Exploring the cause of the dual allosteric targeted inhibition attaching to allosteric sites enhancing SHP2 inhibition

Yang-Chun Ma[1], Wen-Yu Yang[2], Liang Zhou[1], Jing-Wei Wu[1], Wei-Ya Li[1], Shan Du[1], Ying Ma[1], and Run-Ling Wang[1]

[1]Tianjin Medical University
[2]State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

July 23, 2020

Abstract

Protein Tyrosine Phosphatase non-receptor type 11 (PTPN11; encoding protein SHP2) is an important protein tyrosine phosphatase (PTP) in the human body and plays an important role in regulating cell proliferation and differentiation. Overexpression of SHP2 will promote the development of cancer diseases, so research on SHP2 inhibitors has become one of the popular targets for the treatment of cancer. Recent studies have shown that combining SHP099 (an allosteric site 1 inhibitor) with SHP844 (an allosteric site 2 inhibitor) will enhance pharmacological pathway inhibition in cells. This study uses molecular dynamics simulations to explore the inhibition mechanism of SHP099 and SHP844 on SHP2 protein. The result shows that the interactions of allosteric site 1 (THR108-TRP112, LEU236-GLN245), allosteric site 2 (GLN79-GLN87, LEU262-GLN269), P-loop (HIS458-ARG465), and Q-loop (ARG501-THR507) are obviously enhanced in SHP2-SHP099-SHP844 system, which makes the fluctuation of residues more stable and the active site more difficult to be exposed. Meanwhile, residue GLU110 (allosteric site 1), ARG265 (allosteric site 2), and ARG501 (Q-loop) are speculated to be the key residues that led to the SHP2 protein in auto-inhibition conformation. This study provides an idea that help people to understand the mechanism of inhibition of the combining SHP099 with SHP844 on the SHP2.

1. Introduction

Cancer, medically known as malignant neoplasm, has become a major problem that plagues the world. It has the characteristics of rapid increase in morbidity and mortality, which seriously threatens the health and safety of people [1]. Malignant tumors often have biological characteristics, such as abnormal cell differentiation and proliferation, uncontrolled growth, invasion and metastasis [2]. The occurrence is a complex process with multiple factors and multiple steps. At present, surgery, chemotherapy, and radiotherapy are the main methods of treating malignant tumors, but the curative ratio for patients is extremely low [3]. In search of a better treatment, targeted drug therapy has gradually become the focus of tumor research.

Some studies have reported that Protein tyrosine phosphatase non-receptor type 11 (PTPN11; encoding protein SHP2) is related to cancer types, such as lung cancer, breast cancer, leukemia [4]. Therefore, SHP2 has become an important potential target for discovering cancer inhibitors. SHP2 is a protein tyrosine phosphatase (PTP) widely expressed in the human body. It can co-regulate the phosphorylation of tyrosine proteins in the body with protein tyrosine kinases (PTK) [5]. Reversible phosphorylation of tyrosine residues in proteins is not only the key to regulate the activation of cell signal transduction pathways, but also to regulate cell growth, differentiation, metabolism, gene transcription and immune response[6-8]. SHP2 plays an important role in multiple cell signal transduction pathways (RAS-ERK, JAK-STAT, PI3K-AKT, PD-

1

1/PD-L1)[9-11]. SHP2 is the upstream positive regulator of Ras-Raf-MEK-ERK signaling pathway, and ERK can promote the development of cancer diseases. By inhibiting SHP2 to block the cancer-promoting function of Ras-Raf-MEK-ERK pathway is particularly important in the treatment of cancer diseases [12]. In addition, the expression of SHP2 is positively correlated with PD1 in T cells of patients with neck squamous cell carcinoma. When PD-1 is combined with PD-L1/L2, it can recruit SHP2 and release its auto-inhibition state, thereby blocking the activation of T cells. Therefore, it is very important to block the binding of PD-1 to SHP2 and releases the tumor-specific T-cell killing ability that is suppressed [13]. In short, SHP2 is a very attractive target for treating cancer.

Since the first discovery of SHP2 as an important signaling site and a potential anti-cancer target, the identification of pharmacologically relevant inhibitors has aroused widespread interest in the scientific community [14]. SHP2 has two states: "closed state" (SHP2 in auto-inhibited conformation) and "active state" (SHP2 in an open conformation). In closed state, N-SH2 of SHP2 is inserted into PTP domain, which can prevent the substrate from entering the catalytic site. In addition, bis-phosphotyrosyl proteins or peptides can interrupt the binding of N-SH2 to the PTP domain so that auto-inhibiting state released [15]. At present, SHP2 protein allosteric inhibitors mainly strengthen the combination of N-SH2 and PTP domain, which can make SHP2 protein keeping in auto-inhibition state.[16]. However, the mutation of SHP2 protein makes the auto-inhibition state easier to be released. Dual allosteric targeted protein inhibition can stabilize the auto-inhibition state of SHP2 protein through binding to different allosteric sites at the same time, which may improve the inhibition on SHP2 protein. Combining two distinct but compatible SHP2 inhibitors may offer advantages in enhancing SHP2 inhibition and overcoming resistance. Michelle Fodor and his team members tested the pharmacological activity of the SHP2 protein using a dual allosteric targeted protein inhibition method to inhibit the SHP2 protein [17]. Combining SHP099 (allosteric site 1) with SHP844 (allosteric site 2) interact with different allosteric sites in SHP2 protein shown in FIG.1. The structure of SHP099 and SHP844 can be seen in FIG.S1. SHP099 is an effective, selective, highly soluble, SHP2 inhibitor with oral bioactivity. It can inhibit the corresponding signaling pathways and has antitumor activity in a concentration-dependent manner in animal models of transplanted tumors. SHP099 simultaneously binds the interface of n-terminal SH2, c-terminal SH2 and protein tyrosine phosphatase domain, inhibits the activity of SHP2 through an allosteric mechanism, and stabilizes SHP2 in an auto-inhibiting conformation [18]. IC_{50} value is 0.07 μM. SHP844 is a weak SHP2 inhibitor with IC_{50} of 18.9 μM. SHP844 enhanced the inhibition of SHP099 in biochemical phosphatase experiments, and SHP844 enhanced down regulation of the MAPK pharmacodynamic marker DUSP6 by 1 in cellular experiments [19]. Although SHP844 has less inhibitory activity on SHP2 than SHP099, the binding site of SHP844 is a new site. The crystal structure shows that SHP844 binds at the gap formed at the interface between the N-SH2 and the PTP domain, stabilizing the SHP2 concept of the auto-inhibited state, and thus plays a certain inhibitory role. It is precisely because of the differences in binding sites that it is possible to perform dual targeted inhibition of SHP2 protein. Research showed that when the inhibitors (SHP099 and SHP244) acted simultaneously with the allosteric sites 1 and site 2 showed a certain synergistic effect on the pharmacological pathway. Therefore, the double target inhibition of SHP2 is feasible. However, the mechanism of the combining SHP099 and SHP844 inhibiting SHP2 at the molecular level is unclear. In order to promote the development of dual allosteric targeted SHP2 protein inhibitor, a series of studies are performed to explore the changes of residues at allosteric sites and catalytic area.

