Discovery of new structural hits as KRAS G12C inhibitor targeting switch-II pocket using computational screening approach

Kefei Xu [1], Xiaoyun Wu [2], Shuguang Wu [2]

Abstract—KRAS oncogene mutation is one of the most frequent mutations in human cancers. This mutation can cause overactive signaling in the cell and ultimately lead to cancer. Although different types of inhibitors have been found, no effective KRAS inhibitors are approved for clinical use. Recently, a new switch II pocket on the surface of KRAS G12C mutant protein (KRAS G12C S-IIP), in which the twelfth amino acid glycine is replaced by cysteine, has been found to be the most promising allosteric binding site which can be targeted directly by small molecular KRAS inhibitor. The purpose of the present study was to screen new structural types of KRAS inhibitor targeting S-IIP of KRAS G12C from the compound database using computational screening approach. The study results support our hypothesis that five new structural hit compounds with good docking scores, that are different from KRAS inhibitor ARS-1620, were found against the crystal structure of S-IIP of KRAS G12C protein (PDB code 5V9U). The study also found that hit compounds can form hydrogen-bonds and π-π stacking interaction with the amino acid residues around the S-IIP binding site, suggesting that hydrogen-bond and π–π stacking mediate the interactions between the hit compounds and target protein. Moreover, none of the five new molecular scaffold compounds have been reported so far to have inhibitory activity on KRAS. Further structural optimization needs to adjust the potency and physicochemical properties of these hit compounds.

Index Terms—allosteric binding site, ARS-1620, computational screening approach, hit compounds, inhibitor, KRAS G12C, molecular scaffold.

1 INTRODUCTION

Cancer is a leading cause of mortality throughout the globe. The occurrence and development of cancer are often associated with gene mutation or activation in the body. The abnormal genes that induce cancer are called oncogenes. The oncogenes can lead to the production of permanently activated proteins. As a result, this can cause overactive signaling in the cell, result cell proliferation, and ultimately lead to cancer. Inhibition of mutated or activated oncogene is clinically effective in the treatment of cancers [1].

KRAS is an oncogene member of the RAS gene family. It is called KRAS because it was first identified as an oncogene in Kirsten Rat Sarcoma virus and the cause of rat sarcomas [2]. The protein encoded by the KRAS gene is the KRA protein. When KRAS gene mutation occurs, it changes the gene's instruction for making a protein, thus causing the abnormal activation of KRAS protein to induce cancers [3].

- [1] *Kefei Xu, The Hill School, Pottstown, PA19464, USA, PH-6102472197. E-mail: clydekefei.xu@gmail.com*
- [2] *Southern Medical University, China,*
  Xiaoyun Wu, E-mail: xywu@smu.edu.cn
  Shuguang Wu, E-mail: shuguang@smu.edu.cn

KRAS mutation is one of the most frequent human mutations in human cancers. For instance, KRAS mutation is present in nearly 90% of pancreatic cancers, 34.6% of colon cancers and 16.5% of lung cancers [4]. KRAS G12C mutant form, in which the twelfth amino acid, glycine, is replaced by cysteine, is one of the common mutation types of KRAS mutations. The KRAS G12C mutation causes inactive KRAS to become highly active KRAS, thereby activating the signaling pathway inducing cancers [4].

Despite more than 30 years of effort, no effective KRAS inhibitor has been approved for clinical use [5]. Why is it so difficult to develop KRAS inhibitors? This is because the KRAS protein surface lacks deep hydrophobic pockets that would allow for tight binding of small ligands or inhibitors, which make the direct inhibition of KRAS face a lack of efficacy [6]. In view of the KRAS inhibitors that have been discovered to date, although many different types of inhibitors have been studied, the most exciting development is still the discovery of direct inhibitors targeting on the G12C mutants [7]. Of particular importance is the recent discovery of a new unknown allosteric site near G12C switch II pocket region of KRAS protein, known as KRAS G12C S-IIP [8-9]. Inhibitors targeting on G12C S-IIP are called specific KRAS G12C S-IIP inhibitors. The presence of G12CS-IIP may enable design of more potent KRAS mutant inhibitors. The discovery of KRAS G12C S-IIP represents the landmark discovery of KRAS mutant inhibitors, because it changes the perception about KRAS as a drug target.

