

RESEARCH ARTICLE

Open Access



Docking-based virtual screening of T β R1 inhibitors: evaluation of pose prediction and scoring functions

Shuai Wang[†], Jun-Hao Jiang[†], Ruo-Yu Li and Ping Deng^{*†}

Abstract

To improve the reliability of virtual screening for transforming growth factor-beta type 1 receptor (T β R1) inhibitors, 2 docking methods and 11 scoring functions in Discovery Studio software were evaluated and validated in this study. LibDock and CDOCKER protocols were performed on a test set of 24 T β R1 protein–ligand complexes. Based on the root-mean-square deviation (RMSD) values (in Å) between the docking poses and co-crystal conformations, the CDOCKER protocol can be efficiently applied to obtain more accurate dockings in medium-size virtual screening experiments of T β R1, with a successful docking rate of 95%. A dataset including 281 known active and 8677 inactive ligands was used to determine the best scoring function. The receiver operating characteristic (ROC) curves were used to compare the performance of scoring functions in attributing best scores to active than inactive ligands. The results show that Ludi 1, PMF, Ludi 2, Ludi 3, PMF04, PLP1, PLP2, LigScore2, Jain and LigScore1 are better scoring functions than the random distribution model, with AUC of 0.864, 0.856, 0.842, 0.812, 0.776, 0.774, 0.769, 0.762, 0.697 and 0.660, respectively. Based on the pairwise comparison of ROC curves, Ludi 1 and PMF were chosen as the best scoring functions for virtual screening of T β R1 inhibitors. Further enrichment factors (EF) analysis also supports PMF and Ludi 1 as the top two scoring functions.

Keywords: Docking methods, Scoring functions, Molecular docking-based virtual screening, Transforming growth factor β , Inhibitor

Background

Due to the vast improvement of computing power and the rapid development of computational chemistry and biology, computer-aided drug design (CADD) technology has been successfully applied in drug discovery, accelerating the research process and reducing the associated costs and risks [1]. Virtual screening is a common CADD technology used for the identification of hit molecules from large-scale compound libraries, providing a highly efficient approach to discover new compounds in

the early stage of drug development [2]. Virtual screening methods, including molecular docking [3], pharmacophore model [4], and similarity searching [5], are widely used in drug research and development. With an increasingly large number of target protein structures being resolved, the molecular docking-based virtual screening (DBVS), has become one of the most common virtual screening methods, having achieved a great success [6].

However, due to its low accuracy, DBVS suffers from constant criticism. This method involves two basic processes: positioning ligands into a protein active site, scoring and ranking docked ligands. Both steps are imperfect. In the last decade, several studies on docking and scoring methods demonstrated that most of the docking methods and scoring functions are sensitive to the receptor,

*Correspondence: 100865@cqmu.edu.cn

[†]Shuai Wang and Jun-Hao Jiang contributed equally to this work
College of Pharmacy, Chongqing Medical University, Chongqing 400016, China



and conclusions on different receptors may be inconsistent with each other [7–10]. Different docking programs may show different performances in the same target protein [8]. There is often a poor correlation between the scores given by scoring functions and experimental binding affinities. New forms of examination are necessary to maximize the success rate of docking and scoring methods, to determine the best parameters for individual protein targets.

Transforming growth factor- β (TGF β) signaling plays a key role in cell growth, migration, differentiation, epithelial–mesenchymal transition (EMT), and extracellular matrix remodeling [11]. TGF β is reported to be critical in the later stages of tumorigenesis by increasing immunosuppressive Treg cells and facilitating EMT [12]. TGF β type 1 receptor (T β R1) is a key member of the TGF β signaling pathways, and its inhibition has the potential to enhance the immune response against tumors. Thus, T β R1 has attracted wide attention, as a potential drug target [13].

Different virtual screening models can be complementary in the search for new inhibitors. In our previous studies, several inhibitors were screened and designed by the pharmacophore model and structure–activity relationship study [14, 15]. In this study, we aimed to obtain new T β R1 inhibitors using DBVS. For that, we assessed different docking methods and scoring functions for T β R1, before large-scale database screening. We conclude that the CDOCKER protocol with a successful docking rate of 88% and the PMF and Ludi 1 scoring functions with high reliability may be efficiently applied in the virtual screening of new inhibitors of T β R1.

Materials and methods

Database preparation

The Protein Data Bank (PDB; <http://www.rcsb.org>) funded by the National Science Foundation, the National Institutes of Health, and the US Department of Energy is the single worldwide archive of 3D structural data of biological macromolecules, such as proteins, nucleic acids, and complex assemblies [16]. Up to now, 22 crystal structures of T β R1 with various small molecule inhibitors have been reported in PDB, including 1RW8, 1PY5 [17], 1VJY [18], 2WOU, 2WOT [19], 2X7O [20], 3GXL, 3HMM [21], 3FAA [22], 3KCF [23], 3TZM [24], 4X2E, 4X2G, 4X2J [25], 5E8W, 5E8Z [26], 5FRI [27], 5QIK, 5QIL, 5QIM [28], 5USQ [29], and 6B8Y (Additional file 1: Table S1) [12]. Water molecules and co-crystallized ligands were deleted and then the receptor protein was subsequently prepared by the Prepare Protein protocol using the Discovery Studio 2017R2 software package (DS; <https://www.3ds.com/>). The following tasks were performed: insertion of missing atoms in incomplete

residues, modeling of missing loop regions, deleting of alternate conformations (disorder), standardizing of atom names, and protonating of titratable residues.

