HHS Public Access

Author manuscript

J Comput Aided Mol Des. Author manuscript; available in PMC 2016 May 01.

Published in final edited form as:

J Comput Aided Mol Des. 2015 May ; 29(5): 421–439. doi:10.1007/s10822-015-9831-x.

Pharmacophore modeling improves virtual screening for novel peroxisome proliferator-activated receptor-gamma ligands

Stephanie N. Lewis[1,2,4], **Zulma Garcia**[3], **Raquel Hontecillas**[1,4], **Josep Bassaganya-Riera**[1,4,5], and **David R. Bevan**[1,2]

[1]Genetics, Bioinformatics, and Computational Biology Program, Virginia Tech, Blacksburg, VA, USA

[2]Department of Biochemistry, Virginia Tech, Blacksburg, VA, USA

[3]Virginia College of Osteopathic Medicine, Virginia Tech, Blacksburg, VA, USA

[4]Nutritional Immunology and Molecular Medicine Laboratory, Center for Modeling Immunity to Enteric Pathogens, Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA USA

[5]Department of Biomedical Sciences and Pathobiology, Virginia-Maryland Regional College of Veterinary Medicine, Virginia Tech, Blacksburg, VA USA

Abstract

Peroxisome proliferator-activated receptor-gamma (PPARγ) is a nuclear hormone receptor involved in regulating various metabolic and immune processes. The PPAR family of receptors possesses a large binding cavity that imparts promiscuity of ligand binding not common to other nuclear receptors. This feature increases the challenge of using computational methods to identify PPAR ligands that will dock favorably into a structural model. Utilizing both ligand- and structure-based pharmacophore methods, we sought to improve agonist prediction by grouping ligands according to pharmacophore features, and pairing models derived from these features with receptor structures for docking. For 22 of the 33 receptor structures evaluated we observed an increase in true positive rate (TPR) when screening was restricted to compounds sharing molecular features found in rosiglitazone. A combination of structure models used for docking resulted in a higher TPR (40%) when compared to docking with a single structure model (less than 20%). Prediction was also improved when specific protein-ligand interactions between the docked ligands and structure models were given greater weight than the calculated free energy of binding. A large-scale screen of compounds using a marketed drug database verified the predictive ability of the selected structure models. This study highlights the steps necessary to improve screening for PPARγ ligands using multiple structure models, ligand-based pharmacophore data, evaluation of protein-ligand interactions, and comparison of docking datasets. The unique combination of methods presented here holds potential for more efficient screening of compounds with unknown affinity for PPARγ that could serve as candidates for therapeutic development.

Corresponding author: DRBevan drbevan@vt.edu phone: 540-231-5040.

Author Contributions: The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. The manuscript with written by SNL, with editorial and content support from RH, JBR, and DRB. Project development was shared by SNL, RH, and JBR, DRB. Data collection and analysis were conducted by SNL with assistance from ZG.

Keywords

virtual screening; computational molecular docking; PPARγ; pharmacophore modeling; drug discovery and design

Introduction

Peroxisome proliferator-activated receptor-gamma (PPARγ) is a ligand-activated transcription factor within the nuclear hormone receptor superfamily [1]. It is one of three PPAR subtypes: -α -β/δ and -γ [2], and it is involved in transcriptional regulation of numerous biological processes including glucose and lipid homeostasis [1-4], adipocyte differentiation [1,5], cell proliferation, and inflammation regulation [3,4,6-8]. PPARγis highly expressed in various cells that make up the immune system [9], as well as other cell types throughout the body [10-13]. Given the regulatory roles of PPARγ, this protein is an established therapeutic target for type 2 diabetes and holds promise as a novel target for treating chronic inflammatory and immune-mediated diseases, such as inflammatory bowel disease.

Compounds that bind to PPARγ can be categorized as either agonists or antagonists [14]. Agonists are further classified as full, partial, dual, or pan agonists depending on the degree of activation and the number of PPAR subtypes activated. This variation in agonism is due to a relatively large binding cavity within the ligand-binding domain (LBD) that can accommodate a variety of naturally occurring and synthetic compounds [2,15]. The PPARγ drug discovery and development literature has established that agonists should possess a small polar region and an extended hydrophobic region that form hydrogen bonds and hydrophobic interactions, respectively, within the LBD [2,15,16]. Typically, PPARγ ligands have a three-module structure consisting of a reactive polar head group, a linker, and an effector segment [17,18]. The polar head group tends to interact with residues near helix H12, while the effector segment interacts with residues near H3 and the opening to the binding cavity.

Within the agonist category, we see further distinction between compounds that activate PPARγ How the ligand interacts with the residues of the binding cavity determines the activity type. PPARγ-specific full agonists will activate PPARγ only, while dual agonists will activate more than one PPAR (PPARγ and -α or PPARγ and −β/δ), and pan agonists will activate all three PPAR isoforms. Tolerability and safety issues often arise with the use of synthetic compounds that fall into the dual and pan agonist categories [19-21]. Another category of agonists are partial agonists, which are distinguished from full agonists based on the region in which the binding process influences conformational changes: partial agonists influence conformational change through H3, while full agonists influence H12 dynamics. A final, more recently established category, is the selective PPAR modulators, for which one sees a divergence in efficacy of responses. A selected response may follow the full agonist efficacy increase, while other responses, which are often associated with side effects under full agonist regulation, would only occur at low efficacy or require a higher dosage to induce a response with full agonist efficacy [19]. One can also see tissue-specific differences in

gene regulatory responses where a compound will drive full agonism in one tissue where specific co-activators are present, but function as an antagonist in another tissue where that co-activator is not present [20]. The factor that most influences ligand classification is points of contact in the binding cavity, which in turn drives conformational variations and binding partner recruitment [21]. Selective PPARγ modulators (SPPARMs) have progressively become a greater focus in PPARγ-targeted drug development because of the potential for a reduced incidence of side effects when activation can be tailored toward specific gene regulatory outcomes.

After binding and activation occur, the interactions between the agonist and H12 and/or H3 drive conformational changes that influence activity. Agonist binding induces a conformational change in the activation function-2 (AF-2) domain that governs co-repressor release and co-activator recruitment [12,16,22]. The structure of the bound ligand influences the degree of conformational change and/or the conformational dynamics, and therefore influences which co-activator proteins are recruited. The active PPARγ-agonist-co-activator complex, which is heterodimerized with retinoic acid-bound retinoid X receptor-alpha, can associate with PPAR response elements on DNA to induce transcription [2,12,23]. Antagonist binding differs in that binding does not induce co-repressor release. Instead, the inactive conformation is stabilized by binding of the antagonist [24].

High-throughput screening (HTS) methods to identify PPARγ modulators are available [25,26], though none of the HTS methods has been widely adopted. Moreover, experimental methods can be costly if one wants to screen a diverse collection of large numbers of compounds, and the process would require a significant time commitment. Contrasted with experimental HTS, virtual screening is a computational technique for streamlining the drug discovery and development process [27]. Using three-dimensional (3D) representations of test compounds and targets, one can identify potential drug hits based on user-defined binding criteria relevant to the system of interest. In the context of docking, a binder is a compound that is deemed to fit well within a 3D structural model given a pre-defined list of criteria that must be satisfied. Thus, virtual screening involves high-throughput docking to reduce the number of compounds tested experimentally by excluding compounds for which interactions with the target are expected to be unfavorable or nonexistent.

Standard large-scale virtual screening studies with PPARγ structures generally result in disappointing outcomes in terms of predictive value [28-32]. In the context of docking, the large binding cavity can accommodate different types of ligands, many of which would not serve as agonists. An approach that can bring additional information to improve success in virtual screening is pharmacophore modeling. Structure-based pharmacophore modeling utilizes receptor structures to detect features within the binding cavity that would favor interactions between the target receptor and the atoms of potential ligands [33,34]. In contrast, ligand-based pharmacophore modeling can be done in the absence of receptor structure models and involves aligning a set of known ligands to observe common features that may be key to biological activity [35].

