Article

Mechanistic Insights into Side Effects of Troglitazone and Rosiglitazone Using a Novel Inverse Molecular Docking Protocol

Katarina Kores [1], Janez Konc [1,2] and Urban Bren [1,3,*]

1 Laboratory of Physical Chemistry and Chemical Thermodynamics,
Faculty for Chemistry and Chemical Technology, University of Maribor, Smetanova 17,
SI-2000 Maribor, Slovenia; katarina.kores@um.si (K.K.); janez.konc@ki.si (J.K.)
2 Laboratory for Molecular Modeling, Theory Department, National Institute of Chemistry, Hajdrihova 19,
SI-1000 Ljubljana, Slovenia
3 Department of Applied Natural Sciences, Faculty of Mathematics,
Natural Sciences and Information Technologies, University of Primorska, Glagoljaška 8,
SI-6000 Koper, Slovenia
* Correspondence: urban.bren@um.si

Abstract: Thiazolidinediones form drugs that treat insulin resistance in type 2 diabetes mellitus. Troglitazone represents the first drug from this family, which was removed from use by the FDA due to its hepatotoxicity. As an alternative, rosiglitazone was developed, but it was under the careful watch of FDA for a long time due to suspicion, that it causes cardiovascular diseases, such as heart failure and stroke. We applied a novel inverse molecular docking protocol to discern the potential protein targets of both drugs. Troglitazone and rosiglitazone were docked into predicted binding sites of >67,000 protein structures from the Protein Data Bank and examined. Several new potential protein targets with successfully docked troglitazone and rosiglitazone were identified. The focus was devoted to human proteins so that existing or new potential side effects could be explained or proposed. Certain targets of troglitazone such as 3-oxo-5-beta-steroid 4-dehydrogenase, neutrophil collagenase, stromelysin-1, and VLCAD were pinpointed, which could explain its hepatotoxicity, with additional ones indicating that its application could lead to the treatment/development of cancer. Results for rosiglitazone discerned its interaction with members of the matrix metalloproteinase family, which could lead to cancer and neurodegenerative disorders. The concerning cardiovascular side effects of rosiglitazone could also be explained. We firmly believe that our results deepen the mechanistic understanding of the side effects of both drugs, and potentially with further development and research maybe even help to minimize them. On the other hand, the novel inverse molecular docking protocol on the other hand carries the potential to develop into a standard tool to predict possible cross-interactions of drug candidates potentially leading to adverse side effects.

Keywords: thiazolidinediones; rosiglitazone; troglitazone; inverse molecular docking; side effects; CANDOCK

1. Introduction

Thiazolidinediones or glitazones (TZDs) represent established drugs that treat insulin resistance in type 2 diabetes mellitus [1,2], which is a chronic metabolic disorder caused by defects in insulin secretion and insulin action [3]. People with type 2 diabetes mellitus also experience a higher risk for developing cardiovascular diseases [4]. TZDs activate peroxisome proliferator-activated receptor gamma (PPARγ), which is a nuclear receptor that regulates the expression of several genes connected with metabolism. PPARγ represents one of the three PPARs, which are nuclear receptors that regulate lipid metabolism and glucose homeostasis, the latter having control over adipocyte differentiation, lipid storage, and insulin sensitization. The primary effects of TZDs through the activation of

PPARγ are seen in adipose tissues (body fat) by reducing triglycerides and increasing HDL cholesterol [1,2,5,6]. Initially, there were a few TZD drugs developed, but only three have been approved by the FDA [2]. To this day, only rosiglitazone (Avandia, Avandaryl, and Avandamet) and pioglitazone (Actos) remain in use. In this study, we focused on two TZDs, troglitazone (Rezulin), which represents the first glitazone-type drug approved by the FDA, but is no longer in use due to its critical side effects, and rosiglitazone, which has been under the watchful eye of the FDA for some time due to suspicion of the potential development of cardiovascular diseases [7].

Troglitazone (TGZ) represents the first glitazone antidiabetic agent approved for clinical use. However, due to its hepatotoxicity, Rezulin was withdrawn from the market after a few years [8,9]. TGZ contains a chroman ring of the vitamin E moiety (Figure 1), which could provide TGZ with effective antioxidative properties, but in contrast, it has the potential to undergo metabolic activation, in which several reactive intermediates are produced. These intermediates are toxicologically active [9,10]. Overall, there are multiple potential mechanisms of TGZ hepatotoxicity [8,11] proposed, but to this day, there is no scientific evidence, regarding which is the correct one. TGZ remains a widely researched drug not only in connection with diabetes but also cancer [12,13].
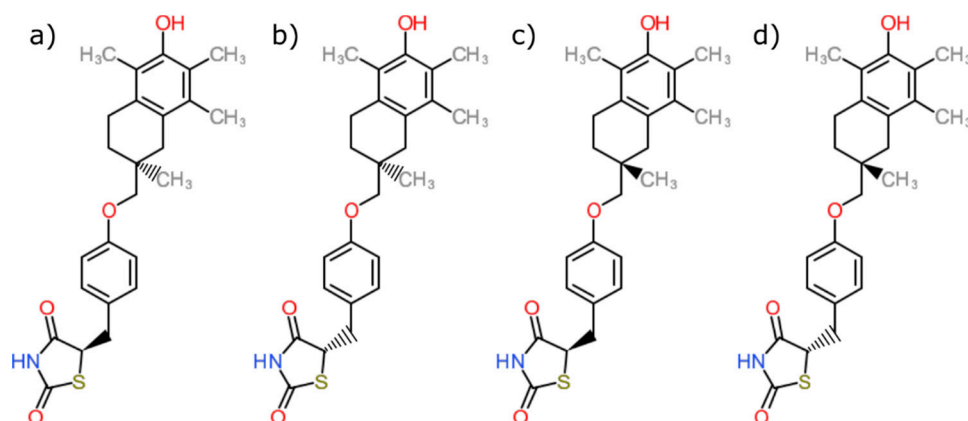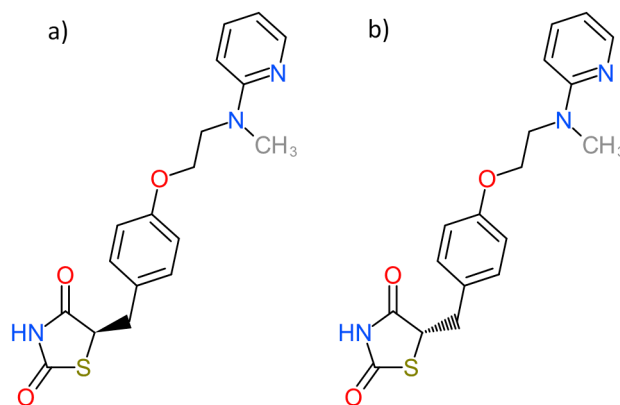

Figure 1. Chemical structures of troglitazone stereoisomers (**a**–**d**). The molecule possesses two chiral centers.

Rosiglitazone (RSG) was designed as an alternative to TGZ since there was no evident connection with hepatotoxicity (Figure 2) [14]. Later, the concern about possible cardiovascular risks was exposed [15,16], and the FDA issued various restrictions on the prescription and application of RSG-containing medicines [17]. After a few years, the FDA lifted the restrictions, as the data collected showed no significant increase in the risk of cardiovascular diseases compared to other drugs [18,19]. The main role of rosiglitazone is to lower insulin resistance, but it also decreases plasma levels of LDL cholesterol, increases levels of HDL cholesterol, creates small changes in triglyceride levels, and decreases blood pressure. These RSG effects could even reduce the increased risk of cardiovascular diseases in patients with type 2 diabetes mellitus [20,21].

Inverse molecular docking represents a new concept in drug discovery, which was now used to identify new potential targets of two thiazolidinediones, RSG and TGZ. Our main goal was to relate these targets with diseases and to use these links to explain the molecular mechanisms of known side effects and, consequently, to anticipate yet unknown, but potential ones. The application of this protocol for identifying possible side effects of drugs in the initial steps of the drug design process has been already reported [22–24] and could increase the success rate of drug development as well as its safety [25]. We used an inverse molecular docking protocol coupled with the CANDOCK algorithm [26] and a database of predicted binding sites from the entire Protein Data Bank (PDB) [27] with the ProBiS web server [28]. The novelty of our computational protocol lies in defining the binding sites for a large fraction of the proteins in the PDB and using them for inverse molecular

docking [28–30]. To the best of our knowledge, this has not been achieved to such an extent ever before. The docking algorithm itself [26] is also new and takes into account aspects such as protein flexibility and protein-ligand scoring using knowledge-based generalized statistical potential functions.

Figure 2. Chemical structures of rosiglitazone enantiomers (**a**,**b**). The molecule possesses one chiral center and is present as a racemate in medicinal applications.

2. Materials and Methods

2.1. Inverse Molecular Docking

Our goal was to provide mechanistic insights into potential side effects of RSG and TGZ using the CANDOCK inverse molecular docking on more than 67,000 protein structures from various organisms including more than 15,000 human protein structures derived from the PDB. The binding sites for small molecules on the proteins were identified and prepared for inverse molecular docking using the method described in [31] and successfully used for the discovery of new potential targets of the polyphenol curcumin [32]. Focusing the docking search space on the binding sites significantly reduced the time and complexity of the inverse molecular docking. Here, we provide a brief description of the steps involved in preparing this database [31,32].

Step 1: For each protein chain (PDB/Chain identifier), the presumed biological structural form was constructed using the data in the header of the corresponding PDB file. Next, the *co*-crystallized ligands of the protein chain identified by HETATM records with >7 heavy atoms, with at least one atom less than 3 Å away from the respective protein chain, were considered. Binding sites were determined from each extracted ligand, i.e., protein atoms < 5 Å away from all ligand atoms. Finally, binding site surface files were generated that served as an input to the ProBiS algorithm in a subsequent step.

Step 2: The ~310,000 protein chains in the PDB were clustered using a sequence identify cutoff at 100%, resulting in 70,000 protein chain clusters. The binding sites obtained in Step 1 were assigned to the corresponding sequence clusters. Each binding site was assigned to a cluster containing the protein (identified by PDB/Chain identifier) from which it was extracted. Water molecules and cofactors within the binding pocket were removed prior to inverse molecular docking simulation.

