

Article

Not peer-reviewed version

In Silico Search for Drug Candidates Targeting the PAX8–PPAR γ Fusion Protein in Thyroid Cancer

[Kaori Sakaguchi](#)^{*}, [Yoshio Okiyama](#)^{*}, [Shigenori Tanaka](#)^{*}

Posted Date: 3 April 2024

doi: 10.20944/preprints202404.0199.v1

Keywords: follicular thyroid neoplasia; PAX8–PPAR γ fusion; fusion protein; structure-based drug design



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

In Silico Search for Drug Candidates Targeting the PAX8–PPAR γ Fusion Protein in Thyroid Cancer

Kaori Sakaguchi ^{1,*}, Yoshio Okiyama ^{1,*} and Shigenori Tanaka ^{1,*}

Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe 657-8501, Japan; tanaka2@kobe-u.ac.jp

* Correspondence: 150x702x@stu.kobe-u.ac.jp (K.S.); okiyama@eagle.kobe-u.ac.jp (Y.O.); tanaka2@kobe-u.ac.jp (S.T.)

Abstract: The PAX8/PPAR γ rearrangement, producing the PAX8–PPAR γ fusion protein (PPFP), is thought to play an essential role in the oncogenesis of thyroid follicular tumors. To identify PPFP-targeted drug candidates and establish an early standard of care for thyroid tumors, we performed ensemble docking-based compound screening. Specifically, we investigated which pocket structure should be adopted to search for a promising ligand compound for the PPFP; the position of the ligand binding pocket on the PPAR γ side of the PPFP is common to that of PPAR γ ; however, the shape is slightly different between them due to environmental factors. We developed a method for selecting a PPFP structure with a relevant pocket and high prediction accuracy for ligand binding. This method was validated using PPAR γ , whose structure and activity values are known for many compounds. Then, we performed docking calculations to PPFP for 97 drug or drug-like compounds with a thiazolidine backbone, which is one of the characteristics of ligands that bind well to PPAR γ , registered in the DrugBank database. Furthermore, the binding affinities of promising ligand candidates were estimated more reliably using the molecular mechanics-Poisson–Boltzmann surface area method. Thus, we propose promising drug candidates for the PPFP with a thiazolidine backbone.

Keywords: follicular thyroid neoplasia; PAX8–PPAR γ fusion; fusion protein; structure-based drug design

1. Introduction

Thyroid cancer is a malignant tumor that originates in thyroid cells and can lead to abnormalities in hormone production and nodule formation. Most thyroid cancers contain one of several known driver mutations, such as the V600E substitution in BRAF, RAS mutation, RET gene fusion, or PAX8/PPAR γ gene fusion. Various drugs targeting these mutants are currently under development and are expected to become new treatment options. For example, if a patient has a BRAF V600E mutation or an ALK fusion gene, they can receive alectinib in an investigator-initiated clinical trial under patient-directed care [1]. Clinical trials on RET inhibitors for RET gene fusion and medullary thyroid cancer with a positive RET gene mutation are ongoing [2,3]. For unresectable thyroid cancer, the multi-targeted kinase inhibitors sorafenib, lenvatinib, and vandetanib are used as molecular-targeted drugs [4–6]. For solid tumors, including NTRK fusion gene-positive thyroid cancer, the ROS1/TRK inhibitor entrectinib has been approved [7]. However, these therapies simultaneously face limitations of side effects and treatment resistance.

The PAX8/PPAR γ rearrangement, producing the PAX8–PPAR γ fusion protein (PPFP), is rare in follicular adenomas. It has been reported that follicular adenomas with PAX8/PPAR γ rearrangement are likely to be follicular carcinomas, as these genetic abnormalities are presumed to be involved in the progression from follicular adenoma to follicular carcinoma [8]. The PAX8/PPAR γ fusion gene is a translocation between chromosomal regions 2q13 and 3p25, which contain PAX8 and PPAR γ , respectively [9]. Focusing on the PPAR γ side of the PPFP, several small molecules have been

considered as candidates for PFPF-targeted therapeutics [10–12]. The PFPF has been described as an inhibitor of PPAR γ activity or a transcription factor with proto-oncogene activity [13]; however, in mice, pioglitazone reportedly binds to the PFPF and induces it to behave like PPAR γ , resulting in loss of malignancy as thyroid cancer cells differentiate into adipocyte-like cells [9,10,13]. Furthermore, recent clinical trials have demonstrated that pioglitazone induces a significant decrease in thyroglobulin levels and a mild decrease in tumor size [12]. Therefore, it is possible to design a unique anticancer strategy, namely, suppressing malignant transformation by differentiation into noncancerous cells through enhancing the activity of the aberrant protein PFPF. Despite its importance as an anticancer drug target, there is no effective molecular targeted therapy for the PFPF. The development of small-molecule therapeutics targeting thyroid tumors with PFPF is expected to establish a standard of care for early stages.

The difficulty in drug discovery of fusion proteins, such as PFPF, is that their three-dimensional (3D) structure and complete sequence are not yet known; however, recent attempts have emerged to bridge this critical gap [14,15]. Fusion proteins naturally produced by chromosomal translocations, such as oncogenes, have many residues, and other environmental factors, including neighboring proteins, must also be considered. In our previous study [16], we attempted to construct the entire 3D structure of the PFPF using homology modeling based on expected sequential information [17–19] and X-ray crystallographic data of the substructures. The PFPF was assumed to be a 901-residue protein consisting of 396 residues of PAX8 on the N-terminal side and 505 residues of PPAR γ on the C-terminal side. When 3D modeling the PFPF, the crystal structure of the PPAR γ –rosiglitazone complex was used. This PPAR γ sequence was the least deficient. The modeled PFPF ligand-binding sites are common in position but have slightly different shapes. Since the PFPF consists of PAX8 and PPAR γ , we also used information on the primary, secondary, and tertiary structures of PPAR γ during its modeling. Using the modeled structure, we analyzed the fluctuations using molecular dynamics (MD) simulations to validate the structural stability. Furthermore, the molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) study of the binding affinities between PFPF/PPAR γ and rosiglitazone revealed different binding free energies for the same binding pocket. These results suggest that there may be optimal compounds for the PFPF that are different from PPAR γ , indicating the need for drug discovery targeting this fusion protein.