Drug action usually involves the interaction between drug ligands and target proteins. Therefore, target selection and the identification of hit compounds are the source of innovation in early stages of drug development. Hit compounds are identified largely through high-throughput

screen (HTS), where an experimental assay is used to screen hundreds of thousands or even millions of compounds from compound library [10]. Such experiments are expensive and time consuming. In recent years, there has been a growing interest in using automated computational approach to screen hit compounds. Computational screening or virtual screening consists of virtually docking collections of millions of compounds into a biological target, followed by an evaluation of the fit scoring. It offers a quick assessment of huge libraries and reduces the number of compounds that need testing in identifying early hits. It can reduce cost, increase efficiency and improve rate of success [11].

In the present study, in order to screen new structural types of KRAS G12C S-IIP inhibitor from the database containing a large number of compounds, we selected the 3D crystal structure of KRAS G12C S-IIP protein (from Protein Data Bank, PDB code 5V9U) as the target protein [8]. The reason why we chose this crystal protein for screening is that it can be combined with the most potent KRAS inhibitor ARS-1620[8]. Therefore, our hypothesis is that new structural types of KRAS G12C S-IIP inhibitor that is different from ARS-1620 may be contained in our selected compound library containing one million and four hundred thousand different structures. On the basis of the target model of 5V9U crystal structure, our research focused on screening new structural types of KRAS G12C S-IIP inhibitor from the database containing a large number of compounds using computational molecular screening approach, aiming to identify new molecular scaffold hits targeting S-IIP of KRAS G12C as well as to examine the binding interactions between each ligand and its target protein.

2 Results

2.1 Validation of Docking

The docking study provides an insight into the understanding of protein–inhibitor interactions and provides the importance of the structural features of active site of the protein and ligand in the interaction process. In order to validate the docking protocols used to predict the binding orientation of the KRAS inhibitors, the ABS-1620 co-crystal ligand was redocked with Glide with default parameters. The calculated heavy-atom root mean squared deviation (RMSD) values between the crystal ligand and redocked ligand were 0.311, 0.190 and 0.340 Å when HTVS, SP and XP scoring functions were used, respectively (Table 1 and Fig 1), which confirms that the docking protocol used in the present investigation is suitable for the prediction of the bioactive conformation of all the inhibitors.