Step 3: Each pair of binding sites was compared within each cluster with 100% sequence identity in an "all versus all" fashion using the ProBiS algorithm [28,29]. The ProBiS algorithm allows for the detection of structurally similar protein binding sites as well as for the local pairwise alignment of X-ray or NMR determined protein structures from the PDB [29].

Step 4: Then, the binding sites within each cluster with 100% sequence identity were then clustered. In this second clustering, the similar binding site pairs (with z-score ≥ 2.0) were assigned to the same cluster. This clustering of binding sites within the existing sequence identity clusters ensures that if a protein has two or more distinct binding sites, each can be assigned to a distinct binding site cluster.

Step 5: The representative binding site surfaces (approximately 35,000 surfaces) were selected along with the co-crystallized ligands that were transferred to the representative binding sites.

Step 6: Then, the pre-calculated database of binding sites and ligands was then processed for each existing protein structural chain in the PDB (~310,000) by comparing that particular chain to the binding site surfaces using the ProBiS algorithm (Step 5). The identified structural similarities allowed for the prediction of binding sites for protein chains (even non-representative ones) for which binding sites had not yet been identified. Binding sites were defined as the union of centroids, i.e., the total space occupied by multiple spheres following the contours of the binding site.

Step 7: Then, we then filtered the obtained database of binding sites of small molecule ligands (consisting of 67,053 binding sites from different organisms) for the human proteins, resulting in a total of 15,482 binding sites of small molecule ligands to human proteins, which were subsequently applied as input for the inverse molecular docking using the newly developed CANDOCK program [26].

The CANDOCK algorithm applies a hierarchical approach to reconstruct small molecules from the atomic grid using graph theory and generalized statistical potential functions. Therefore, the docking scores represent approximations of the relative binding free energies and have arbitrary units. It docks fragments of small molecules into binding sites of investigated proteins using knowledge-based scoring methods. CANDOCK finds the best-docked poses of fragments and applies a fast-maximum clique algorithm [33] to link them together. In reconstructing the molecule, the algorithm uses iterative dynamics for a better placement of the ligand in the binding pocket. A second step represents the minimization of the docking score, using different methods that model the flexibility of the ligand and amino acid residues in the binding site [26,30,32,34].

RSG and TGZ structures, prepared with chemical editor Avogadro [35], served as input to the algorithm. These structures were subsequently optimized with Gaussian 16 [36], an electronic structure program, in conjunction with the *Hartree-Fock* method and *6-31G* basis set. Then, a variety of multi-organism proteins along with the coordinates of their binding sites, defined as centroids, was used as targets. In each target, we docked all enantiomers of RSG and TGZ and took into consideration poses with the lowest predicted knowledge-based score.

TGZ and RSG were successfully docked to more than 21,000 proteins from various organisms, of which slightly less than 5000 formed human protein targets. From the calculated conformations of all stereoisomers of RSG and TGZ we prepared two ranked lists for each molecule, one of multi-organism targets and the other of human targets. We selected the highest-ranking proteins from all four lists based on the 99.7% confidence interval criteria (Figure 3). As potential targets, we explicitly considered all proteins with docking scores below the selected confidence interval.

2.2. Binding Site Comparisons Using ProBiS Web Server

In the PDB, one finds a few known targets of troglitazone and rosiglitazone. At this stage, we searched for protein structures that have structurally similar binding sites to those where RSG or TGZ are already co-crystallized. To achieve this, we utilized the ProBiS web server [29] (http://probis.cmm.ki.si/, accessed on 28 January 2021).

We used all PDB protein structures containing TGZ or RSG as queries for the ProBiS web server. The structural similarity was determined with Z-score, the number of standard deviations from the mean data point. Z-score is used for the calculation of similarity between the selected and the query protein compared with a randomly selected protein. The higher the Z-score, the more similar the two proteins. We considered similarities at distances of 7.0 Å around the bound TGZ or RSG. We took only human protein structures with the highest similarity to each of the already known targets into consideration and selected them according to Z-score ($\geq$2.00).
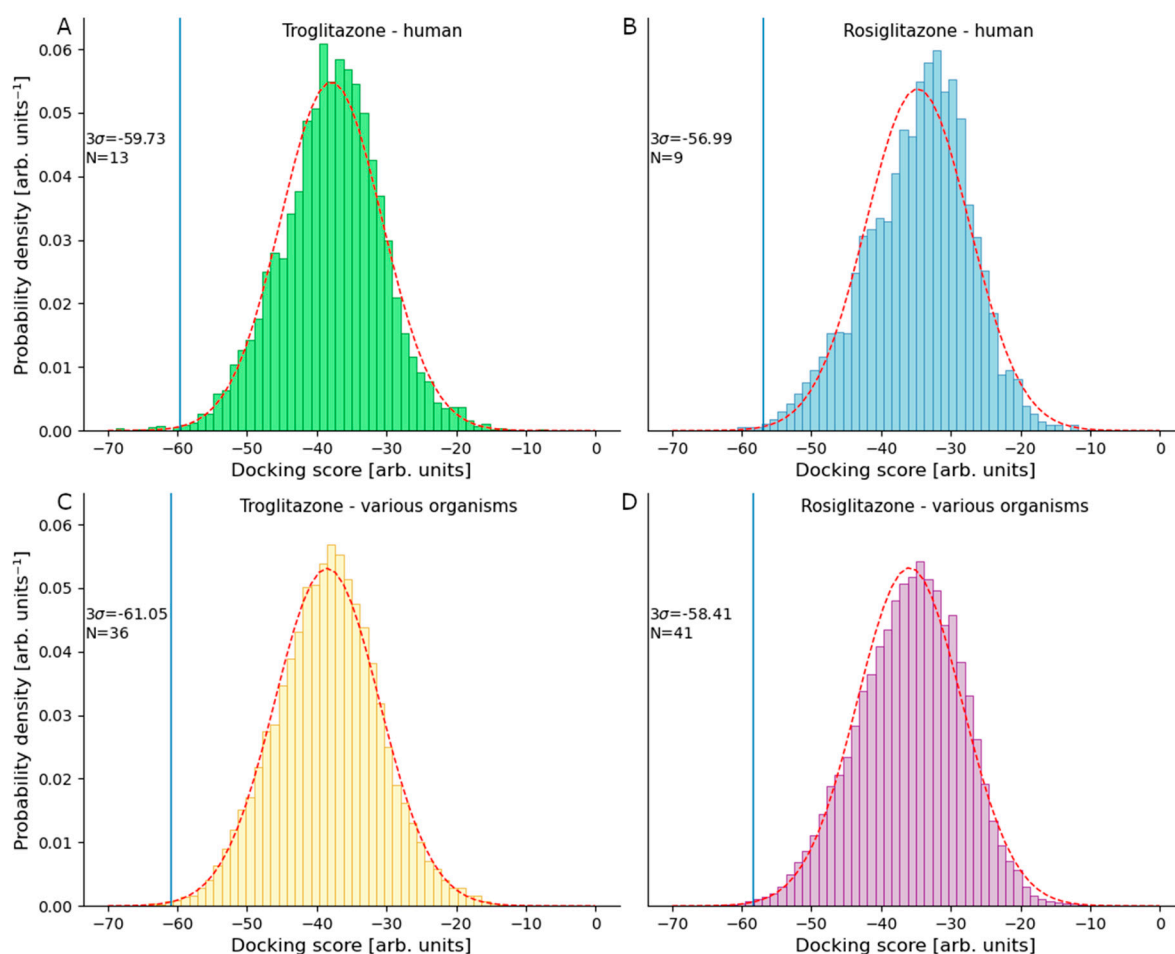
Figure 3. Normal distribution fitting graphs of inverse docking scores for (**A**) troglitazone docked into human targets, (**B**) rosiglitazone docked into human targets, (**C**) troglitazone docked into targets from various organisms and (**D**) rosiglitazone docked into targets from various organisms. 99.7% confidence interval is marked as 3σ and the number of proteins that fit the selected criteria is denoted with N.

2.3. Method Validation

To validate our method, we used a redocking procedure, receiver operating characteristics (ROC) curves [37], enrichment curves [38], and predictiveness curves (PC) [39]. We obtained 12 crystal protein structures from the PDB, which have a co-crystallized ligand TGZ or RSG bound in their binding sites. We removed the co-crystallized ligands but left the potential cofactors. Then we redocked TGZ or RSG to their respective protein structures. The validity of the method was determined with root-mean square deviation (RMSD) of atomic positions between co-crystallized and docked ligands [34]. Figures of TGZ and RSG in in their respective binding sites were prepared using UCSF Chimera [40]. The ROC metric plot shows a correlation between the true positive fraction (TPF) on the y-axis and the false positive fraction (FPF) on the x-axis. In our case, TPF represents experimentally confirmed protein targets from the ChEMBL database [41] for TGZ or RSG with the corresponding PDB entry, while FPF represents protein targets that successfully bound TGZ or RSG. The area under the ROC curve (ROC AUC) is used as a practical method to measure the overall performance. The larger the ROC AUC, the more effective the inverse docking is in distinguishing active from inactive compounds [37]. The enrichment curve represents the early detection quantification of target proteins from TPF [38]. PC also provides the early detection quantification of target proteins from TPF. In addition, the PC can be used to define a score threshold for potential targets from the inverse docking to be tested experimentally [39]. For quantifying early detection, the enrichment factor of 1% of screened

compounds (EF) [42], Boltzmann-enhanced discrimination of ROC (BEDROC) [38], and the robust initial enhancement (RIE) [42] are used. Applying PC, the standardized total gain (TG) can be calculated. The TG summarizes the contribution of the inverse docking protocol scores in explaining the activity of compounds over the entire protein dataset [39]. We used a web-based interactive application Screening explorer [43] that covers all aspects of the presented analysis.

3. Results

3.1. Targets of Troglitazone in the Human Proteome

For the analysis of potential disease treatments and potential side effects of TGZ, we selected the 13 highest-ranking human target proteins according to the 99.7% confidence interval. In our cut off, one could find some protein targets that were represented by two (PDB IDs: 3uzxA, 5fyyA) or even three (PDB ID: 4gs4A) PDB entries and had more than 95% sequence similarity. The only differences between these similar proteins were SNPs as well as initial and ending tails. Since our investigation was focused on finding potential protein targets regardless of small details, we took into consideration only the highest-ranking representative for each match. Table 1 lists the human proteins in which the predicted docking score of TGZ was the most favorable.

