Project Title:

Computational structure-based design of protein inhibitors

Name :○Kam Zhang, Arnout Voet, Ashutosh Kumar, Kamlesh Sahu, Muhammad Muddassar

Affiliation :Zhang Research Initiative Unit Advanced Science Institute, Wako Institute

1. Background and purpose of the project, relationship of the project with other projects

Fragment based drug design has developed significantly over the past decade and is now recognized as successful method of hit identification and lead generation. The method was proposed in 1996 and has gained more and more interest in following years to finally become a tangible alternative to high throughput screening. Fragment based drug design focuses on low molecular weight compounds that target sub-pockets within the overall active site. It samples chemical space more efficiently because the number of fragments needed to cover all reasonable chemotypes is many orders of magnitude smaller than that for more drug-like compounds, where multiple chemotypes are combined. Fragment hits are expected to be more suitable starting point for hit to lead optimization due to their reduced complexity, which leaves more freedom for multidimensional property optimization of the fragment hits, usually by addition of new chemical functions or alternatively by linking of two fragment hits binding in adjacent pockets.

In the past, this strategy has led to the discovery of new scaffolds that were later combined or grown into high affinity inhibitors. However, the small size of fragments and low binding affinity makes it particularly difficult to detect them in standard biochemical assays. Instead biophysical methods such as NMR, protein crystallography and surface Plasmon resonance are used to test up to hundreds or several thousands of compounds. These methods additionally benefit from the fact that they yield structural information about fragment binding poses but despite their utility, there are significant time, labor, and materials costs associated with biophysical screening. Also, considering that estimated size of chemically relevant fragment space is in the order of 10^8 and the number of commercially available fragments is over 300000, the throughput and the associated cost limit the application of biophysical screening.

Structure based virtual screening of drug like libraries has already been proved to be an efficient technology in hit discovery. Furthermore structure based virtual screening was integrated with experimental screening providing drug like hits for large variety of targets. Based on positive experiences one can conclude structure based virtual screening could also support fragment based drug design by molecular docking and prioritizing fragments using various ranking and scoring methodologies such as molecular dynamics simulations. Despite broad application of structure based virtual approaches, *in silico* fragment screening still plays a minor role. Although there have been some case studies reported that demonstrate virtual screening to be a powerful tool to discover potent fragments, this approach is still often considered to be too unreliable using popular computer tools. Indeed, facing the shortcomings of computational protocols with respect to ranking and affinity prediction of small fragments, more elaborate protocols have to be evolved departing from routine virtual screening strategies.

The objective of this project is to develop a virtual screening pipeline integrating structure based virtual screening approaches and utilize them in the rational identification of low molecular weight compounds with weak binding affinity towards a

variety of therapeutically relevant protein targets. A key theme is the development and utilization of computational approaches that incorporate protein flexibility considerations into structure based discovery by using flexible docking and molecular dynamics simulation to identify desired fragments can be later combined or grown into high-affinity inhibitors.

In this project, we proposed to develop a virtual screening protocol and to use it for the discovery of inhibitors of a protein lysine methyltransferase G9a and Sumoylation E2-conjugating enzyme Ubc9. However, the approach is inherently generic and can be applied to many fragment based drug discovery programs for disease like cancer, inflammation, neurodegenerative diseases.

Histone methylation has important roles in regulating transcription, genome integrity and epigenetic inheritance. G9a is a protein lysine methyltransferase that catalyzes methylation of lysine 9 of histone 3 (H3K9) and lysine 373 of tumor suppressor p53. G9a inihibition is an attractive therapeutic strategy for cancer as G9a is over expressed in human cancers and its knockdown inhibits cancer cell growth.

Sumoylation is a post-translational modification that plays an important role in a wide range of cellular processes including DNA replication and repair, chromosome packing and dynamics, genome integrity, nuclear transport, signal transduction and cell proliferation. Among the proteins involved in sumoylation pathway, Ubc9 is the sole E2-conjugating enzyme required for sumoylation and plays a central role by interacting with almost all the partners required for sumoylation. Ubc9 has been implicated in a variety of human malignancies such as ovarian carcinoma, melanoma and lung adenocarcinoma, suggesting that Ubc9 inhibition could be a potential therapeutic approach to control tumorigenesis.

