510, formed π-π conjugate interaction with residue LYS366. In addition, core C was placed in the hairpin structure of the β5-β6-loop region and formed π-π conjugate interaction with ARG362/LYS364 which were likely responsible for the enhanced potency and selectivity. On this basis, fragment A and fragment B were obtained by searching the ZINC-fragment database, with the purpose of making more interactions between the inhibitor and receptor residues LYS366, ARG465, GLN506 and GLV510. Two cores were generated for the core hopping operation to core A part, namely, core A1 and core A2, to replace core A; two cores were generated for the core hopping operation to core B part, namely, core B1 and core B2, to replace core B. Fragment C was obtained by searching the aromatic database with the aim of making the inhibitor interact with the ARG362/LYS364 in a stable conjugated contact, enhancing the potency and selectivity of the inhibitor. Four cores were generated for the core hopping operation to core C part, namely, core C1, core *C*2, core C3, and core C4, to replace core C. Consequently, a total of $2 \times 2 \times 4 = 16$ different combinations for the thiophene [2,3-d] pyrimidine derivatives were thus generated.

The docking score are used to evaluate the compounds obtained by core hopping method. In the first instance, re-docking method was preformed to assess the docking accuracy. The root-mean-square deviation (RMSD) between the docked conformation and co-crystallization conformation was calculated with the value of 1.698 Å (no>2 Å), which proved that the methods and the parameters was satisfactory in this study. Subsequently, the candidate compounds with docking scores

Table 1
Docking score of ten candidate molecules by molecular docking.

Compound	Docking score (kJ/mol) SHP2	SHP1
Comp#1	−10.5038	−6.7001
Comp#2	−11.9772	−7.6455
Comp#3	−20.6249	−11.0057
Comp#4	−19.6624	−10.3425
Comp#5	-22.5625	−12.3831
Comp#6	−19.3184	−10.3663
Comp#7	−17.9907	−8.8527
Comp#8	−14.2884	−7.4418
Comp#9	−11.3207	−7.2914
Comp#10	−10.4864	−6.5926

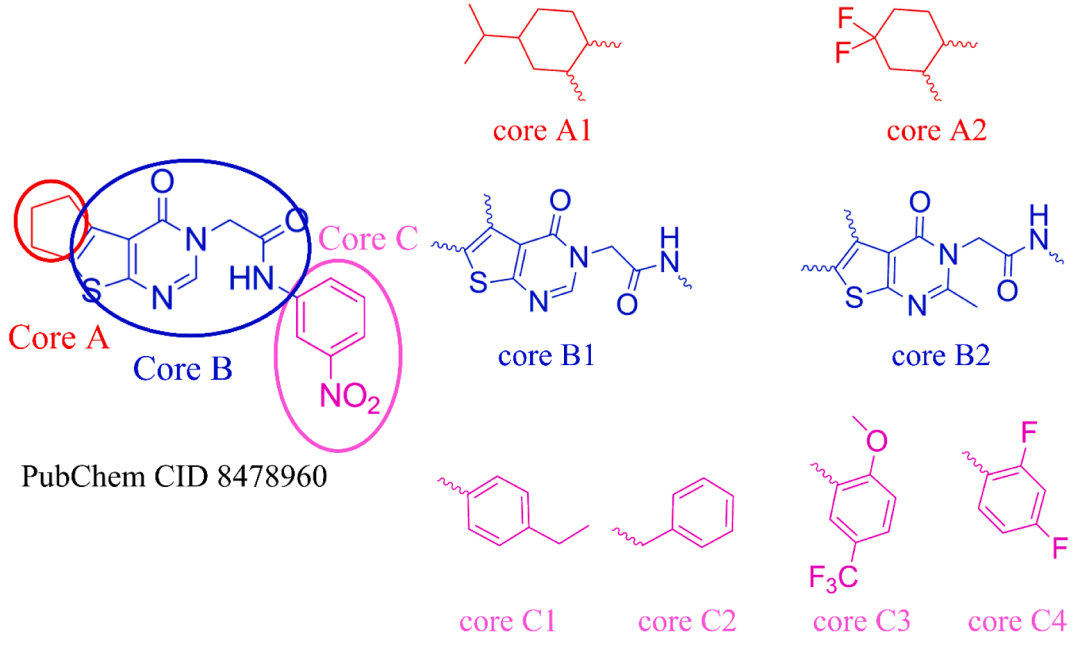

Fig. 1. Generation of the core fragments from thiophene [2,3-d] pyrimidine ring by core hopping method.

in the top ten were calculated as shown in Table 1. It was obviously that most compounds had good affinity with the SHP2 system, among which Comp#5 ranked the first in docking score (–22.5625 kJ/mol). Interestingly, their affinity with SHP1 protein was weak. The synthesis of these ten compounds were then described in details (Fig. 2). The structures of the synthesized compounds were characterized by ^{1}H NMR and ^{13}C NMR (Supplementary material).

3.2. Synthesis of thiophene [2,3-d] pyrimidine derivatives

3.2.1. A general method for preparing thiazolamide from substituted cyclohexanone. The synthesis of isopropyl 2-amino-6-isopropyl-4,5,6,7-tetrahydrobenzo[b]thiophene-3-carboxylate

Under N_2 atmosphere, a mixture of cyclohexanone derivatives-1a, 1b (5.0 g, 35.7 mmol), isopropyl cyanoacetate (5.0 g, 39.2 mmol), sulphur powder (1.7 g, 53.6 mmol) in absolute ethanol (100 mL) were heated and stirred in an oil bath. Morphine was then slowly added dropwise to the mixture, and the temperature was maintained at 35° C for 24 h. The reaction liquid changed from a light yellow to a dark yellow when most of the starting materials were converted into the target compound. TLC (ethyl acetate/petroleum ether = 1/3) detected that the raw material completely disappeared.

At room temperature, the reaction solution was diluted with 100 mL of water. Most of the ethanol was removed by vacuum distillation at 45 °C. The mixture was extracted by 150 mL ethyl acetate, and the combined organic layer was washed with saturated brine, dried over anhydrous magnesium sulfate, filtered, and concentrated by rotary evaporation.

3.2.2. General procedure for the synthesis of thiazolopyrimidinone from thiazolamide

Under N_2 atmosphere, the compound of isopropyl 2-amino −4,5,6,7-tetrahydrobenzo[*b*]thiophene-3-carboxylate-2a, 2b (3.5 g, 12.4 mmol) in formamide (50 mL) was heated and stirred at 150° C for 12 h when the reaction system changed to dark brown. LC-MS monitors the complete conversion of the starting materials to the product. The reaction mixture was cooled to room temperature, and 50 mL water was introduced into the system with stirring for 10 min. Subsequently, the precipitate was suction filtrated and washed with water. Yellow powder (1.5 g) of product was obtained.

3.2.3. A general method for preparing thiazolomethyl substituted pyrimidinone from thiazolamide

A mixture of isopropyl 2-amino −4,5,6,7-tetrahydrobenzo[*b*]thiophene-3-carboxylate-2a, 2b (3.0 g, 12.4 mmol) and acetonitrile (5.1 g, 124 mmol) in dioxane hydrochloride (60 mL, 1.5 mol/L) was heated and stirred at 90° C for 12 h. The reaction mixture was cooled to room temperature, and 60 mL water was introduced into the system with stirring for 10 min. Subsequently, the precipitate was suction filtered, and washed twice with water and petroleum ether (20 mL*2), respectively, and the white solid was obtained.

3.2.4. General method for synthesizing amide from bromoacetyl bromide and amine

A mixture of amines-5a, 6a (10 g, 117 mmol) and pyridine (10.2 g, 128.7 mmol) in dichloromethane (500 mL) was stirred to mix. Bromoacetyl bromide was added dropwise to the mixture, followed by stirring at room temperature for about 2 h. A 10% aqueous solution of copper sulfate was added to the reaction liquid, and the liquid phase was separated to leave the lower organic phase. The organic phase was washed twice with 200 mL of a saturated sodium chloride solution, dried, and concentrated to dryness by vacuum distillation. To the obtained solid was added 100 mL of a petroleum ether / dichloromethane solution (10:1), stirred and filtered to give a white solid.