Table 1 The result of redocking of ABS-1620 into the switch-II pocket of KRAS

	HTVS	SP	XP
G-score (Kcal/mol)	−11.465	−12.850	−11.584
RMSD (Å)	0.311	0.190	0.340

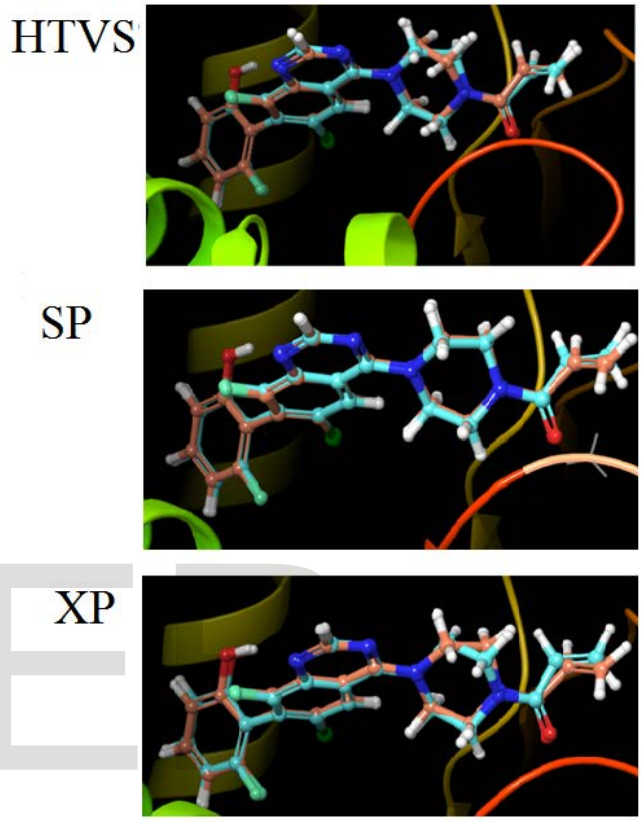


Fig. 1 Superimposed X-ray crystal structure (orange) and docked structure (cyan) of ABS-1620 into the switch–II pocket of KRAS G12C (PDB ID: 5V9U)

2.2 Binding Energies between Ligand and Target by Docking-Based Computational Screening

The proposed hierarchical multistage VS method was applied. Docking process was done in three stages in our analysis. Initially, HTVS was used as the initial docking protocol in VS. Consequently, 1% of the ligands were passed on to the next stage of SP docking, which gave more accurate results than HTVS. Once docking with SP is done, 10% of the total ligands or the top scoring ligands were preceded to XP docking. Then, XP docking was performed. The best hits with good glide score are shown in Table 2. As illustrated, five hit compounds exhibit the strong binding affinity against KRAS G12C S-IIP. ZINC19324064 exhibit the strongest binding affinity (−10.941 kcal/mol), followed by ZINC13388179 (−10.806 kcal/mol), ZINC05196521 (−10.733 kcal/mol), ZINC03908593 (−10.681 kcal/mol) and finally ZINC00041855 (−10.575 kcal/mol). The score of these five hits is very close to

ARS-1620 (−11.584 kcal/mol). These glide scores represent how well the hit compound and target protein are suitable for binding to each other. The more negative the score, the more suitable the hit compound is for binding with its target.

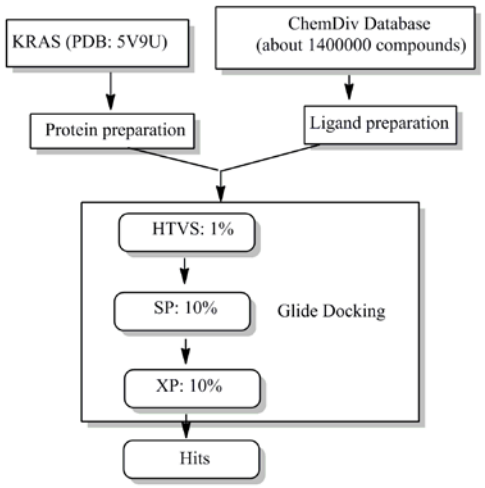
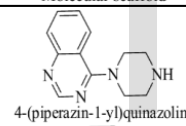
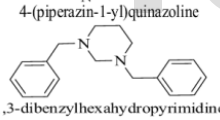
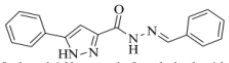
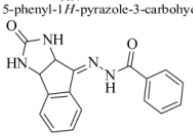
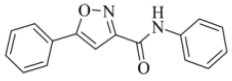
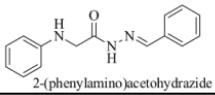


Fig. 2 Workflow of computational screening of KRAS G12C S-II inhibitors

Table 2 The molecular scaffold and Glide score of the identified top 5 hit molecules

	Title	Molecular scaffold	XP G-score (kcal/mol)
1	ABS-1620	 4-(piperazin-1-yl)quinazoline	−11.584
2	ZINC19324064	 1,3-dibenzylhexahydropyrimidine	−10.941
3	ZINC13388179	 5-phenyl-1*H*-pyrazole-3-carbohydrazide	−10.806
4	ZINC05196521	 indeno[1,2-*d*]imidazol-8(2*H*)-ylidene) benzohydrazide	−10.733
5	ZINC03908593	 5-substituted isoxazole-3-carboxamide	−10.681
6	ZINC00041855	 2-(phenylamino)acetohydrazide	−10.575

2.3 Analysis of Binding Model Between Ligand and Receptor

Besides looking into the binding affinities and energies involved in the interaction of five hits and target protein, the bond formed between the ligands and the S-IIP binding on the target were also be analyzed. The two-dimensional and three-dimensional conformations of the identified top 5 hit molecules and reference compound ARS-1620 are shown in Fig 3 and Table 3, respectively.