The impact of molecular dynamics (MD) simulations on drug discovery is increasing [20]. MD methods simulate dynamic biological and chemical events at a molecular level, which can pave the way for drug development. For instance, Ali Rahimi's research group used MD simulation to simulate all atoms of Poly (L-lysine) (PLL) dendrimer of different generations [21]. Magnus Lundborg's group uses MD simulations to predict the permeability of drugs through the skin [22]. MD simulation can provide information about protein molecular levels on a reasonable time scale. Therefore, MD simulation was applied to explore the inhibition mechanism of combining SHP099 inhibitor with SHP844 inhibitor on SHP2 protein, which could provide information on the conformational changes of the SHP2 protein at the molecular level. Based on the information obtained from MD simulation, RMSD, RMSF, PCA, DCCM, ANM and RIN analyses were

performed to investigate the influence of combining SHP099 with SHP844 on SHP2 allosteric site.

Materials and methods

The crystal structure of the complex of SHP2 protein and ligand (SHP099 and SHP244) applied in this study is ID: 6BMY that come from RCSB Protein Data Bank. The topology parameters files of the compounds SHP099 and SHP244 are obtained from the ACPYPE Portal website [23], which was designed to generate topology parameters files for unusual organic chemical compounds. SHP2 protein and ligand preparation, 2D interaction analysis was performed in Discovery studio software. Protein and dual-ligand MD simulations were performed from the AMBER force field in GROMACS 4.5.5.

System preparation

The 6BMY protein structure file downloaded from the RCSB PDB library needs a series of processing before MD simulation. The protein preparation of 6BMY crystal structure is performed under the Prepare Protein protocol of Discovery Studio V3.5. The steps of protein preparation include assigning bond orders, adding hydrogen, treating disulfides, finding overlaps, deleting waters, and filling loop. 6BMY protein preparation can not only remove the poly conformation of the protein, but also supplement the protein structure with incomplete amino acid residues. Ramachandran plot is a visualization method applied to describe whether the dihedral angles ψ and φ of amino acid residues in the protein structure are in a reasonable region [24]. It can also reflect whether the conformation of the protein is reasonable. First, 10 ns molecular dynamics simulation of the original protein structure file is carry out by GROMACS4.5.5, and then frame the files are extracted by molecular dynamics simulation every 20 ps. Finally, the PDB file generated by the frame extraction is imported into DS v3.5 to generate a Ramachandran plot, which is accustomed to identify the optimal conformer of protein model for subsequent MD analysis.

MD simulations

The MD simulation process is carried out in GROMACS 4.5.5. The changes of SHP2 protein in real solvent can be mapped by MD simulation. MD simulation is helpful to analyze the effect of dual allosteric targeted protein inhibitor on the conformation of SHP2 protein [25]. First, pretreatment of SHP2 crystal structure (6BMY), which can remove protein multi-conformation, supplement non-intact amino acid residues, hydrogenate proteins, etc. Second, the coordinate file (gro) and topology file of the SHP2 protein are generated by the pdb2gmx tool in AMBEER99SB [26]. This topology file contains complete information on all interactions in the peptide or protein, mainly including these contents: force field, bond interaction parameters, non-bonding interaction parameters, restrictive parameters, defining some names, etc. The purpose of this study is to analyze the effect of ligand on SHP2 protein, so it is necessary to merge the topology file and gro file of protein and ligand [27]. The coordinate file of the ligand is obtained by conversion from the Acpype Server-Bio2Byte website. Third, the volume of the box is related to the machine time of the MD simulation operation .The smaller the volume is, the shorter the running times are. Therefore, a regular cube box containing proteins (volume less than cubic or triclinic) is established through the editconf tool and the distance from the edge of the protein to the box is less than 1 nm. Fourth, inject the SOL (3104) solvent into the cube box through the genbox tool. We add the appropriate type (NA, CL) and the number of charges according to the system to make the charge of the protein system in a balanced state. Fifth, before simulation, we must ensure that the structure of the system is normal, the distance between atoms is not too close, and the combination configuration is reasonable. This requires relaxation of the structure. This process is called energy minimization (EM) and is a very important step in MD simulation. There are two important conditions in energy minimization: maximum force ¿10.0 KJ/mol, the maximum reaction step is 50000. After EM, NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature) balance is an important step in MD simulation, the core lies in the mdp file. The protein system is briefly simulated by the grompp tool to keep the entire system in a stable state of density and temperature. Whether the density and temperature of the system reach the ideal state is observed by drawing a physical fluctuation curve. Finally, after the system is equilibrated at the required temperature and pressure, the MD simulation can be officially carried out.