Ensemble docking is an efficient and cost-effective technique for compound screening to identify therapeutic drug candidates and has recently received much attention [20,21]. When preparing multiple protein structures for docking screening, it is important to know in advance which protein structure will likely yield in the most favorable results, that is, the structure with the highest predictive accuracy. If the active compounds are known, structures that correctly predict their active performance can be employed for screening; however, if not, such a policy cannot be followed. Based on the idea that proteins tend to bind to similar molecules even if they are not truly active compounds, Fukunishi et al. [22] proposed a solution to this problem using the following protocol: a relatively small molecular assembly of existing active compounds can be collected as universal active probes (UAPs), and the protein structure with the pocket with the highest sum of docking scores can be adopted as the screening structure. Because there are no known active compounds for the PFPF except for those targeting PPAR γ , we can take advantage of the UAP set for docking structure selection.

In this study, we first discuss the selection of the most suitable PFPF pocket for docking simulations from a large number of PFPF structures generated in the MD simulation of the PFPF apo-form and PFPF-rosiglitazone complex structures. We used PPAR γ , for which there is already a large amount of data on activity values and ligands, to verify whether the methodology using the UAP set properly performs and, in that case, what conditions are necessary. Verification included the predictive ability of the docking simulation using the Molecular Operating Environment (MOE) tool [23]. To the methodology using the UAP set, we added the condition that the ligand has a thiazolidinedione backbone, which is one of the characteristics of ligands that bind well to PPAR γ , and its similar structures. Based on these studies, we selected a PFPF structure with a pocket and high activity prediction potential. The selected PFPF structures were docked with 97 compounds

with a thiazolidine (TZD) backbone obtained from DrugBank Online (<https://go.drugbank.com/>) [24]. The most favorable drug candidate ligands for the PPFPP were also investigated. Finally, we propose the most promising candidate ligands for this fusion protein through re-evaluation using a more reliable MM-PBSA method.

2. Results and Discussion

The goal of this study was to predict promising binders for the PPFPP using a database of drugs or drug-like compounds through docking calculations. To this end, it was necessary to select PPFPP structures suitable for docking from the MD trajectories of apo-PPFP and PPFPP-rosiglitazone complexes modeled in our previous study [16]. Although the position of the PPAR γ -derived ligand-binding pocket is the same for all of them, their shapes are slightly different. Here, we first established a protocol to select structures suitable for docking using PPAR γ , for which there is abundant knowledge on co-crystal structures and binding affinities with active compounds. We then applied this protocol to PPFPP, for which there is no prior information on its activity or structure, to propose binders as promising target drug candidates.

2.1. Protocol for Structure Search for Reliable Docking

To establish a protocol for selecting docking structures with good prediction performance using PPAR γ , we first prepared 11 PPAR γ structures from two origins, the MD trajectories and the protein data bank (PDB) (Table 1). The former is based on the PPAR γ -rosiglitazone complex and apo-PPAR γ structures modeled in a previous study [16], giving six representative structures (A01–A06) from the equilibrium states in the 300 K MD simulations. The latter provided five representative crystal structures (A07–A11).

Table 1. Score and ranking of UAP docking to prepared PPAR γ structures. The scores are in kcal/mol and UAP is a small-molecule population consisting of 175 drug-like molecules. The sum of the docking scores was obtained by adding all scores obtained for a single structure. Preparation of PPAR γ structures and the origin and ID name of each structure.

Score Rank	Sum of Scores	Structure ID	Origin of structure
1	-1339.0	A11	X-ray crystal structure with PDB ID: 3VJI
2	-1319.3	A09	X-ray crystal structure with PDB ID: 3ADS
3	-1291.9	A06	Centroid of 2nd cluster for MD trajectory of complex structure with rosiglitazone
4	-1288.0	A07	X-ray crystal structure with PDB ID: 1I7I
5	-1283.8	A10	X-ray crystal structure with PDB ID: 3H0A
6	-1258.5	A01	Centroid of 1st cluster for MD trajectory of apo structure
7	-1236.8	A08	X-ray crystal structure with PDB ID: 2ZK1
8	-1231.1	A05	Centroid of 1st cluster for MD trajectory of complex structure with rosiglitazone
9	-1186.2	A03	Centroid of 3rd cluster for MD trajectory of apo structure
10	-1184.4	A04	Centroid of 4th cluster for MD trajectory of apo structure
11	-1075.7	A02	Centroid of 2nd cluster for MD trajectory of apo structure

We next docked UAPs consisting of 175 small molecules (Table S1) to the aforementioned 11 PPAR γ structures to rank their screening performance by the sum of docking scores (Table 1). The results show that structure ID A11 from the PDB had the best UAP with a docking score sum of -1339.0 kcal/mol, and according to a previous study [22], this structure can be judged as the most suitable for docking screening.

To verify the performance of the UAP protocol on the PPAR γ target, we prepared a compound set consisting of 41 PPAR γ binders with co-crystal structures registered with PPAR γ in the PDB and at least one of the activity values of IC₅₀, K_d, or K_i reported in ChEMBL. Then, the sums of the docking

scores for the compound set with the 11 PPAR γ structures were compared to those of the UAP set. Figure 1 shows the correlation between the sum of docking scores for the 11 structures in the UAP- and PDB-derived compound sets. The high correlation coefficient, $R^2=0.84$, suggests that the UAP protocol would be applicable to PPAR γ for the selection of good pockets for screening with high probability.

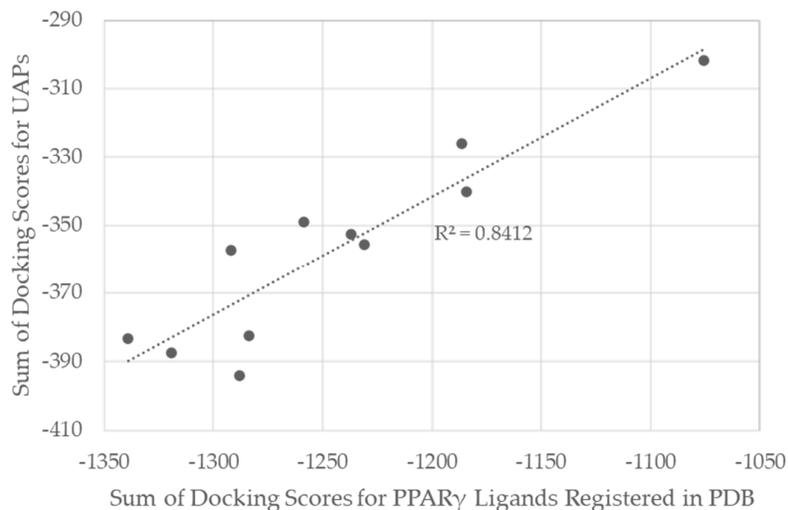


Figure 1. Correlation between the sum of scores obtained by docking the UAP and PDB PPAR γ ligands to the 11 PPAR γ structures. The UAP is a drug-like small molecule population consisting of 175 compounds. The PDB PPAR γ ligands are 41 small molecules with crystal structures and active values. Units are in kcal/mol.