Table 1. Potential human target proteins of troglitazone.

PDB ID with Chain	Docking Score [Arbitrary Units [1]]	Protein Name	Protein Function and Connection with Diseases	References
3uzxA [2]	−68.2	3-oxo-5-beta-steroid 4-dehydrogenase	Catalyzes the reduction of the Δ4 double bond of bile acid intermediates and steroid hormones carrying the Δ4-3-one structure in the A/B cis configuration.	[44]
5fyyA [2]	−68.0	Lysine-specific demethylase 5B	A member of the KDM5 sub-family, acting as a histone 3 lysine 4 trimethyl demethylase, regulating proliferation, stem cell self-renewal, and differentiation. Connected with various forms of cancer.	[45,46]
2wtvA	−66.1	Aurora kinase A	Involved in the regulation of centrosomes and segregation of chromosomes. Associates with the centrosome and the spindle microtubules during mitosis and plays a critical role in several mitotic events. Known connections to various forms of cancer.	[47,48]
4gs4A [3]	−64.7	Alpha-tubulin *N*-acetyltransferase 1	Acetylates K40 on α-tubulin at the lumenal side of microtubules. Connections with neurological disorders, cancer, heart diseases.	[49,50]
1mncA	−64.0	Neutrophil collagenase	A member of the matrix metalloproteinase family. Degrades fibrillar type I, II, and III collagens. Implicated in several degenerative diseases with a slow matrix degradation rate, such as osteoporosis and Alzheimer's disease.	[51,52]

Table 1. *Cont.*

PDB ID with Chain	Docking Score [Arbitrary Units[1]]	Protein Name	Protein Function and Connection with Diseases	References
5hh1A	−63.5	*N*-alpha-acetyl transferase 60	Specifically catalyzes acetylation of the N-terminal α-amine in a majority of transmembrane proteins, mediates lysine acetylation of free histone H4. Possible connections with different forms of cancer.	[53]
3znnA	−62.7	D-amino-acid oxidase	Catalyzes the oxidative deamination of D-isomers of neutral and polar amino acids. The physiological role in the kidney and liver represents the detoxification of accumulated D-amino acids. Known connection with schizophrenia.	[54–56]
1c8tA	−62.3	Stromelysin-1	A member of the matrix metalloproteinase family. It degrades numerous ECM substrates, including collagens III, IV, V, IX, X, and XI, laminins, and elastin. Known connections with fibrosis, neovascularization, and potentially cancer progression.	[57,58]
3b96A	−60.3	Very long-chain specific acyl-CoA dehydrogenase, mitochondrial (VLCAD)	Catalyzes the initial, rate-limiting step of the mitochondrial fatty acid β-oxidation, exerts optimal chain length specificity for fatty acyl-CoAs having 16–24 carbons in length. Connected with hypoketotic hypoglycemia, liver damage, and hypertrophic myocardiopathy.	[59,60]

[1] Predicted knowledge-based docking scores with arbitrary units reflect relative binding free energies of a ligand to a protein. [2] The highest-ranking representative of a protein possessing two PDB entries in our list. [3] The highest-ranking representative of a protein possessing three PDB entries in our list.

The enzyme 3-oxo-5-beta-steroid 4-dehydrogenase controls the expression and activity of cytochrome P450 family. With its inhibition, the expression of P450s is reduced, which directly affects drug and toxin metabolisms in the human liver [61,62]. Lysine-specific demethylase 5B forms a potential tumor suppressor, and with its inhibition, this function could be lost, potentially leading to the development of cancer [46]. Aurora kinase A is known for its overexpression in a variety of cancer types, where its inhibition represents a potential treatment. However, in healthy cells, its inhibition could interfere with normal mitosis [47,48]. With binding of TGZ into alpha-tubulin *N*-acetyltransferase 1, the dysregulation of protein function could occur leading to abnormal levels of tubulin acetylation, which have been already linked to a variety of neurological disorders, heart diseases, and cancer [50]. The inhibition of neutrophil collagenase could successfully treat degenerative diseases with a slow matrix degradation rate. However, its inhibition in healthy cells could generate disorders of collagen degradation, causing a variety of diseases, such as liver cirrhosis. One could expect an analogous behavior for another protein from our set, Stromelysin-1 [52,58,63]. The binding of TGZ into *N*-alpha-acetyl transferase 60 could cause abnormal chromosome segregation, which triggers the development of cancer [64]. As D-amino-acid oxidase plays a detoxification role in the liver, the toxicity of the accumulated D-amino-acids would increase in the case of its inhibition [56]. The VLCAD protein is implicated in the oxidation of fatty acids, which are used to synthesize ketone bodies in the liver, the so-called fuel for the brain and muscle. Inhibition of VLCAD could lead to liver damage, other serious problems, and even death [60]. Its inhibition in the neonate or toddler groups could cause a variety of diseases, ranging from muscle weakness, over developmental regression and behavioral disorders, until attention deficit disorder [65].

Since molecular docking was performed separately for all stereoisomers of TGZ, the highest ranked human targets according to the 99.7% confidence interval of each isomer are presented in the Supplementary Material, Tables S1–S4. As can be seen from these tables, the diversity of protein targets is isomer specific, but there are a few common targets, such as 3-oxo-5-beta-steroid 4-dehydrogenase, lysine-specific demethylase 5B, and aurora kinase A. A table of the highest ranked protein targets of TGZ from different organisms is also shown in the Supplementary Material (Table S7).

3.2. Targets of Rosiglitazone in the Human Proteome

Nine of the highest-ranking human target proteins of RSG were selected (based on the 99.7% confidence interval), and potential diseases treatments and potential side effects were analyzed. Table 2 lists the nine human protein targets in which the predicted docking score of RSG was the most favorable.

Table 2. Potential human target proteins of rosiglitazone.

PDB ID with Chain	Docking score [Arbitrary Units [1]]	Protein Name	Protein Function and Connection with Diseases	References
4jijA	−60.5	Matrix metalloproteinase 9	Degrades components of the extracellular matrix, mostly collagens. Known connections with cancer, neurological disorders, and cardiovascular diseases.	[66–69]
4lk3A	−59.9	UDP-glucuronic acid decarboxylase 1	Catalyzes the formation of UDP-xylose from UDP-glucuronate. No known connections with human diseases.	[70,71]
3f9zA	−59.8	*N*-lysine methyltransferase KMT5A	Monomethylates lysine 20 of histone H4 (H4K20), proliferating cell nuclear antigen (PCNA) and p53. Potential connections with cancer tumorigenesis and neurodegenerative diseases.	[72,73]
4u7pA	−58.8	DNA (cytosine-5)-methyltransferase 3A	Responsible for the de novo methylation of 5-methylcytosine during embryogenesis for the establishment of the somatic methylation pattern, which is crucial for embryonic development. Known connections with hematological malignancies, that is tumors of the hematopoietic and lymphoid tissues, as well as with defects in brain development or neurological defects.	[74–77]
4fvlA	−58.6	Collagenase 3	One of the matrix metalloproteinases. Plays a key role in tissue remodeling and repair. Known connections with cancer, aneurysm, fibrosis, and other diseases.	[78–80]
1fduA	−58.4	Estradiol 17-beta-dehydrogenase 1	Catalyzes the reversible transformation of estrone into biologically active estradiol. Connected with cancer, the biology of reproduction, and neuronal diseases.	[81,82]

Table 2. *Cont.*

PDB ID with Chain	Docking score [Arbitrary Units [1]]	Protein Name	Protein Function and Connection with Diseases	References
5tdhA	−58.0	Guanine nucleotide-binding protein G(i) subunit alpha-1	Transduces intracellular signals from membrane-bound receptors to downstream effector molecules. Cooperates in different biological activities, such as vision or synaptic nerve signal transmissions. Connected with neurological problems, cardiovascular defects, and certain forms of cancer.	[83–85]
1i3kA	−57.6	UDP-glucose 4-epimerase	Catalyzes the interconversion of UDP-galactose and UDP-glucose during normal galactose metabolism, as well as interconversion of UDP-N-acetyl galactosamine (UDP-galNAc) and UDP-N-acetylglucosamine (UDP-glcNAc). Loss of this protein results in the spectrum disorder epimerase deficiency galactosemia.	[86–88]
2pxxA	−57.6	EEF1A lysine methyltransferase 4	Catalyzes methylations on Lys36 in eukaryotic translation elongation factor 1 alpha. No known connections with human diseases.	[89]

[1] Predicted knowledge-based docking scores with arbitrary units reflect relative binding free energies of a ligand to a protein.

RSG possesses a known connection only with the first protein target from our list, matrix metalloproteinase 9 [90], which is essential for collagen degradation. Disorders of collagen degradation cause a variety of diseases, such as arthritis and liver cirrhosis [63]. One could expect similar side effects for collagenase 3 from our list. UDP-glucuronic acid decarboxylase 1 catalyzes the formation of UDP-xylose, and its inhibition could reduce concentrations of xylose and consequently increase expressions of the proteins involved in the formation of hyaluronan, which could led to the development of cancer [91]. RSG could also be a new inhibitor of KMT5A protein, whose inhibition could lead to the dysregulation of protein methylation, which is directly connected with neurodegenerative diseases and cancer [73]. By a selective inhibition of DNA (cytosine-5)-methyltransferase 3A (DNMT3A), one could regulate DNA methylation. However, in the case of absence or errors in DNA methylation, several disorders occur, such as certain cancer types, defects in embryonic and brain development, or neurological diseases [77]. Estradiol is known for its anti-inflammatory, neuroprotective, and neuroregenerative actions in neurological disorders such as Alzheimer's disease, Parkinson's disease, multiple sclerosis, and cerebrovascular stroke [92]. The inhibition of estradiol 17-beta-dehydrogenase 1 by RSG could lead to a reduced production of estradiol thereby decreasing its anti-inflammatory actions. The dysregulation of guanine nucleotide-binding protein by RSG could interfere with the production of cAMP. Decreased cAMP concentrations could lead to hepatocellular carcinoma while sustained elevation results in cell death [85,93]. Inhibition of UDP-glucose 4-epimerase could result in a reduced concentration of D-galactose, which is important for the maintenance of cell surface polysaccharide patterns [94], and EEF1A lysine methyltransferase 4 inhibition could cause alterations in the translation of several mRNAs.