22 X-ray structures of T β R1 deposited in the Protein Data Bank (PDB) were analyzed using some parameters to select the most suitable T β R1 structure for next studies, such as information on the resolution and organisms of crystallographic structures, as well as the similarity values and physicochemical properties of all ligands (Additional file 1: Table S2). The overlay of the crystal structure (Additional file 1: Fig. S1) and comparing the amino acid sequence showed that the 22 crystal structure has a high consistency and a complete binding pocket. Due to the Induced Fit Theory, the T β R1 will change its conformation on binding a ligand. Therefore, there are slight differences in the conformation of 22 co-crystal structures (Fig. 1).

The ligand in 6B8Y (resolution 1.65 Å) with the IC₅₀ value of 0.55 nM was the most active inhibitor of 22 molecules in co-crystal structures. Therefore, 6B8Y may be more representative of the active conformation of the target protein. 6B8Y was selected for the evaluation of scoring functions from up to 22 co-crystal structures.

The root-mean-square deviation (RMSD) value (in Å) between the best docking poses of ligands and the conformations in co-crystal structures was selected as the evaluation criterion to check the accuracy of docking program in the reproduction of the binding poses of ligands in crystal structures of T β R1–ligand complexes. Inhibitors in the above-mentioned crystal structures are shown in Fig. 1. The conformations of the inhibitors confirmed by experimental methods. Thus, 22 active conformations of inhibitors in the crystal structure were



Fig. 1 Overlay of crystal structures between 4X2F (gold ribbon) and 6B8Y (blue ribbon). The ligand in 6B8Y was shown with the purple stick model

selected as the reference ligands. These 22 active conformations were re-docked back to the active site of T β R1 by molecular docking. In previous studies, RMSD values within 2.0 Å were acceptable for the molecular docking-based virtual screening [30]. Thus, in this investigation, the docking results with RMSD \leq 2.0 Å were considered successful, and the docking results with RMSD > 2.0 Å were considered to be failed.

Receiver operating characteristic (ROC) curves were used to distinguish between “active” and “inactive” samples based on the scores. For the evaluation and validation of the sensitivity (SE) and specificity (SP) of scoring functions, a test set covering active and inactive ligands was established by downloading 8958 ligands from the DUD-E database (<http://dude.docking.org/>). Namely, this test set includes 281 known active ligands and 8677 ligands assumed to be inactive (referred to as decoys). The decoy compounds have similar physicochemical properties but dissimilar 2D topology (Additional file 1: Table S3) [31]. The prepared active or decoy ligands were docked into the binding pocket of T β R1, respectively and the top-scoring pose of each ligand along with its docking score was used for further analysis.

Molecular docking and scoring functions

Two molecular docking methods, LibDock and CDOCKER in DS software, were performed in this study. LibDock is a docking program developed by Diller and Merz, which uses protein site features referred to as Hot-Spots [32], while CDOCKER is a grid-based molecular docking method based on CHARMM [33]. Before molecular docking, the ligands in the test set were prepared using the Prepare Ligands protocol to remove duplicates, generate 3D conformations, and remove compounds with undesirable properties. For LibDock, the parameter “Conformation Method” was set to “BEST”, “Max Hits to Save” was set to “1” and “Minimization Algorithm” was set to “Smart Minimizer”. All other parameters were set to default. For CDOCKER, “Top Hits” was set to “1”, “Pose Cluster Radius” was set to ‘0.5’, and others were set to default. Only one top docking pose for each molecule was reported and saved for further analysis. The RMSD values (in Å) between the best docking poses of ligands and the conformations in co-crystal structures were calculated.

The top docking pose for each molecule was re-scored using the Score Ligand Poses protocol in the DS software. The scoring functions, including CDOCKER Scores (CDOCK), LigScore1, LigScore2, PLP1, PLP2, Jain, PME, PMF04, Ludi Energy Estimate 1 (Ludi 1), Ludi Energy Estimate 2 (Ludi 2), Ludi Energy Estimate 3 (Ludi 3), were used to calculate ligand binding in an active receptor site. The calculated scores were reported as positive values

to ensure that higher scores indicated more favorable binding.