3.2.5. A general method for preparation of thiazolpyrimidone derivatives

3a, 3b, 4a, 4b (200 mg, 0.7 mmol), 6a (175 mg, 0.84 mmol) and cesium carbonate (273.5 mg, 0.84 mmol) were added in this order in acetone (20 mL), and stirred in an oil bath at 60 °C for 2–3 h until the TLC monitoring material disappeared completely. The reaction solution was naturally cooled to room temperature, and 20 mL of water was slowly added with stirring. The reaction system gradually clarified and precipitated solids. After stirring for about 10 min, it was filtered, and the precipitate was rinsed twice with water/acetone solution (1:1) (5

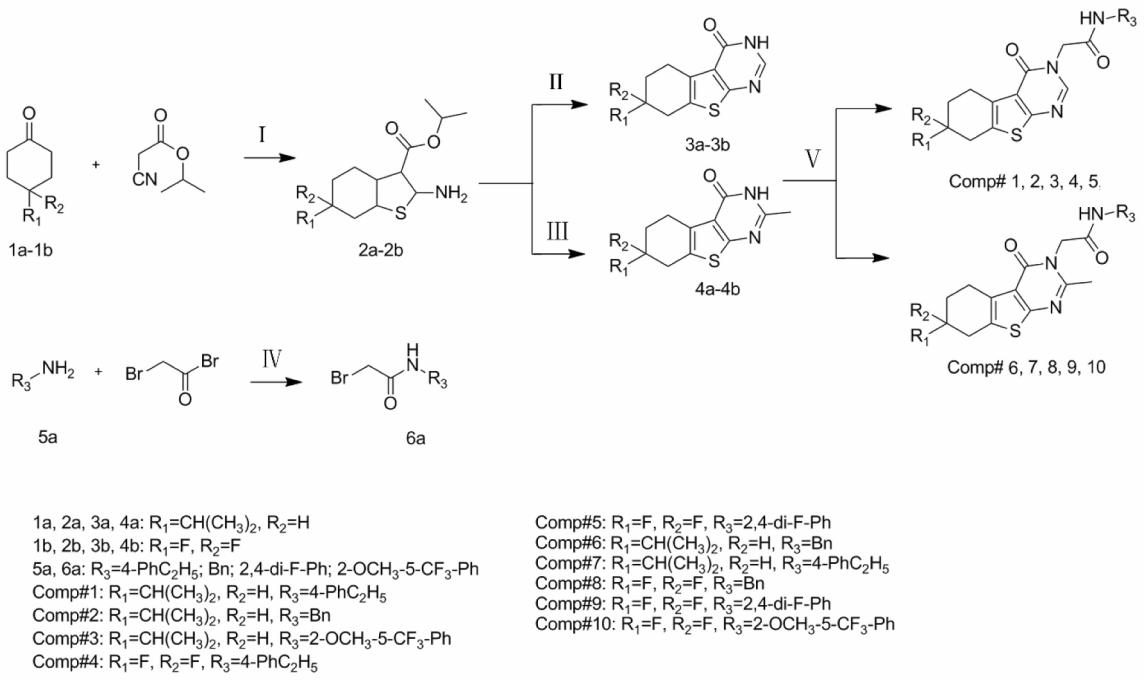

Fig. 2. The synthetic route of comp#1–10. Reagents and conditions: (I) morpholine, sulphur powder, stir, 35 ℃, 24 h; (II) formamide, stir, 150 ℃, 12 h; (III) dioxane hydrochloride, stir, 90 ℃, 12 h; (IV) stir, room temperature, 2 h; (V) cesium carbonate, stir, 60 ℃, 2–3 h.

mL*2) and twice with petroleum ether solution (20 mL*2) to give a white solid.

3.3. ADMET analysis

ADMET characteristics were important indicators to test whether the candidates could meet the required standards, and lead compounds were selected and optimized according to their ADMET properties. In this study, the "Calculate Molecular Properties'' module was employed to calculated a several of molecular characteristics and pharmacokinetic properties of thiophene [2,3-d] pyrimidine derivatives by the DS v3.5 software. The test results of the ten compounds were shown in Table 2. For oral bioavailability, two principal factors affecting oral bioavailability included human intestinal absorption (HIA) and solubility. The poor intestinal absorption level mean that the effect of drug treatment would be significantly reduced. The effect of solubility on the pharmacological activity of compounds involved many aspects, including their absorption, distribution and ultimate bioavailability. Moreover, the plasma protein binding level was used to predict the binding ability of compounds to plasma proteins, which affected the distribution of drugs. The results indicated that all compounds had good affinity to plasma proteins. And the blood–brain barrier permeability model showed that all 10 compounds could penetrate the blood–brain barrier. In addition, the cytochrome P450 2D6 enzyme encoded by human CYP2D6 gene converted the drug into a substance that was easily metabolized for excretion. Inhibition of enzyme activity constituted the majority cases of drug-drug interaction. The results showed that none of the 10 compounds had cytochrome P450 CYP2D6 enzyme inhibitory activity. For hepatotoxicity prediction, all but comp#8 were predicted to be non-toxic. In conclusion, the ADME property values of the compounds in Table 2 (expect Comp#8) were all within the acceptable range of human beings, indicating that these compounds complied with the drug rules.

3.4. The analysis of activity and structure–activity relationship (SAR) of candidate compounds

The compounds were subjected to SHP2-PTP and SHP1-PTP enzymatic activity assays with the pNPP as the substrate. The results were displayed in Table 3. Comp#1, Comp#2, Comp#9 and Comp#10 had no inhibitory activity for SHP2-PTP. Comp#3, Comp#4, Comp#5, Comp#6, Comp#7 and Comp#8 inhibited SHP2-PTP-catalyzed hydrolysis of the pNPP substrate with an IC_{50} value of 1.02 μM, 1.38 μM, 0.4 μM, 2.88 μM, 18.57 μM and 37.87 μM, respectively. While they showed little inhibition for the SHP1 enzyme. Analysis of conformation characteristics of compounds showed that when the substituents of R1 and R2 on Core A were F atoms, it was beneficial to improve the inhibitory activity of compounds (Comp#4 and Comp#5). This indicated that the electron-withdrawing groups on R1 and R2 were beneficial to the inhibitory activity of the candidate compounds. While when R1 was isopropyl, the activity of the compounds decreased significantly (except Comp#3 and Comp#6). Comp#5 and Comp#9 had similar structures, but the activity of the latter was much lower than that of the former, possibly due to methyl groups attached to the pyrimidine ring of the parent nucleus destroying the interaction between the inhibitor 9 and the receptor protein. A similar situation could occur with Comp#8 and Comp#10. In addition, electron-withdrawing groups attached to the benzene ring in the Core C helped to enhance the activity of the inhibitor (Comp#3 and Comp#5).

Table 3
The actual activities of ten candidate molecules by SHP2 and SHP1 phosphatase assay.

Compound	Combination of cores[1]	SHP2 IC_{50} (μM)	SHP1 IC_{50} (μM)
Na_3VO_4	–	29.55	>100
Comp#1	A1-B1-C1	>100	>100
Comp#2	A1-B1-C2	>100	>100
Comp#3	A1-B1-C3	1.02	>100
Comp#4	A2-B1-C1	1.38	>100
Comp#5	A2-B1-C4	0.4	>100
Comp#6	A1-B2-C2	2.88	>100
Comp#7	A1-B2-C1	18.57	>100
Comp#8	A2-B2-C2	37.87	>100
Comp#9	A2-B2-C4	>100	>100
Comp#10	A2-B2-C3	>100	>100

[1] See Fig. 1 for the structure of cores.

3.5. Analyze the activity and selectivity of the inhibitors by molecular docking

To verify and evaluate the putative binding pose of the inhibitor Comp#5 in the active site of SHP2, molecular docking experiments were implemented and visualized by DS v3.5 software. Fig. 3A showed the key amino acids interacted with the ligand. To describe how the Comp#5 bound to the active pocket, the "analyze single complex" module was used to generate a three-dimensional graph of the interactions between receptor and ligand (Fig. 3B). As expected, the six-membered ring of Comp#5 was found in the PTP active pocket and formed intimate interactions with the amino acids in P-loop region, Q-loop region and pTyr-loop region. Specifically, the strong polar substituent F atom on the six-member ring forms hydrogen bond interaction with the residue GLN506 in Q-loop region. The inhibitor formed hydrophobic interaction with residues CYS459-GLY464 in the P-loop region and TYR279 in the pTyr-loop region, which may be related to substrate recognition. Furthermore, the thiophene [2,3-d] pyrimidine of Comp#5 penetrated deeply into the active pocket and formed polar interaction with LYS366 located in β5-β6 loop neighboring the PTP-domain, ARG465 located in P-loop, and GLN510 located in the

Table 2
The ADME prediction for the thiophene [2,3-d] pyrimidine derivatives.