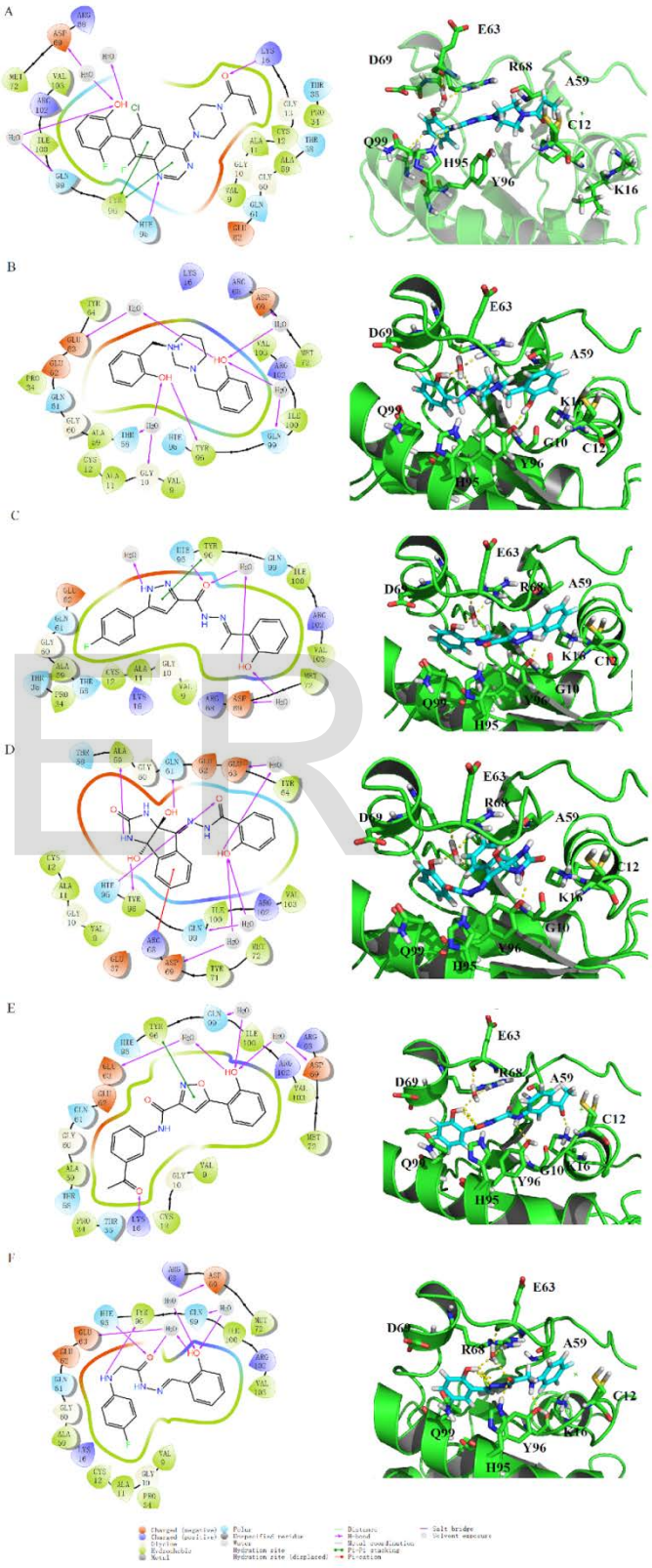


Fig. 3 The two-dimensional and three-dimensional binding mode of

ABS-1620 (A) and the obtained top 5 hit molecules: ZINC19324064 (B), ZINC 13388179(C), ZINC 05196521(D), ZINC 03908593(E), and ZINC 00041855(F) into the switch–II pocket of KRAS G12C (PDB ID: 5V9U)