3

2.3 Root mean square deviation and root mean square fluctuation analysis

Root Mean Square Deviation (RMSD) is a method of measuring the coordinate deviation of a certain atom relative to a reference structure [28]. RMSD is generally applied to imply the difference between protein structures. It is the RMSD of the atomic position between two structures, and is often utilize in biochemical protein analysis [29]. By calculating the RMSD value of SHP2 protein, it is helpful to observe the effect of dual allosteric targeted protein inhibitor on the overall stability of the protein. The calculation formula of RMSD is as follows:

$$RMSD = \sqrt{\frac{\sum_{i=0}^{N}\left[m_i * (X_i - Y_i)^2\right]}{M}}$$

Where N is the total number of atoms analyzed, m_i is the ith atomic mass, X_i and Y_i respectively represent the coordinate vector of the target atom i and the coordinate vector of the atom i in the reference structure in the same coordinate system, and M is the total mass. If RMSD is calculated regardless of mass weighting, all mi=1 and M=N.

Root Mean Square Fluctuation (RMSF) refers to the root mean square displacement of each amino acid in a frame of conformation compared to the average conformation, to determine the flexibility of a certain region of the protein [30]. In this study, RMSF is utilized to analyze the effect of inhibitors on the flexibility of local residues 1 and 2 of SHP2 protein allosteric sites[31]. The calculation formula of RMSF is as follows:

RMSF=

In the formula, tj is the time point, T is the MD simulation time, x ˜ is the average value, and xi (ti) is the coordinate of tj timetable.

2.4 Binding free energy

The calculation of free energy between ligand and protein has always been an important component in computer-aided drug design, and has been widely used in drug design [32, 33]. The molecular binding free energy is a very important physical parameter for describing the affinity of a protein and a ligand [34]. In order to understand the binding between protein and ligand, binding Free energy is applied in this study to investigate the binding strength of protein to amino acid. The calculation formula combining free energy is as follows:

$$G^{A \to B} = RT(\ln \frac{\langle f(V_A - V_B + \quad C)\rangle_B}{\langle f(V_B - V_A - \quad C)\rangle_C})$$

In the formula, A and B represent different states, and$G^{A \to B}$ represents different changes in free energy from the A state to the B state. R, T, respectively, represent the gas constant, temperature 310 K in this research. C is a constant (if the number of samples in the two simulations are different, and n A [?] n B, a correction factor ln should be added to the right (n A / n B)).

2.5 Principal Component Analysis

PCA (Principal Component Analysis), the principal component analysis method, is one of the most widely accepted data dimensionality reduction algorithms [35]. In research in different fields, it is usually necessary to observe multiple data variables and analyze the laws. To a certain extent, multivariable increases the difficulty of data analysis and increase the complexity of problem analysis. If the information in the data is not fully utilized, a lot of useful information will be lost, resulting in erroneous conclusions [36]. PCA is a reasonable method. While reducing the indicators that need to be analyzed, the loss of information contained in the original indicators is minimized to achieve the purpose of comprehensive analysis of the collected data [37]. PCA is a reasonable method, while reducing the indicators that need to be analyzed, try to reduce

the loss of information contained in the original indicators, so as to achieve the purpose of comprehensive analysis of the collected data [38]. Therefore, this study adopts PCA to analyze the conformational change information of SHP2 protein in MD simulation.

$$\Sigma_{ab} = \langle (x_a - \langle x_a \rangle)(x_a - \langle x_b \rangle) \rangle$$

The Cartesian coordinates of the ith and jth C_α atoms were represented by x_a and x_b, respectively. Mean value during molecular dynamics simulation was represents by $\langle \rangle$.

2.6 Domain Cross-Correlation Map Analysis

DCCM can calculate the correlation between each amino acid $C\alpha$ atom and other amino acid $C\alpha$ atoms, which can effectively provide information on the movement of proteins in molecular dynamics simulations [39, 40]. Therefore, DCCM is utilized to analyze the movement of SHP2 double allosteric residues. The calculation formula of DCCM is as follows:

Cij represents the strength of the correlation between atoms in the protein chain, ranging from - 1 to 1. Δri and Δrj represent the vector displacements of atoms I and j, respectively. In the research, the trajectory files of PTP delta are processed by Bio3D library and R-studio software, and the DCCM diagrams are finally obtained, which can show the correlation between atoms.

2.7 Residue interaction network

RIN is a commonly adopted method to analyze various interactions between residues and residues (VDW, H-bond, etc.) [41]. At present, RINs have been widely applied to analyze internal mutations, folding, and catalytic activities of proteins [42, 43]. In order to better analyze the interaction between protein and ligand, RINs analysis is performed on the files generated by MD simulation in this study.

Results and discussion

3.1 The reliability of the initial SHP2 structures

Ramachandran plot is often be accustomed to explain the rotation degree (psi) of the bond between the $C\alpha$ atom and the carbonyl C atom, the rotation degree (phi) of the bond between the $C\alpha$ atom and the N atom in the protein or peptide chain [44, 45]. At the same time, it can also reflect whether the conformation of the protein is reasonable. Therefore, ramachandran plot analysis of SHP2 protein could selected out a reasonable protein conformation for MD simulation. In the ramachandran plot, it could be divided into three types of regions: the best region (blue region), the allowed region (purple region), and the not allowed region. In addition, if the amino acid conformation of the optimal region and the allowable region accounts for more than 90% of the total amino acids, it illustrates that the protein model obeys the rules of stereochemistry. The ramachandran plot of the 133th simulated conformation of the SHP2 protein was shown in FIG.S2, with 505 amino acids distributed in the allowed region and 18 amino acids distributed in the allowed region where amino acids distributed in the allowed region account for 96.6% of the total amino acids. Therefore, the protein model was applied and subsequent analysis and research.

3.2 Stability analysis

In this study, the MD simulation trajectory files were analyzed, respectively, for SHP2 protein system and SHP2-SHP099-SHP844 system by RMSD. The RMSD analysis could be used as a method to assess the overall fluctuations of the main chain $C\alpha$ atoms. The high the RMSD value was, the more flexible the main chain $C\alpha$ atom was. In FIG.2A, it could be found that the SHP2 system changes to be stable in 6 ns, therefore, MD files after 6 ns were utilized for post-MD analysis. In the SHP2 system, the average RMSD of 6-100 ns was 0.315 nm. While in the system SHP2-SHP099-SHP844 system, the average RMSD of 6-100 ns was 0.281 nm. It could be clearly found in FIG.2A that the overall fluctuation of the RMSD of the SHP2-SHP099-SHP844 system was small than that of the SHP2 system, suggesting that the SHP2-SHP099-SHP844 system was more stable. Therefore, ligands SHP099 and SHP844 play a stable role in the overall conformation of SHP2 protein.