To examine the predictive ability of docking for a wide range of PPAR γ known active compounds, we performed docking for 11 PPAR γ structures of 1696 compounds with activity values (IC_{50} , K_d , K_i) for PPAR γ registered in ChEMBL. Table 2 shows the correlation coefficient between docking scores and activity values for the corresponding compounds for each activity type and for each PPAR γ structure (denoted as “All” in Table 2). Almost no correlation was observed for IC_{50} , and weak correlations were observed for K_d and K_i . We then limited our analysis to compounds with TZD and phenylacetic acid (PA) backbones that play important roles in binding to PPAR γ ; an existing marketed drug, pioglitazone, shares the TZD backbone. Even with these restrictions, we found little improvement in the correlation with IC_{50} ; however, we obtained noteworthy correlation coefficients (R^2) with K_i : for the TZD compounds, 0.482 with the structure ID A03, 0.481 with A05, and 0.491 with A06, and for the PA compounds, greater than 0.5 with many structures. As hypothesized, a binding mode similar to that of TZD, including PA, is expected to improve the prediction performance of K_i .

Table 2. Correlation coefficient R^2 between docking scores of 11 prepared PPAR γ structures with PPAR γ ligands obtained from ChEMBL and their activity values. “All” represents all of the respective activity value data; “TZD” represents compounds with a thiazolidine backbone; “PA” represents compounds with a phenylacetic acid backbone.

Structure ID	IC ₅₀			K _d	K _i		
	All	TZD	PA	All	All	TZD	PA
A01	0.057	0.000	0.0115	0.165	0.176	0.321	0.584
A02	0.052	0.010	0.0019	0.033	0.081	0.201	0.026
A03	0.065	0.000	0.0371	0.376	0.169	0.482	0.500
A04	0.006	0.226	0.1212	0.014	0.091	0.160	0.560
A05	0.067	0.003	0.0543	0.055	0.118	0.481	0.519
A06	0.070	0.000	0.0588	0.378	0.171	0.491	0.571
A07	0.052	0.019	0.0066	0.063	0.052	0.390	0.576
A08	0.050	0.001	0.0553	0.242	0.113	0.243	0.482
A09	0.021	0.002	0.0606	0.103	0.147	0.334	0.513
A10	0.013	0.029	0.0016	0.002	0.134	0.232	0.672
A11	0.049	0.022	0.0809	0.372	0.048	0.309	0.492

Since the PPAR γ pocket is softer, the chemical space of possible binders is wider. The PPAR γ compound group registered in ChEMBL has a wide variation. The selection criteria for the pocket that is expected to give the best docking results are as follows: It can be expected that good results will be obtained if the condition is focused on PPAR γ . Among the PPAR γ ligands, it seems best to focus on the TZD skeleton characteristic of pioglitazone, which is already a marketed drug. In the case of the PPF γ , it is not possible to know the activity value in advance; therefore, it would be beneficial to select based on some molecular characteristics. In terms of the position of the pocket, PPAR γ and PPF γ have a pocket in the same position. Therefore, the basic idea of drug discovery is to change the side chain based on the mother skeleton, which is expected to have some activity. First, we focused our discussion on the analogous structure of pioglitazone. Therefore, we propose that the best approach is to have the TZD backbone characteristic of pioglitazone, which is already known to bind well among PPAR γ ligands, in addition to previously used methods. Following only the methods of previous studies, the structure selected for PPAR γ was PDB-derived A11. The correlation between the docking score and activity value showed that the R^2 value was not necessarily good. Based on the proposed method, the best structure was A06, which was derived from MD simulations. It had sufficient K_i data and a larger R^2 value than A11. Thus, an even better docking score prediction was expected.

2.2. Application to the PPF γ

Based on the PPAR γ considerations above, a screening protocol that provided the highest predictive accuracy was applied to the PPF γ . As in the case of PPAR γ , but without any crystal structures, 19 representative structures with structural IDs B01–B19 were extracted from the MD trajectories for the rosiglitazone complex and apo structure in our previous study (Table 3). After evaluation of binding capacities by the UAP set, structure ID B14, showing the highest value of the sum of docking scores (-1303.2 kcal/mol) was selected as the best pocket structure for screening the PPF γ (Figure 2).

Table 3. Docking scores and ranking of UAP docking to prepared PPF structures. The scores are in kcal/mol and UAP is a small-molecule population consisting of 175 drug-like molecules. The sum of the docking scores was obtained by adding all scores obtained for a single structure. Preparation of PPF structures and the origin and ID name of each structure.

Score Rank	Sum of Scores	Structure ID	Origin of structure
1	-1303.2	B14	Centroid of 14th cluster for MD trajectory of apo structure
2	-1278.6	B06	Centroid of 6th cluster for MD trajectory of apo structure
3	-1276.8	B02	Centroid of 2nd cluster for MD trajectory of apo structure
4	-1276.0	B16	Centroid of 2nd cluster for MD trajectory of complex structure with rosiglitazone
5	-1271.3	B04	Centroid of 4th cluster for MD trajectory of apo structure
6	-1253.0	B12	Centroid of 12th cluster for MD trajectory of apo structure
7	-1250.8	B10	Centroid of 10th cluster for MD trajectory of apo structure
8	-1242.3	B15	Centroid of 1st cluster for MD trajectory of complex structure with rosiglitazone
9	-1240.9	B08	Centroid of 8th cluster for MD trajectory of apo structure
10	-1236.3	B11	Centroid of 11th cluster for MD trajectory of apo structure
11	-1229.1	B17	Centroid of 3rd cluster for MD trajectory of complex structure with rosiglitazone
12	-1225.7	B18	Centroid of 4th cluster for MD trajectory of complex structure with rosiglitazone
13	-1220.5	B19	Centroid of 5th cluster for MD trajectory of complex structure with rosiglitazone
14	-1219.4	B05	Centroid of 5th cluster for MD trajectory of apo structure
15	-1217.0	B13	Centroid of 13th cluster for MD trajectory of apo structure
16	-1213.6	B09	Centroid of 9th cluster for MD trajectory of apo structure
17	-1213.3	B07	Centroid of 7th cluster for MD trajectory of apo structure
18	-1177.6	B01	Centroid of 1st cluster for MD trajectory of apo structure
19	-1170.1	B03	Centroid of 3rd cluster for MD trajectory of apo structure