2. Specific usage status of the system and calculation method

The basic goal of our virtual screening protocol is to computationally screen millions of commercially available small molecules against a specific target protein to prioritize small number of compounds for biological testing and for their further development to high affinity inhibitors. However, *in silico* structure based virtual screening requires intensive computing especially in case of large database with millions of chemical compounds, which is in the range of few Tflops per day on one target protein. Our virtual screening protocol involves the screening of small molecule library using flexible molecular docking simulation. We have used RosettaLigand program for this purpose. The hits which ranked higher were further prioritized by molecular dynamics simulation based binding free energy calculations. AMBER program and MM-GB(PB)SA approach was employed for this purpose. Our procedure incorporate full protein and ligand flexibility which greatly increases the computation time due to the vast number of conformations need to be explored; however, this allows more accurate treatment of protein ligand interactions.

All molecular docking simulations were performed using GOLD. GOLD has been used to dock virtual libraries of commercial available compounds to the different protein receptor structures relevant for our research. The result of this simulation is an estimated affinity of a compound for a given receptor as well as a putative binding mode. Based on this information desired compounds can be selected for further evaluation.

Our molecular dynamics simulations were performed using GROMACS. These simulations are important to gain insights in the dynamics of a protein and to sample conformations. These conformations are important to assess drugability of protein interfaces and to determine the optimal strategy to develop a drug using computational methods.

In our quest for potent G9a inhibitor as

anti-cancer agent, 108,866 CPU hours of RICC were used for docking (using GOLD), molecular dynamics calculations (using AMBER) and MMPBSA/GBSA. Docking results were post filtered by a Pharmacophore model that was developed using ligandscout (a tool for 3D pharmacophore modeling) and MOE (Molecular operating environment from chemical computing group) based on crystal structure of G9a (a protein methyl transferase overexpressed in cancer) and it's known inhibitors (2,4-Diamino-7-aminoalkoxy-quinazolines (DXQ) analogues).

3. Result

We have developed a virtual screening workflow integrating various structure based virtual screening approaches that can help in the prioritization of virtual screening hits to support early stages of drug discovery. The workflow exploits the use of full flexible molecular docking simulations and MM-GB(PB)SA based binding free energy calculations to prioritize potential virtual screening hits. We have used the workflow for the identification of small molecule inhibitors of SUMO E1 and SUMO E2 (Ubc9), sumoylation pathway proteins regulating diverse functions like cell proliferation, signal transduction, DNA replication and repair, genome integrity etc.

To identify potential small molecule inhibitor of SUMO E1 that could be used in structural and biochemical studies, we have used our virtual screening workflow to prioritize hits from Maybridge small molecule library for biological assay. About 24 hits were acquired and assayed using *in situ* sumoylation assay. Out of 24 hits, four hits showed more than 85 % inhibition of sumoylation reaction with most active compound showed the IC_{50} of 5.4 μM. With a goal to improve the inhibitory potency of these compounds, we have carried out similarity search to identify commercially available derivatives of most potent compounds. Based on predicted potency and structural and chemical diversity, about 50 derivatives were identified to be tested further. In order to exploit the therapeutic potential of Ubc9 against cancer and neurodegenerative diseases like Huntington's and Alzheimer's disease, we have identified the potential site to target for rational drug design using molecular docking and molecular dynamics simulation. The structural information derived was then used to identify a library of focused candidates targeting Ubc9 using our hybrid structure-based virtual screening protocol. Nineteen compounds were acquired from different chemical vendors and were tested first using in situ sumoylation assay and then by using E1 and E2 SUMO intermediate formation assay. About five compounds showed inhibitory activity against Ubc9. Similarity search was then carried out to identify commercially available derivatives of most potent compound which will be tested further. We have also optimized our virtual screening protocol so that it can be used in fragment based virtual screening campaigns. With this protocol we have participated in SAMPL3 virtual screening challenge and on the basis of our results and retrospective analyses from SAMPL3 fragment screening challenge we anticipate that chances of success in a fragment screening process could be increased significantly with careful selection of receptor structures, protein flexibility, sufficient conformational sampling within binding pocket and accurate assignment of ligand and protein partial charges.

We have also analyzed all the crystal structures of sumoylation enzymes to identify protein-protein interaction hotspots amenable for small molecule inhibition. The major interesting targets were selected and a virtual screening experiment resulted in the selection of several molecules which are ordered and currently being evaluated in biological assays

We have performed several docking, molecular dynamics and binding energy calculations using crystal structure of G9a (from PDB entry 3k5k) in complex with docked inhibitors to establish

correlation between published IC50 values and calculated binding free energies. The correlation has improved by selecting post filtered poses (using pharmacophore) for MD and by using co-crystallized water as co-factor but it has to be further improved to have more confidence in the protocol that will be used to predict binding energies of pharmacophore hits.