The effect of dual allosteric targeted protein inhibitors on the side chain atoms of SHP2 protein were study by RMSF, which was applied to analyze the trajectory files. FIG.2B was the RMSF that was generated by the SHP2 system and the SHP2-SHP099-SHP844 system at 6-100 ns. The smaller the RMSF value was, the more stable the side chain atoms were. The areas with large fluctuations in the two systems were marked with yellow and black boxes, respectively. In addition, there were the six residue regions (GLN79-GLN87, THR108-TRP112, LEU236-GLN245, LEU262-GLN269, HIS458-ARG465, and ARG501-THR507) marked by the black box and three regions (THR153-SER165, ASN200-GLN211, and HIS293-VAL301) marked by the yellow box, respectively. The area marked by the yellow ellipse was not in the allosteric site area or in the active area, so it was not within the scope of this study. Residues THR108-TRP112 and LEU236-GLN245 belong to allosteric site 1, residues GLN79-GLN87, LEU262-GLN269 belong to allosteric site 2. Residues HIS458-ARG465 and ARG501-THR507 belong to P-loop and Q-loop, respectively. In SHP2 system, the average RMSF of residue GLN79-GLN87, THR108-TRP112, LEU236-GLN245, LEU262-GLN269, HIS458-ARG465, and ARG501-THR507 were 0.168 nm, 0.128 nm, 0.206 nm, 0.142 nm, 0.113 nm and 0.099 nm, respectively. While in the SHP2-SHP099-SHP844 system, the average RMSF of residue GLN79-GLN87, THR108-TRP112, LEU236-GLN245, LEU262-GLN269, HIS458-ARG465, and ARG501-THR507 was 0.101 nm, 0.100 nm, 0.150 nm, 0.099 nm, 0.067 nm and 0.076 nm, respectively. Obviously, the RMSF fluctuation of the SHP2 system in the region marked by the black box was higher than that of the SHP2-SHP099-SHP844 system, indicating that the protein side chain atoms become more stable at the above regions. The reduction of side chain fluctuations at allosteric site 1 (residues THR108-TRP112, LEU236-GLN245), allosteric site 2 (GLN79-GLN87, LEU262-GLN269) and catalytic area (residues HIS458-ARG465, ARG501-THR507) in the SHP2-SHP099-SHP844 system implies that the protein structure at allosteric site 1, allosteric site 2 and catalytic area were stabilized, respectively. In conclusion, the difference in RMSF fluctuations reflects that the flexibility of the allosteric site 1, allosteric site 2 and catalytic area was decreased in the SHP2-SHP099-SHP844 system, which might be caused by the tight interactions between the ligand and the residues in allosteric site 1, allosteric site 2 and catalytic area.

3.3 Conformation transitions of the SHP2 and SHP2-SHP099-SHP844 system

The biological function of a protein is determined by its specific structure, and the intervention of the ligand will cause the change of conformations of the SHP2 protein and affect its function. PCA can be utilized to obtain the information on the conformation status of SHP2 system and SHP2-SHP099-SHP844 system within 6-100 ns, respectively. Based on the calculation and diagonalization of the covariance matrix, principal component analysis was to project the protein dynamics information into the form of eigenvectors and eigenvalues. As shown in FIG.3, the top 20 PC values account for 77.5% and 79.3% of the total variation in the SHP2 system and the protein-ligand complex system, respectively. In the SHP2 system, the first two PCs accounted for 26.4% and 15.5% respectively, while the highest value of other PCs was only 8.7%. At the same time, the first two PCs in the protein-ligand complex system accounted for 34.8% and 14.2% of the total variance, respectively, while the maximum value of the other PCs was not higher than 5.1%. Because the first two feature vectors (PC1, PC2) account for a large proportion of the overall protein conformation, the conformational transitions of the SHP2 and SHP2-SHP099-SHP844 could be reflected by projecting the PC1 and PC2 into a two-dimensional (2D) map, respectively. The two conformational states marked as red dots and blue dots and the intermediate state was shown as white dots in the figure5. It was observed that in the FIG.3A most of dots were in the transition state, illustrating that the conformation of the SHP2 system was in an unstable state. The red dots and blue dots were distributed on both sides of the midline in the SHP2-SHP099-SHP844 system, indicating the SHP2-SHP099-SHP844 system was in a stable state.

The fluctuations of the protein backbone of the SHP2 system and the SHP2-SHP099-SHP844 system were displayed in the Anisotropic Network Model (ANM) (FIG.4) [46]. ANM could analyze how the SHP099 and SHP844 together affected the motions of allosteric sites 1 and 2 and the catalytic sites of SHP2, respectively[47]. The magnitude of the motion was represented by the length of the arrow and the direction of the motion was represented by the direction of the arrow. As shown in FIG.4, the overall length of the arrows in the SHP2-SHP099-SHP844 system was significantly reduced compared to the SHP2 system, implying that the overall protein backbone in the SHP2-SHP099-SHP844 system was in a more stable state. In addi-

tion, in the SHP2-SHP099-SHP844 system, the arrows fluctuations of residues THR108-TRP114 (red) and LEU236-GLN245 (yellow) in allosteric site 1, the arrows fluctuations of residues GLN79-GLN87 (blue) and LEU262-GLN269 (green) in allosteric site 2 have become smaller than those in the SHP2 system, which indicates that residues fluctuations become more stable. After the ligand binding to the receptor, the residues in the allosteric site were stable. Moreover, the fluctuations of residue at the catalytic region HIS458-ARG465 (P-loop purple) and ARG501-THR507 (Q-loop cyan) were reduced, meanwhile, the P-loop and Q-loop were close to each other, which causing that the active site in the PTP domain hardly be exposed to the substrate and further results in the decreased activity of SHP2 protein. The ANM study was consistent with RMSD and RMSF analysis results.