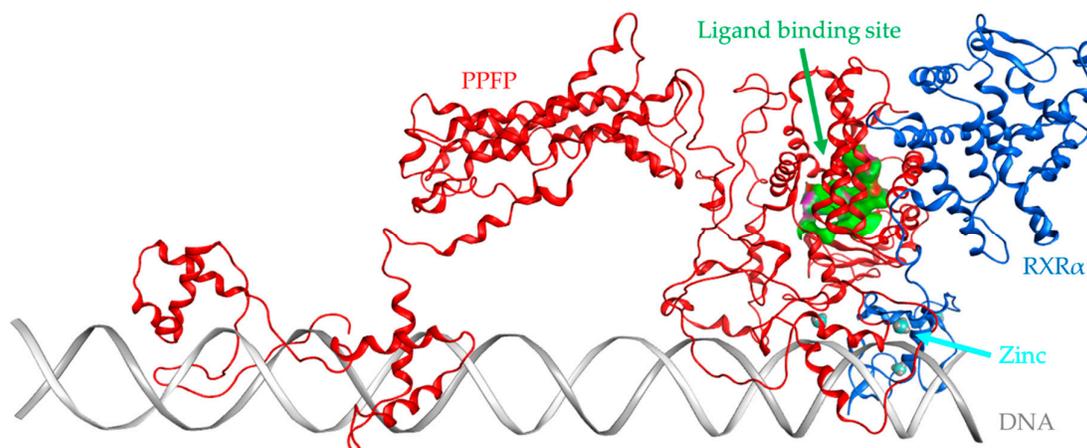
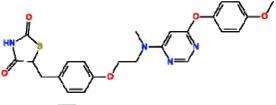
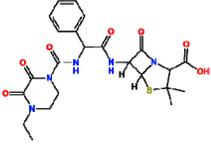
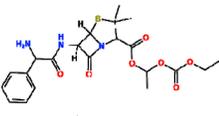
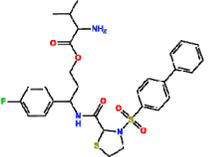
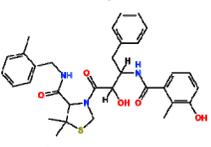
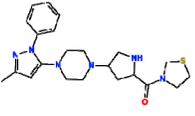
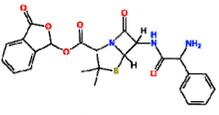
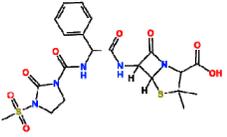
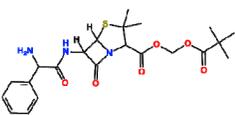
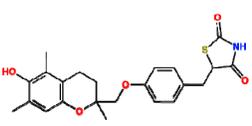
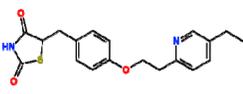
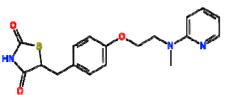


Figure 2. Ribbon drawing of PPF/RXRα on DNA. Red denotes the PPF and blue RXRα. The PPARγ used for modeling PPF is known to form a dimer with RXRα. Zinc ions are shown as balls. The ligand binding pocket in PPF is shown in green.

In this study, we attempted to identify PPF target drug candidates from the perspective of drug repositioning. The compounds to be screened were obtained from DrugBank [24], an online database of drug and drug candidate compounds and their targets, which includes 9213 compounds available as of November 2022. Based on the scope of the proposed protocol, 97 compounds with TZD backbones were identified (Table S2). Because we focused on comparing the drug candidates with

existing drugs, such as pioglitazone and rosiglitazone, candidates with a PA backbone were not considered in this study. The 97 compounds were docked to the PPF structure with an ID of B14. Table 4 shows the top 10 scoring compounds, along with pioglitazone and rosiglitazone. Pioglitazone and rosiglitazone are widely known PPAR γ drugs; however, their binding scores with PPF were not very high in this study.

Table 4. Data on top 10 docking scores, ligand ID names, generic names, and 2D structures.

Score Rank	Docking Score*	Ligand ID	Generic Name	2D
1	-9.837	L62	Lobeglitazone	
2	-9.387	L03	Piperacillin	
3	-9.346	L21	Bacampicillin	
4	-9.248	L96	Ebopirant	
5	-9.162	L29	JE-2147	
6	-9.133	L72	Teneligiptin	
7	-9.119	L87	Talampicillin	
8	-9.080	L13	Mezlocillin	
9	-8.931	L23	Pivampicillin	
10	-8.693	L58	(5R)-5-(4-((2R)-6-HYDROXY-2,5,7,8-TETRAMETHYL-3,4-DIHYDRO-2H-CHROMEN-2-YL)METHOXY)BENZYL)-1,3-THIAZOLIDINE-2,4-DIONE	
28	-8.009	L18	Pioglitazone	
43	-7.643	L04	Rosiglitazone	

*Score values are in kcal/mol.

For a more reliable binding-affinity evaluation, multi-sampling MM-PBSA calculations were performed to rescore the 12 docked compounds, including pioglitazone and rosiglitazone. Table 5 shows that teneligliptin (L72), a DDP-4 inhibitor, was the most promising binder for the PPF. Rosiglitazone (L4) and pioglitazone (L18) had high binding affinities. For ligand candidate L13, the docking score was good; however, the binding free energy was -11.98 kcal/mol, excluding it from the candidate list. Although pioglitazone and rosiglitazone are naturally good candidates, we also determined better binders, such as teneligliptin (see Figure 3), using our computational protocol.

Table 5. Calculation of MM/PBSA binding free energies for PPF candidates.