These two pharmacophore modeling methods independently provide useful structural information for establishing binder criteria in a search for potential agonists. Pharmacophore

modeling has been applied to virtual screening of ligands for PPARγ as well as the other PPAR subtypes, −α and −β/δ [36]. A combination of the two pharmacophore methods may narrow the search to facilitate compound screening with respect to PPARγ [30-32]. Through application of this concept, pharmacophore modeling can be an important tool in the discovery and design of novel drugs for PPARγ that are more efficacious and safer than existing drugs [37].

In this study, we set out to establish criteria that could be applied in a virtual screen of a diverse ligand database using multiple structure models for PPARγ. The goal was to reduce the amount of docking necessary to find binders while improving the frequency with which true binders were identified. Using pharmacophores as filters [38], we sought to cluster ligands by feature similarity, pair ligand clusters to structure models in which the compounds could be appropriately accommodated during the docking process, and develop a scoring system where pharmacophores and potential interactions are considered when proposing binders for validation. A test of the methods with a dataset of compounds with unknown PPARγ binding ability indicated that identification of compounds that should bind to PPARγ is possible given a diverse compound database, our combined knowledge-based approach, and selected criteria for recognizing binders of the full agonism activity type.

Methods

Catalog of PPARγ 3D structures

PPARγ structures from the RCSB protein data bank (PDB) [39,40] were cataloged. As of December 2014, 122 structures were available, of which most were in complex with a ligand. In some cases, the same ligand was present in multiple structures. An activity type was assigned to each structure based on the bound ligand and activity reported in the published experiments (see references in Tables 1 and 2). The activity classes explored in this study were full agonist, fatty acid agonist, partial agonist, and antagonist. Although the fatty acid ligands exhibit full agonism, they were placed in a separate group because they are the endogenous ligands for PPARγ and are structurally different from synthetic agonists.

Selection of structures for pharmacophore modeling

An evaluation of the 122 structures determined some would not be suitable for analysis in this study. Crystal structures possess atom location variations that can influence structural assessment. No single crystal structure is perfect, but a combination of structures can be used to evaluate protein-ligand interactions under the assumption that multiple structures will allow for sampling of residue side chain variability in a rigid macromolecule docking study. Structures without co-crystallized ligands, with large groups of residues missing, with multiple side chain positions for binding cavity residues, with covalently bound ligands, with co-crystallized ligands that fit other ligand categories (e.g., pan agonist), with multiple distinct locations for ligand binding (i.e., opposite ends of the binding cavity), and with ambiguous co-crystallized ligand binding patterns (i.e., both partial and full activity reported) were excluded because these characteristics would introduce less reliable or less appropriate atom coordinate data. A subset of 33 crystal structures (Table 1) remained for the structure-based pharmacophore analysis and the docking component of the study

(detailed below). Each selected structure contained a ligand with a clearly defined activity type. The collection of structures represented four activity types: 11 full agonist-, 5 fatty acid-, 15 partial agonist-, and 2 antagonist-containing structures.

For the ligand-based pharmacophore modeling step, filters were applied to the original 122 structure list to find known active binders. In this case, ligands with ambiguous binding activity, activity beyond the four types to be evaluated, and ligands that would prevent clear delineation between the ligand types were excluded. Duplicate ligands were also removed. The final ligand counts for each activity type were 22 full agonists, 5 fatty acids, 21 partial agonists, and 2 antagonists (Table 1).

Identification of structure-based pharmacophores

Generation of structure-based pharmacophores involved application of the AutoLigand program, which uses a grid-based representation of binding potentials to identify sites of favorable interactions in a protein binding site [41]. This approach calculates atom-specific grids within the binding cavity, with the resulting affinity potential grid being called an envelope [41]. The atom types considered for the envelopes were carbon, oxygen, nitrogen, hydrogen, and sulfur. Prior to generating the envelopes, UCSF Chimera [42] was used to overlay the structures and save protein coordinates to the same coordinate frame, using the 1FM6 PDB structure [2] as the reference. The grid box dimensions were set to 46 Å by 42 Å by 46 Å, with 1.0 Å grid point spacing and the box center at (15.178, -22.133, 9.98) for all structures. The grid box contained 94,987 total grid points. This box covered the entire binding cavity and the exterior surfaces of the ligand-binding domain. Some structures that contained the same bound ligand were included because the distance-based protein-ligand interactions varied among these crystal structures.

Ten 300-point envelopes were generated for each structure. The most energetically favorable contiguous envelope for each structure was used for comparison among structures within the same activity class. USCF Chimera [42] was used to visualize the representative envelopes, in which tallies of identical points from multiple envelopes were used as modifiers for grid point representation. Two visualization techniques were applied: (1) tallies were radius multipliers for sphere size [41], and (2) frequencies of atom-independent grid points were used as a gradient for sphere color for all envelopes in a class. Binding cavity arms were designated as arms I through III using standard naming conventions for visualization and evaluation of the binding cavity (Figure 1) [43,44]. Arm I is the main arm of the binding cavity extending from the cavity opening to H12. Arm II is positioned near the cavity opening, while arm III is at the rear of the binding cavity near H12.

Structures were then grouped based on envelope similarity according to the distribution of percent similarity values. Percent similarity was calculated for each pair of envelopes as a percentage of common grid points. This value was the number of common grid points divided by the total number of points (300 grid points). Activity class was not considered for grouping, which translated to groups of structures in which potentially more than one activity class could be evaluated.

Identification of ligand-based pharmacophores

The ligand-based pharmacophores were determined using PharmaGist [37,79], which is a freely available web server for ligand feature assessment. The key features recognized by PharmaGist include aromatic, hydrogen bond acceptor, hydrogen bond donor, positively charged, negatively charged, and hydrophobic groups. Feature weights were set to 0.3 for hydrophobic groups, 1.0 for aromatics, 3.0 for hydrogen bond groups (acceptors and donors), and 1.5 for charged groups (positive and negative). Hydrogen bonding is necessary for full agonism in PPARγ and therefore weighted the heaviest. Files containing coordinates for groups of ligands based on activity class were uploaded to PharmaGist. As only two antagonists were identified, a full and a partial agonist with similar molecular composition were included to meet the server minimum of three ligands. Only the pairwise results for the two antagonist compounds were evaluated to determine potential pharmacophores. Models, ranked by the PharmaGist feature group score, were generated for each activity class (full agonist, fatty acid agonist, partial agonist, and antagonist) based on the highest scoring feature group and the group representing the most ligands. The high scoring group for each was composed of only a few compounds that possessed the combination of features that yielded the highest score. The feature group representing the most ligands reflects the minimum number of ligand features required for a compound to fall within the specified ligand class.

Three-dimensional pharmacophore models were generated from the lists of PharmaGist features with the Molecular Operating Environment (MOE) software suite [80]. The pivot molecule was used as the representative ligand for each activity class. These models facilitated quick screening of ligand databases for ligands that matched the pharmacophores. Databases of PDB-derived PPARγ ligands and PPARγ known active and decoy compounds obtained from the Directory of Useful Decoys (DUD) [30] were compared to the models for validation (Online Resource 2). A test of the pharmacophore models against the known PPARγ actives indicated that the fatty acid and full agonist pharmacophores did not match a large enough proportion of the training set. Specifically, the fatty acid and full agonist models missed some fatty acids and all the thiazolidinedione (TZD) compounds, respectively. Using the PharmaGist features for full and fatty acid agonists, additional pharmacophore models were made. The full agonist pharmacophore model was based on rosiglitazone (BRL; PDB ID 1FM6), and the fatty acid model was based on 9-hydroxyoctadeca-10,12-dienoic acid (9HO; PDB ID 2VSR). These models recovered the missing ligands. As some compounds did match the initial full agonist and fatty acid agonist models, they were maintained in the ligand pharmacophore model set. The final number of ligand-based pharmacophore models used was eight.