3.4 The correlation motion of SHP2 and SHP2-SHP099-SHP844 system

To further study the effects of SHP099 and SHP844 on the interactions between residues and residues of SHP2, DCCM analysis was performed on the SHP2 system and the SHP2-SHP099-SHP844 system, respectively. In FIG.5, the green and red regions represent the positive and negative correlation of movement of specific amino acids, respectively. The darker the color was, the stronger the correlation was. While the white areas (0.25 to - 0.25) represent no significant correlation. It can be seen from FIG.5, that the red and green regions were darker in the protein-ligand system than that in SHP2 system, which points that the combination of SHP099 and SHP844 in the allosteric site has enhanced the correlation movement of SHP2 residues. The area with obviously differences in the figure between the SHP2 and SHP2-SHP099-SHP844 system were marked by the black boxes. In the SHP2-SHP099-SHP844 system, residues GLN79-GLN87 (N-SH2) form strong positive correlation with residues LEU262-GLN269 (αB-αC link loop), HIS458-ARG465 (P-loop) and ARG501-THR507 (Q-loop), respectively, while in the SHP2 system there was a weak positive correlation movement. Moreover, in the SHP2-SHP099-SHP844 system, residues THR108-TRP112 and residues LEU236-GLN245 also have a strong positive correlation motion, while in the SHP2 system there was no or weak correlation motion. Meanwhile, the ARG421-ASP431 (WPD-loop) region forms strong positive correlation with Q-loop (HIS458-ARG465) and ARG501-THR507 (P-loop), respectively, while in the SHP2 system, there were almost no correlations. The positive correlation of the above-mentioned residue regions suggests that the residues in these regions moved in the same direction. From the above-mentioned analysis, the correlation movement of the SHP2-SHP099-SHP844 system was significantly raised compared to the SHP2 system, which implies that the interaction between residue and residue was significantly enhanced. In conclusion, the applications of the dual allosteric targeted protein inhibitors for SHP2 protein not only increases the interaction between C-terminal SH2, N-terminal SH2 and PTP domains, but also enhance the correlation of catalytic regions, which pulled residues tight and made the active site not easily exposed. The analysis results of DCCM are consistent with the RMSD, RMSF, PCA, and ANM analysis results.

3.5 Study on the binding stability of ligand to SHP2 in MD simulation

In order to study the interaction of ligands SHP099 and SHP844 with residues on allosteric sites 1 and 2 during MD simulation, respectively. The interaction between the ligand and the residue was shown in Figure.6 [48]. First, we observe the interactions between residues in allosteric site 1 and SHP099. Residues THR108, GLU110, ARG111, PHE113 and HIS114 located in the loop region connects N-SH2 and PTP domains (FIG. 6A). The residues THR108, GLU110 and PHE113 form H-bond interactions with the N22 amino group of SHP099; Cl2 and benzene rings of SHP099 form hydrogen bonds and P-Π conjugate interactions with residues HIS111, respectively. Residual LEU216, ASN217, THR218, and THR219 located in the loop region connecting C-SH2 and PTP domains form VDW interactions with SHP099. Residues GLU249, THR253, LEU254, GLN495, GLU250, GLN257, PRO491, ASP489, and LYS492 are located in the PTP domains of SHP2 protein. SHP099 form electrostatic interactions with residues GLU249, THR253, LEU254, and GLN495; N7 amino group of SHP099form an H-bond interaction with the residue GLU250. Observing the above data, it could be found that SHP099 and the residue at the allosteric site 1 of the SHP2 protein form a strong interaction. Residues were mainly distributed in the loop area connecting N-SH2 and PTP domains, the loop area connecting C-SH2 and PTP domains. This suggests that when the allosteric inhibitors SHP099 act on the SHP2 protein, which strengthens the connection between the C-SH2, N-SH2, and PTP domains

7

(allosteric site 1).

Then observe the allosteric site 2 in the interaction formed by SHP844 (FIG. 6B). Residues GLN79, TYR80, GLN87, GLU83, and HIS84 were all located in the N-SH2 region. The O2 atom of SHP844 could form H-bond interaction with residue GLN79; O4 and O5 of SHP844 could form H-bond interaction with residue TYR80; O4 of SHP844 could form H-bond interaction with residue GLN87; Ligand SHP844 can form VDW and electrostatic interaction with residues GLU83 and HIS84respectively. Meanwhile, H (HO2) of SHP844 could form H-bond interaction with residue LEU262 (αB and αC link of PTP domain); Residue ARG265(αB and αC link of PTP domain) could not only form H-bond with O2, but also form P-Π conjugate interaction with triazole ring; N3 and N4 form H-bond interaction with residue GLN269 (αB and αC link of PTP domain), respectively; O5 of SHP844 forms charge interaction with residue LYS280 (αD and βB link of PTP domain). In addition, SHP844 could form electrostatic interaction with TYR263 (αB and αC link of PTP domain), ASN281 (αB and αC link of PTP domain); SHP844 can form VDW interaction with residues SER264 (αB and αC link of PTP domain), LYS266 (αB and αC link of PTP domain), LYS274 (αD of PTP domain), and LEU283 (αB and αC link of PTP domain). Obviously, there was a strong interaction between the ligand SHP844 and the SHP2 protein N-SH2 and PTP domain (allosteric site 2), which indicates that SHP844 binding to N-SH2 and PTP domain at the allosteric site 2 was more conducive to maintaining SHP2 protein in self-inhibiting state.