ROC curve analysis

All ligands in the test set (281 active and 8677 inactive) were docked into the binding pocket of T β R1 (PDB ID: 6B8Y). The top docking pose for each ligand was re-scored used the above-mentioned 11 scoring functions. Using a diagnostic trial approach, the SE and SP of each scoring function for distinguishing between “active” and “inactive” samples, were calculated. In the ROC curve, the SE is plotted against 100 – SP for every cut-off point of the variable (score calculated by scoring function). The SE is defined as the ratio of true positive (TP) to the sum of TP and false negative (FN): [TP/(TP + FN)]; and the SP is defined as the ratio of true negative (TN) to the sum of TN and false positive (FP): [TN/(TN + FP)] [14]. The reliability of the scoring function was estimated using the area under the curve (AUC), together with the Chi-square test. A statistically significant difference was defined when $P < 0.05$. Typically, a ROC curve has an AUC baseline of 0.5, which demonstrates an evenly distributed system. An AUC value closer to 1 indicates good selectivity, whereas a value close to 0.5 indicates random selection.

Enrichment factor analysis

Taking enrichment factor (EF) as the indicator, the performance of the scoring function as a virtual screening tool applied to T β R1 was evaluated. EF is defined as:

$$EF = \frac{Hits_{sampled}^{X\%}}{N_{sampled}^{X\%}} \cdot \frac{N_{total}}{Hits_{total}}$$

where $Hits_{sampled}^{X\%}$ is the number of hit compounds at X% of the database, $N_{sampled}^{X\%}$ is the number of compounds screened at X% of the database, N_{total} is the number of compounds in the database, and $Hits_{total}$ is the number of active compounds in the database. Among the top-rank screened database compounds, the enrichment ability of active compounds is the most noteworthy. Therefore, we mainly focus on the enrichment factor at 0.5%, 1%, and 2% of the ranked database, which are defined as $EF^{0.5\%}$, $EF^{1\%}$, and $EF^{2\%}$.

Results and discussion

Assessment of docking methods

Properly docking the ligands to the active site is the most critical step in the virtual screening. The LibDock and CDOCKER docking programs were performed in this study. The results require evaluation to determine the best docking method for the T β R1 protein target and to maximize the probability of success. The performance of

the docking programs for predicting the binding mode between ligand and T β R1 was evaluated with a success rate. 22 known active conformations of T β R1 inhibitors were re-docked into the corresponding binding pockets. As virtual screening projects always involve thousands to tens of thousands of ligands, only the top-scoring pose of each ligand was considered as the possible active conformation. The RMSD values between the docked poses with the highest score and those in co-crystal structures are listed in Table 1.

Results demonstrate that the accuracy of the Lidock program (59%) is much lower than that of the CDOCKER program (95%). Thus, the CDOCKER program was chosen as the docking programs in the subsequent study, based on its reliability.

Assessment of scoring functions

Scoring and ranking the docked ligands play an important role in virtual screening. The best scoring function for T β R1 should be selected to maximize the chances of

Table 1 RMSD values (in Å) between the best docking poses of ligands and the conformations in co-crystal structures for all retrieved actives ligands

NO.	Ligand	MW	RMSD values (Å)	
			LibDock	CDOCKER
1	1RW8	293.338	0.2397	1.5416
2	1PY5	272.304	1.1255	1.1468
3	1VJY	287.319	0.2443	0.3263
4	2WOU	369.418	2.0171	0.3847
5	2WOT	458.509	0.5475	0.2087
6	2X7O	494.627	0.9240	0.6066
7	3GXL	352.392	0.3867	0.4265
8	3FAA	362.36	5.6841	1.5150
9	3HMM	313.356	0.2322	0.1976
10	3KCF	360.409	0.6144	0.2904
11	3TZM	390.435	4.9755	4.8933
12	4X2F	253.259	5.7255	0.4702
13	4X2G	253.259	5.2960	1.1824
14	4X2J	253.259	6.2834	0.3497
15	5E8W	466.531	0.7055	1.2770
16	5E8Z	420.464	8.5303	1.1120
17	5FRI	332.742	4.9761	0.3629
18	5QIK	365.361	0.9489	0.2173
19	5QIL	359.381	0.9670	0.2635
20	5QIM	359.381	6.9569	0.7050
21	5USQ	438.928	1.8278	1.9753
22	6B8Y	374.295	0.2779	0.2861
Success rate			59%	95%

Ligands were named with the PDB ID; MW, molecular weight; docking results with RMSD \leq 2.0 Å are considered successful, and docking results with RMSD $>$ 2.0 Å are considered to be failed

success. In this study, 281 known active and 8677 inactive ligands were used to determine the best scoring functions before a larger-scale screening of unknown compounds. In this study, the ROC curves are used to compare the performance of scoring functions in attributing better scores to active than inactive ligands. The docking results showed that, among 8959 ligands, only one active molecule and 22 inactive molecules could not be docked into the active cavity. The best docking poses for each docked ligand were recalculated with 11 scoring functions, respectively. Based on the information of known active and inactive ligands, the ROCs (Fig. 2) were drawn using statistical software, and relevant statistical parameters are listed in Table 2.