Computational Docking

The training portion of virtual screening was conducted with a total of 3308 ligands: 96 known active compounds and binding partners from PDB structures, 85 known actives from the DUD, and 3127 decoys from the DUD, which are structurally similar to actives but do not activate PPARγ. The compiled ligand database included duplicates of known agonists, which provided checks for reproducibility and consistency in pose sampling for known agonists. Docking was performed against the 33 superimposed PPARγ structure models with

identical grid boxes set to 30 Å by 40 Å by 40 Å, the center at (18.656, -21.929, 7.715), and grid spacing of 1.0 Å, which yielded 52,111 grid points and a total grid volume of 48,000 $Å^3$.AutoDock Vina [81] (Vina) was used for docking against the PDB structure models. Vina is a next-generation update to the AutoDock 4 platform for computational docking. A knowledge-based scoring function is used instead of the force field-based scoring algorithm implemented in AutoDock 4. The knowledge-based scoring agrees with the interaction searching method used for separating the ligand types, which indicated Vina was the better software package to use for this study. As in a previous study, re-docking and cross-docking were performed to establish predictability of the structure models [82].

Perl [83] scripts written in-house were used for post-processing the docking results. This analysis included retrieving lowest energy poses and the associated energy scores, measuring distances between interacting atoms, and formatting output for ease of reading and assessment. Only the most energetically favorable pose of up to nine poses for each ligand was used for analysis. Root-mean-squared deviation (RMSD) calculations and prediction of interactions were conducted as previously described [82]. The terms “successful pose” and “success” were applied to docked poses with an RMSD value less than or equal to 2.0 Å [84], or for poses with hydrogen-bond interactions observed between the ligand and all four critical residues (Ser289, His323, His449, and Tyr473) of the receptor model.

Screening statistics indicated the probability of success and failure at accurately identifying active compounds while excluding decoys. The primary values considered were specificity, which is the proportion of negatives/non-binders recognized as such, and sensitivity, which is the probability that a randomly selected ligand is a true positive/binder. Ideal specificity values should be greater than 0.95, or 95 percent. Sensitivity, which is also called the true positive rate (TPR), improves as the value approaches 1.00, or 100 percent. Specificity was used to calculate the false positive rate (FPR), which should approach zero. Specificity and sensitivity are calculated by counting the number of true positives, false positives, true negatives, and true positives.

Distinctions between true positives (TP; binders identified as binders) and false negatives (FN; binders inaccurately categorized as non-binders) were determined based on the RMSD values for docking the PDB ligands, and the presence of key interactions with docked poses for the DUD active compounds. TPs would possess RMSD values less than or equal to 2.0 Å for PDB docked poses, and interactions with the four key residues for the DUD actives. FNs would possess RMSD values over 2.0 Å for the PDB ligands, and no hydrogen-bonding interactions for the DUD actives. False positives (FP; non-binders inaccurately categorized as binders) and true negatives (TN; non-binders identified as non-binders) were assessed by evaluating key interactions for docked poses for the DUD decoy compounds. FPs would erroneously possess interactions with the four key residues, while TNs would not possess interactions with all four key residues. TPRs and FPRs were used to assess the receiver operating characteristic (ROC) space for the structures (Online Resource 2). The ROC method is a statistic that indicates the ability of a test to separate data into two populations [85]. Evaluation of the ROC space allowed for creating a shortened list of structures for which binders were more accurately identified while correctly excluding non-binders.

Validation of the docking protocol

Compounds (a total of 2319) from the MicroSource U.S. Drugs (MSUSDrugs) dataset of the ZINC online chemical catalog [86] were used as unknown compounds for a validation test of the structure models selected for docking. These compounds are all commercially available drugs, which potentially included known agonists like rosiglitazone. No pre-screening was performed, however, in order to maintain a level of blind testing. The MSUSDrugs compounds were docked into the four selected PPARγ structure models from the structure-based pharmacophore analysis described above to test the validity of the selected structures as representative models for screening. The docking procedure used for the training step (described above) was also used for the validation step.

Additional filters were applied to the MSUSDrugs docking results in order to rank and discover potential binders. Poses that possessed interactions with the four hydrogen-bonding residues were ranked according to free energy of binding as calculated by Vina. The ligand-based pharmacophore models were applied to the compounds as well to hypothesize potential activity types for each ligand. Compounds also were evaluated for toxicity and existing therapeutic patents that target PPARγ. Patents were identified using key word searching on the internet. Toxicity was assessed using MOE [80] and the ToxAlerts server [87]. MOE references two publications in which potentially toxic and reactive molecular groups were reported [88,89]. ToxAlerts included these two publications and 23 others dating from 1976 to 2013.

Scheme 1 illustrates how all the components of the methods come together to make up the virtual screening method composed for this study.

Results

Shared features were identified for each ligand type

In identifying ligand-based pharmacophores, the PharmaGist online server reported several lists of features shared by compounds within each activity class. The highest overall scoring feature set and the set representative of the most ligands were assessed (Table 2 and Figure 2). The highest scores ranged from 27.2 to 48.5 for the full agonist, fatty acid agonist, and partial agonist ligand groups. The number of features ranged from 7 to 10. Two groups of partial agonists were assessed with Pharmagist because the initial group contained three compounds that were molecularly similar and biased the pharmacophore results toward those three compounds. The second group of features for partial agonists was noticeably different from the initial feature set, and it shared features seen with the full agonists. A greater number of hydrogen bond acceptor groups and a reduction in the number of aromatic groups for this highest scoring feature set resulted in a slightly higher score compared to the first feature set generated using all 21 partial agonists. The feature groups for the fatty acids included the carboxylate moiety, which was also seen with the first partial agonist highest scoring feature group. Seven of the ten features identified for fatty acids were hydrophobic features that characterized the aliphatic chains, which distinguished this activity class from the others.

The feature clusters that represent the most ligands for each category possessed three features each for the full and partial agonists, and ten features for the fatty acids. It was anticipated that similarities in the most common features would exist because all the ligands bind to PPARγ. The three features for the full agonists and second partial agonist set were two hydrogen bond acceptor groups and an aromatic moiety (Figure 2). Interestingly, the placement of these groups relative to the binding cavity was different for each of these two activity classes. The full agonist features were closer to H12, while the partial agonist features were closer to H3. The initial partial agonist feature set possessed two aromatic groups and one hydrogen bond acceptor group, which may suggest greater feature diversity within the partial agonist set compared to the full agonists.

The antagonist feature sets were relatively similar, with each containing one aromatic moiety, one hydrogen bond acceptor, and one hydrogen bond donor. The difference between the two was an additional aromatic group when the ligand denoted NSI (PDB ID 2HFP) was the pivot molecule versus an additional hydrogen bond acceptor group when GW9 (PDB ID 3E00) was the pivot molecule. Overall, the pairwise antagonist features suggested the two compounds were similar to each other but distinct from the other ligands. However, given the variation between the two feature sets and the limited number of available structures, the feature sets for both NSI and GW9 were used to generate antagonist pharmacophore models. For the other activity classes, the feature set representing the most ligands was used to make pharmacophore models for classifying ligands in the compound database used for virtual screening.