In order to further study the binding SHP099 and SHP844 to the internal residues of the SHP2 protein, the binding free energy of the SHP2-SHP099-SHP844 system was calculated by Gromacs.4.5.5. The calculated binding free energies were decomposed into the individual residue contributions using the MM/GBSA method to identify the key residues and their contribution for elucidating the SHP2 interactions with the SHP099 and SHP844. The binding energy generated by the ligand and the protein mainly include (Table.S1): van der Waal energy(-492.983 kJ/mol), Electrostattic energy(-661.303 kJ/mol), Polar solvation energy(520.524 kJ/mol)and SASA energy(-41.324 kJ/mol). The overall binding energy value was -675.086 kJ/mol. Residues THR108, GLU110, ARG111, PHE113, , GLU250, GLU252, THR253, PRO491, and LYS492 form strong interactions with SHP099 (FIG.7A), with binding energies of -7.2500 kJ/mol, -24.2746 kJ/ mol, -10.6974 kJ/mol, -5.6764 kJ/mol, -24.6568 kJ/mol, -25.5066 kJ/mol, -8.0207 kJ/mol, -7.2894 kJ/mol, and -23.0974, respectively. Meanwhile, SHP844 stably binds to the residues GLN79, TYR80, GLN87, TYR263, ARG264, ARG265, GLN269, LYS280, and LYS266 (FIG.7B), and the binding energy were -2.8856 kJ/mol, -27.9870 kJ/mol, -3.1199 kJ/mol, -58.7989 kJ/mol, -2.1882 kJ/mol, -5.6728 kJ/mol, -8.1144 kJ/mol, -36.6236 kJ/mol, -79.7206 kJ/mol,respectively. The stronger the interactions (H-bond, hydrophobic, conjugation interactions) between protein and ligand were, the more negative the calculated binding free energy value becomes. Therefore, according to the performance of the binding energy of the residues at the allosteric sites 1 and 2 with the ligands, it was show that SHP099 and SHP844 bound tightly in the allosteric sites 1 and 2, respectively.

3.6 The ligand SHP-099-844 influencing the interactions of the residues in SHP2

By analyzing the RIN of SHP2 system and protein-ligand system, the effect of ligands on the structure and function of protein can be further explained. RIN was generated based on the average trajectory file, and could display various interactions between amino acids (for example: H-bond, VDW, electrostatic). There were two important parameters (shortest path betweenness and Closeness Centrality) in RIN analysis. Residues with high shortest path betweenness value can stabilize the structure of the protein, and residues with high Closeness Centrality value could affect the function of the protein [49-51]. The values of the Shortest Path Betweenness and Closeness Centrality for the SHP2 system and SHP2-SHP099-SHP844 system were shown in Table.S2. The notable residues was the residues with the value of Shortest Path Betweenness greater than 0.1. The notable residues include GLU110, GLN255, ARG501, ARG498, GLU258, PHE113, GLU250, VAL137, THR253, ASN306, GLU252 and ARG265 in the SHP2-SHP099-SHP844 system. The Shortest Path Betweenness value of notable residue GLU110 (allosteric site 1), ARG501 (Q-loop), and ARG265 (allosteric site 2) are 0.196, 0.144, and 0.101in SHP2-SHP099-SHP844 system, while the value were 0.004, 0.047, and 0.035 in SHP2 system. Compared to SHP2 system, the Shortest Path Betweenness value of notable residue

GLU110, ARG501, and ARG265 had obviously increased. Therefore, GLU110, ARG501, and ARG265 may be were key residues that kept SHP2 in auto-inhibition conformation after that combining SHP099 with SHP844 inhibiting on SHP2.

The results of the interaction between residues and residues at the allosteric site 1, allosteric site 2 and the binding site of the SHP2 system and the SHP2-SHP099-SHP844 system were shown in FIG.8. In RINs, the network edge line styles corresponded to no covalent and covalent residue interactions that were preserved in both two systems (black solid lines), presented only in the SHP2 system (green dotted lines) [52], or only in the SHP2-SHP099-SHP844 system (red dashed lines). From FIG.8, in binding site 1 of the SHP2-SHP099-SHP844 system, it could be found that the residue GLU110 (N-SH2 and C-SH2 link loop) with residues THR253 (αB at PTP domains) and GLN256 (αB of PTP domains) could formed VDW interactions and H-bond, respectively; residues GLU250 (αB of PTP domains) could formed VDW, H-bond, and H-bond interactions with residues THR219 (C-SH2 and PTP domain link loop), PHE247 (αA and αB link loop of PTP domains), THR253 (αB of PTP domains), respectively. However, these interactions do not exist in the SHP2 system. Obviously, in the SHP2-SHP099-SHP844 system, the interactions between the residues of C-SH2, N-SH2, and PTP domain around the binding site 1 had been significantly strengthened, which indicate the C-SH2, N-SH2, and PTP domain combine more tightened and stable. In addition, in the SHP2-SHP099-SHP844 system, residues around the binding site 2 residue GLN79 (N-SH2) could form VDW interactions with LEU262 (αB and αC link loop of PTP domains) and ARG265 (αB and αC link loop of PTP domains); residue LEU77 (N-SH2) forms an H-bond interaction with the residue TYR81 (N-SH2); Residue HIS84 can form VDW interactions with residues TYR80 (N-SH2) and GLN87 (N-SH2); Residue ARG265 (αB and αC link loop of PTP domains) could not only form one H-bond with residue GLY268(αB and αC link loop of PTP domains), but also can form VDW and H-bond interactions with residue GLN269 (αB and αC link loop of PTP domains), however, these interactions do not exist in the SHP2 system. The interaction between the N-SH2 and PTP domains residues increased significantly in the SHP2-SHP099-SHP844 system, which implies that SHP844 stabilizes the combination of N-SH2 and PTP domains at the allosteric site 2. Moreover, compared with the SHP2 system, there are many newly formed interactions between residues around the catalytically active pocket of the SHP2-SHP099-SHP844 system: Residues CYS459 (P-loop) forming H-bond and VDW interaction with ILE463 (P-loop) , residues GLY467 (P-loop) and residue MET504 (Q-loop) form H-bond interaction, residues ARG501 (Q-loop) and residues ILE463 (P-loop) and ARG498 (near Q-loop) could form VDW interactions, residue SER502 (Q-loop) forms VDW interaction with residue PHE285 (near pTy-loop). The increased interactions in the catalytically active region of the SHP2-SHP099-SHP844 system could make the active pocket tighten making it difficult to expose the active site. RIN analysis was consistent with RMSD, RMSF, DCCM, and ANM analysis results.