	Molecular weight	ALogP98[a]	Absorption Level[b]	Solubility Level[c]	PPB Level	BBB Level[d]	CYP2D6 Prediction	Hepatotoxic Prediction
comp#1	409.54	4.851	0	2	TRUE	1	FALSE	TRUE
comp#2	395.52	3.916	0	2	TRUE	1	FALSE	TRUE
comp#3	479.52	4.835	0	1	TRUE	1	FALSE	TRUE
comp#4	403.45	3.533	0	2	TRUE	2	FALSE	TRUE
comp#5	411.37	3.002	0	2	TRUE	2	FALSE	TRUE
comp#6	409.54	3.517	0	2	TRUE	2	FALSE	TRUE
comp#7	423.57	4.93	0	1	TRUE	1	FALSE	TRUE
comp#8	403.45	2.676	0	2	TRUE	2	FALSE	FALSE
comp#9	425.4	3.081	0	2	TRUE	2	FALSE	TRUE
comp#10	487.44	3.595	0	2	TRUE	2	FALSE	TRUE

a: The logarithmic coefficient of *n*-octanol and water is taken as the logarithm, the value of AlogP98 $\leq$ -2.0 or AlogP98 $\geq$ 7.0 means very poor;
b: Human intestinal absorption level, where 0 stands for good, 1 for moderate, 2 for poor, and 3 for very poor;
c: Aqueous solubility level, where 0 stands for extremely low, 1 for very low, but possible, 2 for low, and 3 for good;
d: Blood-brain barrier permeability, where 1 stands for good, 2 for moderate, and 3 for poor.

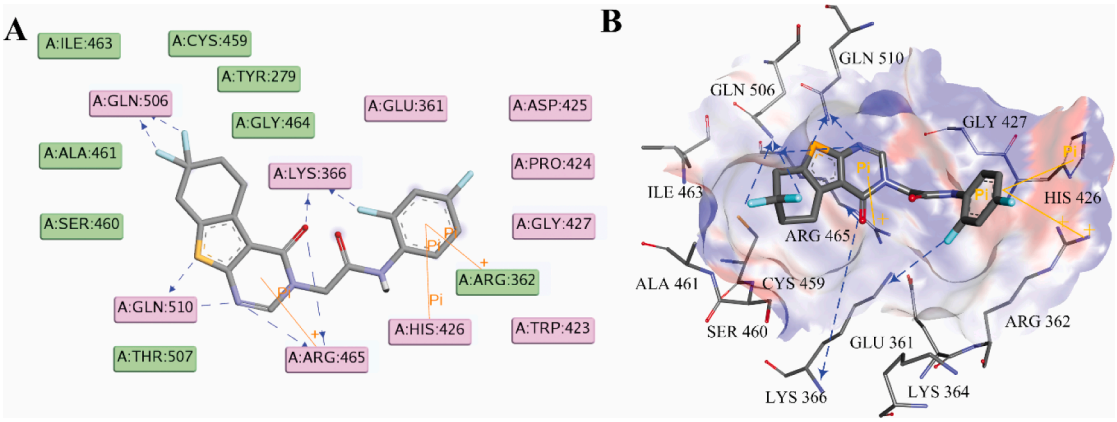

Fig. 3. Docking complex between inhibitor Comp#5 and SHP2 protein. (A) The 2D map of the interaction between inhibitor Comp#5 and key amino acids in SHP2 protein. (B) Three-dimensional map of the binding pattern for inhibitor Comp#5 to the active pocket in SHP2 protein. The pink frame represented polar interactions, including hydrogen bonds and charge interactions, and the green frame represented nonpolar interactions, including hydrophobic interactions and van der Waals forces. The dotted lines connecting the compound to the residues represent the H bond interaction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

surrounding of Q-loop region. In particular, the sulfur atoms of thiophene formed hydrogen-bond interaction with the side chains of residue GLN510, nitrogen atoms of pyrimidine formed hydrogen-bond interaction with the side chains of residue ARG465 and GLN510, oxygen atoms of pyrimidine formed hydrogen-bond interaction with the main chain of residue LYS366 and ARG465, respectively. In addition to the hydrogen-bond interactions, the pyrimidine ring of Comp#5 also formed π-π conjugate interaction with the residue ARG465 located in P-loop. There were also additional polar contacts between Comp#5 and the residues located in the WPD-loop region (ARG421-ASP431). Additionally, the distal phenyl ring of Comp#5 was sandwiched between the side chains of ARG362 located in the β5-β6 loop (residue 362–365) and HIS426 located in WPD-loop, and formed π-π conjugate interaction with the two residues. Interestingly, the β5-β6 loop (residue 362–365) was highly diverse in the PTPs , especially in SHP1 and SHP2[44]. The conjugation between the benzene ring and residues ARG362 might be the main cause of the potency and selectivity enhancement of inhibitor Comp#5. Fig. S1 showed that the interaction between Comp#5 and SHP1 was weak. Especially, the alicyclic ring and benzene ring of Comp#5 were located outside the active pocket. This made the F atoms on the alicyclic ring and benzene ring in Comp#5 form weak interaction with SHP1, resulting in the decrease of inhibitor activity.

For comparison, the docking conformation of Comp#9 and SHP2 protein was visualized (Fig. S2). Obviously, the Comp#9 was located outside the active pocket (Fig. S2A), and it only formed hydrogen bonds with the residues LYS366 and THR507, with the other major interactions disappearing (Fig. S2B). The affinity between Comp#9 and SHP2 protein was much lower than that between Comp#5 and SHP2 protein, which explained why the lower inhibitory activity of Comp#9 than that of Comp#5. The results showed that the inhibitor Comp#5 had strong interactions with the active pocket of PTP protein, including polar and non-polar interactions, indicating that the inhibitor played an important role in stabilizing protein conformation. In summary, the intimate interaction between the inhibitor and the SHP2 protein enhanced the stability of SHP2 protein, thereby negatively affecting the catalytic activity of the PTP region. Meanwhile, the benzene ring of Comp#5 formed conjugated interaction with ARG362 in the β5-β6 loop region, which enhanced the selectivity of the inhibitor.

3.6. Fluorescence titrations

The excitation wavelength of SHP2 was 295 nm, and the fluorescence monitoring range was from 360 nm to 500 nm. Combining with the compound will inhibit the activity of the protein, thereby attenuating the fluorescence value. Fig. 4 illustrated the quenching of SHP2 fluorescence intensity by Comp#3 and Comp#5 in a dose-dependent manner. Obviously, fluorescence experiments demonstrated that Comp#3 and Comp#5 bound directly to the SHP2 protein, thus weakening its activity on the substrate.

3.7. Evaluation of the binding energy

The MM-PB/SA methods were employed to calculated the complete binding free energy aimed at evaluating the binding affinity of the docking analysis. The snapshots were extracted from the last 95 ns MD simulation at equal intervals of 100 ps and served as input for the calculation. Table 4 displayed the binding energies for SHP2-Comp#5 and SHP2-Comp#9 complex, respectively. The predicted binding free energy of SHP2-Comp#5 and SHP2-Comp#9 was −107.937 kJ/mol and −48.252 kJ/mol, respectively, which verified that the binding affinity between Comp#5 and SHP2 was higher than that between Comp#9 and SHP2. The binding free energy was composed of Van der Waal energy, electrostatic energy, polar solvation energy and SASA energy. It was rather remarkable that the Van der Waals and nonpolar solvation energy were major contributors to the binding energy between inhibitor to SHP2, because the inhibitor has hydrophobic contact with non-polar residues around the active site. While polar solvation was not conducive to the stable of the complex of the protein and the ligand, which could largely cut down the total contribution. In short, the high binding affinity between Comp#5 and SHP2 protein was due to the significant contribution of the van der Waals interaction between the inhibitor and the non-polar surface of the residues at the active site and the hydrogen bond interaction with the side chains at the active site. High binding affinity should be one of the main reasons for the effective inhibition of inhibitors. On the other hand, the active site was precisely occupied by Comp#5, which hindered the entry of substrates and thus exerted an inhibitory effect. However, Comp#9 did not occupy the active site, so it was difficult to exert an inhibitory effect.

To understand the details of the interaction of SHP2-inhibitor Comp#5 with the receptor, the binding energy of the interaction between ligands and key amino acids was calculated. The differences between the interaction energies of SHP2-Comp#5 and SHP2-Comp#9 were compared, as shown in Fig. S3. In SHP2-Comp#5 system, the inhibitor formed hydrogen bonds with residues, including LYS366 in the β5-β6-loop, ARG465 in P-loop, GLN506 and GLN510 in Q-loop of the active region, contributed significantly to the binding free energy. Besides these hydrogens, the additional conjugate interaction of the thiophene ring of Comp#5 with ARG465 was formed, making ARG465

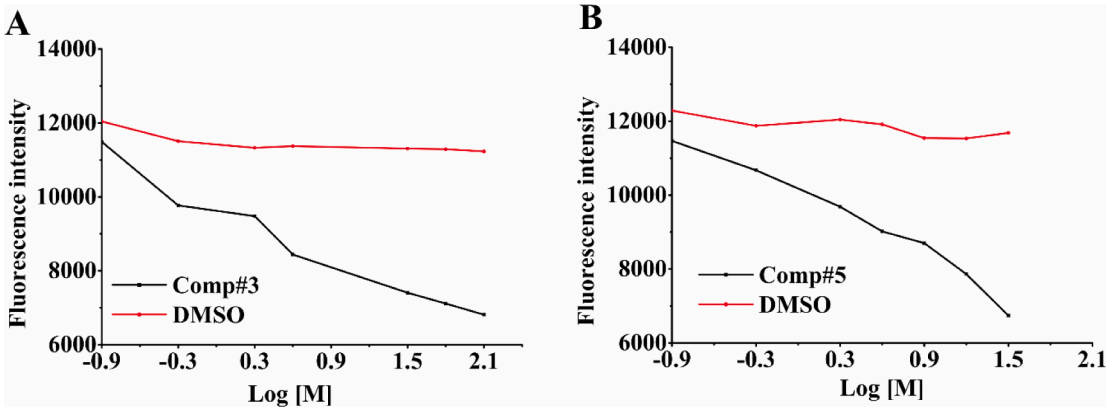

Fig. 4. The relationship between the Log concentration of the Comp#3/Comp#5 and the fluorescence intensity of SHP2 protein. The data displayed the mean ± SEM of triplicate representative experiments. Analogical results were obtained for all compounds tested.