Binding cavity differences validated the need for docking with multiple structure models

The envelopes generated using AutoLigand are comparable to structure-based pharmacophores. Some similarities were observed in the envelopes across all activity classes (Figure 3). Points representing carbon atoms occupied a large portion of the cavities (Figure 3A). The full agonist and fatty acid containing structures both showed higher densities for sulfur and oxygen atoms at the region of the binding cavity closest to H12 and near the entrance to the binding cavity. This characteristic was not seen in the composite envelope for the partial agonist-containing structures. The clustering of hydrogen bond acceptors near H12 is indicative of the binding interactions seen with full agonists and fatty acids that are not seen with partial agonists. The full agonist- and fatty acid-containing structures showed a more defined shape to the pocket, whereas the consensus envelopes for the partial agonist-containing structures contained points spread over more of the pocket. This difference agrees with the greater variability in interaction points seen with the partial agonist-containing structures compared to those containing full and fatty acid agonists. For the antagonist group, points were clustered near H12, which was seen in the envelopes for the full- and fatty acid-containing structures, but these envelopes are less complete than the others due to the small number of antagonist-containing structures.

Counts of common grid points independent of atom type were also calculated and visualized in a heat-map-like representation for each activity class (Figure 3B). Such a representation indicated patterns in grid points favored for each activity type, which would reflect where ligand atoms should sit most favorably within the cavity for a given activity. The fatty acid-,

full agonist- and antagonist-containing structures all showed favorable ligand atom positions near H12. The fatty acid- and full agonist-bound structures additionally showed favorable points along arm I, with the fatty acid structures showing greater favorability near the entrance to the binding cavity compared to the full agonist structures. The partial agonist-bound structures suggested atoms around H3 were preferred. The antagonist class also showed points under and beyond His449 that are not as prevalent in the envelopes for the other structures. The side chain for His449 sits at a different angle in the antagonist crystal structures when compared to structures of the other activity types, which may account for the difference in atom placement favorability for that area of the structures.

The percent similarity between each pair of envelopes was calculated to divide the structure models into groups with similar grid-based docking pockets. Structure models with similar docking pockets should accommodate particular groups of ligands in a similar fashion. A distribution of these values showed a major peak around 0.4 or 40 percent similarity between pairs of envelopes with a second less predominant peak at 0.6 (60 percent) (Figure 4). The envelope groupings were not distinct with 0.4 cutoff because of the high degree of overall binding cavity similarity between the crystal structures. The 0.6 cutoff resulted in groupings that were more distinct and six structure model groups were established (Table 3).

Full agonist-bound structures were present in Group 1 and Group 2, fatty acid-containing structures were in Group 1 and Group 3, and partial agonist-bound structures were in Group 2, Group 3, Group 4, and Group 5. Group 6 was composed of the remaining crystal structures that returned low similarity to structures in groups 1 through 5 and low similarity to each other. All the activity types are represented in this group. Group 6 possessed the largest number of structures and suggested there is a degree of specificity within the cavity of these crystal structures that precludes use of them in a docking study with a diverse ligand database. Each of the first five groups was paired with ligand pharmacophore models (Table 3) based on the types of ligands that should be best accommodated by the structure model. Ideally, a single structure within each group could be established as the representative model because the binding cavities within the group possess a higher than average degree of similarity. Docking provided a test of this assumption as will be described below.

Docking of crystal structure-derived ligands into the PPARγ structure models

Re-docking and cross-docking were conducted to determine which of the PPARγ crystal structures would be most suitable for virtual screening. With 2.0 Å as the threshold for acceptable RMSD values, re-docking results were favorable for 10 out of the 33 analyzed structures, including representatives of full and partial agonist-containing structures (Table 4). During cross-docking, all the structures, including those for which re-docking was not successful, showed successful docking of more than one non-native ligand. A table containing all the RMSD values for the re- and cross-docking poses can be found in Online Resource 1. The re-and cross-docking data provided an initial evaluation of how well the structure models could be used to accurately identify binders. We next hypothesized that incorporating the pharmacophore screening would improve the predictive ability of structure models based on the activity type represented by the pharmacophore model. This concept required additional statistical measures of docking potential.

Positive predictive rates for multiple PPARγ structure models

The information pertaining to successful or unsuccessful pose prediction was used to calculate the frequency at which binders were accurately recognized by the structure models. The initial calculation was conducted with just the ligands from PDB structures so RMSD values could be used to assess true positive identification. The overall per-model TPRs were low (2.08 to 18.75 percent success rates) for the 33 structure models (Table 5, original TPRs). The low rates may derive from the number and variety of ligands screened and the use of rigid receptor model structures for docking. Some structures served better for docking one type of ligand over the other types. For example, docking of full and dual agonists to 2PRG (PPARγ-rosiglitazone) was successful, and only one partial agonist docked well. Antagonists in general did not dock well, though two structures, 1FM6 and 3DZY, returned favorable docking of the two antagonist ligands tested.

In an attempt to improve predictability as assessed through true positive rates, pharmacophore filtering was applied to the docked poses using the rosiglitazone ligand-based pharmacophore model. The results of this filtering were used to calculate adjusted TPRs. A subset of 22 out of the 33 structures showed some increase in TPR (Table 5, adjusted TPRs). Filtering reduced the total number of compounds for calculating the rate, which suggested that pre-screening of test compounds with our pharmacophore models should maintain the true binders while reducing the amount of screening that needs to be carried out. For cases where the rate decreased, successful binders were lost from the list suggesting the rosiglitazone pharmacophore model missed ligands that docked well into those structures. Ligands that did not match the rosiglitazone model were presumably not TZD-like full agonists. Therefore, the structures with reduced success rates would not be ideal for screening of TZD-like full agonists. Taken together, these results supported the use of multiple ligand-based pharmacophore models and structure models representative of different activity types in a diverse, large-scale virtual screen for potential full agonists.

It was possible that the structures for which the TZD ligand model did not work were ideal for screening other types of full agonists, or the structures were not appropriate for screening of full agonists at all. Eleven structures had reduced success rates of which nine were partial agonist-containing structures and two were full agonist-containing structures. Of the compounds that docked well into these structures, most were partial agonists or compounds that matched multiple pharmacophore models. Therefore, it was determined that the structures that cannot accommodate TZDs were not appropriate for full agonist-specific screening.

Structures for which the TPR increased upon application of the rosiglitazone-based pharmacophore model would indicate either a structure most appropriate for TZD-like full agonist docking or an improvement in rate due solely to the reduction in the total number of active compounds considered for the TPR calculation. Structures with a high increase in TPR (>75%) showed a low number of successfully docked compounds (Table 5, percent change) and therefore would not be ideal models despite the significant rate increase. The structures with a moderate increase (∼50%) in rate contained ligands that fell into two categories: active identified as such that matched the pharmacophore models deemed to fit the structure model binding cavity, or actives identified as such but did not match the

assigned pharmacophore models. This suggests that the structure model is ideal for screening compounds matching the selected pharmacophore while possessing enough promiscuity in the binding cavity landscape to screen for other compounds that potentially possess similar activity but may not completely match the selected pharmacophore. Given these data, a structure model showing moderate increase in success rate with the inclusion of pharmacophore filtering might prove better suited for diverse ligand screening. The next steps were to evaluate filtering using the proposed ligand pharmacophore models and assess the rate at which false positive predictions occurred.

Selection of structure models for validation docking

It was clear that the individual positive predictive rates were low for each structure model, but it was hypothesized that selection of multiple structure models for docking and considering multiple sets of docking results would improve predictability. Completeness of structure, the relative docking success of each structure, and any improvements in TPR based on pharmacophore screening were considered for identification of the representative structure models for the other structure groups. Structures considered to be the most successful were those within the group that had the highest TPRs. The most complete models were those containing all the protein residues of the binding cavity necessary for accurate sampling of binding poses.

Structure 1ZGY from Group 1 performed well in the TPR assessment analysis. This structure possessed all atoms for all residues surrounding the binding cavity and was therefore selected as the representative structure for this group. The IK74 and 1RDT structures, both of which were in Group 2, were determined to be the most likely to predict binders based on the ROC space graph (Online Resource 2). Because the TPR for 1K74 was higher than the TPR for 1RDT, the former was better suited for full agonist and fatty acid agonist prediction.