Conclusion

The purpose of this study is to investigate the effects of dual allosteric targeted protein inhibitors on SHP2 protein and to provide new ideas for the development of SHP2 protein inhibitors. First of all, 100 ns molecular dynamics simulations are performed for the SHP2 system and the SHP2-SHP099-SHP844 system, respectively. Then, post-MD post-simulation analysis (RMSD, RMSF, PCA, ANM, DCCM, and RIN) is applied to explore the mechanism of dual allosteric targeted protein inhibition on allosteric sites and active area. According to the results of RMSD, RMSF, it shows that SHP099 binding to residues THR108, GLU110, ARG111, PHE113, GLU250, GLU252, THR253, PRO491, LYS492 and SHP844 binds to residues GLN79, TYR80, GLN87, LEU262, TYR263, ARG265, GLN269, cause the decrease in fluctuations of residues THR108-TRP112 (N-SH2 and C-SH2 link loop), LEU236-GLN245 (C-SH2 and PTP link loop), GLN79-GLN87 (N-SH2), LEU262-GLN269 (αB and αC link loop of PTP domain), HIS458-ARG465 (P-loop), ARG501-THR507 (Q-loop) . Furthermore, the PCA, ANM, DCCM analysis suggest that the conformations of SHP2 are more stable and the interactions between residue areas (allosteric site 1, allosteric site 2, P-loop, and Q-loop) has visibly enhanced due to the combining SHP099 with SHP844,which might be crucial for the inhibition on allosteric sites and active area of the SHP2 . Therefore, the substrate cannot easily act on the active site, which can hardly be exposed. Finally, RIN result speculated that the residues GLU110, ARG265, and ARG501 are key residues, which make the residues of allosteric site 1, allosteric site 2, and catalytic

area tighten. Above all, those results provided the comprehensive structural information to understand the mechanism of combining SHP099 with SHP844 inhibiting SHP2 protein.

Supporting Information Available

Table.S1 The binding energy of ligands (SHP099, SHP844) with SHP2.

Table.S2 The value of Shortest Path Betweenness and Closeness Centrality in SHP2 system. B: The value of Shortest Path Betweenness and Closeness Centrality in SHP2-099-844 system.

References

1. Li, M., L.D. Maso, and S. Vaccarella, *Global trends in thyroid cancer incidence and the impact of overdiagnosis.* Lancet Diabetes Endocrinol, 2020. 8(6): p. 468-470.2. Khazaei, S., et al.,*Effects of Human Development Index and Its Components on Colorectal Cancer Incidence and Mortality: a Global Ecological Study.*Asian Pac J Cancer Prev, 2016. 17(S3): p. 253-6.3. Nolan, G.S., et al.,*Global incidence of incomplete surgical excision in adult patients with non-melanoma skin cancer: study protocol for a systematic review and meta-analysis of observational studies.* Syst Rev, 2020. 9(1): p. 83.4. Idrees, M., et al.,*PTPN11 (SHP2) Is Indispensable for Growth Factors and Cytokine Signal Transduction During Bovine Oocyte Maturation and Blastocyst Development.* Cells, 2019. 8(10).5. Grossmann, K.S., et al.,*The tyrosine phosphatase Shp2 in development and cancer.* Adv Cancer Res, 2010. 106: p. 53-89.6. Griger, J., et al.,*Loss of Ptpn11 (Shp2) drives satellite cells into quiescence.*Elife, 2017. 6.7. Spalinger, M.R., et al.,*Protein tyrosine phosphatase non-receptor type 2 and inflammatory bowel disease.* World J Gastroenterol, 2016. 22(3): p. 1034-44.8. Tiganis, T. and A.M. Bennett, *Protein tyrosine phosphatase function: the substrate perspective.* Biochem J, 2007. 402(1): p. 1-15.9. Bunda, S., et al.,*Inhibition of SHP2-mediated dephosphorylation of Ras suppresses oncogenesis.* Nat Commun, 2015. 6: p. 8859.10. Ehrman, L.A., et al.,*The protein tyrosine phosphatase Shp2 is required for the generation of oligodendrocyte progenitor cells and myelination in the mouse telencephalon.* J Neurosci, 2014. 34(10): p. 3767-78.11. Agazie, Y.M. and M.J. Hayman, *Molecular mechanism for a role of SHP2 in epidermal growth factor receptor signaling.* Mol Cell Biol, 2003. 23(21): p. 7875-86.12. Easton, J.B., A.R. Royer, and D.S. Middlemas, *The protein tyrosine phosphatase, Shp2, is required for the complete activation of the RAS/MAPK pathway by brain-derived neurotrophic factor.* J Neurochem, 2006. 97(3): p. 834-45.13. Yuan, X., et al.,*Recent Advances of SHP2 Inhibitors in Cancer Therapy: Current Development and Clinical Application.* J Med Chem, 2020.14. LaRochelle, J.R., et al.,*Structural and Functional Consequences of Three Cancer-Associated Mutations of the Oncogenic Phosphatase SHP2.* Biochemistry, 2016. 55(15): p. 2269-77.15. Xu, D. and C.K. Qu,*Protein tyrosine phosphatases in the JAK/STAT pathway.* Front Biosci, 2008. 13: p. 4925-32.16. Garcia Fortanet, J., et al., *Allosteric Inhibition of SHP2: Identification of a Potent, Selective, and Orally Efficacious Phosphatase Inhibitor.* J Med Chem, 2016. 59(17): p. 7773-82.17. Fodor, M., et al.,*Dual Allosteric Inhibition of SHP2 Phosphatase.* ACS Chem Biol, 2018. 13(3): p. 647-656.18. Sun, X.J., et al.,*Mutation selectivity of the allosteric SHP2 inhibitor SHP099.*Cancer Research, 2018. 78(13).19. Fodor, M., et al.,*Dual Allosteric Inhibition of SHP2 Phosphatase.* Acs Chemical Biology, 2018. 13(3): p. 647-656.20. Ravindran, A., S. Anishetty, and G. Pennathur, *Molecular dynamics of the membrane interaction and localisation of prodigiosin.* Journal of Molecular Graphics & Modelling, 2020. 98.21. Rahimi, A., S. Amjad-Iranagh, and H. Modarress, *Molecular dynamics simulation of coarse-grained poly(L-lysine) dendrimers.* Journal of Molecular Modeling, 2016. 22(3).22. Lundborg, M., et al.,*Predicting drug permeability through skin using molecular dynamics simulation.* Journal of Controlled Release, 2018. 283: p. 269-279.23. Sousa da Silva, A.W. and W.F. Vranken, *ACPYPE - AnteChamber PYthon Parser interfacE.* BMC Res Notes, 2012. 5: p. 367.24. Momen, R., et al.,*Exploration of the forbidden regions of the Ramachandran plot (varphi-psi) with QTAIM.* Phys Chem Chem Phys, 2017. 19(38): p. 26423-26434.25. Gruebele,