Since molecular docking was performed separately for both stereoisomers of RSG, the highest ranked human targets according to the 99.7% confidence interval of each isomer are shown in Supplementary Material, Tables S5 and S6. As can be observed from these tables, most of the protein targets differ between isomers. A table of the highest ranked protein

targets of RSG from different organisms is also shown in the Supplementary Material (Table S8).

3.3. Binding Site Comparison Using ProBiS Web Server

3.3.1. Troglitazone

In the PDB, one finds a single human protein structure with co-crystallized troglitazone, Cytochrome P450 2C8 (PDB ID: 2vn0, chain A). We compared the binding site of troglitazone from this known structure against the entire PDB using the ProBiS web server and obtained a list of human proteins with structurally similar (Z-score larger than 2) binding sites to the crystallized target (Table 3).

Table 3. Cytochrome P450 2C8 structure with co-crystallized troglitazone as query protein and the identified human proteins with structurally similar binding sites.

Query Protein	Proteins with Similar Binding Sites		
PDB ID with Chain	PDB ID with Chain	Protein	Z-Score
2vn0A	1r9oA	Cytochrome P450 2C9	3.83
	4gqsA	Cytochrome P450 2C19	3.82
	3tbgA	Cytochrome P450 2D6	3.21
	3ibdA	Cytochrome P450 2B6	3.13
	3czhA	Cytochrome P450 2R1	3.11
	4nkyA	Steroid 17-Alpha-Hydroxylase/17,20 Lyase	3.00
	4y8wA	Cytochrome P450 21-Hydroxylase	2.93
	3e6iA	Cytochrome P450 2E1	2.54
	3ld6A	Lanosterol 14-Alpha Demethylase	2.06

All proteins found with the ProBiS web server belong to the same Cytochrome P450 family. All possess very similar 3D structures, but different sequences (from 30–70% sequence similarity with the notable exception of lanosterol 14-alpha demethylase: 3ld6A).

3.3.2. Rosiglitazone

In the PDB, one finds 11 human protein structures with co-crystallized rosiglitazone. All PDB entries represent the same peroxisome proliferator-activated receptor gamma (PPARγ) protein. PPARγ is a member of the hormone receptor family, but its role in the heart is not well understood, besides the fact that its overexpression in the heart tissue can lead to cardiac dysfunction [95]. We compared these structures to the entire PDB, using the ProBiS web server and obtained the list of human proteins with structurally similar (Z-score larger than 2) binding sites to query proteins with co-crystallized rosiglitazone (Table 4).

All query structures with bound rosiglitazone are within 95–100% sequence similarity, with the notable exception of 3dzyD, which has around 50% sequence similarity with the others because of its longer chain. However, the binding site possesses an almost identical sequence in all query proteins, 3dzy included.

As one can see from Table 4, the identified proteins are virtually the same for all known binders, with the addition of two notable exceptions, Oxysterols receptor LXR-alpha (PDB ID: 3ipq) and MIT domain-containing protein 1 (PDB ID: 4a5x).

3.3.3. Method Validation

Redocking procedure. We obtained a structure with co-crystallized troglitazone from PDB (PDB ID: 2vn0, chain A), which we used in the redocking procedure. We took into the consideration three highest-scoring poses for all four stereoisomers of troglitazone (Table 5). We also obtained 11 human protein structures with co-crystallized rosiglitazone. For each structure, we considered three highest-scoring poses for each enantiomer of rosiglitazone (Table 6). For each pose, we calculated the RMSD between the docked and the native positions from the PDB. We reviewed all considered poses and selected the best ones

(denoted in green) based on a combination of RMSD and correct chirality according to the native poses of co-crystallized rosiglitazone or troglitazone. RMSD values are a little higher than expected [96], but that could well be a direct consequence of molecular size and its chirality. In Figure 4 comparison of the poses of native (blue) and docked (brown) troglitazone in cytochrome P450 2C8 is depicted (PDB ID: 2vn0, chain A). Figure 5 presents a comparison of the native (blue) and docked (brown) rosiglitazone poses in peroxisome proliferator-activated receptor gamma protein (PDB ID: 2prg, chain A).

Table 4. Peroxisome proliferator-activated receptor gamma structures with co-crystallized rosiglitazone as query proteins and the identified human proteins with structurally similar binding sites.

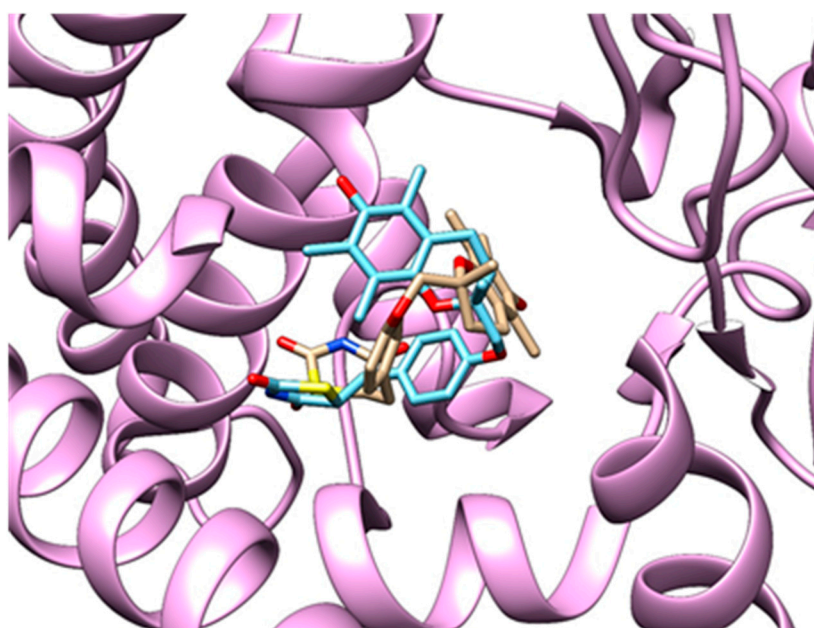| Query Proteins	Proteins with Similar Binding Sites		
PDB ID with Chain	PDB ID with Chain	Protein	Z-Score
1fm6DX	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	3.84
	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	3.77
	3ipqA	Oxysterols Receptor LXR-Alpha	2.04
1zgyA	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	3.78
	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	3.66
2prgAB	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	3.60
	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	3.59
3cs8A	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	3.30
	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	2.98
3dzyD	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	3.25
	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	3.05
4emaA	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	3.73
	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	3.14
4o8fAB	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	3.63
	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	3.45
4xldA	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	3.85
	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	3.44
5ji0D	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	3.94
	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	3.86
5ycpA	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	3.15
	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	2.73
	4a5xA	MIT Domain-Containing Protein 1	2.04
6md4A	3tkmA	Peroxisome Proliferator-Activated Receptor Delta	3.76
	3vi8A	Peroxisome Proliferator-Activated Receptor Alpha	3.36

Table 5. Results of the redocking procedure for troglitazone.

| PDB ID with Chain (Crystal Structure)	Stereoisomer (Figure 1)	Docking Score [Arbitrary Units]	RMSD [Å]
2vn0A	a	−24.77	4.51
	b	−14.1	5.45
	c	−26.5	5.84
	d	−25.1	6.12

Validation with ROC, EF, and PC. We performed inverse molecular docking using the CANDOCK algorithm on a set of 15,482 human proteins, including 128 confirmed target proteins of TGZ (and 133 of RSG) whose measured IC_{50} values for binding TGZ (or RSG) were <10 μM [41]. The inverse docking was successfully performed on 69 (TGZ) and 74 (RSG) of those confirmed targets. The ability of CANDOCK to distinguish the confirmed protein targets of TGZ and RSG from non-target proteins was assessed using established metrics, which are shown in Figure 6 (TGZ) and in Figure 7 (RSG).

**Table 6.** Results of the redocking procedure for rosiglitazone.

PDB ID with Chain (Crystal Structure)	Stereoisomer (Figure 2)	Docking Score [Arbitrary Units]	RMSD [Å]
1fm6D	a	−30.6	2.15
	b	−32.1	2.25
1zgyA	a	−32.9	2.26
	b	−38.2	2.47
2prgA	a	−37.1	0.85
	b	−46.2	2.70
3cs8A	a	−29.0	2.89
	b	−28.2	8.07
3dzyD	a	−41.4	2.22
	b	−32.0	1.09
4emaA	a	−43.6	1.96
	b	−21.0	3.20
4o8fA	a	−32.2	1.12
	b	−23.8	2.03
4xldA	a	−39.3	1.78
	b	−35.8	2.62
5ji0D	a	−37.2	1.44
	b	−39.7	2.14
5ycpA	a	−34.3	1.76
	b	−44.9	1.90
6md4A	a	−31.8	2.06
	b	−37.6	2.62

Figure 4. Native (blue) and redocked (brown) poses of troglitazone in the cytochrome P450 2C8 protein (light purple). RMSD: 4.51 Å.