The DUD dataset proved most effective for assessment of full agonist and fatty acid agonist binding, which were not the predominant activity type represented in Groups 3, 4, and 5. For these three groups, the most complete structures were identified and used for docking under the assumption that screening for full agonist binders with these structures might be less successful and therefore serve as comparative datasets rather than true identifiers of full agonist hits. Both structures in Group 5 were missing residues within the binding cavity and were excluded. The four PDB models selected for the virtual screening validation step were 1ZGY, 1K74, 2I4J, and 2I4Z. The combined TPR calculated from the RMSD data of these four structure models was calculated to be approximately 0.4, which was higher than the individual TPRs shown in Table 5 for any single structure.

Full agonist predictions improved with pharmacophore and interaction filtering methods

As a means of validating the screening methods, screening was performed with 2,319 established drug compounds from the MSUSDrugs database. These compounds were docked into the four representative structure models and the proposed filters were applied to identify full agonist binders. Some compounds docked successfully to more than one of the structure models suggesting some agreement in the docking results. For instance, folic acid

showed up as a hit in the 1ZGY, 1K74, and 2I4Z lists. Overall, the composition and order of hits was slightly different for all the lists. Most of the hits matched at least one of the pharmacophore models proposed to fit the binding cavity. The only exceptions were compounds that docked into 2I4Z. Five of the ten hits in the 2I4Z list did not match either partial agonist model. Compounds identified as hits for 2I4J and 2I4Z were not present in the 1ZGY and 1K74 lists. The 2I4J and 2I4Z structure models favored partial agonist binding, but full agonist criteria were used for identifying binders in this study for consistency. Therefore, compounds deemed as binders to these two structures may be either partial agonists or false positives. Compounds unique to the 1ZGY and 1K74 lists, like methotrexate, would most likely be true full agonists.

Data from the 1ZGY docking is shown here (Table 7) with additional docking data listed in Online Resource 3. Only the data for the lowest energy pose for each docked compound is shown in Table 7. Two stereoisomers for rosiglitazone appeared in the ninth and tenth position of the top-ten list for 1ZGY with similar free energy of binding values so only one of those results is listed. Folic acid was deemed the best binder. Rosiglitazone, which is a TZD known to target PPARγ, held a position at the bottom of the ranked list. Pioglitazone hydrochloride, which was sixth on the list, is also a TZD that has been used to target PPARγ in the treatment of disease. One-half of the members of the rank list matched the rosiglitazone-based pharmacophore model, with more compounds matching the fatty acid-based models. The patent search indicated that seven of the nine compounds identified have been considered for binding to PPARγ. These results suggest that the proposed progress flow and filters applied in this study were successfully capable of accurately predicting true PPARγ binders.

Discussion

The progress flow proposed for this study worked well for identifying PPARγ binders of the full agonist activity type. Our analysis suggested pre-screening with ligand-based pharmacophore models and matching test compounds with structures for which the binding cavity can accommodate key features is a useful method to screen large, diverse compound databases for PPARγ binders. Furthermore, pharmacophore pre-screening helped with classification of the type of agonism that a binder might possess. As a result, the predicted interactions necessary for binding were weighed more heavily than the free energy of binding score, as the interactions with key residues are less influenced by the number of atoms that constitute a compound. The free energy of binding remains as a means of ranking the results, but did not necessarily correlate to the rank of affinity for the compounds. For example, rosiglitazone, pioglitazone, and benzofibrate possess EC50 values of 0.009 μM, 0.28 μM, and 60 μM, respectively [47,90-92]. Affinity would translate to rosiglitazone > pioglitazone > benzofibrate, but this order is reversed when ranked by calculated free energy of binding. Beyond the criteria outlined in the results, other factors to consider when identifying binders include how well the features for the test compound align to the pharmacophore models and if compounds match more than one ligand model.

In a 2004 review of docking and scoring methods, Kitchen mentioned considerations for improving the hit identification process [93]. In the context of energy scoring, Kitchen

suggested that scoring should derive from knowledge-based structural comparisons in which binding interactions are considered [93]. Rather than developing an energy score where specific interactions are considered as terms or a numerical interaction-derived docking score, we took the approach of a binary indication of absence or presence of interactions to complement existing energy scoring methods. Kitchen also discussed pre-screening of ligand databases with 3D filters [93]. Shape-dependent filters can be limiting as he states, but our implementation of multiple pharmacophore models appears to address that limitation when considering diverse ligand databases. The filtering allows one to assess compounds that share features with established active compounds, but also indicates which compounds do not resemble known actives. Such compounds would not be the primary focus of a docking study given they typically would be classified as non-binders. A subsequent analysis, however, of the non-matching compounds for which key interactions were seen might indicate novel pharmacophores or hits for further study. Alternatively, the non-matching compounds may share features within that set that should be excluded when screening compounds with unknown activity.

All of the compounds that appeared in the top-ranked list for 1ZGY either directly interact with PPARγ or function as a therapeutic in conjunction with other drugs for treating a metabolic or inflammatory disease through activation of PPARγ. Rosiglitazone and pioglitazone are well known PPARγ agonists and established therapeutics for type 2 diabetes. The presence of these compounds in the list is encouraging and immediately strengthens the reliability of our virtual screening process. Additionally, flufenamic acid is a non-steroidal anti-inflammatory drug that activates PPARγ [94]. The other four compounds that appeared in patents were listed as compounds used in conjunction with PPARγ agonists. It was not clear in the patents if the compounds directly or indirectly activate PPARγ, but the addition of the compounds appeared to facilitate the activity of other agonists included in the patent documentation. Furthermore, the disease types listed in the patents included cancers, metabolic disorders, and inflammation-related diseases that have been linked to PPARγ-mediated regulatory processes. Only two of the compounds did not immediately return patent information for PPARγ-related disease therapeutics: balsalazide disodium and ceftriaxone sodium trihydrate. Balsalazide disodium is a compound that has been suggested to treat active ulcerative colitis, which is a disease in which PPARγ and PPARδ play a role [95]. Clinical trials have shown that rosiglitazone has been effective at reducing the severity of ulcerative colitis [96]. The literature suggests balsalazide disodium directly binds to and activates PPARγ and thereby ameliorates the disease [97]. The only unclear hit was ceftriaxone sodium trihydrate. No patents or published papers appeared to list the compound and PPARγ. This compound could be a promising hit, but the presence of toxic molecular groups could prove problematic for experiments beyond the *in vitro* stage. Nevertheless, this compound and derivatives of it would warrant experimental consideration.

Further, the trends seen in the rated unknown docking data indicates that compounds matching more structure-preferred pharmacophores may be better binders. This follows the ranking trend seen with the free energy of binding. Pharmacophore matching could also inform predictions on downstream effects. Higgins and DePaoli suggest efficacy and potency trends create a distinction between simple agonists and selective modulators [19].

Simple agonists exhibit a consistent trend for multiple responses where an increase in potency translates to an increase in efficacy for multiple biological responses [19]. Selective modulators, however, show uncoupled dose responses for either potency or efficacy, which results in perturbation of a single response, or subset of responses, rather than all of the biological responses controlled by the targeted regulated gene [19]. It is possible that compounds that match multiple pharmacophores might bind to multiple regulators and affect more genes, whereas compounds that match fewer pharmacophores would be more selective and trigger uncoupled regulatory responses.

