Usage Report for Fiscal Year 2018

Project Title:

Computational Structure-based Design of Protein Inhibitors

Name:

○Kam Y. J. Zhang (1), Ashutosh Kumar (1), Dileep Vijayan (1), Matej Janezic (1) and Erma Fatiha Mohammad (1)

Laboratory at RIKEN:

(1) Laboratory for Structural Bioinformatics, Centre for Biosystems Dynamics Research

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

Shape complementarity between receptor and ligand has been widely recognized as the key determinant for the activity of small molecules. Compound filtering based on ligand 3D shape similarity has become a method of choice in many virtual screening studies. Apart from utilizing ligand 3D shape similarity as compound rank-ordering approach, ligand 3D shape similarity has been used to escort ligand sampling towards the important binding pocket regions. It has also been employed to select reliable poses from the many docking generated ones. Ligand 3D shape similarity was also combined with ensemble molecular docking to improve both pose prediction and virtual screening performance.

We have previously shown that shape similarity between query and crystal ligands could be used as a viable scoring approach to rank-order docking generated decoys. Ligand 3D shape similarity with crystallographic ligands was also utilized to select the suitable receptor for each ligand in a docking library. This cross-docking based virtual screening pipeline (CDVS) was suitable for docking big screening libraries while taking advantage of multiple protein receptors. Prospective evaluation of CDVS demonstrated improved performance over single or multiple-receptor docking. Moreover, in order to predict the binding pose of a query ligand, we have developed the Pose Prediction using Shape Similarity method (PoPSS). This method utilizes shape similarity with known crystal ligands to predict binding mode of query ligands. PoPSS first identifies a conformation of a query ligand which has the highest 3D shape similarity with the existing crystallographic ligands. After placing the selected conformation of a query ligand in the binding pocket, clashes with the protein are removed by performing side-chain repacking and Monte Carlo energy minimization. Although PoPSS demonstrated excellent performance in both retrospective and prospective scenarios, some drawbacks limit the wide applicability of PoPSS. Apart from its limitation of requiring at least one suitable crystal ligand for predicting the binding pose of a query ligand, PoPSS sometimes fail to predict the correct orientation of functional groups even though the position and orientation of the core scaffold is correctly predicted. In order to avoid clashes with the target protein binding pocket, PoPSS only performs small rigid body rotation and translation moves to the shape similarity selected ligand conformation during side-chain repacking and Monte Carlo energy minimization. This is sometimes not sufficient as binding of the query ligand require formation of critical interactions with the protein that can only be satisfied by good interaction geometry of the functional groups.

To address this limitation, we have slightly modified the original PoPSS approach. These modifications include the replacement of side-chain repacking and Monte Carlo energy minimization with a simple grid-based energy minimization that allows sampling of terminal functional groups while

keeping the core scaffold fixed. Other modifications include increasing the number of query ligand conformations and a different similarity metric for shape comparisons. As this is the simplification of the original PoPSS implementation, we have named it as PoPSS-Lite.

2. Specific usage status of the system and calculation method

PoPSS-Lite method places a ligand conformation of the highest 3D shape similarity with known crystal ligand into the target protein binding pocket and then refines the pose by performing energy minimization. All available ligand bound crystal structures of target protein and other homologous proteins from PDB were downloaded and were superimposed with the target protein. Relevant ligands were extracted and ligand 3D shape similarity calculations between ligand conformations and crystal structure ligands were performed. The top scoring ligand conformation for each ligand was placed into the target protein binding site and refined using the grid-based energy minimization protocol. Receptor structures for pose prediction were also selected based on the shape similarity calculation. We have tested PoPSS-Lite in Drug Design Data Resource Grand Challenge 3 (D3R GC3) which provides a platform for high-quality dataset exchange and prospective environment for testing methods in computer-aided drug design area (https://drugdesigndata.org/about/grand-challenge-3). D3R GC3 included the binding pose prediction of 24 Cathepsin S ligands for which previously unreleased crystal structures were available with D3R GC3 organizers. For an unbiased assessment, we have also compared it with PoPSS and CDVS that revealed excellent pose prediction performance of PoPSS-Lite.

3. Result

Evaluation of the RMSD between the predicted and crystal structures of Cathepsin S ligands revealed superior performance of PoPSS-Lite than both CDVS and PoPSS with mean and median RMSD of 2.39 and 2.05 Å respectively (Fig. 1A and Table 1).