M.,*Protein dynamics in simulation and experiment.* J Am Chem Soc, 2014. 136(48): p. 16695-7.26. Nava, M.,*Implementing Dimer Metadynamics Using Gromacs.* Journal of Computational Chemistry, 2018. 39(25): p. 2126-2132.27. Rakhshani, H., E. Dehghanian, and A. Rahati, *Enhanced GROMACS: toward a better numerical simulation framework.* Journal of Molecular Modeling, 2019. 25(12).28. Maiorov, V.N. and G.M. Crippen, *Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins.* J Mol Biol, 1994. 235(2): p. 625-34.29. Hildebrandt, A.K., et al., *Efficient Computation of Root Mean Square Deviations Under Rigid Transformations.* Journal of Computational Chemistry, 2014. 35(10): p. 765-771.30. Margreitter, C. and C. Oostenbrink, *MDplot: Visualise Molecular Dynamics.* R J, 2017. 9(1): p. 164-186.31. Assadollahi, V., et al.,*Interaction and molecular dynamics simulation study of Osimertinib (AstraZeneca 9291) anticancer drug with the EGFR kinase domain in native protein and mutated L844V and C797S.* J Cell Biochem, 2019. 120(8): p. 13046-13055.32. Wang, L., J. Chambers, and R. Abel, *Protein-Ligand Binding Free Energy Calculations with FEP.* Methods Mol Biol, 2019. 2022: p. 201-232.33. Irwin, B.W.J. and D.J. Huggins, *Estimating Atomic Contributions to Hydration and Binding Using Free Energy Perturbation.* J Chem Theory Comput, 2018. 14(6): p. 3218-3227.34. Aldeghi, M., J.P. Bluck, and P.C. Biggin, *Absolute Alchemical Free Energy Calculations for Ligand Binding: A Beginner's Guide.* Methods Mol Biol, 2018. 1762: p. 199-232.35. Giuliani, A., *The application of principal component analysis to drug discovery and biomedical data.* Drug Discovery Today, 2017. 22(7): p. 1069-1076.36. Latry, P., et al.,*Use of principal component analysis in the evaluation of adherence to statin treatment: a method to determine a target population for public health intervention.* Fundamental & Clinical Pharmacology, 2010. 24: p. 70-70.37. Liu, Y., et al.,*Adaptive robust principal component analysis.* Neural Networks, 2019. 119: p. 85-92.38. Lin, Z., L. Wang, and J. Cao, *Interpretable functional principal component analysis.*Biometrics, 2016. 72(3): p. 846-54.39. Chuang, T.W., et al.,*Cross-Correlation Map Analyses Show Weather Variation Influences on Mosquito Abundance Patterns in Saginaw County, Michigan, 1989-2005.*Journal of Medical Entomology, 2012. 49(4): p. 851-858.40. Sadaie, W., et al.,*Quantitative In Vivo Fluorescence Cross-Correlation Analyses Highlight the Importance of Competitive Effects in the Regulation of Protein-Protein Interactions.* Molecular and Cellular Biology, 2014. 34(17): p. 3272-3290.41. Liu, R. and J. Hu,*Computational prediction of heme-binding residues by exploiting residue interaction network.* PLoS One, 2011. 6(10): p. e25560.42. Hu, G., et al.,*Residue interaction network analysis of Dronpa and a DNA clamp.*Journal of Theoretical Biology, 2014. 348: p. 55-64.43. Xia, Q. and Y.R. Ding,*Thermostability of Lipase A and Dynamic Communication Based on Residue Interaction Network.* Protein and Peptide Letters, 2019. 26(9): p. 702-716.44. Mannige, R.V., J. Kundu, and S. Whitelam, *The Ramachandran Number: An Order Parameter for Protein Geometry.* Plos One, 2016. 11(8).45. Ho, B.K. and R. Brasseur,*The Ramachandran plots of glycine and pre-proline.* BMC Struct Biol, 2005. 5: p. 14.46. Eyal, E., G. Lum, and I. Bahar, *The anisotropic network model web server at 2015 (ANM 2.0).* Bioinformatics, 2015. 31(9): p. 1487-9.47. Sarkar, R.,*Stiffening of flexible SUMO1 protein upon peptide-binding: Analysis with anisotropic network model.* Math Biosci, 2018. 295: p. 67-72.48. Farrokhzadeh, A., F.B. Akher, and M.E.S. Soliman, *Probing the Dynamic Mechanism of Uncommon Allosteric Inhibitors Optimized to Enhance Drug Selectivity of SHP2 with Therapeutic Potential for Cancer Treatment.* Appl Biochem Biotechnol, 2019. 188(1): p. 260-281.49. Singh, S.S., et al.,*Scaling in topological properties of brain networks.* Sci Rep, 2016. 6: p. 24926.50. Jiao, X., et al.,*Construction and application of the weighted amino acid network based on energy.* Phys Rev E Stat Nonlin Soft Matter Phys, 2007. 75(5 Pt 1): p. 051903.51. Xue, W., et al.,*Molecular modeling and residue interaction network studies on the mechanism of binding and resistance of the HCV NS5B polymerase mutants to VX-222 and ANA598.* Antiviral Res, 2014. 104: p. 40-51.52. Jiao, X. and S. Ranganathan, *Prediction of interface residue based on the features of residue interaction network.* J Theor Biol, 2017. 432: p. 49-54.

A　　　　　　　**Residue Cross Correlation**

B　　　　　　　**Residue Cross Correlation**