The promiscuity of PPARγ makes identifying all possible pharmacophores challenging but it has proven possible given what is known about binders. For instance, many groups have proposed both ligand-based and structure-based pharmacophore models from available structure data [44,62,90,98-101]. Models from other studies were more complex or more general, were specific to a certain activity type, or required a large number of known actives to elucidate models compared to those proposed here. The models presented by Markt et al. and Tanrikulu et al. were complex, with ten or more features used to identify binders [44,90]. The Guasch et al. and Petersen et al. models were specific to partial agonists, while the goal of the Paliwal et al. study was searching for antagonists [98-100]. A study by Sohn et al. yielded an agonist pharmacophore profile with three hydrogen bond acceptors and one hydrophobic group [101]. This study possessed an outcome similar to ours, and like ours, was limited in that only agonists, presumably of the full agonist activity class, could be profiled for binder identification. While these models are excellent for identifying binders that fit a specific binding mode, the use of individualized models can limit diverse ligand screening and binder prediction. Our inclusion of multiple models pushes the work of these groups to a larger realm of compound evaluation.

An examination of the ligands used to generate the pharmacophore models relative to the cumulative envelopes for each activity type indicated sampling of various key binding cavity regions (Figure 5). A bulky or aromatic group and proximal hydrogen bond acceptor group appeared to be necessary near H12 for full agonists. This matched what is seen with the TZD compound family. The full agonists also filled in more of arm III compared to fatty acids. With the fatty acids there are fewer bulky features influencing binding within arm I of the binding cavity. Additionally, the length of these compounds can fill the cavity while driving the acid group to the region where hydrogen bonding must occur for full agonism. The flexible nature of the aliphatic tail combined with the region of negative atom types near the entrance matched what is seen with activation induced by binding of two fatty acid chains and substituted fatty acids [14]. Partial agonists possess diverse feature group arrangement, and they appear to fill in the area of the binding cavity near the entrance along H3. This molecular localization agrees with published crystal structure data suggesting that H3 is involved in partial agonism and is independent of H12 [52]. The crystal structures available for antagonists bind in such a way that the area of the pocket underneath the hydrogen-bonding site was open. The antagonists appeared to bind in the same area of the binding cavity as the head group of full agonists, but differences in cavity shape and atom type prevalence may provide insight into how these two binder classes diverge. Unfortunately, the amount of structural information regarding binding of antagonists is not sufficient to draw firmer conclusions.

What we can determine by comparing binder location relative to the envelopes is the specific regions in the cavity where interactions occur. This information has the potential to clarify larger dynamic influences seen for specific ligand types, and would require molecular dynamics simulations. Without these simulations on large collections of structures, it is difficult to conclude how binding of each ligand type would influence conformational rearrangement for co-regulator recruitment. Recent advances in course-grained molecular dynamics might make a study of that scale and magnitude feasible.

A limitation exists when using interactions as a filter for activity type: lack of information about protein-ligand interactions that are truly indicative of a particular activity for all the activity types. Full agonist interaction information, which is the most established, only allows one to screen for compounds that are potential full agonists. Within this activity type, there are variations that can further limit which agonists are recognized without an extensive evaluation of differences in binding within the class. The same would be true for each of the other activity classes. This idea, though limiting, is also a positive factor when the issue of trying to predict activity type is considered. Only full agonists can be identified with full agonist key interaction information, which makes distinction from other activity types easier. Therefore, the true limitation is how much information is known about the interactions that dictate a specific activity type. What is not clear is which residues are required for partial agonists to bind and are unique to the partial agonism activity type. The same can be said for antagonists. For all the additional agonism types (dual agonist, pan agonist, partial agonist, and antagonist), more analyses would be needed to sift through the long list of known protein-ligand interactions taken from crystal structure data to establish specific residue lists for key interactions with each activity class.

Another valuable next step would be exploration of potential recruitment events that occure after the ligand binding event. This can be recruitment of binding partners like retinoid X receptor-α, or co-activators and co-repressors. It has been suggested that a single compound can induce different activation patterns depending on the tissue environment in which PPARγ is expressed in the presence of that ligand [19-21]. Ligand binding and the resulting interactions are the trigger for the recruitment event. The interactions that arise drive the conformational changes that influence H12 rearrangement and charge clamp formation to prep the surface of the protein for recruitment. Such a computational study would take the much needed exploration of the recruitment process from “proof of concept” [21] to an integrated part of the virtual screening process. Examination of mutation studies and structures, plus an in-depth mining of the literature to catalog biological outcomes of ligand-induced activation patterns would be the first step. Implementation of molecular dynamics techniques would provide theories as to why some ligands are capable of inducing co-activator recruitment, while others prevent release of or recruit co-repressors, and even more are capable of both depending on the tissue environment. The charge clamp on the surface of the protein that facilitates this process would be the focus of such a study and require simulation data for numerous ligand-bound PPARγ structures. Using conclusions from this sort of data in the lead selection process would help with the tailoring of selective modulators for therapeutic development.

There is a need to add to the binder identification criteria for screening partial agonists. This could be accomplished with a technique called steered molecular dynamics, in which interactions necessary to hold a ligand within the binding cavity are assessed based on the force required to liberate the ligand from the binding cavity. Examining more partial agonist-containing crystal structures might prove useful as well. Additional structures are available and more envelopes could be generated to see if other structures exist that might prove helpful for finding partial agonists. As a large number of these structures are missing areas that are potentially significant for the dynamics of the protein, rebuilding of the missing regions and sampling for favorable starting conformations would be necessary prior to any dynamic testing or additional envelope screening.

Conclusions

The prediction of PPARγ binders is a prominent area of research in recent years given therapeutic possibilities for numerous chronic diseases including type 2 diabetes, inflammatory bowel disease, and colorectal cancer. Often in the disease state, endogenous PPARγ ligands are present in low, ineffective physiological concentrations and are lower affinity binders compared to synthetic, high-affinity counterparts like rosiglitazone [102,103]. Additionally, use of the predominant synthetic compounds, thiazolidinediones, have been associated with severe adverse side effects that can cause additional negative medical outcomes and/or reduce quality of life [96,104-108]. Therefore, there is a need to find alternative therapeutic options that will improve health outcomes. Identifying binders as a first step toward therapeutic development encompasses the use of various traditional bench-top methods like high-throughput screening, isothermal titration calorimetry, and surface plasmon resonance, and computational methods like molecular docking and pharmacophore modeling. A combined and iterative approach of applied and theoretical methods is ideal for timely screening and discovery of compounds for therapeutic development. Further advancement of virtual screening scoring and ranking is necessary to improve predictive ability and potentially expedite the lead identification process with a smaller margin of error. In order to address this issue appropriately, structural information that translates to activity has to be considered and incorporated into the scoring process. We have shown that adding pharmacophore and interaction filtering to the binder criteria improved PPARγ binder identification while reducing unnecessary screening of compounds that would not fit the models or are unlikely to bind. Additionally, screening using multiple structure models that can accommodate different molecular shapes and comparing docking results for those structures improved the positive predictive rate. Overall, this study shows how accessible and efficient the virtual screening process can be for anyone interested in identifying small molecule leads for therapeutic development, while providing a novel collection of protocols for accomplishing that goal.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Lera Brannan for her initial efforts with identifying ligand-based pharmacophore screening methods to use for this study. The authors would also like to thank Dr. David Goodsell for his help with visualization techniques for the AutoLigand data.

The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources and technical support, which contributed to the results reported within this paper. URL: http://www.arc.vt.edu

Funding was provided by the National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health (NIH) (Grant number 1F31DK091186-01A1).