Table 3 The amino acids involved in the interactions with KRAS for the identified top 5 hit molecules

	Title	Amino acids involved in Hydrogen-bond	Amino acids involved in Water-mediated Hydrogen-bond	Amino acids involved in π-π stacking
1	ABS-1620	LYS16, HIS95	ASP69, GLN99	TYR96
2	ZINC19324064	TYR96	GLY10, GLU63, GLN99	-
3	ZINC13388179	HIS95	ASP69	TYR96
4	ZINC05196521	ALA59, GLN61, HIS95, TYR96	ASP69, GLN99	-
5	ZINC03908593	LYS16	GLN63, ASP69, GLN99	TYR96
6	ZINC00041855	HIS9, TYR96	GLN63, ASP69, GLN99	-

The amino acid residues of hydrogen bonds formed by ARS-1620 and target protein were LYS16 and HIS95 (hydrogen-bond) as well as ASP69 and GLN99 (water-mediated hydrogen-bond) (Table 3 and Fig 3-A). A π-π stacking is present between the ligand and the TYR96 residues of target protein [8].

The docking mode shows that the hit molecule ZINC19324064 formed one hydrogen-bond with residues of TYR96, and water-mediated hydrogen-bonds with GLY10, THR58, GLU63 and GLN99 of its target protein, respectively (Table 3 and Fig 3-B). ZINC13388179 formed one hydrogen-bond with HIS95 and water-mediated hydrogen-bonds with ASP69 of the target, respectively (Table 3 and Fig 3-C). The π-π stacking is also present between ZINC13388179 and the TYR96 residue of the target protein. ZINC05196521 formed four hydrogen-bonds with the residue of GLN61, ALA59, HIS95, TYR96, and water-mediated hydrogen-bonds with ASP69 and GLN99 of target, respectively (Table 3 and Fig 3-D). ZINC03908593 formed one hydrogen-bond with the amino acids of LYS 16, and water-mediated hydrogen-bonds with GLN63, ASP69 and GLN99 of the target, respectively (Table 3 and Fig 3 E). The π-π stacking is also present between ZINC03908593 and the TYR96 residues of the target protein. ZINC00041855 formed two hydrogen-bonds with the amino acids of HIS9, TYR96, and water-mediated hydrogen-bonds with GLN63, ASP69 and GLN99 of the target, respectively (Table 3 and Fig 3-F).

3 DISCUSSION

In present study, the computational screening approach was used to screen the large chemical database to find new molecular scaffold inhibitors targeting the KRAS G12C S-IIP. The most encouraging result of this study is the discovery of five new structural hit compounds with good docking scores against the crystal structure of S-IIP of KRAS G12C protein. None of these five hit compounds has been reported to have inhibitory activity on KRAS.

The small molecular ligands in this study were obtained from the commercial Chemdiv database, in which the compounds have different molecular scaffolds. The compounds in this library may have or have no biological activity reported. Among the five newly discovered molecules, ZINC19324064 was reported to be a myeloperoxidase inhibitor [12], and ZINC00041855 is a pfENR inhibitor [13]. The biological properties of the other three compounds, ZINC13388179, ZINC05196521 and ZINC03908593, have not been studied to date. The molecular scaffolds of these five hits were shown in Table 2. These new scaffolds may be considered as potential KRAS G12C S-IIP inhibitors for structural modification and optimization.

The computational screening approach based on the receptor structure is most commonly used when a three-dimensional structure of the receptor is available. However, even if the crystal structure of target protein from Protein Databank is available, its quality needs to be carefully evaluated [14]. In this study, a crystal structure of KRAS (PDB code 5V9U) was used to screen hit compounds from the compound library. Prior to docking, the following questions must be taken into consideration. First, the crystal structure of 5V9U must be an experimentally resolved receptor which is a co-crystal complex of the inhibitor ARS-1620 and KRAS protein. Second, the Switch –II pocket of KRAS needs to be defined for docking. Finally, the binding site presents the area around pocket (Table 3). The 5V9U crystal structure selected in our study has reached the above standards, thus ensuring the reliability and success of the experimental results.