Table 4
MM/PBSA based binding free energies (kJ/mol).

Complex	3O5X-Comp#5	3O5X-Comp#9
Van der Waal Energy (kJ/mol)	−237.936	−96.197
Electrostatic Energy (kJ/mol)	−31.217	−59.354
Polar Solvation Energy (kJ/mol)	179.799	119.103
SASA Energy (kJ/mol)	−18.583	−11.803
Binding Free Energy (kJ/mol)	−107.937	−48.252

contribute the highest binding energy. In SHP2-Comp#9 system, however, the residues above contributed slightly to the binding free energy. Although the residue LYS366 and Comp#9 formed hydrogen bond interaction, the binding free energy of residue LYS366 was still low. In addition, the van der Waals interactions formed between the inhibitor and residues CYS459, SER460, ALA461, and ILE463 in the P-loop region, which was particularly important in catalytic reactions and was widely recognized as binding target of inhibitors[44], were another major reason for the high affinity in SHP2-Comp#5 system, which was verified by van der Waals' significant contribution. As mentioned above, residues ARG362 and HIS426 formed π-π conjugate interactions with the inhibitor Comp#5, and their affinity values were −11.095 kJ/mol and −6.301 kJ/mol, respectively. The strong interactions between Comp#5 and residue ARG362 highlighted the contribution of the β5-β6-loop region, which was the main reason for the high selectivity of the compound. However, due to the weak interactions between Comp#9 and key amino acids, binding free energy was generally low. Collectively, the inhibitor Comp#5 had high affinity with the active site residues, thereby stabilizing the conformation of the protein. These results were consistent with the above structural observation that Comp#5 achieved its potency and specificity by targeting to the active regions divergent in PTPs.

3.8. MD simulation analysis

Dynamic characteristics and structural stability of SHP2 protein, SHP2-Comp#5 protein and SHP2-Comp#9 protein was explored through MD simulation. The stability and convergence of these systems were determined by RMSDs of the backbone Cα atoms relative to the initial structure. In the SHP2-without ligands system, the RMSD values experienced a rapid rise from 0 to 0.55 nm, reflecting the unstable state of the initial simulation of the protein. The trajectory converged and the RMSD values retained below 0.5 nm after approximately 5 ns, which indicated that the model reached an equilibrium state (Fig. 5A). In the SHP2-Comp#5 complex system, although RMSD also experienced an initial increase, its value converged after 5 ns simulation and stabilized below 0.42 nm (Fig. 5A). In addition, RMSD analysis showed that SHP2-Comp#9 system also reached a stable state after 5 ns MD simulation and the average RMSD value was similar to that of the SHP2-without ligands system (Fig. S4A). The results showed that all the systems reached an equilibrium state after a short simulation of 5 ns. In particular, the RMSD value of the SHP2-Comp#5 complex system was slightly lower than that of the SHP2-without ligands system and SHP2-Comp#9 system, presumably because the inhibitor Comp#5 weakened the flexibility of the system and made the system more stable than the other two systems [45].

Root mean square fluctuation (RMSF) evaluated dynamic characteristics by recording the average atomic fluctuation of side chain atoms (N, Cα, and C), which played an important role in the indication of the conformational features of a dynamic system. Overall, the RMSF fluctuations of the SHP2-Comp#5 protein was very similar to that of the SHP2-without ligands and SHP2-Comp#9 protein, except for some crucial regions like the β5-β6 loop, P-loop and Q-loop region. As a matter of fact, the magnitude of RMSF curve fluctuations was closely related to the interaction of inhibitor with SHP2 protein. For the catalytic activity areas (marked with solid black boxes), the magnitude of RMSF value in SHP2-Comp#5 complex system was smaller than that in SHP2-without ligands system and SHP2-Comp#9 system (Fig. 5B and Fig. S4B). Specifically, the RMSF values of residues CYS459-ARG465 were 0.26 nm, 0.14 nm and 0.25 nm in SHP2-without ligands, SHP2-Comp#5 system and SHP2-Comp#9 system, respectively. Besides, the RMSF values of catalytic activity areas residue GLN506-GLN510 were calculated to be 0.3 nm, 0.14 nm and 0.26 nm in SHP2-without ligands, SHP2-Comp#5 system and SHP2-Comp#9 system, respectively. The binding of the inhibitor Comp#5 to the key amino acids of SHP2-PTP protein limited its fluctuation and increased the overall stability of the system. However, the binding affinity between Comp#9 and SHP2-PTP protein was weak, which made it have little impact on the overall flexibility. The low flexibility of complex mainly depended on the tight packed protein arrangement, as well as strong interactions between the amino acids in proteins. The results indicated that the inhibitor Comp#5 might increase the stability of SHP2-Comp#5 protein. For the region located at the entrance of catalytic active region (β5-β6 loop), the average RMSF value of residues GLU361-LYS366 was 0.26 nm in SHP2-without ligands system and SHP2-Comp#9 system, whereas the average RMSF value of the regions was 0.22 nm in the SHP2-Comp#5 system. The results suggested that the inhibitor 5 binding to the β5-β6 loop region of the protein reduced the fluctuation of this region, while Comp#9 had slight effect on this region, which might explain the selectivity of Comp#5.

3.9. Principal component analysis

The conformation of a protein has been considered as one of the principal characteristics for the determination of biological function.