Ligand ID	Delta Total Mean (s.d.)	Score Rank
L72	-25.21 (3.05)	6
L29	-22.82 (4.44)	5
L21	-22.73 (4.49)	3
L03	-21.76 (3.69)	2
L62	-21.15 (3.36)	1
L96	-21.09 (3.49)	4
L04	-20.89 (3.10)	43
L87	-20.32 (3.66)	7
L58	-20.22 (3.25)	10
L23	-17.79 (4.57)	9
L18	-17.22 (2.99)	28
L13	-11.98 (4.40)	8

The unit of binding free energy (Delta Total) is kcal/mol. "s.d." means the standard deviation.

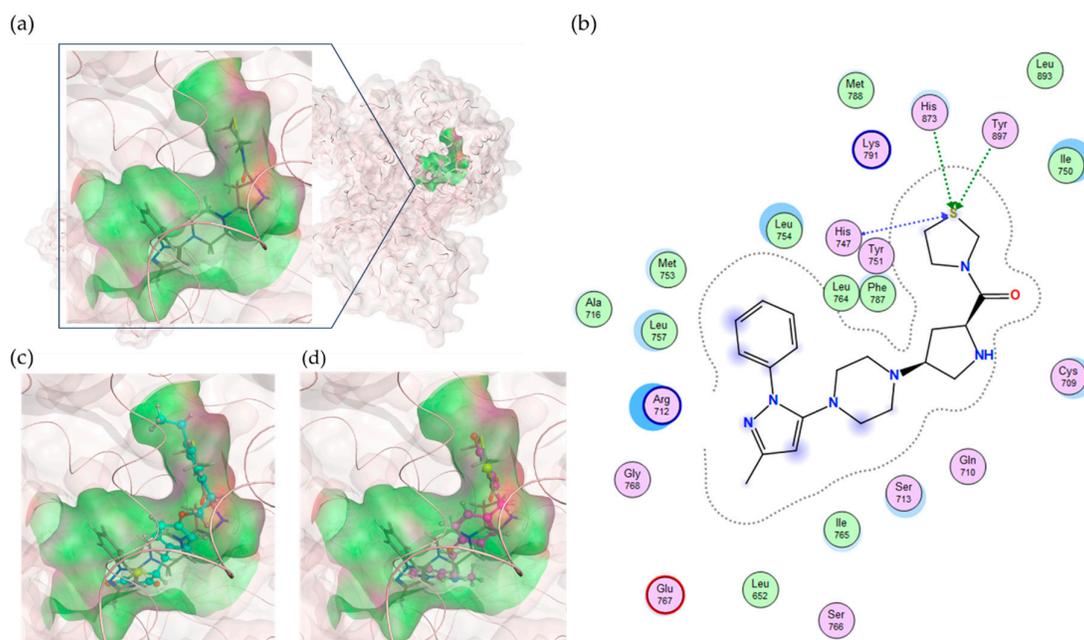


Figure 3. Binding pose of tenegliptin (L72) in PPFp. (a) Three-dimensional view, where L72 is represented by stick. (b) Two-dimensional view. (c) Superposition of L72 (stick) to pioglitazone (cyan) in binding pocket. (d) Superposition of L72 (stick) to rosiglitazone (magenta) in binding pocket.

3. Materials and Methods

3.1. Protein Structures for Docking

3.1.1. PPAR γ Structures

We prepared 11 PPAR γ structures categorized into two structure types. Numerous structures were generated by MD simulations during modeling the PPAR γ apo-body and the PPAR γ -rosiglitazone complex structures in our previous study [16]. These were clustered based on the shape of the ligand-binding sites using the density-based spatial clustering of applications with noise (DBSCAN) method of the Amber tool in the MD analysis program. Finally, six structures were selected. On the other hand, the Site Finder application in MOE software was used to detect pocket shapes in 256 PPAR γ -ligand complex crystal structures downloaded from the PDB. The features of the pocket shapes were quantified using Sterimol parameters [20,21]. The numerical values were used to cluster the molecules using the k-means method; the Sterimol parameters are a set of vectors describing the steric occupancy of a molecule. They have been extensively used in quantitative structure-activity relationship (QSAR) studies for drug discovery [22,23]. To this end, five structures were selected PDB ID: 1I7I, 2ZK1, 3ADS, 3H0A, and 3VJI. Structures downloaded from the PDB were preprocessed using MOE for docking simulations [23]. Hydrogen atoms were added using Protonate3D. AMBER10:EHT was applied to the force field. The hydrogen atoms were structurally optimized.

3.1.2. PPFp Structures

We created 19 PPFp structures. In our previous study [16], we proposed a homology modeling construction of the entire 3D structure of the PPFp, which is still unresolved, based on X-ray crystal structure data for PPAR γ and PAX8. To validate these model structures, we analyzed the fluctuations using MD simulations and predicted the physical properties based on the structures and sequences. Many trajectories extracted from these MD simulations were clustered using the DBSCAN method of the Amber tool in the MD analysis program [25]. The parameter radius distance was set to $\epsilon=0.9$ and the threshold with neighborhood density was set to $\text{minPts}=25$. The clustering criterion was the root-mean-square deviation (RMSD) of the residues around the PPFp pocket. The sampled temporal range was between 10 and 100 ns, which was determined to be the equilibrium state, and 18,000 frames were read every five frames. The top clusters, which accounted for more than 1% of all frames, were extracted, and their centroid was used as the representative structure in the equilibrium state. Numerous PPFp-rosiglitazone complex and PPFp apo-body structures, which were previously modeled [16], were generated using MD simulations. These were also clustered based on the shape of the ligand-binding sites. Because the crystal structure of the PPFp was not available, only the MD simulation-derived structures were used. The same procedure as for PPAR γ was performed for the 19 prepared PPFp structures.

3.2. Compound Sets for Docking

3.2.1. UAPs

Fukunishi et al. [22] proposed the UAP set, which is a collection of relatively small molecules among active compounds against various targets, to evaluate the screening ability of the pocket. This evaluation is based on the idea that similar molecules tend to bind to a pocket, even if it is not the actual active compound of the pocket. Therefore, the pocket with the highest score obtained by docking UAPs to multiple pockets is considered the binding pocket suitable for screening active

compounds. In this study, the UAP set consisting of 175 small drug-like compounds registered in the PDB was obtained from MyPresto version 5 (<https://www.mypresto5.jp>).