References

1. Tontonoz P, Spiegelman BM. Annual Review of Biochemistry. 2008; 77:289.
2. Gampe RT, Montana VG, Lambert MH, Miller AB, Bledsoe RK, Milburn MV, Kliewer SA, Willson TM, Xu HE. Molecular Cell. 2000; 5:545. [PubMed: 10882139]
3. Bassaganya-Riera J, Guri A, King J, Hontecillas R. Current Nutrition and Food Science. 2005; 1:179.
4. Guri AJ, Hontecillas R, Bassaganya-Riera J. Clinical Nutrition. 2006; 25:871. [PubMed: 17052808]
5. Rosen ED, Hsu CH, Wang X, Sakai S, Freeman MW, Gonzalez FJ, Spiegelman BM. Genes and Development. 2002; 16(1):22. [PubMed: 11782441]
6. Martin H. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2009; 669:1.
7. Shao D, Rangwala S, Bailey S, Krakow S, Reginato M, Lazar M. Nature. 1998; 396(6709):377. [PubMed: 9845075]
8. Wahli W. Journal of Internal Medicine. 2008; 263:613. [PubMed: 18479261]
9. Straus DS, Glass CK. TRENDS in Immunology. 2007; 28(12):551. [PubMed: 17981503]
10. Bassaganya-Riera J, Ferrer G, Casagran O, Sanchez S, de Horna A, Duran E, Orpi M, Guri AJ, Hontecillas R. e-SPEN, European e-Journal of Clinical Nutrition and Metabolism. 2009; 4:e90.
11. Bassaganya-Riera J, Mikyak S, Guri AJ, Hontecillas R. Cellular Immunology. 2009; 258:138. [PubMed: 19423085]
12. Lewis SN, Bassaganya-Riera J, Bevan DR. PPAR Research. 2010; 2010
13. Huang THW, Teoh AW, Lin BL, Lin DSH, Roufogalis B. Pharmacological Research. 2009; 60:195. [PubMed: 19646659]
14. Itoh T, Fairall L, Amin K, Inaba Y, Szanto A, Balint BL, Nagy L, Yamamoto K, Schwabe JWR. Nature Structural & Molecular Biology. 2008; 15(9):924.
15. Pochetti G, Godio C, Mitro N, Caruso D, Galmozzi A, Scurati S, Loiodice F, Fracchiolla G, Tortorella P, Laghezza A, Lavecchia A, Novellino E, Mazza F, Crestani M. Journal of Biological Chemistry. 2007; 282(23):17314. [PubMed: 17403688]
16. Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, Kurokawa R, Rosenfeld MG, Willson TM, Glass CK, Milburn MV. Nature. 1998; 395:137. [PubMed: 9744270]
17. Casimiro-Garcia A, Bigge CF, Davis JA, Padalino T, Pulaski J, Ohren JF, McConnell P, Kane CD, Royer LJ, Stevens KA, Auerbach BJ, Collard WT, McGregor C, Fakhoury SA, Schaum RP, Zhou H. Bioorganic & Medicinal Chemistry. 2008; 16(9):4883. [PubMed: 18394907]
18. Farce A, Renault N, Chavatte P. Current Medicinal Chemistry. 2009; 16(14):1768. [PubMed: 19442144]
19. Higgins LS, DePaoli AM. The American Journal of Clinical Nutrition. 2010; 91:267S. [PubMed: 19906796]
20. Shearer B, Billin A. Biochimica et Biophysica Acta. 2007; 1771(8):1082. [PubMed: 17602866]
21. Gelman L, Feige JN, Desvergne B. Biochimica et Biophysica Acta. 2007; 1771(8):1094. [PubMed: 17459763]
22. Kallenberger BC, Love JD, Krishna V, Chatterjee K, Schwabe JWR. Nature. 2003; 10(2):136.

23. Rosen ED, Spiegelman BM. The Journal of Biological Chemistry. 2001; 276(41):37731. [PubMed: 11459852]
24. Xu HE, Stanley TB, Montana VG, Lambert MH, Shearer BG, Cobb JE, McKee DD, Galardi CM, Plunket KD, Nolte RT, Parks DJ, Moore JT, Kliewer SA, Willson TM, Stimmel JB. Nature. 2002; 415:813. [PubMed: 11845213]
25. DeGrazia MJ, Thompson J, Vanden Heuvel JP, Peterson BR. Bioorganic & Medicinal Chemistry. 2003; 11:4325. [PubMed: 13129568]
26. Wu B, Gao J, Wang Mw. Acta Pharmacologica Sinica. 2005; 26(3):339. [PubMed: 15715931]
27. Klebe G. Drug Discovery Today. 2006; 11(13-14):580. [PubMed: 16793526]
28. Brozell SR, Mukherjee S, Balius TE, Roe DR, Case DA, Rizzo RC. Journal of Computer-Aided Molecular Design. 2012; 26(6):749. [PubMed: 22569593]
29. Irwin J, Shoichet B, Mysinger M, Huang N, Colizzi F, Wassam P, Cao Y. Journal of Medicinal Chemistry. 2009; 52(18):5712. [PubMed: 19719084]
30. Huang N, Shoichet BK, Irwin JJ. Journal of Medicinal Chemistry. 2006; 49(23):6789. [PubMed: 17154509]
31. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Journal of medicinal Chemistry. 2013; 55:6582. [PubMed: 22716043]
32. Zhang X, Wong SE, Lightstone FC. Journal of Computational Chemistry. 2013; 34:915. [PubMed: 23345155]
33. Cavasotto CN, Abagyan A. Journal of Molecular Biology. 2004; 337:209. [PubMed: 15001363]
34. Lyne PD. Drug Discovery Today. 2002; 7(20):1047. [PubMed: 12546894]
35. Stahura FL, Bajorath J. Current Pharmaceutical Design. 2005; 11:1189. [PubMed: 15853666]
36. Maltarollo V, Honorio K. Chemical Biology and Drug Design. 2012; 80(4):533. [PubMed: 22672760]
37. Schneidman-Duhovny D, Dror O, Inbar Y, Nussinov R, Wolfson H. Nucleic Acids Research. 2008; 36:W223. [PubMed: 18424800]
38. Yang SY. Drug Discovery Today. 2010; 15(11-12):444. [PubMed: 20362693]
39. Berman HM, Battistuz T, Bhat TN, Bluhm W, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Thanki N, Weissig H, Westbrook JD, Zardecki C. Acta Crystallographica Section D: Biological Crystallography. 2002; D58:899.
40. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyolov IN, Bourne PE. Nucleic Acids Research. 2000; 28:235. [PubMed: 10592235]
41. Harris R, Olson AJ, Goodsell DS. Proteins. 2007
42. Pettersen E, Goddard T, Huang C, Couch G, Greenblatt D, Meng E, Ferrin T. Journal of Computational Chemistry. 2004; 25(13):1605. [PubMed: 15264254]
43. Zoete V, Grosdidier A, Michielin O. Biochimica et Biophysica Acta. 2007; 1771:915. [PubMed: 17317294]
44. Markt P, Schuster D, Kirchmair J, Laggner C, Langer T. Journal of Computer-Aided Molecular Design. 2007; 21:575. [PubMed: 17960326]
45. Hopkins CR, O'Neil SV, Laufersweiler MC, Wang Y, Pokross M, Mekel M, Evdokimov A, Walter R, Kontoyianni M, Petrey ME, Sabatakos G, Roesgen JT, Richardson E, Demuth TP Jr. Bioorg Med Chem Lett. 2006; 16(21):5659. [PubMed: 16919947]
46. Chandra V, Huang P, Hamuro Y, Raghuram S, Wang Y, Burris TP, Rastinejad F. Nature. 2008; 456:350. [PubMed: 19043829]
47. Xu HE, Lambert MH, Montana VG, Plunket KD, Moore LB, Collins JL, Oplinger JA, Kliewer SA, Gampe RT Jr, McKee DD, Moore JT, Willson TM. Proceedings of the National Academy of Science. 2001; 98(24):13919.
48. Haffner C, Lenhard J, Miller A, McDougald D, Dwornik K, Ittoop O, Gampe R Jr, Xu H, Blanchard S, Montana V, Consler T, Bledsoe R, Ayscue A, Croom D. Journal of Medicinal Chemistry. 2004; 47:2010. [PubMed: 15056000]
49. Li Y, Choi M, Suino K, Kovach A, Daugherty J, Kliewer SA, Xu HE. Proceedings of the National Academy of Science. 2005; 102(27):9505.