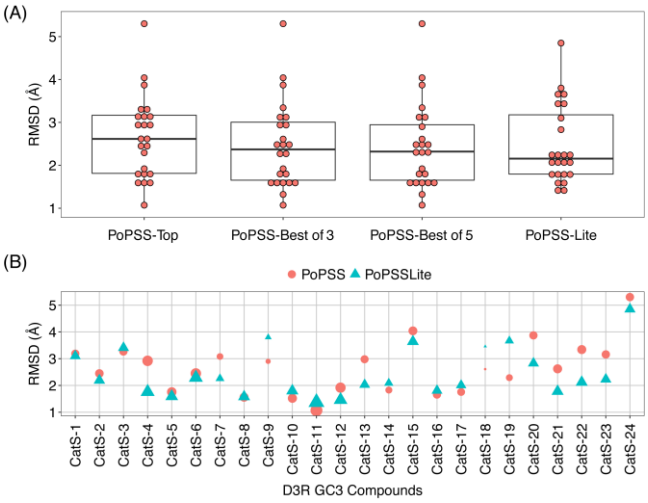

Fig. 1 (A) Boxplots showing the distribution of RMSDs between crystal structure ligands and predicted poses for 24 Cathepsin S ligands in D3R GC3 for PoPSS and PoPSS-Lite methods. (B) RMSDs shown as dot plots.

Table 1 Summary of pose prediction performance of CDVS, PoPSS and PoPSS-Lite in D3R GC3.

Method		Top pose RMSD (Å)	Best of 3 pose RMSD (Å)	Best of 5 pose RMSD (Å)	D3R GC3 rank
CDVS	Mean	3.14	3.13	3.13	12
	Median	2.31	2.35	2.35	8
PoPSS	Mean	2.60	2.46	2.43	2
	Median	2.61	2.37	2.32	12
PoPSS-Lite	Mean	2.39	-	-	1
	Median	2.05	-	-	4

Seventeen out of 24 Cathepsin S ligands were predicted with RMSD values lower than 3 Å (Fig. 1B) while about half of them were predicted within 2 Å. PoPSS-Lite is a simplification of PoPSS where side-chain repacking and subsequent Monte Carlo energy minimization were replaced with simple grid-based energy minimization. When compared with PoPSS, this modification resulted in the improvement of RMSD values in 14 out of 24 cases (Fig. 1B) with more than 0.5 Å or more than 25% improvement in half of these 14 Cathepsin S ligands. Small deterioration in RMSD values was detected in

10 out of 24 cases, however, the drop of more than 0.5 Å was only observed in 3 out of 24 Cathepsin S ligands (Fig. 1B). We further analyzed the reason behind the improvement in prediction performance of PoPSS-Lite. PoPSS predicted binding mode of CatS-21 ligand revealed more or less correct pose (Fig. 2A). However, the orientation of methyl amino benzyl ring is far from optimal to make stacking contacts with Phe-70. As PoPSS doesn't do any further ligand sampling so geometry of this ring could not be improved. PoPSS-Lite on the other hand performs grid-based energy minimization allowing the sampling of terminal functional groups. Thus, PoPSS-Lite achieved better pose for CatS-21 ligand by improving the stacking contacts with Phe-70 (Fig. 2B).

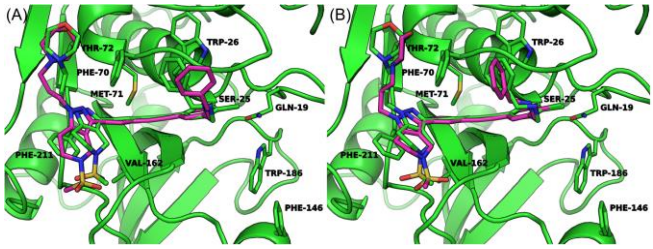

Fig. 2 Comparison of (A) PoPSS and (B) PoPSS-Lite predicted poses of CatS-21 (shown in magenta atom color) with crystal conformation (shown in green atom color)

Comparison of three pose prediction methods (CDVS, PoPSS and PoPSS-Lite) revealed superior performance of PoPSS-Lite for D3R GC3 pose prediction dataset (Table 1). Moreover, the comparison of PoPSS-Lite with 44 other D3R participants revealed top performance for PoPSS-Lite in mean RMSD based ranking (Fig. 3A). PoPSS-Lite was also among the top performers in median RMSD based ranking with 4th overall rank (Fig. 3B).