As a higher score indicates a more favorable binding, the result of the CDOCK is unreliable. The ROC curves of Ludi 1, PMF, Ludi 2, Ludi 3, PMF04, PLP1, PLP2, LigScore2, Jain and LigScore1 all tend to the upper left corner, with AUCs of 0.864, 0.856, 0.842, 0.812, 0.776, 0.774, 0.769, 0.762, 0.697 and 0.660, respectively ($P < 0.0001$). All the AUCs of scoring function are significantly larger than those of random distribution (Reference line in Fig. 2, AUC = 0.5), which indicates that the prediction capacity of these 10 scoring functions is better than the random distribution model. The AUC of Ludi 1 is the largest, and the curve is closest to the upper left corner, indicating that Ludi 1 can efficiently distinguish active and inactive molecules. Pairwise comparison of ROC curves shows a significant difference in AUCs between Ludi 1 and other scoring functions ($P < 0.0001$), except for PMF ($P = 0.6170$). Given the high accuracy of Ludi 1 and PMF for T β R1, we strongly recommend Ludi 1 and PMF as scoring functions for virtual screening of new inhibitors of T β R1.

Large-scale bioactivity testing is an extremely expensive process. Therefore, it is essential to minimize the number of virtual screening false positives, before a large database search of active molecules and their experimental validation. Therefore, achieving a high positive predictive value ($PPV = TP / (TP + FP)$) is mandatory. In general, large databases exhibit a low frequency of active ligands. This suggests that increasing SP is more helpful to improve the PPV than altering SE. For example, in this study, active ligands accounted for 3% (280/8958) of the test set. Based on the results from Ludi1, with the increase in SP from 80 to 90%, 95% and 99%, the PPV value increases from 12% (222/1904) to 17% (171/1028), 22% (118/547) and 28% (34/120), respectively. If PMF instead of Ludi 1 as a scoring function, the PPV value increases from 12% (219/1862) to 18% (166/948), 24% (136/569) and 43% (66/151), respectively. As illustrated in Fig. 2, the PMF ROC curve is above the Ludi1 ROC curve in the higher