Our protocol was successful in discriminating between true targets of TGZ (ROC AUC of 0.631) and RSG (ROC AUC of 0.606). Early detection of protein targets with BEDROC of 0.110 and 0.078, RIE of 2.030 and 1.431, and EF1% of 1.2095 and 1.1641, for TGZ and RSG, respectively, was satisfactory. The protocol produced score variations in the detection of true target proteins (TG 0.236 and 0.189 for TGZ and RSG, respectively), which, in combination with the ROC AUC above 0.6, demonstrates that the protocol is expected to provide a satisfactory agreement with experiments [39].
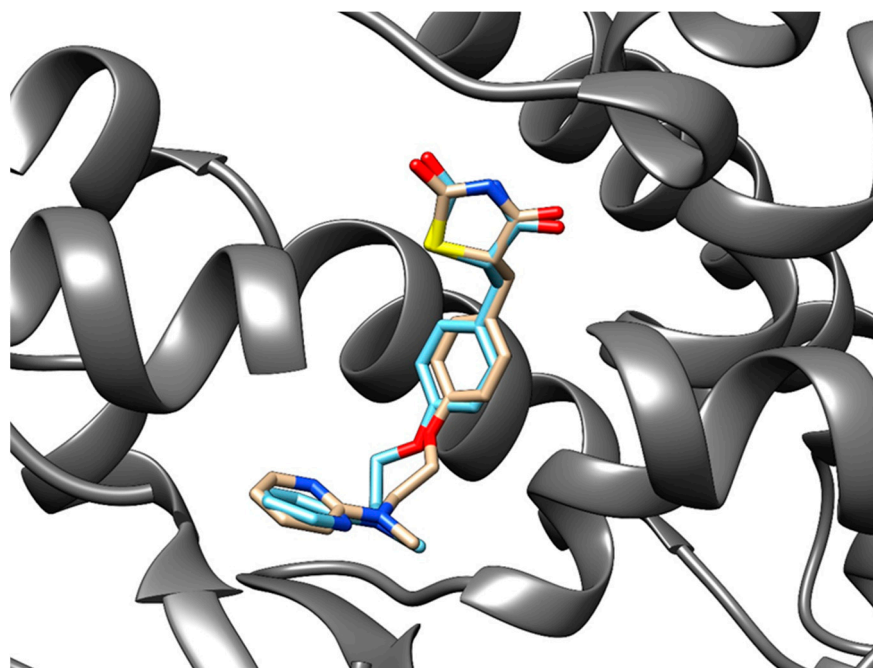
Figure 5. Native (blue) and redocked (brown) poses of rosiglitazone in the peroxisome proliferator-activated receptor gamma protein (gray). RMSD: 0.85 Å.
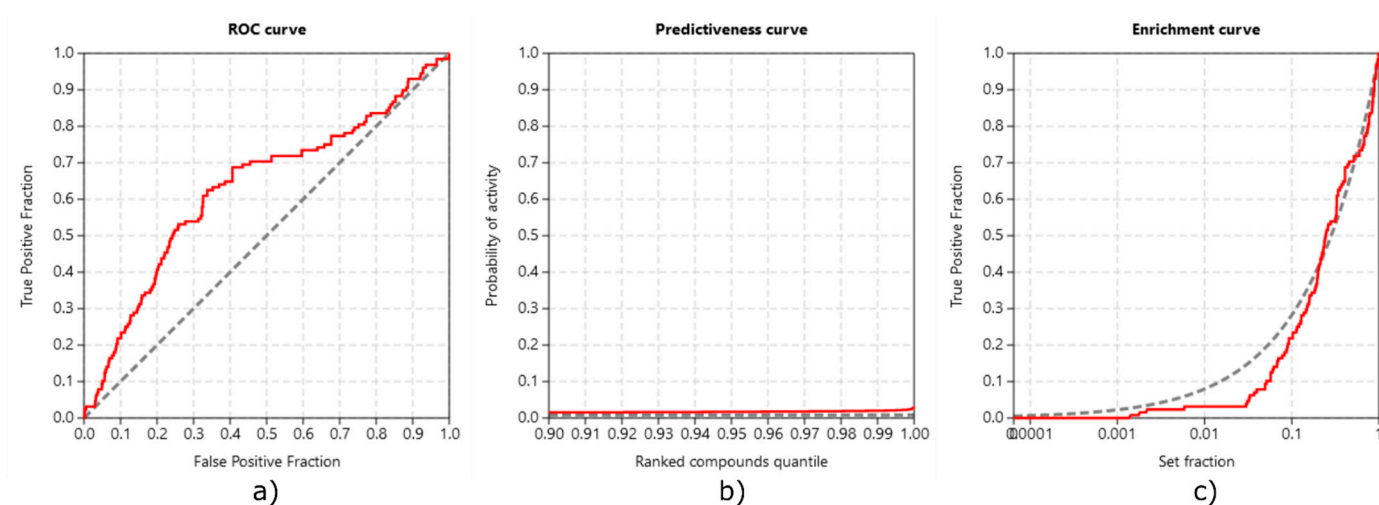
Figure 6. Validations of inverse molecular docking protocol of troglitazone against all human protein targets from the PDB: (**a**) the receiver operating characteristics (ROC) curve; (**b**) the predictiveness curve; and (**c**) the enrichment curve.

Our inverse docking protocol was previously extensively validated by Fine and Konc et al. [26], Furlan et al. [32], and Kores et al. [34].
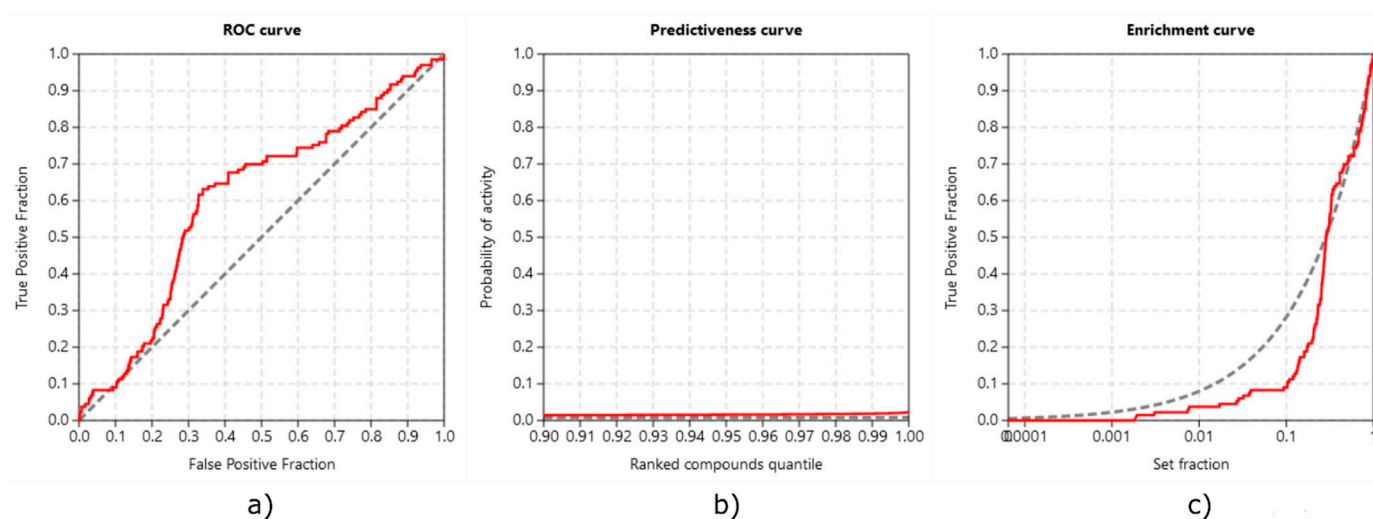
Figure 7. Validations of inverse molecular docking protocol of rosiglitazone against all human protein targets from the Protein Date Bank (PDB): (**a**) the receiver operating characteristics (ROC) curve; (**b**) the predictiveness curve; and (**c**) the enrichment curve.

4. Discussion

Predicting potential side effects of new drugs represents one of the key stages in drug development [22]. This is also important for existing drugs that have been on the market for a long time. Here, it is important to explain the molecular mechanisms behind why the already known side effects occur, and thus potentially predict new ones that may still occur or have not yet been associated with this drug. The mechanistic understanding may even afford directions on how to alleviate these side effects.

TGZ and RSG are known agonists of human PPARγ. In our set of targets where both drugs were successfully docked, there are four structures of the said protein (PDB IDs: 3dzuD, 4r6sA, 4oj4A, 3u9qA). The docking scores for TGZ ranged from −43 to −58, while the scores for RSG ranged from −37 to −45. These results indicate that the docking scores of TGZ and RSG docked to the PPARγ are higher and thus less favorable than our proposed targets within the 99.7% confidence interval, implying that these new potential targets may indeed contribute to the understanding of the side effects of both drugs. In addition, the redocking scores of RSG are similar to those from inverse docking for human PPARγ.

Troglitazone was withdrawn from use due to potential liver defects [8]. In our computational study, troglitazone indeed successfully docked into certain protein structures, which have direct or indirect impacts on the liver [52,60,61,63]. The highest-ranking targets included 3-oxo-5-beta-steroid 4-dehydrogenase [44], neutrophil collagenase [51], stromelysin-1 [57], and VLCAD [59], which could all be involved in the development of hepatotoxicity. Troglitazone also successfully docked into certain protein structures such as lysine-specific demethylase 5B [45], aurora kinase A [47], and alpha-tubulin *N*-acetyltransferase 1 [50], which are connected to cancer. As the latest findings show, troglitazone could indeed be used to treat various cancer types [97,98], but our results indicate that troglitazone could also trigger carcinogenesis [46,64]. None of our highest-ranking targets has known connections with troglitazone. Binding site comparison using ProBiS webserver returned targets with similar binding sites, which all belonged to the same cytochrome P450 family. Enzymes from this family are very important in drug metabolism, and in the case of troglitazone, because of its higher dosages, the covalent bonding to different CYP enzymes could indeed result in an increased risk of hepatotoxicity [99].

Rosiglitazone became accepted as an alternative to troglitazone, since there were no indications that it causes hepatotoxicity [14]. Rosiglitazone is known to have an experimental connection with the matrix metalloproteinase family (MMPs), and two of its members are indeed among our top targets. MMPs play an important role in tissue remodeling and repair. So, any irregularity could lead to various diseases, such as cancer, neurological

disorders, or cardiovascular diseases [63,69,79]. This could also represent the explanation behind the possible cardiovascular risk, which caused multiple restrictions on the application of rosiglitazone-containing drugs [15,17]. Dysregulation or inhibition of certain proteins with rosiglitazone could potentially lead to cancer [77,91,93] and neurodegenerative diseases [73,77,92] as well. Binding site comparison was performed on 11 PDB entries representing the same PPARγ protein, which is known as the main target of glitazones [6]. The results discern the potential binding of rosiglitazone to two additional proteins from the PPAR family. Therefore, we could talk about PPAR co-agonists. PPAR co-agonists were indeed tested in the past with several promising leads, but there is very little known about the possible side effects of these kinds of drugs [6,100].

The main limitation of the inverse docking approach is that not all human proteins are available in the PDB yet, as this is the requirement of our protocol. However, proteome-wide structural coverage is already available for most pharmaceutically relevant protein classes, especially for humans [101]. We strongly believe that the application of our approach can provide important results to improve not only the drug design and discovery but also the management of potential side effects in most known human diseases and disorders.