3.2.2. PPAR γ Ligands Registered in ChEMBL

One hundred and ninety five (195) PPAR γ -ligand complex crystal structures were available from the PDB. One hundred and forty eight (148) PPAR γ ligand activity values (IC_{50} , K_d , K_i) were in ChEMBL. Among them, 63 PPAR γ ligands had both crystal structures and activity values. Excluding duplicates, 41 PPAR γ ligands were obtained. These 41 PPAR γ ligands were docked against 11 structures of PPAR γ .

The target-associated bioactive values for PPAR γ were extracted from ChEMBL29 (<https://www.ebi.ac.uk/chembl/>) with the target ID ChEMBL235. This resulted in 1696 compounds with SMILES (Table S3).

The computational analysis procedure was the same even when a different type of ligand was docked to PPAR γ . The median and logP values for the activity were used in the calculation and analysis; K_i had 20 small activity values with $\log P \leq 2$, which were excluded as outliers (Figure S1). First, when all IC_{50} values were used (All), there was little correlation; the 1696 PPAR γ ligands obtained from ChEMBL included those with large molecular weights. Lipinski's rule of five [26,27] summarizes the chemical properties of compounds that are likely to become oral drugs and suggests that one of the properties of the drug is a molecular weight of less than 500.

TZDs are a class of heterocyclic compounds with thioether and amine groups attached at the first and third positions of the saturated five-membered ring, respectively (Figure 4). The TZD skeleton is characteristic of pioglitazone. Pioglitazone is a ligand for PPAR γ , and is already a marketed drug; a two-dimensional view of the ligand binding sites of PPAR γ and pioglitazone is displayed in Figure 4. Positional relationships and interactions of the peripheral residues are seen, where hydrogen bonds, ionic bonds, π - π interactions, CH- π interactions, cation- π interactions, halogen bonds, and coordination bonds are anticipated. The TZD backbone plays an important role in the bonding; thus, we extracted 101 compounds with this backbone from the aforementioned 1696 bioactive compounds (Table S3).

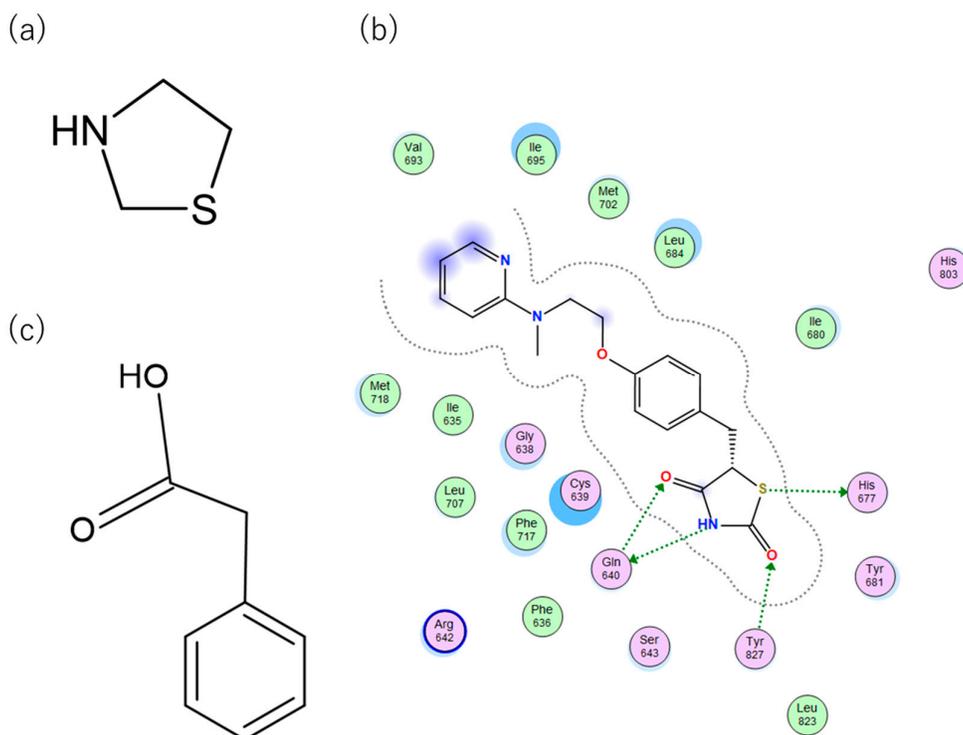


Figure 4. (a) Two-dimensional chemical structure of TZD. (b) Two-dimensional view of the characterization of the ligand binding sites of PPAR γ and pioglitazone. Dotted green arrows indicate interactions between TZD and Gln640, Tyr827, and His677. (c) Two-dimensional chemical structure of PA.

PA is another common skeleton of PPAR γ binding compounds with a similar binding mode to TZD (Figure 4). To investigate the predicted activity performance of compounds with such similar binding modes, we extracted 31 compounds with a PA backbone from the aforementioned 1696 bioactive compounds of PPAR γ (Table S3). Using these experimental values (IC_{50} , K_d , and K_i), we confirmed that they could be reproduced in the order of increasing activity and correlation. This method was applied to the PPF γ . In the case of PPAR γ , we established a procedure to predict the superiority of activity values from docking calculations, determining the ligand that binds well to the PPF γ . All activity values for PPAR γ were used for calculation and analysis and were obtained by evaluating median and logP values. For K_i , 366 original values were obtained, of which 20 extremely low activity values were excluded as outliers.

3.2.3. TZD-Backbone Compounds Registered in the DrugBank Database

As of November 2022, there were 9213 ligands registered in the DrugBank database [24]. The TZD backbone portion is a characteristic structure of pioglitazone and plays an important role in the binding of pioglitazone to the ligand binding site of PPAR γ . Similarly, the TZD backbone is thought to play an important role in the binding to the ligand-binding site of the PPF γ . A total of 97 ligands with the TZD skeleton were identified. In this study, we did not consider the case of PA because we wanted to compare the candidates with existing commercial drugs, pioglitazone and rosiglitazone.

3.3. Computational Methods

3.3.1. Docking

Docking simulations were performed using MOE software [23]. The final binding free energies and docking scores were evaluated using the London dG and GBVI/WSA dG scoring functions.