About 1 million compounds from ChemDiv databases were docked on G9a with and without co-crystallized water molecules. Hit compounds will be selected for further biochemical assays.

4. Conclusion

In this project, using our structure based virtual screening protocol we have identified small molecule inhibitors of sumoylation protein SUMO E1 and SUMO E2 (Ubc9) that could serve as initial hits which can be optimized into highly potent and selective lead molecules for the development of drugs against cancer and other diseases. We have also found that our virtual screening protocol can be used to identify low molecular weight compounds for further optimization in a fragment based drug discovery program. The chances of success in a fragment screening process could be increased significantly with careful selection of receptor structures, protein flexibility, sufficient conformational sampling within binding pocket and accurate assignment of ligand and protein partial charges.

We have also used docking and molecular dynamics protocol to screen a large collection of chemicals for actives against G9a. Many top hits were selected and waiting to be tested by biochemical assays.

5. Schedule and prospect for the future

We plan to optimize the initial hits for sumoylation enzyme E1 and E2 that have been identified by virtual screen and confirmed by biochemical assays. Various computation tools will be used to optimize these initial hits into more potent inhibitors. The computation time in RICC is critical for us to achieve this goals.

Shape similarity and electrostatic potential similarity search have been shown to be an effective virtual screening strategy for the identification of initial hits. Shape query and color queries were made using ROCS tool of OpenEye and we plan to screen databases using RICC to find a match to the shape and electrostatic potential of existing G9a inhibitor (DXQ) present with G9a as co-crystallized structure in pdb entry 3k5k.

We also plan to discover small molecule protein-protein interaction inhibitors (SMPPII) of the sumoylation pathway. SMPPIIs have been considered as high-hanging fruits in drug discovery. They offer a clear advantage over traditional enzymatic site inhibitors. We have established a pharmacophore based protocol for the discovery of SMPIIs. We will use the RICC time to discover SMPPIIs for sumoylation. These inhibitors can potentially be developed into drugs for treating various diseases such like cancer.

6. If you wish to extend your account, provide usage situation (how far you have achieved, what calculation you have completed and what is yet to be done) and what you will do specifically in the next usage term.

We have achieved the majority of our objectives in this fiscal year. Our virtual screen has identified many active inhibitors for SUMO E1 and E2 as confirmed by biochemical assays. The most potent compound identified using our virtual screen protocol is at single digit μM IC50. This is very encouraging. Our next objective is to optimize these initial hits into more potent and selective inhibitors for drug development. For this purpose, the RICC computational resources will be critical for our computational drug design and optimization approaches. The specific plans are briefly described

in the previous section.

7. If you have a "General User" account and could not complete your allocated computation time, specify the reason.

We have utilized ~82% of our allocated computation time as of early March. We anticipate that we will be very close to complete our allocated computation time by the end of this month.

8. If no research achievement was made, specify the reason.

N/A.

Fiscal Year 2011 List of Publications Resulting from the Use of RICC

[Publication]

1. Kumar, A., Zhang, K. Y. J. (2012) Computational Fragment-based Screening Using RosettaLigand: the SAMPL3 Challenge. *J. Comput-Aided Mol. Des.*, DOI:10.1007/s10822-011-9523-0.
2. Mishra, V., Kumar, A., Ali, V., Nozaki, T., Zhang, K. Y. J., Bhakuni, V. (2012) Role of conserved active site tryptophan-101 in functional activity and stability of phosphoserine aminotransferase from an enteric human parasite. *Amino Acids*, DOI:10.1007/s00726-011-1105-x
3. Mishra, V., Kumar, A., Ali, V., Nozaki, T., Zhang, K. Y. J., Bhakuni, V. (2012) Glu-108 is essential for subunit assembly and dimer stability of D-phosphoglycerate dehydrogenase from *Entamoeba histolytica, Molecular and Biochemical Parasitology*, **181**, 117-124.

[Proceedings, etc.]

If a publication does not contain the acknowledgement, please provide the reason for the missing of acknowledgement

[Oral presentation at an international symposium]

[Others]