5. Conclusions

In this paper, we focused on two drugs from the thiazolidinedione family, troglitazone, which was the first of its type but was withdrawn from use due to its severe side effects, and rosiglitazone, which was under FDA evaluation for a long time as a potential trigger of cardiovascular diseases [7]. We used the in-house developed inverse docking algorithm CANDOCK [26]. Structures of both drugs were prepared with Avogadro [35] and optimized with Gaussian [36], and the database of protein structures was prepared from the online Protein Data Bank [102]. In addition, a binding site comparison with ProBiS web server [29] was performed.

Our results successfully explained the side effects of troglitazone and rosiglitazone, and we also predicted additional ones, that could occur after the long-term application of these two drugs. With further development of computational methods and in conjunction with experimental research, these results can be further analyzed to better explain the side effects of both drugs and to help mitigate or even eliminate them. A detailed study of the individual isomers on the protein targets could also help to develop TGZ and RSG drugs with only those stereoisomers that have a beneficial effect on humans and do not show negative side effects. On the other hand, the novel inverse molecular docking protocol could be generally applied to discern possible cross-interactions of drug candidates, potentially leading to adverse side effects.

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

References

1. Lebovitz, H.E. Thiazolidinediones: The Forgotten Diabetes Medications. *Curr. Diab. Rep.* **2019**, *19*, 151. [CrossRef] [PubMed]
2. Schoonjans, K.; Auwerx, J. Thiazolidinediones: An Update. *Lancet* **2000**, *355*, 1008–1010. [CrossRef]
3. DeFronzo, R.A. Pharmacologic Therapy for Type 2 Diabetes Mellitus. *Ann. Intern. Med.* **1999**, *131*, 281–303. [CrossRef] [PubMed]
4. Martín-Timón, I.; Sevillano-Collantes, C.; Segura-Galindo, A.; del Cañizo-Gómez, F.J. Type 2 Diabetes and Cardiovascular Disease: Have All Risk Factors the Same Strength? *World J. Diabetes* **2014**, *5*, 444–470. [CrossRef] [PubMed]
5. Lehmann, J.M.; Moore, L.B.; Smith-Oliver, T.A.; Wilkison, W.O.; Willson, T.M.; Kliewer, S.A. An Antidiabetic Thiazolidinedione Is a High Affinity Ligand for Peroxisome Proliferator-Activated Receptor γ (PPARγ). *J. Biol. Chem.* **1995**, *270*, 12953–12956. [CrossRef] [PubMed]
6. Fanny, L. Staels Bart Fibrates, Glitazones, and Peroxisome Proliferator—Activated Receptors. *Arter. Thromb. Vasc. Biol.* **2010**, *30*, 894–899. [CrossRef]
7. Nanjan, M.J.; Mohammed, M.; Prashantha Kumar, B.R.; Chandrasekar, M.J.N. Thiazolidinediones as Antidiabetic Agents: A Critical Review. *Bioorganic Chem.* **2018**, *77*, 548–567. [CrossRef] [PubMed]
8. Smith, M.T. Mechanisms of Troglitazone Hepatotoxicity. *Chem. Res. Toxicol.* **2003**, *16*, 679–687. [CrossRef]
9. Yokoi, T. Troglitazone. In *Adverse Drug Reactions*; Uetrecht, J., Ed.; Handbook of Experimental Pharmacology; Springer: Berlin/Heidelberg, Germany, 2010; pp. 419–435, ISBN 978-3-642-00663-0.
10. Ogimura, E.; Nakagawa, T.; Deguchi, J.; Sekine, S.; Ito, K.; Bando, K. Troglitazone Inhibits Bile Acid Amidation: A Possible Risk Factor for Liver Injury. *Toxicol. Sci.* **2017**, *158*, 347–355. [CrossRef] [PubMed]
11. Yu, L.; Li, H.; Zhang, C.; Zhang, Q.; Guo, J.; Li, J.; Yuan, H.; Li, L.; Carmichael, P.; Peng, S. Integrating in Vitro Testing and Physiologically-Based Pharmacokinetic (PBPK) Modelling for Chemical Liver Toxicity Assessment—A Case Study of Troglitazone. *Environ. Toxicol. Pharm.* **2020**, *74*, 103296. [CrossRef]
12. Przybelski, S.; Ulloa, M.E.; Rudie, M.; Cornett, M.; Strom, D. Characterization of the Troglitazone Induced Apoptosis in a Human Colon Cancer Cell Line. *FASEB J.* **2020**, *34*, 1. [CrossRef]
13. Jung, K.-H.; Lee, J.H.; Park, J.-W.; Moon, S.-H.; Cho, Y.S.; Lee, K.-H. Troglitazone Exerts Metabolic and Antitumor Effects on T47D Breast Cancer Cells by Suppressing Mitochondrial Pyruvate Availability. *Oncol. Rep.* **2020**, *43*, 711–717. [CrossRef] [PubMed]
14. Nissen, S.E. The Rise and Fall of Rosiglitazone. *Eur. Heart J.* **2010**, *31*, 773–776. [CrossRef]
15. Home, P.D.; Pocock, S.J.; Beck-Nielsen, H.; Gomis, R.; Hanefeld, M.; Jones, N.P.; Komajda, M.; McMurray, J.J.V. Rosiglitazone Evaluated for Cardiovascular Outcomes—An Interim Analysis. *N. Engl. J. Med.* **2007**, *357*, 28–38. [CrossRef] [PubMed]
16. Nissen, S.E.; Wolski, K. Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes. *N. Engl. J. Med.* **2007**, *356*, 2457–2471. [CrossRef]
17. Research, C.D.E.; FDA. *Drug Safety Communication: Updated Risk Evaluation and Mitigation Strategy (REMS) to Restrict Access to Rosiglitazone-Containing Medicines Including Avandia, Avandamet, and Avandaryl*; FDA: Silver Spring, MD, USA, 2019.
18. Research, C.D.E.; FDA. *Drug Safety Communication: FDA Requires Removal of Some Prescribing and Dispensing Restrictions for Rosiglitazone-Containing Diabetes Medicines*; FDA: Silver Spring, MD, USA, 2019.
19. Research, C.D.E.; FDA. *Drug Safety Communication: FDA Eliminates the Risk Evaluation and Mitigation Strategy (REMS) for Rosiglitazone-Containing Diabetes Medicines*; FDA: Silver Spring, MD, USA, 2019.
20. Wagstaff, A.J.; Goa, K.L. Rosiglitazone. *Drugs* **2002**, *62*, 1805–1837. [CrossRef]
21. Wang, C.-H.; Weisel, R.D.; Liu, P.P.; Fedak, P.W.M. Verma Subodh Glitazones and Heart Failure. *Circulation* **2003**, *107*, 1350–1354. [CrossRef]
22. Chen, Y.Z.; Ung, C.Y. Prediction of Potential Toxicity and Side Effect Protein Targets of a Small Molecule by a Ligand–Protein Inverse Docking Approach. *J. Mol. Graph. Model.* **2001**, *20*, 199–218. [CrossRef]
23. Ma, C.; Kang, H.; Liu, Q.; Zhu, R.; Cao, Z. Insight into Potential Toxicity Mechanisms of Melamine: An in Silico Study. *Toxicology* **2011**, *283*, 96–100. [CrossRef]
24. Vallone, A.; D'Alessandro, S.; Brogi, S.; Brindisi, M.; Chemi, G.; Alfano, G.; Lamponi, S.; Lee, S.G.; Jez, J.M.; Koolen, K.J.M.; et al. Antimalarial Agents against Both Sexual and Asexual Parasites Stages: Structure-Activity Relationships and Biological Studies of the Malaria Box Compound 1-[5-(4-Bromo-2-Chlorophenyl)Furan-2-Yl]-N-[(Piperidin-4-Yl)Methyl]Methanamine (MMV019918) and Analogues. *Eur. J. Med. Chem.* **2018**, *150*, 698–718. [CrossRef] [PubMed]
25. Xu, X.; Huang, M.; Zou, X. Docking-Based Inverse Virtual Screening: Methods, Applications, and Challenges. *Biophys. Rep.* **2018**, *4*, 1–16. [CrossRef] [PubMed]

26. Fine, J.; Konc, J.; Samudrala, R.; Chopra, G. CANDOCK: Chemical Atomic Network-Based Hierarchical Flexible Docking Algorithm Using Generalized Statistical Potentials. *J. Chem. Inf. Model.* **2020**, *60*, 1509–1527. [CrossRef]

27. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]

28. Konc, J.; Janežič, D. ProBiS Algorithm for Detection of Structurally Similar Protein Binding Sites by Local Structural Alignment. *Bioinformatics* **2010**, *26*, 1160–1168. [CrossRef] [PubMed]

29. Konc, J.; Janežič, D. ProBiS-2012: Web Server and Web Services for Detection of Structurally Similar Binding Sites in Proteins. *Nucleic Acids Res.* **2012**, *40*, W214–W221. [CrossRef] [PubMed]

30. Konc, J.; Janežič, D. ProBiS-Ligands: A Web Server for Prediction of Ligands by Examination of Protein Binding Sites. *Nucleic Acids Res.* **2014**, *42*, W215–W220. [CrossRef]

31. Štular, T.; Lešnik, S.; Rožman, K.; Schink, J.; Zdouc, M.; Ghysels, A.; Liu, F.; Aldrich, C.C.; Haupt, V.J.; Salentin, S.; et al. Discovery of Mycobacterium Tuberculosis InhA Inhibitors by Binding Sites Comparison and Ligands Prediction. *J. Med. Chem.* **2016**, *59*, 11069–11078. [CrossRef] [PubMed]

32. Furlan, V.; Konc, J.; Bren, U. Inverse Molecular Docking as a Novel Approach to Study Anticarcinogenic and Anti-Neuroinflammatory Effects of Curcumin. *Molecules* **2018**, *23*, 3351. [CrossRef] [PubMed]

33. Konc, J.; Janežič, D. A Branch and Bound Algorithm for Matching Protein Structures. In *Proceedings of the Adaptive and Natural Computing Algorithms*; Beliczynski, B., Dzielinski, A., Iwanowski, M., Ribeiro, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 399–406.

34. Kores, K.; Lešnik, S.; Bren, U.; Janežič, D.; Konc, J. Discovery of Novel Potential Human Targets of Resveratrol by Inverse Molecular Docking. *J. Chem. Inf. Model.* **2019**, *59*, 2467–2478. [CrossRef] [PubMed]

35. Hanwell, M.D.; Curtis, D.E.; Lonie, D.C.; Vandermeersch, T.; Zurek, E.; Hutchison, G.R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminformatics* **2012**, *4*, 17. [CrossRef]

36. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 16 Rev. A.03*; Gaussian, Inc.: Wallingford, UK, 2016.

37. Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. Virtual Screening Workflow Development Guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *J. Med. Chem.* **2005**, *48*, 2534–2547. [CrossRef]

38. Truchon, J.-F.; Bayly, C.I. Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. *J. Chem. Inf. Model.* **2007**, *47*, 488–508. [CrossRef] [PubMed]

39. Empereur-mot, C.; Guillemain, H.; Latouche, A.; Zagury, J.-F.; Viallon, V.; Montes, M. Predictiveness Curves in Virtual Screening. *J. Cheminformatics* **2015**, *7*, 52. [CrossRef] [PubMed]

40. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef] [PubMed]

41. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrián-Uhalte, E.; et al. The ChEMBL Database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954. [CrossRef] [PubMed]

42. Sheridan, R.P.; Singh, S.B.; Fluder, E.M.; Kearsley, S.K. Protocols for Bridging the Peptide to Nonpeptide Gap in Topological Similarity Searches. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1395–1406. [CrossRef] [PubMed]

43. Empereur-Mot, C.; Zagury, J.-F.; Montes, M. Screening Explorer–An Interactive Tool for the Analysis of Screening Results. *J. Chem. Inf. Model.* **2016**, *56*, 2281–2286. [CrossRef] [PubMed]

44. Kondo, K.H.; Kai, M.H.; Setoguchi, Y.; Eggertsen, G.; Sjöblom, P.; Setoguchi, T.; Okuda, K.I.; Björkhem, I. Cloning and Expression of CDNA of Human Delta 4-3-Oxosteroid 5 Beta-Reductase and Substrate Specificity of the Expressed Enzyme. *Eur. J. Biochem.* **1994**, *219*, 357–363. [CrossRef] [PubMed]

45. Tumber, A.; Nuzzi, A.; Hookway, E.S.; Hatch, S.B.; Velupillai, S.; Johansson, C.; Kawamura, A.; Savitsky, P.; Yapp, C.; Szykowska, A.; et al. Potent and Selective KDM5 Inhibitor Stops Cellular Demethylation of H3K4me3 at Transcription Start Sites and Proliferation of MM1S Myeloma Cells. *Cell Chem. Biol.* **2017**, *24*, 371–380. [CrossRef]

46. Roesch, A.; Mueller, A.M.; Stempfl, T.; Moehle, C.; Landthaler, M.; Vogt, T. RBP2-H1/JARID1B Is a Transcriptional Regulator with a Tumor Suppressive Potential in Melanoma Cells. *Int. J. Cancer* **2008**, *122*, 1047–1057. [CrossRef] [PubMed]

47. Dodson, C.A.; Kosmopoulou, M.; Richards, M.W.; Atrash, B.; Bavetsias, V.; Blagg, J.; Bayliss, R. Crystal Structure of an Aurora-A Mutant That Mimics Aurora-B Bound to MLN8054: Insights into Selectivity and Drug Design. *Biochem. J.* **2010**, *427*, 19–28. [CrossRef] [PubMed]

48. Carvalhal, S.; Ribeiro, S.A.; Arocena, M.; Kasciukovic, T.; Temme, A.; Koehler, K.; Huebner, A.; Griffis, E.R. The Nucleoporin ALADIN Regulates Aurora A Localization to Ensure Robust Mitotic Spindle Formation. *Mol. Biol. Cell* **2015**, *26*, 3424–3438. [CrossRef] [PubMed]

49. Friedmann, D.R.; Aguilar, A.; Fan, J.; Nachury, M.V.; Marmorstein, R. Structure of the α-Tubulin Acetyltransferase, ATAT1, and Implications for Tubulin-Specific Acetylation. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 19655–19660. [CrossRef] [PubMed]

50. Li, L.; Yang, X.-J. Tubulin Acetylation: Responsible Enzymes, Biological Functions and Human Diseases. *Cell. Mol. Life Sci.* **2015**, *72*, 4237–4255. [CrossRef] [PubMed]

51. Stams, T.; Spurlino, J.C.; Smith, D.L.; Wahl, R.C.; Ho, T.F.; Qoronfleh, M.W.; Banks, T.M.; Rubin, B. Structure of Human Neutrophil Collagenase Reveals Large S1' Specificity Pocket. *Nat. Struct. Biol.* **1994**, *1*, 119–123. [CrossRef] [PubMed]

52. Matter, H.; Schwab, W.; Barbier, D.; Billen, G.; Haase, B.; Neises, B.; Schudok, M.; Thorwart, W.; Schreuder, H.; Brachvogel, V.; et al. Quantitative Structure−Activity Relationship of Human Neutrophil Collagenase (MMP-8) Inhibitors Using Comparative Molecular Field Analysis and X-Ray Structure Analysis. *J. Med. Chem.* **1999**, *42*, 1908–1920. [CrossRef]

53. Chen, J.-Y.; Liu, L.; Cao, C.-L.; Li, M.-J.; Tan, K.; Yang, X.; Yun, C.-H. Structure and Function of Human Naa60 (NatF), a Golgi-Localized Bi-Functional Acetyltransferase. *Sci. Rep.* **2016**, *6*, 31425. [CrossRef] [PubMed]

54. Hopkins, S.C.; Heffernan, M.L.R.; Saraswat, L.D.; Bowen, C.A.; Melnick, L.; Hardy, L.W.; Orsini, M.A.; Allen, M.S.; Koch, P.; Spear, K.L.; et al. Structural, Kinetic, and Pharmacodynamic Mechanisms of D-Amino Acid Oxidase Inhibition by Small Molecules. *J. Med. Chem.* **2013**, *56*, 3710–3724. [CrossRef]

55. Pollegioni, L.; Piubelli, L.; Sacchi, S.; Pilone, M.S.; Molla, G. Physiological Functions of D-Amino Acid Oxidases: From Yeast to Humans. *Cell. Mol. Life Sci.* **2007**, *64*, 1373–1394. [CrossRef] [PubMed]

56. Smith, S.M.; Uslaner, J.M.; Hutson, P.H. The Therapeutic Potential of D-Amino Acid Oxidase (DAAO) Inhibitors. *Open Med. Chem. J.* **2010**, *4*, 3–9. [CrossRef] [PubMed]

57. Steele, D.L.; El-Kabbani, O.; Dunten, P.; Windsor, L.J.; Kammlott, R.U.; Crowther, R.L.; Michoud, C.; Engler, J.A.; Birktoft, J.J. Expression, Characterization and Structure Determination of an Active Site Mutant (Glu202-Gln) of Mini-Stromelysin-1. *Protein Eng.* **2000**, *13*, 397–405. [CrossRef] [PubMed]

58. Sternlicht, M.D.; Lochter, A.; Sympson, C.J.; Huey, B.; Rougier, J.-P.; Gray, J.W.; Pinkel, D.; Bissell, M.J.; Werb, Z. The Stromal Proteinase MMP3/Stromelysin-1 Promotes Mammary Carcinogenesis. *Cell* **1999**, *98*, 137–146. [CrossRef]

59. McAndrew, R.P.; Wang, Y.; Mohsen, A.-W.; He, M.; Vockley, J.; Kim, J.-J.P. Structural Basis for Substrate Fatty Acyl Chain Specificity: Crystal Structure of Human Very-Long-Chain Acyl-CoA Dehydrogenase. *J. Biol. Chem.* **2008**, *283*, 9435–9443. [CrossRef]

60. Merinero, B.; Pérez-Cerdá, C.; Garcia, M.J.; Gangoiti, J.; Font, L.M.; Garcia Silva, M.T.; Vianey-Saban, C.; Duran, M.; Ugarte, M. Mitochondrial Very Long-Chain Acyl-CoA Dehydrogenase Deficiency with a Mild Clinical Course. *J. Inherit. Metab. Dis.* **1996**, *19*, 173–176. [CrossRef]

61. Chaudhry, A.S.; Thirumaran, R.K.; Yasuda, K.; Yang, X.; Fan, Y.; Strom, S.C.; Schuetz, E.G. Genetic Variation in Aldo-Keto Reductase 1D1 (AKR1D1) Affects the Expression and Activity of Multiple Cytochrome P450s. *Drug Metab. Dispos.* **2013**, *41*, 1538–1547. [CrossRef]

62. Zanger, U.M.; Schwab, M. Cytochrome P450 Enzymes in Drug Metabolism: Regulation of Gene Expression, Enzyme Activities, and Impact of Genetic Variation. *Pharmthera* **2013**, *138*, 103–141. [CrossRef] [PubMed]

63. Pérez-Tamayo, R. Pathology of Collagen Degradation. A Review. *Am. J. Pathol.* **1978**, *92*, 508–566. [PubMed]

64. Van Damme, P.; Hole, K.; Pimenta-Marques, A.; Helsens, K.; Vandekerckhove, J.; Martinho, R.G.; Gevaert, K.; Arnesen, T. NatF Contributes to an Evolutionary Shift in Protein N-Terminal Acetylation and Is Important for Normal Chromosome Segregation. *PLoS Genet.* **2011**, *7*. [CrossRef]

65. Clark-Taylor, T.; Clark-Taylor, B.E. Is Autism a Disorder of Fatty Acid Metabolism? Possible Dysfunction of Mitochondrial β-Oxidation by Long Chain Acyl-CoA Dehydrogenase. *Med. Hypotheses* **2004**, *62*, 970–975. [CrossRef]

66. Tranchant, I.; Vera, L.; Czarny, B.; Amoura, M.; Cassar, E.; Beau, F.; Stura, E.A.; Dive, V. Halogen Bonding Controls Selectivity of FRET Substrate Probes for MMP-9. *Chem. Biol.* **2014**, *21*, 408–413. [CrossRef]

67. Wang, J.; Tsirka, S.E. Neuroprotection by Inhibition of Matrix Metalloproteinases in a Mouse Model of Intracerebral Haemorrhage. *Brain* **2005**, *128*, 1622–1633. [CrossRef] [PubMed]