3.3.2. MM-PBSA Method

The MM-PBSA approach is an efficient method for evaluating the binding free energy (ΔG_{bind}) of protein-ligand complexes [28]. We used the MMPBSA.py module of AmberTools 14 to calculate the binding free energies of 10 ligands considered promising candidates for PPF γ with PPF γ , pioglitazone with PPF γ , and rosiglitazone with PPF γ , and investigated their binding affinities [29]. First, MD simulations were run for 10 ns to investigate the conformational fluctuations at equilibrium in solution. The protocol for this MD simulation was the same as that previously published [16]. Consequently, 100 structures were extracted from the MD trajectories of the production runs. The scores were reevaluated using the MM-PBSA method [30]. The binding free energies were calculated using the MM-PBSA method to determine the binding affinity of the receptor (PPF γ) and ligand (several small molecules of the candidate drugs). The difference between the free energy of the complex and those of the ligand and receptor is the binding free energy (Equation 1):

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad (1)$$

The ΔG_{bind} can also be decomposed into ligand-residue interactions [29]. In this study, the dielectric constant of the solute (protein, internal dielectric constant) was 4.0, while that of the solvent (water, external dielectric constant) was 80.0.

4. Conclusions

A method for selecting protein structures with pockets suitable for ligand docking from numerous protein structures was developed, and correlations between docking scores of PPAR γ and PPAR γ ligands and their activity values were examined. Correlations were checked to determine

whether they could be reproduced in the order of the experimental activity. The method used in a previous study using UAP is promising. To this method, we added the condition that the PPAR γ ligands have a TZD backbone characteristic of the existing drugs, pioglitazone and rosiglitazone. In this way, the correlations improved. We applied this method to the PPF and identified small molecules with a TZD moiety as promising drug compounds. In addition to PPAR γ ligands included in existing drugs, compounds with structures similar to TZDs, such as PA, may also be candidates. This suggests the need to search for drug candidates within the PPF pocket rather than the PPAR γ pocket. The top ten compounds reported herein are expected to be potential candidate drugs for the PPF, for which the ligands are promising.

Supplementary Materials: The following supporting information can be downloaded at website of this paper posted on Preprints.org, Table S1: SMILES of UAP; Table S2: Small molecule SMILES with a thiazolidinedione skeleton registered in DrugBank; Table S3: SMILES of ligands of PPAR γ with activity values registered in ChEMBL and their respective conditional classification.; Figure S1: (a) Outliers in K_i in docking to structure A03 to show correlation (b).

Author Contributions: Conceptualization, K.S., Y.O. and S.T.; methodology, K.S., Y.O. and S.T.; software, K.S., Y.O. and S.T.; validation, K.S., Y.O. and S.T.; formal analysis, K.S., Y.O. and S.T.; investigation, K.S., Y.O. and S.T.; resources, S.T.; data curation, K.S. and Y.O.; writing—original draft preparation, K.S.; writing—review and editing, Y.O. and S.T.; visualization, K.S. and Y.O.; supervision, S.T.; project administration, S.T.; funding acquisition, S.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Grants-in-Aid for Scientific Research (No. 21K06098) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and Japan Society for the Promotion of Science (JSPS), and by MEXT Quantum Leap Flagship Program (Grant No. JPMXS0120330644).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tahara, M. Genomic Medicine in Thyroid Cancer toward Precision Medicine. *Folia Endocrinol. Jpn.* **2020**, *37*, 110–114, doi:https://doi.org/10.11226/jaesjsts.37.2_110.
2. Subbiah, V.; Velcheti, V.; Tuch, B.B.; Ebata, K.; Busaidy, N.L.; Cabanillas, M.E.; Wirth, L.J.; Stock, S.; Smith, S.; Lauriault, V.; et al. Selective RET Kinase Inhibition for Patients with RET-Altered Cancers. *Ann. Oncol.* **2018**, *29*, 1869–1876, doi:10.1093/annonc/mdy137.
3. Subbiah, V.; Gainor, J.F.; Rahal, R.; Brubaker, J.D.; Kim, J.L.; Maynard, M.; Hu, W.; Cao, Q.; Sheets, M.P.; Wilson, D.; et al. Precision Targeted Therapy with BLU-667 for RET -Driven Cancers. *Cancer Discov.* **2018**, *8*, 836–849, doi:10.1158/2159-8290.CD-18-0338.
4. Wells, S.A.; Robinson, B.G.; Gagel, R.F.; Dralle, H.; Fagin, J.A.; Santoro, M.; Baudin, E.; Elisei, R.; Jarzab, B.; Vasselli, J.R.; et al. Vandetanib in Patients with Locally Advanced or Metastatic Medullary Thyroid Cancer: A Randomized, Double-Blind Phase III Trial. *J. Clin. Oncol.* **2012**, *30*, 134–141, doi:10.1200/JCO.2011.35.5040.
5. Schlumberger, M.; Tahara, M.; Wirth, L.J.; Robinson, B.; Brose, M.S.; Elisei, R.; Habra, M.A.; Newbold, K.; Shah, M.H.; Hoff, A.O.; et al. Lenvatinib versus Placebo in Radioiodine-Refractory Thyroid Cancer. *N. Engl. J. Med.* **2015**, *372*, 621–630, doi:10.1056/NEJMoa1406470.
6. Brose, M.S.; Nutting, C.M.; Jarzab, B.; Elisei, R.; Siena, S.; Bastholt, L.; de la Fouchardiere, C.; Pacini, F.; Paschke, R.; Shong, Y.K.; et al. Sorafenib in Radioactive Iodine-Refractory, Locally Advanced or Metastatic Differentiated Thyroid Cancer: A Randomised, Double-Blind, Phase 3 Trial. *Lancet* **2014**, *384*, 319–328, doi:10.1016/S0140-6736(14)60421-9.
7. Doebele, R.C.; Drilon, A.; Paz-Ares, L.; Siena, S.; Shaw, A.T.; Farago, A.F.; Blakely, C.M.; Seto, T.; Cho, B.C.; Tosi, D.; et al. Entrectinib in Patients with Advanced or Metastatic NTRK Fusion-Positive Solid Tumours: Integrated Analysis of Three Phase 1–2 Trials. *Lancet Oncol.* **2020**, *21*, 271–282, doi:10.1016/S1470-2045(19)30691-6.
8. Klemke, M.; Drieschner, N.; Laabs, A.; Rippe, V.; Belge, G.; Bullerdiek, J.; Sendt, W. On the Prevalence of the PAX8-PPARG Fusion Resulting from the Chromosomal Translocation t(2;3)(Q13;P25) in Adenomas of the Thyroid. *Cancer Genet.* **2011**, *204*, 334–339, doi:10.1016/j.cancergen.2011.05.001.