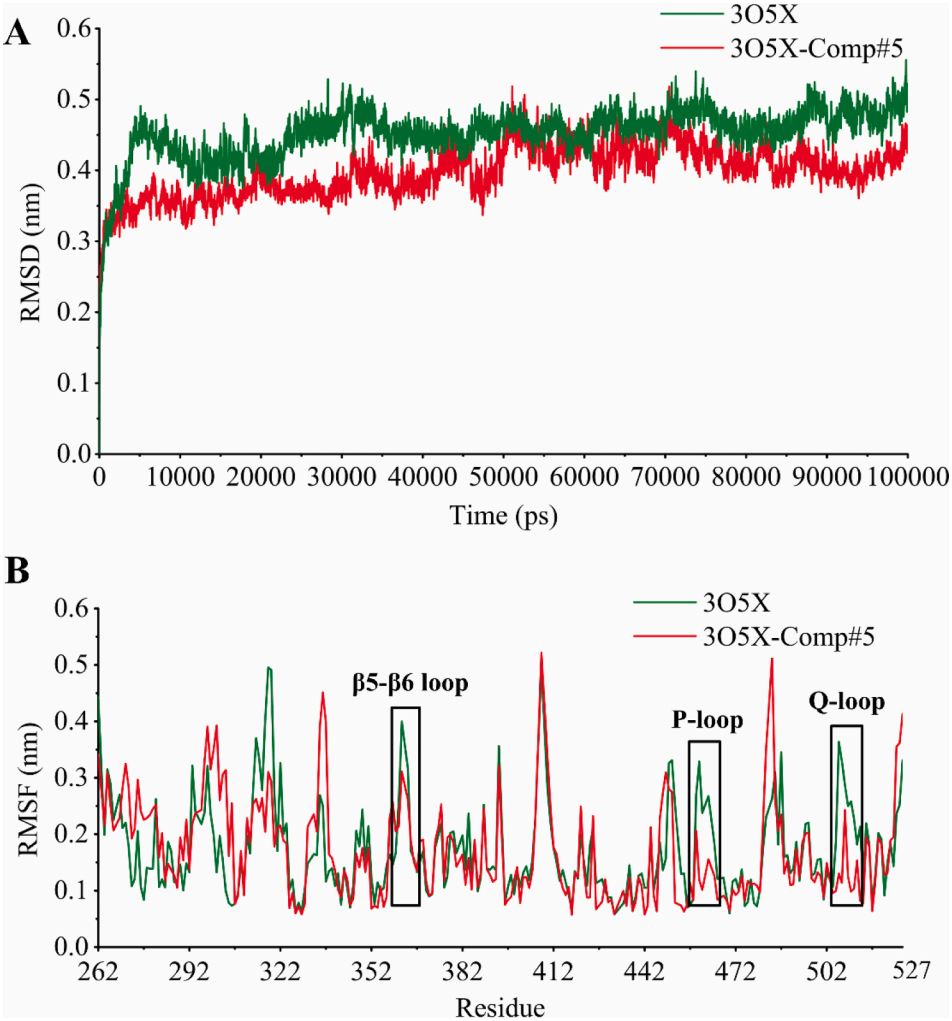

Fig. 5. (A) RMSD of backbone atoms during the simulation time. (B) The per-residue RMSF of side chain versus individual residue number. In the figure, the green curve represents the fluctuations of SHP2-without ligands system, while the red curve represents the fluctuations of SHP2-Comp#5 system. In addition, the areas circled by the black box correspond to the areas with significant differences. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The protein structure was grouped into the subsets according to the conformational similarity by PCA clustering method[46]. PCA analysis were applied using the last 95 ns dynamic trajectories in SHP2-without ligands protein, SHP2-Comp#5 protein and SHP2-Comp#9 protein. The results were shown in Fig. 6 and Fig. S5. The left side represented the states of frame extracted at 2 ps intervals during the whole simulation process, and the right side represented the probability distribution of eigenvalues. The PC analysis suggested that the first two PCs took the percentage of 49.0%, 50.9% and 36.8% of the variance in the dynamic trajectories observed in simulation process of the SHP2-without ligands protein, SHP2-Comp#5 protein and SHP2-Comp#9 protein, respectively, which were predominant among these components. Whereas, other individual components contributed below 7.2%. The results showed that the conformational state distributions of PC1 and PC2 were representative. Subsequently, the PC analysis plots were obtained by projecting along the direction of the first two PCs, which explained the motion characteristics related to the conformational behavior of the main system. In the Fig. 6, the red, white and blue dots were used to show different state, where each point represented the conformation of the two systems. The continuous colors emphasized the periodic transitions between these conformations. The convergent stable conformation and the unstable scattered conformation were described by the aggregation and dispersion states of red/blue dots and light red/blue dots, respectively. In the SHP2-without ligands system (Fig. 6A), it was obviously that the red and blue dots presented loose scattering conformations, and even most of them were in the intermediate transition state, indicating the poor structural stability of the system. Interestingly, in the SHP2-Comp#9 system, the red dots and blue dots showed loosely scattered conformations similar to that of the SHP2-without ligands system (Fig. S5B). Whereas, in the SHP2-Comp#5 compound system (Fig. 6B), the red and blue dots were clustered in a tight distribution, indicating the conformational of the system was stability. In brief, SHP2-Comp#5 complex system was more compacted than SHP2-without ligands and SHP2-Comp#9 system, thus, binding of Comp#5 inhibitor to the protein decreased the conformation flexibility, which lead to a decrease in protein activity. This result was consistent with the RMSF analysis. However, in order to further explore how the Comp#5 would have effect on the residue-residue interactions of SHP2, the additional dynamic analyses were required.

3.10. Dynamic cross-correlation mapping

All paired cross-correlations of amino acid fluctuations served as a reference for understanding how the movement or displacement of one atom in a system was related to another[47]. The functional flexibility and the relative coordination between the atomic groups determined the conformational transformation of the system, which provided a wide range of protein conformational information[48]. Therefore, the dynamic features exhibited by SHP2-without ligands system, SHP2-Comp#5 system and SHP2-Comp#9 system based on the DCCM analysis were performed using the last 95 ns dynamic trajectories. The correlated movements (highly positive) in specific areas were stained dark green, while anti-correlated movements (highly negative) in specific areas were stained orange. The correlated movement of the

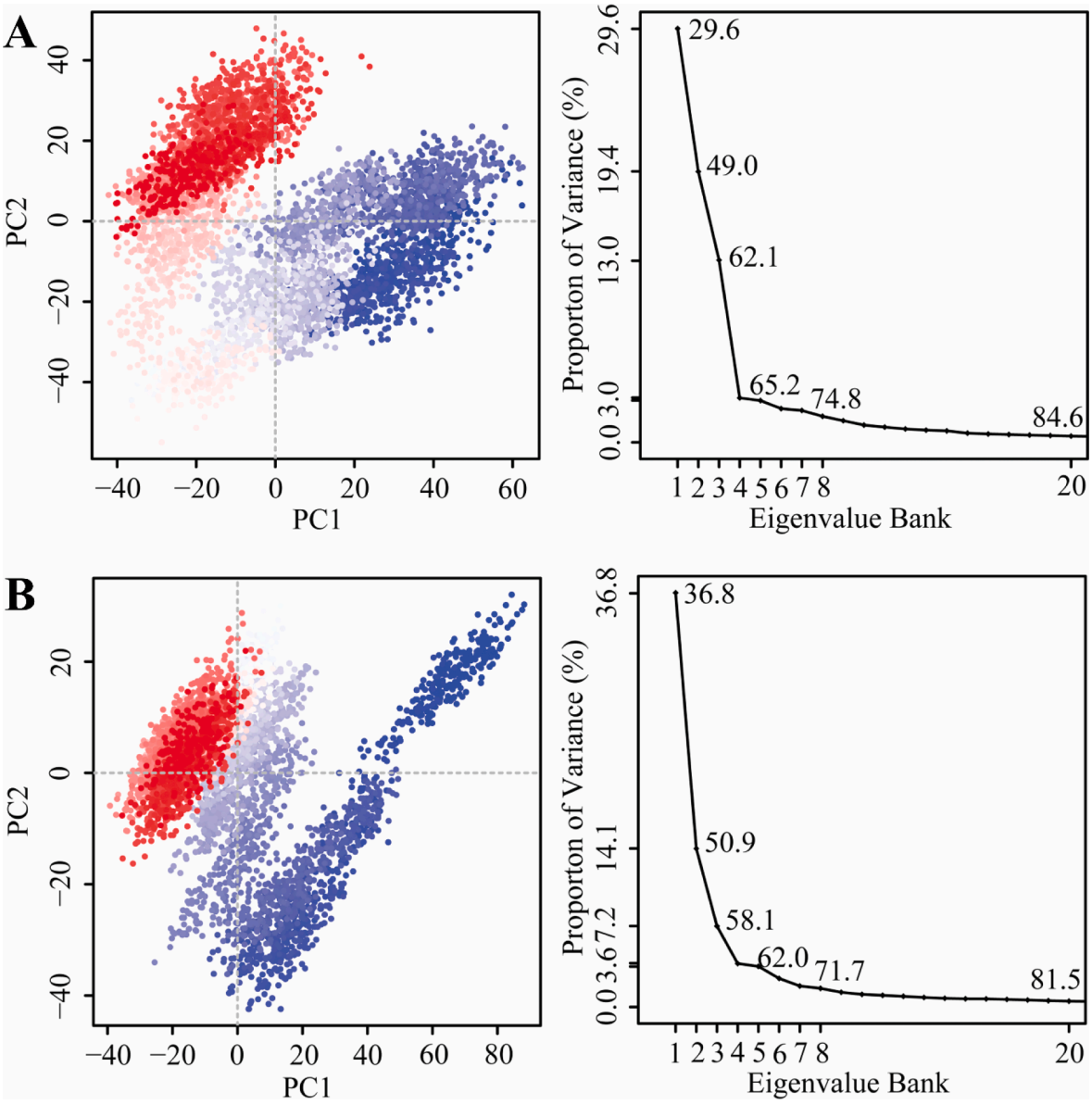

Fig. 6. PCA scatter plots of the trajectories constructed by projecting the first two principal components (PC1 and PC2) in the conformational space of (A) SHP2-without ligands system and (B) SHP2-Comp#5 complex.

diagonal region described the coordinated motion with respect to the residual itself. As could be seen from the Fig. 7 and Fig. S6, the interactions between residues in the SHP2-without ligands system and SHP2-Comp#9 system were weak, while the correlations between residues in the SHP2-Comp#5 system was strengthened, revealing that the binding of inhibitor Comp#5 not Comp#9 to SHP2-PTP protein might increase the stability of the complex system. In the SHP2-Comp#5 complex system, residues TRP423-GLY427 of WPD-loop formed positively correlated with residues GLN506-GLN510 of Q-loop (region c) in catalytic active area, while a weak correlation was observed in the SHP2-without ligands system and SHP2-Comp#9 system. This indicated that the action of the Comp#5 on the catalytic activity region lead to the enhancement of the amino acid interaction in this region. Similarly, the correlation between residues MET355-ARG362 of β5-β6 loop and residues TRP423-GLY427 of WPD-loop (region a), residues MET355-ARG362 of β5-β6 loop and residues CYS459-ARG465 of P-loop (region b) in the active region in the SHP2-Comp#5 complex system were superior to that in the SHP2-without ligands system and SHP2-Comp#9 system. The Comp#5 interacted with the catalytic activity region, which resulted in increased amino acid compactness between residues MET355-ARG362 and TRP423-GLY427, as well as residues MET355-ARG362 and residues CYS459-ARG465. It was noteworthy that in the SHP2-Comp#5 complex system, the residues MET355-ARG362 in the β5-β6 loop interacted strongly with the nearby catalytic active region, which might be the main reason for the high potency and strong selectivity of the inhibitor Comp#5. All in all, the DCCM result analysis suggested that the binding of compound to the catalytic active region lead to stronger residual interactions, and the complex system is more stable than the initial system, which may inhibit the activity of proteins.