4. Conclusion

Here, we have presented the prospective performance of PoPSS-Lite method in D3R GC3. PoPSS-Lite is the simplification of a previously reported PoPSS that utilized shape similarity with crystal ligands to improve pose prediction performance. PoPSS pose prediction consists of two stages: shape similarity conformation selection and placement and refinement of complex structures. In PoPSS-Lite, shape similarity guided conformation selection was improved by increasing the number of ligand conformations and using a different similarity metric. Refinement of complex structures was improved by replacing the side-chain repacking and Monte Carlo energy minimization with a simple grid-based energy minimization. These modifications enabled PoPSS-Lite to perform better than the original implementation with lower mean and median RMSD values. Moreover, the comparison of PoPSS-Lite performance with other D3R GC3 pose prediction submission revealed top performance for PoPSS-Lite.

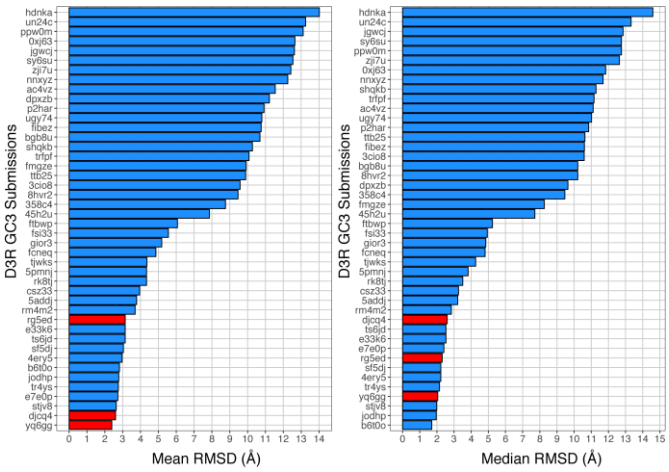

Fig. 3 Comparison of CDVS (rg5ed), PoPSS (djcq4) and PoPSS-Lite (yq6gg) pose prediction performances with 44 submissions for 24 Cathepsin S ligands in D3R GC3 using (A) mean RMSD of the top predicted pose, (B) median RMSD of the top predicted pose. Our submissions are colored in red.

5. Schedule and prospect for the future

In future, we plan to extend the applicability of PoPSS and PoPSS-Lite methods. The current implementation of our methods is limited to cases where several ligand-bound crystal structures of target protein are available. However, utilizing ligand 3D shape similarity to predict binding poses could be a general approach not limited to target

protein. As ligand belonging to the same congeneric series mostly bind to homologous proteins in a similar manner, so conceptually ligands binding to homologous protein structures could also be used for ligand 3D similarity based pose prediction. We plan to use 3D shape similarity of ligands and binding pockets to demonstrate their effectiveness in improving pose prediction and virtual screening performance. In addition to extend the utility of our pose prediction methods, we plan to apply them in various in-house drug discovery projects. We will utilize PoPSS and PoPSS-lite to perform virtual screening of large small molecule libraries to identify small molecule inhibitors of various proteins with potential therapeutic values.

Usage Report for Fiscal Year 2018

Fiscal Year 2018 List of Publications Resulting from the Use of the supercomputer

[Publication]

1. Kumar, A, and Zhang K. Y. J. (2019) Shape similarity guided pose prediction: lessons from D3R Grand Challenge 3. *J Comput-Aided Mol Design,* **33**, 47-59. doi: 10.1007/s10822-018-0142-x.

2. Kumar, A, and Zhang K. Y. J. (2018) Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery. *Frontier in Chem.,* 6:315.

3. Kumar, A, and Zhang K. Y. J. (2018) A cross docking pipeline for improving pose prediction and virtual screening performance. *J Comput-Aided Mol Design*, 32(1), 163-173.

[Oral presentation]

1. Zhang K. Y. J. and Kumar, A, (2018) Exploiting Ligand 3D Shape Similarity for Computational Structure Based Drug Design. Presented in International Conference on Drug Discovery and Translational Medicine (ICDDTM'18), Dec. 4-5, 2018, Putrajaya, Malaysia.

2. Kumar, A, and Zhang K. Y. J. (2018) A Pose Prediction Approach Based on Ligand 3D Shape Similarity: Lesson Learned in D3R Challenge. Presented in D3R Workshop 2018 held on Feb 22-23, 2018 at University of California, San Diego, USA.

3. Kumar, A, and Zhang K. Y. J. (2018) A Pose Prediction Approach Based on Ligand 3D Shape Similarity. Presented in Workshop on Mathematics of Drug Discovery/Design held on June 4-8, 2018 at Field Institute, University of Toronto, Toronto, Canada.