9. Kroll, T.G. PAX8-PPAR γ 1 Fusion in Oncogene Human Thyroid Carcinoma. *Science* (80-). **2000**, *289*, 1357–1360, doi:10.1126/science.289.5483.1357.
10. Dobson, M.E.; Diallo-Krou, E.; Grachtchouk, V.; Yu, J.; Colby, L.A.; Wilkinson, J.E.; Giordano, T.J.; Koenig, R.J. Pioglitazone Induces a Proadipogenic Antitumor Response in Mice with PAX8-PPAR γ Fusion Protein Thyroid Carcinoma. *Endocrinology* **2011**, *152*, 4455–4465, doi:10.1210/en.2011-1178.
11. Xu, B.; O'Donnell, M.; O'Donnell, J.; Yu, J.; Zhang, Y.; Sartor, M.A.; Koenig, R.J. Adipogenic Differentiation of Thyroid Cancer Cells through the Pax8-PPAR γ Fusion Protein Is Regulated by Thyroid Transcription Factor 1 (TTF-1). *J. Biol. Chem.* **2016**, *291*, 19274–19286, doi:10.1074/jbc.M116.740324.
12. Giordano, T.J.; Haugen, B.R.; Sherman, S.I.; Shah, M.H.; Caoili, E.M.; Koenig, R.J. Pioglitazone Therapy of PAX8-PPAR γ Fusion Protein Thyroid Carcinoma. *J. Clin. Endocrinol. Metab.* **2018**, *103*, 1277–1281, doi:10.1210/je.2017-02533.
13. Giordano, T.J. Delineation, Functional Validation, and Bioinformatic Evaluation of Gene Expression in Thyroid Follicular Carcinomas with the Pax8-Pparg Translocation. *Clin. Cancer Res.* **2006**, *12*, 1983–1993, doi:10.1158/1078-0432.CCR-05-2039.
14. Kumar, H.; Tang, L.; Yang, C.; Kim, P. FusionPDB: A Knowledgebase of Human Fusion Proteins. **2023**, 1–16.
15. Shamriz, S.; Ofoghi, H. Design, Structure Prediction and Molecular Dynamics Simulation of a Fusion Construct Containing Malaria Pre-Erythrocytic Vaccine Candidate, PfCelTOS, and Human Interleukin 2 as Adjuvant. *BMC Bioinf* **2016**, *17*, 71, doi:10.1186/s12859-016-0918-8.
16. Sakaguchi, K.; Okiyama, Y.; Tanaka, S. In Silico Modeling of PAX8-PPAR γ Fusion Protein in Thyroid Carcinoma: Influence of Structural Perturbation by Fusion on Ligand-Binding Affinity. *J. Comput. Aided. Mol. Des.* **2021**, *35*, 629–642, doi:10.1007/s10822-021-00381-x.
17. Raman, P.; Koenig, R.J. Pax-8-PPAR- γ Fusion Protein in Thyroid Carcinoma. *Nat. Rev. Endocrinol.* **2014**, *10*, 616–623, doi:10.1038/nrendo.2014.115.
18. Vuttariello, E.; Biffali, E.; Pannone, R.; Capiluongo, A.; Monaco, M.; Sica, V.; Aiello, C.; Matuozzo, M.; Chiofalo, M.G.; Botti, G.; et al. Rapid Methods to Create a Positive Control and Identify the PAX8/PPAR γ Rearrangement in FNA Thyroid Samples by Molecular Biology. *Oncotarget* **2018**, *9*, 19255–19262, doi:10.18632/oncotarget.24995.
19. Pasca di Magliano, M.; Di Lauro, R.; Zannini, M. Pax8 Has a Key Role in Thyroid Cell Differentiation. *Proc. Natl. Acad. Sci.* **2000**, *97*, 13144–13149, doi:10.1073/pnas.240336397.
20. Tian, S.; Sun, H.; Pan, P.; Li, D.; Zhen, X.; Li, Y.; Hou, T. Assessing an Ensemble Docking-Based Virtual Screening Strategy for Kinase Targets by Considering Protein Flexibility. *J. Chem. Inf. Model.* **2014**, *54*, 2664–2679, doi:10.1021/ci500414b.
21. Uehara, S.; Tanaka, S. Cosolvent-Based Molecular Dynamics for Ensemble Docking: Practical Method for Generating Druggable Protein Conformations. *J. Chem. Inf. Model.* **2017**, *57*, 742–756, doi:10.1021/acs.jcim.6b00791.
22. Fukunishi, Y.; Ohno, K.; Orita, M.; Nakamura, H. Selection of In Silico Drug Screening Results by Using Universal Active Probes (UAPs). *J. Chem. Inf. Model.* **2010**, *50*, 1233–1240, doi:10.1021/ci100108p.
23. ULC C.C.G. Molecular Operating Environment (MOE), 2020.09. Chemical Computing Group ULC, McGill University; Montreal, QC, Canada: 2020. **2020**, 2020.
24. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082, doi:10.1093/nar/gkx1037.
25. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining; 1996.
26. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings 1PII of Original Article: S0169-409X(96)00423-1. The Article Was Originally Published in Advanced Drug Delivery Reviews 23 (1997). *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26, doi:10.1016/S0169-409X(00)00129-0.
27. Lipinski, C.A. Lead- and Drug-like Compounds: The Rule-of-Five Revolution. *Drug Discov. Today Technol.* **2004**, *1*, 337–341, doi:10.1016/j.ddtec.2004.11.007.
28. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461, doi:10.1517/17460441.2015.1032936.
29. Miller, B.R.; McGee, T.D.; Swails, J.M.; Homeyer, N.; Gohlke, H.; Roitberg, A.E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. **2012**.
30. Sahakyan, H. Improving Virtual Screening Results with MM/GBSA and MM/PBSA Rescoring. *J. Comput. Aided. Mol. Des.* **2021**, *35*, 731–736, doi:10.1007/s10822-021-00389-3.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s)

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.