3.11. Residue interaction network

RIN analysis was a useful strategy to explore and identify the key residue interaction and structure–function, which could be used to explore how the ligand would change the interactions between residues of proteins. In this study, a representative RIN network was generated by the last 95 ns molecular dynamics trajectories to investigate the mechanism of Comp#5 and Comp#9 effecting on SHP2. The topological

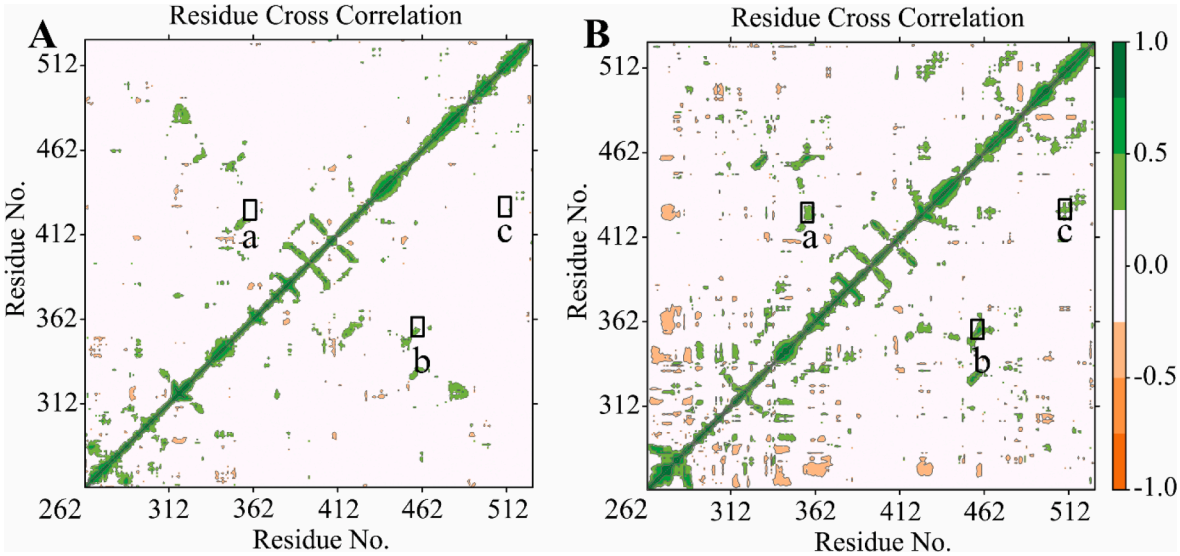

Fig. 7. Description of the correlation movement of all amino acids in the SHP2-without ligands system (A) and the SHP2-Comp#5 system (B). The color distribution between orange (from −1.0 to −0.25), white (from −0.25 to 0.25) and green (from 0.25 to 1.0). The negative values correspond to the negative correlation motion, indicating the opposite direction of the atomic displacement; the positive values correspond to the relevant motion, indicating the same direction of atomic displacement. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

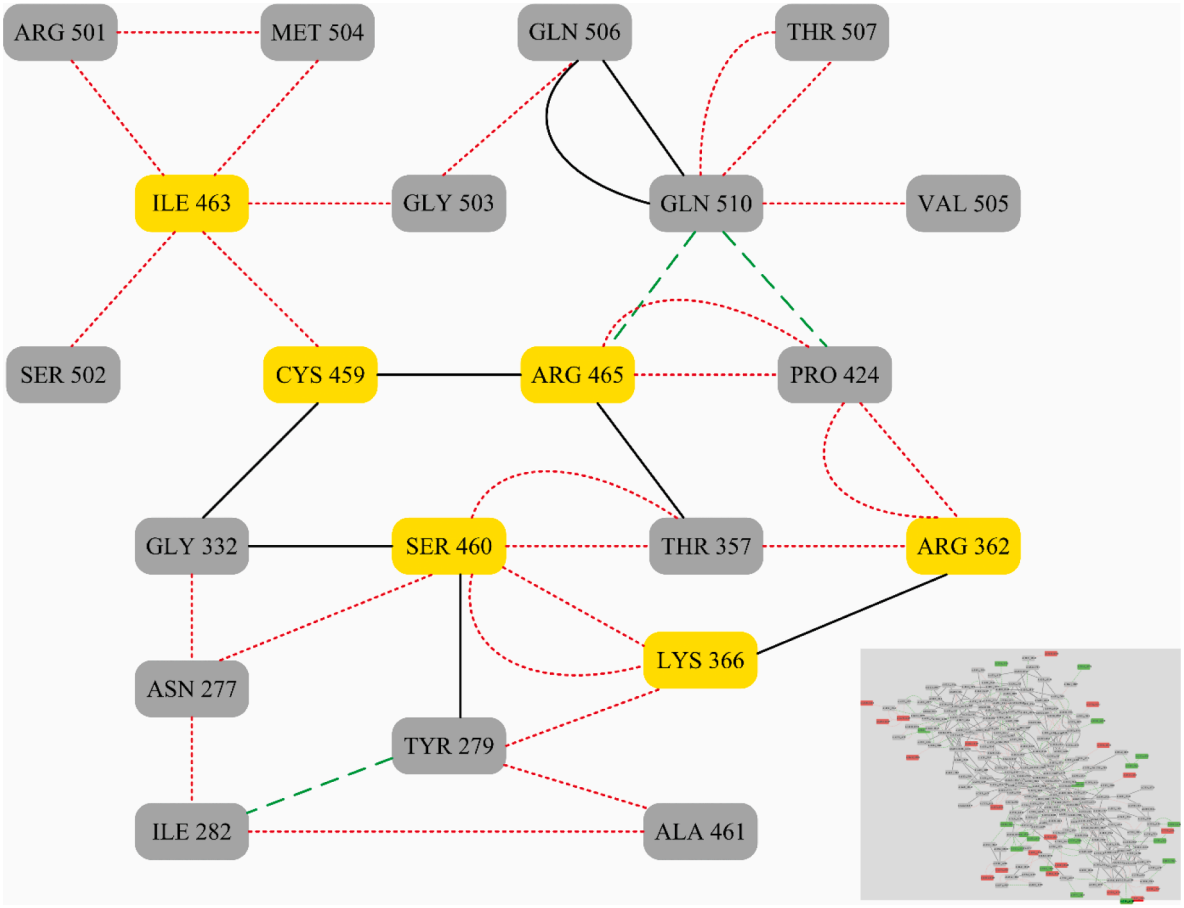

Fig. 8. The distribution of interactions among protein residues in the comparison network. The fringe pattern corresponds to the residual interactions retained in the system, where the green dashed lines, the red dotted line and the black solid line represent the interactions between the amino acids in the SHP2-without ligands system, the SHP2-Comp#5 system and the two systems, respectively. The active areas with significant differences were highlighted by yellow solid coils. the molecular mechanism of Comp#5 inhibiting SHP2 activity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

properties of single amino acid residues, including shortest path interactivity and compact centrality, are characterized by the relationships among the residues within SHP2 protein (Table S1). Interestingly, the residues located at the protease activity sites in the SHP2-Comp#5 complex system had higher betweenness and closeness value than those in the SHP2-without ligands system, such as THR357, ARG362, LYS366, PRO424, CYS459, SER460, ALA461, ILE463, ARG465, THR507 and GLN510. It was evident in the fact that, after the inhibitor bound to the active pocket, information traveled faster from one residue in the protein to another that could reach it. By comparing the differences in the above active site residues in SHP2-without ligands and SHP2-Comp#9 systems, it could be seen that betweenness and closeness value between the protease active site residues in the SHP2-Comp#9 complex system was even lower than that in the SHP2-without ligands system.