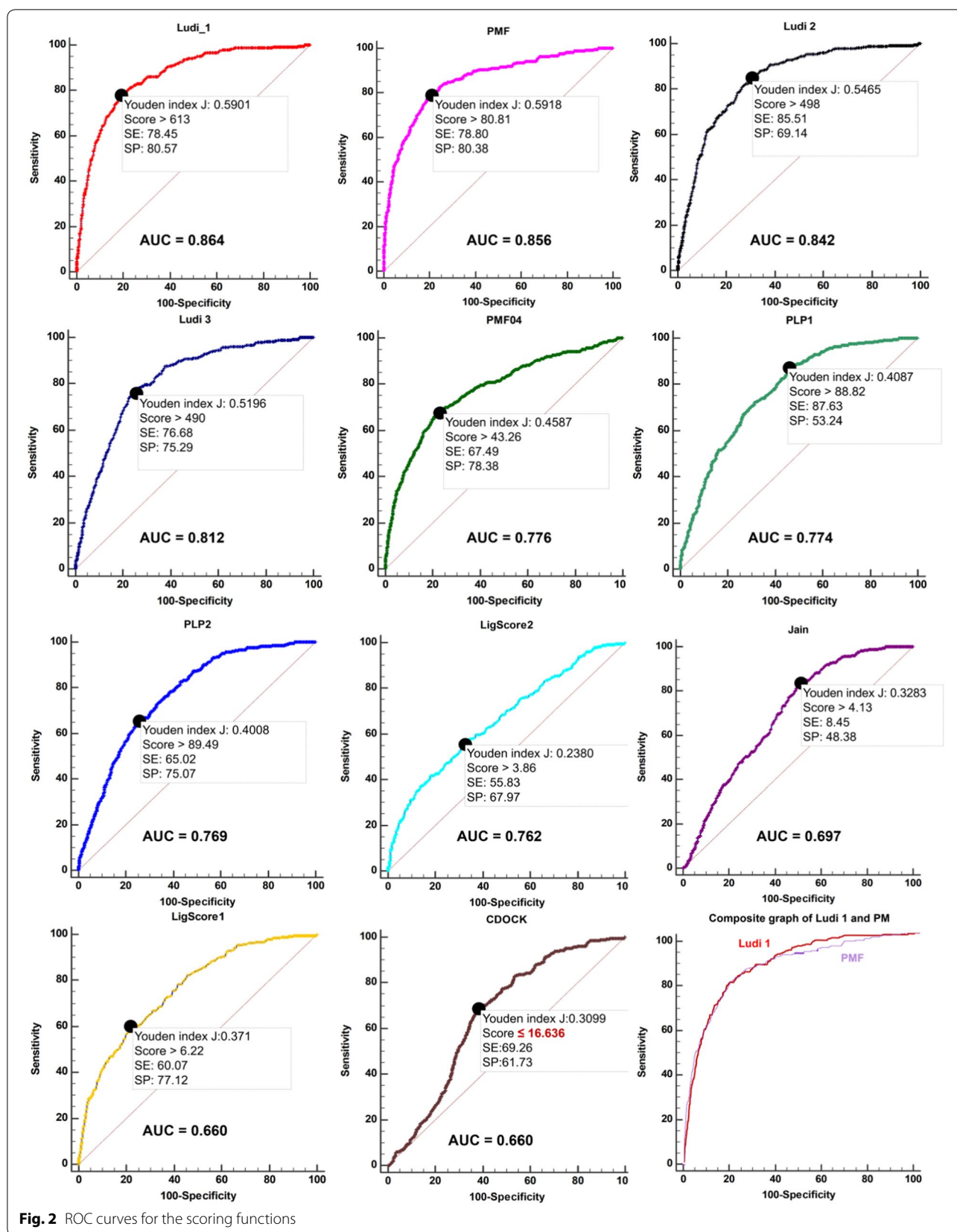


Table 2 ROC curve results for the scoring functions

Scoring function	AUC			Youden's index			
	AUC	95% CI	P	Youden index J	Score	SE	SP
Ludi 1	0.864	0.856 to 0.871	<0.0001	0.5901	>613	78.45	80.57
PMF	0.856	0.849 to 0.864	<0.0001	0.5918	>80.81	78.80	80.38
Ludi 2	0.842	0.834 to 0.849	<0.0001	0.5465	>498	85.51	69.14
Ludi 3	0.812	0.803 to 0.820	<0.0001	0.5196	>490	76.68	75.29
PMF04	0.776	0.767 to 0.784	<0.0001	0.4587	>43.26	67.49	78.38
PLP1	0.774	0.765 to 0.782	<0.0001	0.4087	>88.82	87.63	53.24
PLP2	0.769	0.760 to 0.777	<0.0001	0.4008	>89.49	65.02	75.07
LigScore2	0.762	0.753 to 0.771	<0.0001	0.3719	>6.22	60.07	77.12
Jain	0.697	0.687 to 0.706	<0.0001	0.3283	>4.13	84.45	48.38
LigScore1	0.660	0.650 to 0.669	<0.0001	0.2380	>3.86	55.83	67.97

AUC area under the curve, 95% CI asymptotic 95% confidence interval, P significance level P (area = 0.5), SE sensitivity, SP specificity

Table 3 EFs of the scoring functions and pharmacophore model A02

Scoring function	EF ^{0.5%}	EF ^{1%}	EF ^{2%}
Ludi 1	11.93	9.58	8.42
PMF	19.38	15.61	12.60
Ludi 2	9.12	7.37	5.80
Ludi 3	7.18	5.68	5.12
PMF04	10.53	7.45	7.81
PLP1	8.61	7.10	4.79
PLP2	7.18	6.03	4.44
LigScore2	5.61	5.61	6.42
Jain	0.72	0.71	1.21
LigScore1	4.31	4.16	5.18
Pharmacophore model A02	12.92	8.16	4.44

SP region. Furthermore, for the same SP, the PMF ROC curve has higher SE and PPV. Therefore, the PMF scoring function is suitable for this particular research objective.

To obtaining more novel molecular structures for drug design, higher SE must be considered in the virtual screening. When SE=0.9 (255 active compounds were obtained), a total of 3772 compounds were obtained by PMF, while only 3597 compounds were screened by Ludi1, with fewer false positive ligands. As shown in Fig. 2, the Ludi 1 ROC curve is above the PMF ROC curve in the higher SE region in the upper right corner. For the same SE, the Ludi 1 ROC curve has a higher SP. Therefore, our results indicate that Ludi 1 scoring function has more advantages to this particular objective.

It can be seen from Table 3 that for T β R1, PMF all performs best at EF^{0.5%}, EF^{1%}, and EF^{2%} among all scoring functions, and Ludi 1 ranks the second. Besides, Jain is the worst among the above 10 scoring functions. In line

with ROC curve analysis results, EF evaluation results also support PMF and Ludi 1 as the top two scoring functions.

Different virtual screening models can be complementary in the search for new inhibitors. In our previous studies, based on the crystal structure of T β R1-BMS22 complex, a pharmacophore model was constructed [14]. In this study, EF^{0.5%}, EF^{1%}, and EF^{2%} of the pharmacophore model A02 was also evaluated. It can be seen from Table 3 that EFs comparison indicates that the pharmacophore model A02 has very excellent performance at EF^{0.5%}, only lower than PMF. Meanwhile, it ranks third at EF^{1%}, slightly lower than PMF and Ludi 1. However, the low ranking of the pharmacophore model A02 at EF^{2%} means that the screening results may not be ideal. EF evaluation results show that the pharmacophore model can also obtain excellent virtual screening results in the top-scoring area. Therefore, we recommend that molecular docking and pharmacophore cross-validation can be performed to obtain more accurate virtual screening results.