68. Groblewska, M.; Siewko, M.; Mroczko, B.; Szmitkowski, M. The Role of Matrix Metalloproteinases (MMPs) and Their Inhibitors (TIMPs) in the Development of Esophageal Cancer. *Folia Histochem. Cytobiol.* **2012**, *50*, 12–19. [CrossRef]

69. Dollery, C.M.; McEwan, J.R.; Henney, A.M. Matrix Metalloproteinases and Cardiovascular Disease. *Circ. Res.* **1995**, *77*, 863–868. [CrossRef] [PubMed]

70. Walsh, R.M.; Polizzi, S.J.; Kadirvelraj, R.; Howard, W.W.; Wood, Z.A. Man o' War Mutation in UDP-α-d-Xylose Synthase Favors the Abortive Catalytic Cycle and Uncovers a Latent Potential for Hexamer Formation. *Biochemistry* **2015**, *54*, 807–819. [CrossRef] [PubMed]

71. Moriarity, J.L.; Hurt, K.J.; Resnick, A.C.; Storm, P.B.; Laroy, W.; Schnaar, R.L.; Snyder, S.H. UDP-Glucuronate Decarboxylase, a Key Enzyme in Proteoglycan Synthesis Cloning, Characterization, and Localization. *J. Biol. Chem.* **2002**, *277*, 16968–16975. [CrossRef] [PubMed]

72. Couture, J.-F.; Dirk, L.M.A.; Brunzelle, J.S.; Houtz, R.L.; Trievel, R.C. Structural Origins for the Product Specificity of SET Domain Protein Methyltransferases. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20659–20664. [CrossRef] [PubMed]

73. Milite, C.; Feoli, A.; Viviano, M.; Rescigno, D.; Cianciulli, A.; Balzano, A.L.; Mai, A.; Castellano, S.; Sbardella, G. The Emerging Role of Lysine Methyltransferase SETD8 in Human Diseases. *Clin. Epigenetics* **2016**, *8*, 102. [CrossRef] [PubMed]

74. Guo, X.; Wang, L.; Li, J.; Ding, Z.; Xiao, J.; Yin, X.; He, S.; Shi, P.; Dong, L.; Li, G.; et al. Structural Insight into Autoinhibition and Histone H3-Induced Activation of DNMT3A. *Nature* **2015**, *517*, 640–644. [CrossRef]

75. Kim, G.-D.; Ni, J.; Kelesoglu, N.; Roberts, R.J.; Pradhan, S. Co-Operation and Communication between the Human Maintenance and de Novo DNA (Cytosine-5) Methyltransferases. *Embo J.* **2002**, *21*, 4183–4195. [CrossRef]

76. Yang, L.; Rau, R.; Goodell, M.A. DNMT3A in Haematological Malignancies. *Nat. Rev. Cancer* **2015**, *15*, 152–165. [CrossRef] [PubMed]

77. Rajavelu, A.; Tulyasheva, Z.; Jaiswal, R.; Jeltsch, A.; Kuhnert, N. The Inhibition of the Mammalian DNA Methyltransferase 3a (Dnmt3a) by Dietary Black Tea and Coffee Polyphenols. *BMC Biochem.* **2011**, *12*, 16. [CrossRef]

78. Stura, E.A.; Visse, R.; Cuniasse, P.; Dive, V.; Nagase, H. Crystal Structure of Full-Length Human Collagenase 3 (MMP-13) with Peptides in the Active Site Defines Exosites in the Catalytic Domain. *FASEB J.* **2013**, *27*, 4395–4405. [CrossRef] [PubMed]

79. Woessner, J.F. Matrix Metalloproteinases and Their Inhibitors in Connective Tissue Remodeling. *FASEB J.* **1991**, *5*, 2145–2154. [CrossRef] [PubMed]

80. Knäuper, V.; López-Otin, C.; Smith, B.; Knight, G.; Murphy, G. Biochemical Characterization of Human Collagenase-3. *J. Biol. Chem.* **1996**, *271*, 1544–1550. [CrossRef] [PubMed]

81. Mazza, C.; Breton, R.; Housset, D.; Fontecilla-Camps, J.C. Unusual Charge Stabilization of NADP+ in 17β-Hydroxysteroid Dehydrogenase. *J. Biol. Chem.* **1998**, *273*, 8145–8152. [CrossRef] [PubMed]

82. Mindnich, R.; Möller, G.; Adamski, J. The Role of 17 Beta-Hydroxysteroid Dehydrogenases. *Mol. Cell. Endocrinol.* **2004**, *218*, 7–20. [CrossRef]

83. Liu, P.; Jia, M.; Zhou, X.E.; De Waal, P.W.; Dickson, B.M.; Liu, B.; Hou, L.; Yin, Y.; Kang, Y.; Shi, Y.; et al. The Structural Basis of the Dominant Negative Phenotype of the Gαi1β1γ2 G203A/A326S Heterotrimer. *Acta Pharm. Sin.* **2016**, *37*, 1259–1272. [CrossRef] [PubMed]

84. Coleman, D.E.; Sprang, S.R. Crystal Structures of the G Protein Giα1 Complexed with GDP and Mg2+: A Crystallographic Titration Experiment. *Biochemistry* **1998**, *37*, 14376–14385. [CrossRef] [PubMed]

85. Milligan, G.; Kostenis, E. Heterotrimeric G-Proteins: A Short History. *Br. J. Pharm.* **2006**, *147*, S46–S55. [CrossRef]

86. Thoden, J.B.; Wohlers, T.M.; Fridovich-Keil, J.L.; Holden, H.M. Molecular Basis for Severe Epimerase Deficiency Galactosemia X-RAY Structure of The Human V94m-Substituted Udp-Galactose 4-Epimerase. *J. Biol. Chem.* **2001**, *276*, 20617–20623. [CrossRef]

87. Thoden, J.B.; Wohlers, T.M.; Fridovich-Keil, J.L.; Holden, H.M. Crystallographic Evidence for Tyr 157 Functioning as the Active Site Base in Human UDP−Galactose 4-Epimerase. *Biochemistry* **2000**, *39*, 5691–5701. [CrossRef]

88. Daenzer, J.M.I.; Sanders, R.D.; Hang, D.; Fridovich-Keil, J.L. UDP-Galactose 4′-Epimerase Activities toward UDP-Gal and UDP-GalNAc Play Different Roles in the Development of Drosophila Melanogaster. *PLoS Genet.* **2012**, *8*, e1002721. [CrossRef] [PubMed]

89. Tempel, W.; Wu, H.; Dombrovsky, L.; Zeng, H.; Loppnau, P.; Zhu, H.; Plotnikov, A.N.; Bochkarev, A. An Intact SAM-Dependent Methyltransferase Fold Is Encoded by the Human Endothelin-Converting Enzyme-2 Gene. *Proteins Struct. Funct. Bioinforma.* **2009**, *74*, 789–793. [CrossRef]

90. Liu, H.; Zang, C.; Fenner, M.H.; Possinger, K.; Elstner, E. PPARγ Ligands and ATRA Inhibit the Invasion of Human Breast Cancer Cells in Vitro. *Breast Cancer Res. Treat.* **2003**, *79*, 63–74. [CrossRef] [PubMed]

91. Egger, S.; Chaikuad, A.; Kavanagh, K.L.; Oppermann, U.; Nidetzky, B. Structure and Mechanism of Human UDP-Glucose 6-Dehydrogenase. *J. Biol. Chem.* **2011**, *286*, 23877–23887. [CrossRef] [PubMed]

92. Wise, P.M.; Suzuki, S.; Brown, C.M. Estradiol: A Hormone with Diverse and Contradictory Neuroprotective Actions. *Dialogues Clin. Neurosci.* **2009**, *11*, 297–303. [PubMed]

93. Hickie, R.A.; Walker, C.M.; Croll, G.A. Decreased Basal Cyclic Adenosine 3′,5′-Monophosphate Levels in Morris Hepatoma 5123 t.c.(h). *Biochem. Biophys. Res. Commun.* **1974**, *59*, 167–173. [CrossRef]

94. Robinson, E.A.; Kalckar, H.M.; Troedsson, H.; Sanford, K. Metabolic Inhibition of Mammalian Uridine Diphosphate Galactose 4-Epimerase in Cell Cultures and in Tumor Cells. *J. Biol. Chem.* **1966**, *241*, 2737–2745. [CrossRef]

95. Aubert, G.; Vega, R.B.; Kelly, D.P. Perturbations in the Gene Regulatory Pathways Controlling Mitochondrial Energy Production in the Failing Heart. *Biochim. Biophys. Acta Bba Mol. Cell Res.* **2013**, *1833*, 840–847. [CrossRef]

96. Warren, G.L.; Andrews, C.W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M.H.; Lindvall, M.; Nevins, N.; Semus, S.F.; Senger, S.; et al. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912–5931. [CrossRef] [PubMed]

97. Kubota, T.; Koshizuka, K.; Williamson, E.A.; Asou, H.; Said, J.W.; Holden, S.; Miyoshi, I.; Koeffler, H.P. Ligand for Peroxisome Proliferator-Activated Receptor γ (Troglitazone) Has Potent Antitumor Effect against Human Prostate Cancer Both in Vitro and in Vivo. *Cancer Res.* **1998**, *58*, 3344–3352. [PubMed]

98. Hosokawa, M.; Kudo, M.; Maeda, H.; Kohno, H.; Tanaka, T.; Miyashita, K. Fucoxanthin Induces Apoptosis and Enhances the Antiproliferative Effect of the PPARγ Ligand, Troglitazone, on Colon Cancer Cells. *Biochim. Biophys. Acta BBA Gen. Subj.* **2004**, *1675*, 113–119. [CrossRef] [PubMed]

99. Guengerich, F.P. Cytochrome P450 and Chemical Toxicology. *Chem. Res. Toxicol.* **2008**, *21*, 70–83. [CrossRef] [PubMed]

100. Takada, I.; Makishima, M. Peroxisome Proliferator-Activated Receptor Agonists and Antagonists: A Patent Review (2014-Present). *Expert Opin. Pat.* **2020**, *30*, 1–13. [CrossRef] [PubMed]

101. Somody, J.C.; MacKinnon, S.S.; Windemuth, A. Structural Coverage of the Proteome for Pharmaceutical Applications. *Drug Discov. Today* **2017**, *22*, 1792–1799. [CrossRef]

102. RCSB PDB: Homepage. Available online: http://www.rcsb.org/ (accessed on 5 February 2018).