Even small changes in a single residue site could have extreme effects on protein structure, eventually leading to significant distortions of the active site fragment. Therefore, it was extremely important to study the interactions between individual amino acids. In the RIN analysis, we mainly focused on analyzing the residues in the active site regions residues CYS459-ARG465 in P-loop and residues GLN506-GLN510 in Q-loop and the entrance of catalytic active region residues GLU361-LYS366 in β5-β6 loop, which were consistent with areas of significant difference in RMSF analysis. In the Fig. 8, the boxes corresponded to amino acids, and the lines represented the interactions between amino acids. It was obvious from the comparison network that the binding of inhibitor Comp#5 lead to different residues interaction at the protease active site (P-loop, Q-loop) in these two systems. Specifically, residue ILE463 in P-loop region formed van der Waals interaction with residue ARG501, SER502, GLY503, MET504 in Q-loop region in SHP2-Comp#5 complex system, respectively. However, only van der Waals interaction between residue ARG465 and GLN510 was found in SHP2-without ligands system. Compared with the SHP2-without ligands system, amino acids in the P-loop region and Q-loop region of the SHP2-Comp#5 system had stronger interactions, suggesting that inhibitors may strengthen the connection between these two regions and make the conformation of the complex more stable. Furthermore, residue ALA461 in P-loop region formed van der Waals interaction with residue TYR279 and ILE282 in pTyr loop, residue SER460 in P-loop region formed van der Waals interaction with residue ASN277 and TYR279 in pTyr loop in the SHP2-Comp#5 system, while only van der Waals interaction between SER460 and TYR279 was found in the SHP2-without ligands system. The results showed that the interaction between P-loop region and pTyr-loop region in SHP2-Comp#5 system was stronger than that in SHP2-without ligands system. Interestingly, the P-loop region also strongly interacted with the β5-β6 loop region in the SHP2-Comp#5 system, which was manifested in H-bond formation and van der Waals interaction between residue SER460 and residue THR357, as well as residue SER460 and LYS366, but these interactions had not been found in the SHP2-without ligands system. The interaction of ARG362 residue with surrounding amino acids had a significant effect on the activity and selectivity of SHP2-Comp#5 protein. In addition, strong H-bond interaction and van der Waals interaction between ARG465 in the P-loop and PRO424 of the WPD loop was found in the pTyr-bound PTP structures, which plays an important role in stabilizing the closed conformation. Therefore, the binding of Comp#5 to the SHP2-without ligands receptor affected the internal interaction network of amino acids.

In the Compare RINs analysis between SHP2-without ligands system and SHP2-Comp# 9 system, Comp#9 had little effect on the active site regions residues CYS459-ARG465 in P-loop and residues GLN506-GLN510 in Q-loop, while it weakened the connections between amino acids in the entrance of catalytic active region residues GLU361-LYS366 in β5-β6 loop (Fig. S7).

4. Conclusion

The primary objective of this research was to obtain novel and highly effective and selective SHP2 inhibitors. In this study, the core hopping method was employed to modify the lead compound PubChem CID 8,478,960 (IC_{50} = 45.01 μM), obtaining a variety of thiophene [2,3-d] pyrimidine derivatives (IC_{50} = 0.4–37.87 μM) with improved core properties. SHP2-PTP enzymatic activity assays showed that the Compo# 5 had the strongest potency and selectivity (IC_{50} = 0.4 μM). Then, in order to elucidate the molecular mechanism of Comp#5 inhibiting SHP2 activity, MD simulation and post-kinetic analysis of SHP2-without ligands, SHP2-Comp#5 and SHP2-Comp#9 (serve as a negative control) systems were carried out. The results showed that Comp#5 had a strong binding affinity with SHP2 protein, and it was worth mentioning that its conjugated interaction with ARG362 residues might help improve the selectivity of the system. whereas Comp#9 had a weak binding affinity with SHP2 protein. Moreover, the RMSF analysis suggested that the β5-β6 loop region, P-loop region and Q-loop region of the SHP2-Comp#5 protein was more stable than those of the SHP2-without ligands and SHP2-Comp#9 systems. In addition, the DCCM analysis confirmed that there were positive correlations between the β5-β6 loop region and the WPD-loop region, the P-loop region and the β5-β6 loop region, and the Q-loop region and the WPD-loop region in SHP2-Comp#5 system, while their correlation was extremely weak in SHP2-without ligands and SHP2-Comp#9 system. This indicated that the Comp#5 inhibitor enhanced the interaction of amino acids within the SHP2 protein, enhancing the stability of the system structure, and thus inhibiting the activity of protease. Nevertheless, Comp#9 had little effect on the stability and interaction of these key residues. Specifically, Comp#5 enhanced the interaction of residues THR357, ARG362, LYS366, PRO424, CYS459, SER460, ALA461, ILE463, ARG465, THR507 and GLN510 with the surrounding residues, enhancing the stability of the catalytic active region and the entrance of catalytic active region. These results explained the mechanism of Comp#5 inhibiting SHP2 at the molecular level.

5. Compliance with ethics requirements

This article does not contain any studies with human or animal subjects

Declaration of Competing Interest

The authors declare no competing financial interest.

Acknowledgements

The National Natural Science Foundation of China (Grant No. 81273361);

The Cooperation and Exchange Project of the National Natural Science Foundation of China (Grant No. 81611130090);

The Natural Science Foundation of Tianjin (Grant No.16JCZDJC32500, Grant No. 18JCQNJC13700);

The Science & Technology Development Fund of Tianjin Education Commission for Higher Education (Grant No. 2017KJ229).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bioorg.2020.104391.

References

[1] R.A. Birnbaum, A. O'Marcaigh, Z. Wardak, Y.Y. Zhang, G. Dranoff, T. Jacks, D. W. Clapp, K.M. Shannon, Nf1 and Gmcsf interact in myeloid leukemogenesis, Mol. Cell 5 (1) (2000) 189–195.