The present study also found the hydrogen bond, including water-involve hydrogen bond, between the hit compounds and the receptor protein. A hydrogen bond is an electrostatic attraction between an hydrogen which is bound to a more electronegative atom such as nitrogen or oxygen and another adjacent atom bearing a lone pair of electrons [15]. Because of the retention of water molecules in Glide docking, water mediated hydrogen bonds were observed. In addition, a π–π stacking was also observed in the hit compounds, ZINC13388179 and ZINC03908593 as well as ARS-1620. The π–π stacking refers to attractive, noncovalent interactions between aromatic rings of ligand and receptor, since they contain π bonds. These results suggest that hydrogen-bond and π–π stacking mediate the interaction between the hit compound and target protein [16].

Considering the results of present study, the five hit compounds need to move to the next experimental step, e.g. chemical purchase or synthesis (in case of compounds not available) and in vitro and in vivo studies of enzymatic and pharmacological activity. On the other hand, the hit compounds need to be optimized to covalently react with SH group of Cys12 of the switch II pocket of KRAS by introduction of a cysteine-reactive moiety such as acrylamide

[17].The structural optimization also needed to adjust their potency and selectivity as well as their physicochemical properties.

4 MATERIALS AND METHODS

4.1 PROTEIN PREPARATION

A single set of receptor coordinates of ABS-1620 in complex with KRAS (pdb code 5V9U) was retrieved from the Protein Databank Bank (PDB) (http://www.rcsb.org/).The protein structures were prepared and minimized using the protein preparation wizard of Schrödinger suite (New York, NY, 2017). Hydrogen atoms were added, proper bond orders were assigned, and missing side chains were generated using prime module. The protonation states of each side chain were generated using Epik at pH = 7. The relevance and accuracy of the protein preparation were examined by extracting the ARS-1620 molecule from the co-crystal structure and by re-docking the molecule to the crystal structure. Protein minimization was performed with the default cutoff root mean square deviation (RMSD) value of 0.3 Å using Optimized Potentials for Liquid Simulations (OPLS) 2005 force field. The ligands were prepared using the Ligprep module of the Schrödinger Suite. These prepared structures were used for molecular docking studies.

4.2 LIGAND PREPARATION

All compounds from Chemdiv were retrieved in the .SDF format from the public ZINC library. All ligands were prepared with Ligprep software from the Schrödinger Suite and energy minimized using OPLS-2005 force field. For each ligand, all possible ionization states were assigned at pH 7 ± 1, the chirality was retained, and the tautomers were generated.

4.3 VALIDATION OF DOCKING

The prepared crystal structure was used for grid generation on Glide [18-20], where the center is the active site of the receptor with the co-crystallized ligand. Aiming to validate the molecular docking approach, the co-crystallized ligand ABS-1620 was submitted to a redocking in the target active site using Glide HTVS, SP and XP scoring functions.

4.4 DOCKING-BASED VIRTUAL SCREENING

The ligands were docked in a step-wise manner using high throughput virtual screening (HTVS) followed by standard precision (SP) and extra precision (XP) modes with default parameters. Since HTVS is a fast process, this is used as the initial docking protocol in VS of large databases. After HTVS, 1% of the resulted ligands were taken to the next process called SP docking. Once docking with SP is done, 10% of the total ligands or the top scoring ligands were taken and were preceded to XP docking. XP docking is the final procedure in VS technique. It can differentiate all the minute interaction modes between ligand and the protein. This process increases the docking accuracy and thus gives the best results during the VS procedures as it can eliminate those false positives that had survived in the SP stage.

ACKNOWLEDGEMENTS

We would like to thank Professor Jiajie Zhang at School of Pharmaceutical Sciences, Southern Medical University for his excellent comments and suggestions on the completion of this manuscript.

REFERENCE

[1] Croce, CM. "Oncogenes and cancer." N Engl J Med, vol. 358, no. 5, pp. 502-511, 2008.