In conclusion, our results show that PMF and Ludi1 scoring functions have high reliability in the virtual screening of new inhibitors of T β R1. According to the goal of virtual screening, the scoring function can be flexibly selected. To reduce the cost of experimental activity verification, the PMF scoring function should be chosen first. The Ludi 1 scoring function can be chosen first to obtain more novel inhibitors of T β R1.

Conclusion

Drug target interaction plays an important role in drug discovery based on target [34, 35], and molecular docking can well simulate the interaction between drug and target, which is widely used [36–38]. Molecular docking is

routinely used in virtual screening and generally involves two separate steps. The first step is to position ligands into a protein active site correctly, and the second step is scoring and ranking these docked ligands reasonably. An increasingly large number of studies reported that the docking method and the scoring functions are sensitive to receptors.

In this study, we present a thorough investigation of the performance of the LibDock and CDOCKER protocol in DS software on a test set of 22 T β R1 protein–ligand complexes. The results demonstrate that the CDOCKER protocol based on CHARMM is the optimal docking protocol for T β R1. With a successful docking rate of 95%, which means high reliability, the CDOCKER protocol can be efficiently applied for accurate docking in medium-size virtual screening experiments of T β R1.

A dataset including 281 known active and 8677 inactive ligands was used to determine the best scoring function in DS software for scoring ligands posed by molecule docking. The ROC curves are used to compare the performance of scoring functions in attributing better scores to active than inactive ligands. The ROC curves of Ludi 1, PMF, Ludi 2, Ludi 3, PMF04, PLP1, PLP2, LigScore2, Jain and LigScore1 are better than that of the random distribution model, with AUC of 0.864, 0.856, 0.842, 0.812, 0.776, 0.774, 0.769, 0.762, 0.697 and 0.660, respectively. Based on the pairwise comparison of ROC curves, we strongly recommend that Ludi 1 and PMF should be chosen as scoring functions for virtual screening of new T β R1 inhibitors. The PMF ROC curve is above the Ludi 1 ROC curve in the higher SP region in the lower left quarter, thus for the same SP, the PMF ROC curve has higher SE and PPV values. It is advantageous to reduce the number of false positive molecules and the cost of experimental activity verification. Additionally, the Ludi 1 ROC curve is above the PMF ROC curve in the higher SE region in the upper-right corner; therefore, for the same SE, the Ludi 1 ROC curve has higher SP, which is optimal for obtaining more novel molecules.

Consistent with ROC curve, EF evaluation also supports PMF and Ludi-1 as the top two scoring functions. EF evaluation results also demonstrates that the pharmacophore model can get good virtual screening results in the highest score areas. Therefore, molecular docking and pharmacophore cross validation can be used to obtain more accurate virtual screening results.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13065-020-00704-3>.

Additional file 1: Table S1. Crystal structures of T β R1 with inhibitors reported in PDB. **Table S2.** Physicochemical properties of each

crystallographic ligand. MW, molecular weight; HBD, Number of hydrogen bond donors; HBA, Number of hydrogen bond receptors. **Table S3.** Physicochemical properties of compounds in the test set. NC, Number of compounds; MW, molecular weight; nROT, Number of rotatable bonds; nHBA, Number of hydrogen acceptor donors; nHBD, Number of hydrogen bond receptors; nRings, Number of rings. **Fig. S1.** Overlay of 22 T β R1 proteins in the PDB. Except for the Gly367-Gly374 region (far away from the binding site) of 1RW8 and 1PY5, the rest of the crystal structures were well overlapped.

Abbreviations

TGF- β : Transforming growth factor- β ; T β R1: Transforming growth factor- β type 1 receptor; RMSD: Root-mean-square deviation; CADD: Computer-aided drug design; DBVS: Molecular docking-based virtual screening; ROC: Receiver operating characteristic; SE: Sensitivity; SP: Specificity; TP: True positive; FN: False negative; TN: True negative; FP: False positive; AUC: Area under the curve; 95% CI: Asymptotic 95% confidence interval; P: Significance level; EF: Enrichment factor.

Acknowledgements

Not applicable.

Authors' contributions

Authors WS and JJH have performed evaluation and validation of docking methods and scoring functions and drafted the manuscript. LRY has completed the literature search. DP outlined the research strategy and revised the manuscript. All authors read and approved the final manuscript.