[2] M.G. Mohi, I.R. Williams, C.R. Dearolf, G. Chan, J.L. Kutok, S. Cohen, K. Morgan, C. Boulton, H. Shigematsu, H. Keilhack, K. Akashi, D.G. Gilliland, B.G. Neel, Prognostic, therapeutic, and mechanistic implications of a mouse model of leukemia evoked by Shp2 (PTPN11) mutations, Cancer Cell 7 (2) (2005) 179–191.
[3] P.D. Emanuel, L.J. Bates, R.P. Castleberry, R.J. Gualtieri, K.S. Zuckerman, Selective hypersensitivity to granulocyte-macrophage colony-stimulating factor by juvenile chronic myeloid leukemia hematopoietic progenitors, Blood 77 (5) (1991) 925–929.
[4] S. Stier, Y. Ko, F. Forkert, C. Lutz, T. Neuhaus, E. Grunewald, T. Cheng, D. Dombkowski, L.M. Calvi, S.R. Rittling, D.T. Scadden, Osteopontin is a hematopoietic stem cell niche component that negatively regulates stem cell pool size, J. Exp. Med. 201 (11) (2005) 1781–1791.
[5] M. Tartaglia, C.M. Niemeyer, A. Fragale, X.L. Song, J.C. Buechner, K. Hahlen, H. Hasle, A. Jung, J.D. Licht, B.D. Gelb, Somatic PTPN11 mutations in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia, Pediatr. Res. 53 (4) (2003) 279a.
[6] M.L. Loh, S. Vattikuti, S. Schubbert, M.G. Reynolds, E. Carlson, K.H. Lieuw, J. W. Cheng, C.M. Lee, D. Stokoe, J.M. Bonifas, N.P. Curtiss, J. Gotlib, S. Meshinchi, M.M. Le Beau, P.D. Emanuel, K.M. Shannon, Mutations in PTPN11 implicate the SHP-2 phosphatase in leukemogenesis, Blood 103 (6) (2004) 2325–2331.
[7] W. Liu, W.M. Yu, J. Zhang, R.J. Chan, M.L. Loh, Z. Zhang, K.D. Bunting, C.K. Qu, Inhibition of the Gab2/PI3K/mTOR signaling ameliorates myeloid malignancy caused by Ptpn11 (Shp2) gain-of-function mutations, Leukemia 31 (6) (2017) 1415–1422.
[8] P. Hof, S. Pluskey, S. Dhe-Paganon, M.J. Eck, S.E. Shoelson, Crystal structure of the tyrosine phosphatase SHP-2, Cell 92 (4) (1998) 441–450.
[9] D. Barford, B.G. Neel, Revealing mechanisms for SH2 domain mediated regulation of the protein tyrosine phosphatase SHP-2, Structure 6 (3) (1998) 249–254.
[10] J. Li, H.B. Jie, Y. Lei, N. Gildener-Leapman, S. Trivedi, T. Green, L.P. Kane, R. L. Ferris, PD-1/SHP-2 Inhibits Tc1/Th1 Phenotypic Responses and the Activation of T Cells in the Tumor Microenvironment, Cancer Res. 75 (3) (2015) 508–518.
[11] G. Chan, D. Kalaitzidis, B.G. Neel, The tyrosine phosphatase Shp2 (PTPN11) in cancer, Cancer Metast Rev 27 (2) (2008) 179–192.
[12] J. Xu, L.F. Zeng, W. Shen, J.J. Turchi, Z.Y. Zhang, Targeting SHP2 for EGFR inhibitor resistant non-small cell lung carcinoma, Biochem. Biophys. Res. Commun. 439 (4) (2013) 586–590.
[13] L.F. Zeng, R.Y. Zhang, Z.H. Yu, S. Li, L. Wu, A.M. Gunawan, B.S. Lane, R.S. Mali, X. Li, R.J. Chan, R. Kapur, C.D. Wells, Z.Y. Zhang, Therapeutic potential of targeting the oncogenic SHP2 phosphatase, J. Med. Chem. 57 (15) (2014) 6594–6609.
[14] J. Chen, Z. Cao, J. Guan, SHP2 inhibitor PHPS1 protects against atherosclerosis by inhibiting smooth muscle cell proliferation, BMC cardiovascular disorders 18 (1) (2018) 72.
[15] K. Hellmuth, S. Grosskopf, C.T. Lum, M. Wurtele, N. Roder, J.P. von Kries, M. Rosario, J. Rademann, W. Birchmeier, Specific inhibitors of the protein tyrosine phosphatase Shp2 identified by high-throughput docking, PNAS 105 (20) (2008) 7275–7280.
[16] H.M. Berman, T.N. Bhat, P.E. Bourne, Z.K. Feng, G. Gilliland, H. Weissig, J. Westbrook, The Protein Data Bank and the challenge of structural genomics, Nat. Struct. Biol. 7 (2000) 957–959.
[17] K.C. Chou, D.Q. Wei, W.Z. Zhong, Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS, Biochem. Biophys. Res. Commun. 308 (1) (2003) 148–151.
[18] Y. Ma, S.Q. Wang, W.R. Xu, R.L. Wang, K.C. Chou, Design Novel Dual Agonists for Treating Type-2 Diabetes by Targeting Peroxisome Proliferator-Activated Receptors with Core Hopping Approach, PLoS ONE 7 (6) (2012).
[19] Y.Q. Duan, Y. Ma, X.J. Wang, Y.Y. Jin, R.L. Wang, W.L. Dong, W.R. Xu, D.X. Kong, S.Q. Wang, Design Potential Selective Inhibitors for Treating Cancer by Targeting the Src Homology 2 (SH2) Domain-Containing Phosphatase 2 (Shp2) with Core Hopping Approach, Protein Pept. Lett. 21 (6) (2014) 556–563.
[20] A. Bak, T. Magdziarz, J. Polanski, Pharmacophore-based database mining for probing fragmental drug-likeness of diketo acid analogues, SAR QSAR Environ. Res. 23 (1–2) (2012) 185–204.
[21] D.K. Nevin, M.B. Peters, G. Carta, D. Fayne, D.G. Lloyd, Integrated Virtual Screening for the Identification of Novel and Selective Peroxisome Proliferator-Activated Receptor (PPAR) Scaffolds, J. Med. Chem. 55 (11) (2012) 4978–4989.
[22] D.F. McCain, L. Wu, P. Nickel, M.U. Kassack, A. Kreimeyer, A. Gagliardi, D. C. Collins, Z.Y. Zhang, Suramin derivatives as inhibitors and activators of protein-tyrosine phosphatases, J. Biol. Chem. 279 (15) (2004) 14713–14725.
[23] X.Q. Wu, G. Xu, X.B. Li, W.R. Xu, Q.J. Li, W. Liu, K.A. Kirby, M.L. Loh, J. Li, S. G. Sarafianos, C.K. Qu, Small Molecule Inhibitor that Stabilizes the Autoinhibited Conformation of the Oncogenic Tyrosine Phosphatase SHP2, J. Med. Chem. 62 (3) (2019) 1125–1137.
[24] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J.C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29(7) (2013) 845-854.
[25] S. Gharaghani, T. Khayamian, M. Ebrahimi, Molecular dynamics simulation study and molecular docking descriptors in structure-based QSAR on acetylcholinesterase (AChE) inhibitors, SAR QSAR Environ. Res. 24 (9) (2013) 773–794.
[26] W. Shinoda, M. Mikami, Rigid-body dynamics in the isothermal-isobaric ensemble: A test on the accuracy and computational efficiency, J. Comput. Chem. 24 (8) (2003) 920–930.
[27] B. Hess, P-LINCS: A parallel linear constraint solver for molecular simulation, J. Chem. Theory Comput. 4 (1) (2008) 116–122.
[28] R. Salomon-Ferrer, A.W. Gotz, D. Poole, S. Le Grand, R.C. Walker, Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald, J. Chem. Theory Comput. 9 (9) (2013) 3878–3888.
[29] H. Wang, X.Y. Gao, J. Fang, Multiple Staggered Mesh Ewald: Boosting the Accuracy of the Smooth Particle Mesh Ewald Method, J. Chem. Theory Comput. 12 (11) (2016) 5596–5608.
[30] R. Kumari, R. Kumar, A. Lynn, O.S.D.D. Consort, g_mmpbsa-A GROMACS Tool for High-Throughput MM-PBSA Calculations, J. Chem. Inf. Model. 54 (7) (2014) 1951–1962.
[31] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, MMPBSA.py: An Efficient Program for End-State Free Energy Calculations, Journal of chemical theory and computation 8(9) (2012) 3314-3321.
[32] Y. Yu, J. Wang, Q. Shao, J. Shi, W. Zhu, Effects of drug-resistant mutations on the dynamic properties of HIV-1 protease and inhibition by Amprenavir and Darunavir, Sci. Rep. 5 (2015) 10517.
[33] N. Homeyer, H. Gohlke, Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method, Mol. Inf. 31 (2) (2012) 114–122.
[34] B.J. Grant, J.A. McCammon, L.S. Caves, R.A. Cross, Multivariate analysis of conserved sequence-structure relationships in kinesins: coupling of the active site and a tubulin-binding sub-domain, J. Mol. Biol. 368 (5) (2007) 1231–1248.
[35] E. Papaleo, P. Mereghetti, P. Fantucci, R. Grandori, L. De Gioia, Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case, J. Mol. Graph. Model. 27 (8) (2009) 889–899.
[36] D. Spiliotopoulos, A. Spitaleri, G. Musco, Exploring PHD fingers and H3K4me0 interactions with molecular dynamics simulations and binding free energy calculations: AIRE-PHD1, a comparative study, PLoS ONE 7 (10) (2012) e46902.
[37] G.G. Maisuradze, A. Liwo, H.A. Scheraga, Relation between free energy landscapes of proteins and dynamics, J. Chem. Theory Comput. 6 (2) (2010) 583–595.
[38] P.H. Hunenberger, A.E. Mark, W.F. van Gunsteren, Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations, J. Mol. Biol. 252 (4) (1995) 492–503.
[39] T.J. Hou, J.M. Wang, Y.Y. Li, W. Wang, Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations, J. Chem. Inf. Model. 51 (1) (2011) 69–82.
[40] A.J.M. Martin, M. Vidotto, F. Boscariol, T. Di Domenico, I. Walsh, S.C.E. Tosatto, RING: networking interacting residues, evolutionary information and energetics in protein structures, Bioinformatics 27 (14) (2011) 2003–2005.
[41] Y. Assenov, F. Ramirez, S.E. Schelhorn, T. Lengauer, M. Albrecht, Computing topological parameters of biological networks, Bioinformatics 24 (2) (2008) 282–284.
[42] N.T. Doncheva, Y. Assenov, F.S. Domingues, M. Albrecht, Topological analysis and interactive visualization of biological networks and protein structures, Nat. Protoc. 7 (4) (2012) 670–685.
[43] M. Vendruscolo, N.V. Dokholyan, E. Paci, M. Karplus, Small-world view of the amino acids that play a key role in protein folding, Phys. Rev. E 65 (6) (2002).
[44] X. Zhang, Y. He, S. Liu, Z. Yu, Z.X. Jiang, Z. Yang, Y. Dong, S.C. Nabinger, L. Wu, A. M. Gunawan, L. Wang, R.J. Chan, Z.Y. Zhang, Salicylic acid based small molecule inhibitor for the oncogenic Src homology-2 domain containing protein tyrosine phosphatase-2 (SHP2), J. Med. Chem. 53 (6) (2010) 2482–2493.
[45] O. Carugo, S. Pongor, A normalized root-mean-square distance for comparing protein three-dimensional structures, Protein science : a publication of the Protein Society 10 (7) (2001) 1470–1473.
[46] A. Wolf, K.N. Kirschner, Principal component and clustering analysis on molecular dynamics data of the ribosomal L11 center dot 23S subdomain, J. Mol. Model. 19 (2) (2013) 539–549.
[47] S.L. Williams, J.W. Essex, Study of the Conformational Dynamics of the Catalytic Loop of WT and G140A/G149A HIV-1 Integrase Core Domain Using Reversible Digitally Filtered Molecular Dynamics, J. Chem. Theory Comput. 5 (2) (2009) 411–421.
[48] C. Arcangeli, A.R. Bizzarri, S. Cannistraro, Molecular dynamics simulation and essential dynamics study of mutated plastocyanin: structural, dynamical and functional effects of a disulfide bridge insertion at the protein surface, Biophys. Chem. 92 (3) (2001) 183–199.