[2] Tsuchida, N., Ryder, T., Ohtsubo, E. "Nucleotide sequence of the oncogene encoding p21 transforming protein of Kirsten murine sarcoma virus." *Science*, vol. 217, no. 4563, pp. 937–939, 1982.

[3] Simanshu, DK. Nissley, DV., McCormick, F. "RAS proteins and their regulators in human disease." *Cell*, vol.170, no. 1, pp.17-33, 2017.

[4] Singh, H., Longo, DL., Chabner, BA. "Improving prospects for targeting RAS." *J Clin Oncol , vol.* 33, no. 31, pp.3650-3659, 2015.

[5] Cox, AD., Fesik, SW., Kimmelman, AC.,Luo, J., Der, CJ. "Drugging the undruggable RAS: mission possible?" *Nat Rev Drug Discov*, vol. 13, no.11, pp.828-851, 2014.

[6] Cromm, PM., Spiegel, J., Grossmann, TN., Waldmann, H. "Direct modulation of small GTPase activity and function." *Angew. Chem. Int. Ed. Engl*. Vol. 54, no. 46, pp.13516– 13537, 2015.

[7] Ostrem, JM., Peters, U., Sos, ML., Wells, JA., Shokat, KM. "K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions". *Nature*, vol. 503, no. 7477, pp.548-551, 2013.

[8] Janes, MR., Zhang, J., Li, LS., Hansen, R., Peters, U., Guo, X. *et al*. "Targeting KRAS mutant cancers with a covalent G12C-specific inhibitor." *Cell*, vol. 172, no. 3, pp.578-589, 2018.

[9] Patricelli, MP., *et al*. "Selective inhibition of oncogenic KRAS output with small molecules targeting the inactive state." *Cancer Discov*, vol. 6, no. 3, pp.316-329, 2016.

[10] Lionta, E., Spyrou, G., Vassilatis, DK., Cournia, Z. "Structure-based virtual screening for drug discovery: principles, application and recent advances." *Curr Top Med Chem*, vol. 14, no. 16, pp.1923-1938, 2014.

[11] Guedes, IA., de Magalhães, CS., Dardenne, LE. "Receptor-ligand molecular docking" *Biophys Rev.*, vol. 6, pp.75-87, 2014.

[12] Aldib, I. *et al.* "Evaluation of new scaffolds of myeloperoxidase inhibitors by rational design combined with high-throughput virtual screening." *J Med Chem*, vol. 55, no. 16, pp.7208-7218, 2012.

[13] Samal, RP., *et al*. " Design, synthesis, structural characterization by IR, ^{1}H, ^{13}C, ^{15}N, 2D-NMR, X-ray diffraction and evaluation of a new class of phenylaminoacetic acid benzylidene hydrazines as pfENR inhibitors." *Chemical Biology & Drug Design*, vol. 81, no.6, pp. 715-729, 2013.

[14] Macalino, SJ., Gosu, V., Hong, S., Choi, S. "Role of computer-aided drug design in modern drug discovery." *Arch Pharm Res*, vol. 38, no. 6, pp.1686-1701, 2015.

[15] Davis, AM., Teague, SJ. " Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis." *Angew Chem Int Ed Engl*, vol. 38, no. 6, pp.736-749, 1999.

[16] Phipps, MJ., Fox, T., Tautermann, CS., Skylarism, CK. "Energy

decomposition analysis approaches and their evaluation on prototypical protein-drug interaction patterns." *Chem Soc Rev*, vol. 44, no. 10, pp.3177-211, 2015.

[17] Baillie, TA. "Targeted covalent inhibitors for drug design." *Angew Chem Int Ed Engl*, vol. 55, no. 43, pp.13408-13421, 2016.

[18] Friesner, RA., *et al.* "Extra precision Glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes." *J. Med. Chem,* vol. 49, no. 21, pp. 6177-6196, 2006.

[19] Halgren, TA., *et al*. "Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening." *J Med Chem*, vol. 47. no. 7, pp.1750-1759, 2004.

[20] Friesner, RA., *et al*. "Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy." *J Med Chem,* vol. 47, no. 7, pp. 1739-1749, 2004.