Funding

Research reported in this publication was supported by the Fundamental and Advanced Research Projects of Chongqing City (cstc2017cyjAX0352) and the Research Projects of Chongqing Medical University (NSFYY201824, ZHYX2019022). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 4 March 2020 Accepted: 6 August 2020

Published online: 14 August 2020

References

- Zhong F, Xing J, Li X, Liu X, Fu Z, Xiong Z, Lu D, Wu X, Zhao J, Tan X, Li F, Luo X, Li Z, Chen K, Zheng M, Jiang H (2018) Artificial intelligence in drug design. *Sci China Life Sci* 61:1191–1204
- Liang JW, Wang S, Wang MY, Li SL, Li WQ, Meng FH (2019) Identification of novel PI3K δ selective inhibitors by SVM-based multistage virtual screening and molecular dynamics simulations. *Int J Mol Sci* 20:6000
- Torres PHM, Sodero ACR, Jofily P, Silva-Jr FP (2019) Key topics in molecular docking for drug design. *Int J Mol Sci* 20:4574
- Zeb A, Son M, Yoon S, Kim JH, Park SJ, Lee KW (2019) Computational simulations identified two candidate inhibitors of Cdk5/p25 to sbrogate tau-associated neurological disorders. *Comput Struct Biotechnol J* 17:579–590
- Stumpfe D, Bajorath J (2011) Similarity searching. *WIREs Comput Mol Sci* 1:260–282

6. Huang T, Sun J, Zhou S, Gao J, Liu Y (2017) Identification of direct activator of adenosine monophosphate-activated protein kinase (AMPK) by structure-based virtual screening and molecular docking approach. *Int J Mol Sci* 18:1408
7. Li GB, Yang LL, Wang WJ, Li LL, Yang SY (2013) ID-Score: a new empirical scoring function based on a comprehensive set of descriptors related to protein-ligand interactions. *J Chem Inf Model* 53:592–600
8. Huang S, Song C, Wang X, Zhang G, Wang Y, Jiang X, Sun Q, Huang L, Xiang R, Hu Y (2017) Discovery of new SIRT2 inhibitors by utilizing a consensus docking/scoring strategy and structure-activity relationship analysis. *J Chem Inf Model* 57:669–679
9. Wang R, Lu Y, Wang S (2003) Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* 46:2287–2303
10. Li X, Li Y, Cheng T, Liu Z, Wang R (2010) Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. *J Comput Chem* 31:2109–2125
11. Howley BV, Howe PH (2019) TGF- β signaling in cancer: post-transcriptional regulation of EMT via hnRNP E1. *Cytokine* 118:19–26
12. Hari Krishnan LS, Warriar J, Tebben AJ, Tonukunuru G, Madduri SR, Baligar V, Mannoori R, Seshadri B, Rahaman H, Arunachalam PN, Dikundwar AG, Fink BE, Fargnoli J, Fereshteh M, Fan Y, Lippy J, Ho CP, Wautlet B, Sheriff S, Ruzanov M, Borzilleri RM (2018) Heterobicyclic inhibitors of transforming growth factor beta receptor I (TGF β RI). *Bioorg Med Chem* 26:1026–1034
13. Haque S, Morris JC (2017) Transforming growth factor- β : a therapeutic target for cancer. *Hum Vaccine Immunother* 13:1741–1750
14. Jiang JH, Zhou H, Jiang QH, Sun LL, Deng P (2018) Novel transforming growth factor-beta receptor 1 antagonists through a pharmacophore-based virtual screening approach. *Molecules* 23:2824
15. Jiang JH, Deng P (2019) Discovery of new inhibitors of transforming growth factor-beta type 1 receptor by utilizing docking and structure-activity relationship analysis. *Int J Mol Sci* 20:4090
16. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
17. Sawyer JS, Beight DW, Britt KS, Anderson BD, Campbell RM, Goodson T Jr, Herron DK, Li HY, McMillen WT, Mort N, Parsons S, Smith EC, Wagner JR, Yan L, Zhang F, Yingling JM (2004) Synthesis and activity of new aryl- and heteroaryl-substituted 5,6-dihydro-4-pyrrolo[1,2-*b*]pyrazole inhibitors of the transforming growth factor- β type I receptor kinase domain. *Bioorg Med Chem Lett* 14:3581–3584
18. Gellibert F, Woolven J, Fouchet MH, Mathews N, Goodland H, Lovegrove V, Laroze A, Nguyen VL, Sautet S, Wang R, Janson C, Smith W, Krysa G, Boullay V, De Gouville AC, Huet S, Hartley D (2004) Identification of 1,5-naphthyridine derivatives as a novel series of potent and selective TGF- β type I receptor inhibitors. *J Med Chem* 47:4494–4506
19. Goldberg FW, Ward RA, Powell SJ, Debreczeni JE, Norman RA, Roberts NJ, Dishington AP, Gingell HJ, Wickson KF, Roberts AL (2009) Rapid generation of a high quality lead for transforming growth factor-beta (TGF- β) type I receptor (ALK5). *J Med Chem* 52:7901–7905
20. Roth GJ, Heckel A, Brandl T, Grauert M, Hoerer S, Kley JT, Schnapp G, Baum P, Mennerich D, Schnapp A, Park JE (2010) Design, synthesis, and evaluation of indolinones as inhibitors of the transforming growth factor β receptor I (TGF β RI). *J Med Chem* 53:7287–7295
21. Gellibert F, Fouchet MH, Nguyen VL, Wang R, Krysa G, Gouville ACD, Huet S, Dodic N (2009) Design of novel quinazoline derivatives and related analogues as potent and selective ALK5 inhibitors. *Bioorg Med Chem Lett* 19:2277–2281
22. Bonafoux D, Chuaqui C, Boriack-Sjodin PA, Fitch C, Hankins G, Josiah S, Black C, Hetu G, Ling L, Lee WC (2009) 2-Aminoimidazoles inhibitors of TGF- β receptor 1. *Bioorg Med Chem Lett* 19:912–916
23. Guckian K, Carter MB, Lin EY, Choi M, Sun L, Boriack-Sjodin PA, Chuaqui C, Lane B, Cheung K, Ling L, Lee WC (2010) Pyrazolone based TGF beta R1 kinase inhibitors. *Bioorg Med Chem Lett* 20:326–329
24. Ogunjimi AA, Zeqiraj E, Ceccarelli DF, Siccheri F, Wrana JL, David L (2012) Structural basis for specificity of TGF β family receptor small molecule inhibitors. *Cell Signal* 24:476–483
25. Czodrowski P, Hölzemann G, Barnickel G, Greiner H, Musil D (2015) Selection of fragments for kinase inhibitor design: decoration is key. *J Med Chem* 58:457–465
26. Tebben AJ, Ruzanov M, Gao M, Xie D, Kiefer SE, Yan C, Newitt JA, Zhang L, Kim K, Lu H, Kopcho LM, Sheriff S (2016) Crystal structures of apo and inhibitor-bound TGF β R2 kinase domain: insights into TGF β RI isoform selectivity. *Acta Crystallogr D Struct Biol* 72:658–674
27. Goldberg FW, Daunt P, Pearson SE, Greenwood R, Grist M, Debreczeni JÉ (2016) Identification and optimisation of a series of *N*-(4-anilino-2-pyridyl)acetamide activin receptor-like kinase 1 (ALK1) inhibitors. *Med Chem Comm* 7:1204–1208
28. Zhang Y, Zhao Y, Tebben AJ, Sheriff S, Ruzanov M, Fereshteh MP, Fan Y, Lippy J, Swanson J, Ho CP, Wautlet BS, Rose A, Parrish K, Yang Z, Donnell AF, Zhang L, Fink BE, Vite GD, Augustine-Rauch K, Fargnoli J, Borzilleri RM (2018) Discovery of 4-azaindole inhibitors of TGF beta RI as immunoncology agents. *ACS Med Chem Lett* 9:1117–1122
29. Sabat M, Wang H, Scora N, Lawson JD, Atienza J, Kamran R, Hixon MS, Dougan DR (2017) Design, synthesis and optimization of 7-substituted-pyrazolo[4,3-*b*]pyridine ALK5 (activin receptor-like kinase 5) inhibitors. *Bioorg Med Chem Lett* 27:1955–1961
30. Sutherland JJ, Nandigam RK, Erickson JA, Vieth M (2007) Lessons in molecular recognition. 2. Assessing and improving cross-docking accuracy. *J Chem Inf Model* 47:2293–2302
31. Mysinger MM, Michael C, Irwin JJ, Shoichet BK (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* 55:6582–6594
32. Diller DJ, Merz KM (2015) High throughput docking for library design and library prioritization. *Proteins* 43:113–124
33. Wu G, Robertson DH, Brooks CL, Vieth M (2010) Detailed analysis of grid-based molecular docking: a case study of CDOCKER—a CHARMM-based MD docking algorithm. *J Comput Chem* 24:1549–1562
34. Chinnasamy S, Selvaraj G, Kaushik AC, Kaliamurthi S, Chandrabose S, Singh SK, Thirugnanasambandam R, Gu KR, Wei DQ (2019) Molecular docking and molecular dynamics simulation studies to identify potent AURKA inhibitors: assessing the performance of density functional theory, MM-GBSA and mass action kinetics calculations. *J Biomol Struct Dyn* 37:1–11
35. Chu YY, Kaushik AC, Wang XG, Wang W, Zhang YF, Shan XQ, Salahub DR, Xiong Y, Wei DQ (2019) DTI-CDF: a cascade deep forest model towards the prediction of drug-target interactions based on hybrid features. *Brief Bioinform*. <https://doi.org/10.1093/bib/bbz152>
36. Hashemzadeh S, Ramezani F, Rafii-Tabar H (2019) Study of molecular mechanism of the interaction between MEK1/2 and trametinib with docking and molecular dynamic simulation. *Interdiscip Sci* 11:115–124
37. Li J, Fu A, Zhang L (2019) An overview of scoring functions used for protein-ligand interactions in molecular docking. *Interdiscip Sci* 11:320–328
38. Nguyen PTV, Yu H, Keller PA (2018) Molecular docking studies to explore potential binding pockets and inhibitors for chikungunya virus envelope glycoproteins. *Interdiscip Sci* 10:515–524

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.