50. Mahindroo N, Huang CF, Peng YH, Wang CC, Liao CC, Lien TW, Chittimalla SK, Huang WJ, Chai CH, Prakash E, Chen CP, Hsu TA, Peng CH, Lu IL, Lee LH, Chang YW, Chen WC, Chou YC, Chen CT, Goparaju CM, Chen YS, Lan SJ, Yu MC, Chen X, Chao YS, Wu SY, Hsieh HP. J Med Chem. 2005; 48(26):8194. [PubMed: 16366601]
51. Trump R, Cobb J, Shearer B, Lambert M, Nolte R, Wilson TM, Buckholz R, Zhao S, Leesnitzer L, Iannone M, Pearce K, Billin A, Hoekstra W. Science Direct. 2007; 17
52. Bruning JB, Chalmers MJ, Prasad S, Busby SA, Kamenechka TM, He Y, Nettles KW, Griffin PR. Structure. 2007; 15:1258. [PubMed: 17937915]
53. Montanari R, Saccoccia F, Scotti E, Crestani M, Godio C, Gilardi F, Loiodice F, Fracchiolla G, Laghezza A, Tortorella P, Lavecchia A, Novellino E, Mazza F, Aschi M, Pochetti G. Journal of medicinal Chemistry. 2008; 51(24):7768. [PubMed: 19053776]
54. Li Y, Kovach A, Suino-Powell K, Martynowski D, Xu HE. Journal of Biological Chemistry. 2008; 283(27):19132. [PubMed: 18469005]
55. Mueller J, Schupp M, Unger T, Kintscher U, Heinemann U (2011) No publication
56. Waku T, Shiraki T, Oyama T, Fujimoto Y, Maebara K, Kamiya N, Jingami H, Morikawa K. Journal of Molecular Biology. 2009; 385(1):188. [PubMed: 18977231]
57. Oyama T, Toyota K, Waku T, Hirakawa Y, Nagasawa N, Kasuga J, Hashimoto Y, Miyashi H, Morikawa K. Acta Crystallographica Section D: Biological Crystallography. 2009; 65:786.
58. Waku T, Shiraki T, Oyama T, Maebara K, Nakamori R, Morikawa K. EMBO Journal. 2010; 29:3395. [PubMed: 20717101]
59. Ohashi M, Oyama T, Nakagome I, Satoh M, Nishio Y, Nobusada H, Hirono S, Morikawa K, Hashimoto Y, Miyachi H. Journal of medicinal Chemistry. 2011; 54:331. [PubMed: 21128600]
60. Tomioka D, Hashimoto H, Sato M, Shimizu T (2011) No paper published
61. Artis DR, Lin JJ, Wang W, Mehra U, Perreault M, Erbe D, Krupka HI, England BP, Arnold J, Plotnikov AN, Marimuthu A, Nguyen H, Will S, Signaevsky M, Kral J, Cantwell J, Settachatgull C, Yan DS, Fong D, Oh A, Shi S, Womack P, Powell B, Habets G, West BL, Zhang KYJ, Milburn MV, Vlasuk GP, Hirth KP, Nolop K, Bollag G, Ibrahim PN, Tobin JF. PNAS. 2009; 106(1):262. [PubMed: 19116277]
62. Lin C, Peng Y, Coumar M, Chittimalla SK, Liao CC, Lyn P, Huang C, Lien TW, Lin W, Hsu J, Cheng J, Chen X, Wu J, Chao YS, Lee H, Juo C, Wu S, Hsieh HP. Journal of medicinal Chemistry. 2009; 52:2618. [PubMed: 19301897]
63. Connors RV, Wang Z, Harrison M, Zhang A, Wanska M, Hiscock S, Fox B, Dore M, Labelle M, Sudom A, Johnstone S, Liu J, Walker NPC, Chai A, Siegler K, Li Y, Coward P. Bioorganic & Medicinal Chemistry Letters. 2009; 19(13):3550. [PubMed: 19464171]
64. Fracchiolla G, Laghezza A, Piemontese L, Tortorella P, Mazza F, Montanari R, Pochetti G, Lavecchia A, Novellino E, Pierno S, Conte Camerino D, Loiodice F. Journal of medicinal Chemistry. 2009; 52(20):6382. [PubMed: 19775169]
65. Peng YH, Coumar M, Leou J, Wu J, Shiao H, Lin C, Lyu P, Hsieh HP, Wu SY (2011) To be published
66. Ostberg T, Svensson S, Selen G, Uppenberg J, Thor M, Sundbom M, Sydow-Backman M, Gustavsson A, Jendeberg L. Journal of Biological Chemistry. 2004; 279:41124. [PubMed: 15258145]
67. Burgermeister E, Schnoebelen A, Flament A, Benz J, Stihle M, Gsell B, Rufer A, Ruf A, Kuhn B, Maerki HP, Mizrahi J, Sebokova E, Niesor E, Meyer M. Molecular Endocrinology. 2006; 20(4): 809. [PubMed: 16373399]
68. Lu IL, Huang CF, Peng YH, Lin YT, Hsieh HP, Chen CT, Lien TW, Lee HJ, Mahindroo N, Prakash E, Yueh A, Chen HY, Goparaju CMV, Chen X, Liao CC, Chao YS, Jsu JTA, Wu SY. Journal of Medicinal Chemistry. 2006; 49(9):2703. [PubMed: 16640330]
69. Ambrosio AL, Dias SM, Polikarpov I, Zurier RB, Burstein SH, Garratt RC. J Biol Chem. 2007; 282(25):18625. [PubMed: 17462987]
70. Einstein M, Akiyama TE, Castriota GA, Wang CF, McKeever B, Mosley RT, Becker JW, Moller DE, Meinke PT, Wood HB, Berger JP. Mol Pharmacol. 2008; 73(1):62. [PubMed: 17940191]
71. Pochetti G, Montanari R, Mazza F, Crestani M, Godio C, Loiodice F, Fracchiolla G, Lavecchia A, Novellino E To be published

72. Oberfield J, Collins J, Holmes C, Goreham D, Cooper J, Cobb J, Lenhard J, Hull-Ryde E, Mohr C, Blanchard S, Parks D, Moore L, Lehmann J, Plunket K, Miller A, Milburn M, Kliewer S, Willson T. Proceedings of the National Academy of Sciences, USA. 1999; 96:6102.
73. Wakabayashi K, Hayashi S, Matsui Y, Matsumoto T, Furukawa A, Kuroha M, Tanaka N, Inaba T, Kanda S, Tanaka J, Okuyama R, Wakimoto S, Ogata T, Araki K, Ohsumi J. Biological & Pharmaceutical Bulletin. 2011; 34(7):1094. [PubMed: 21720019]
74. Motani A, Wang Z, Weiszmann J, McGee L, Lee G, Liu Q, Staunton J, Fang Z, Fuentes H, Lindstrom M, Liu J, Biermann D, Jaen J, Walker N, Learned R, Chen J, Li Y. Journal of Molecular Biology. 2009; 386(1301-1311)
75. Li Y, Wang Z, Furukawa N, Escaron P, Weiszmann J, Lee G, Lindstrom M, Liu J, Liu X, Xu H, Plotnikova O, Prasad V, Walker N, Learned R, Chen J. Journal of Biological Chemistry. 2008; 283(14):9168. [PubMed: 18263587]
76. Furukawa A, Arita T, Satoh S, Wakabayashi K, Hayashi S, Matsui Y, Araki K, Kuroha M, Ohsumi J. Bioorganic & Medicinal Chemistry Letters. 2010; 20(7):2095. [PubMed: 20219371]
77. Riu A, Grimaldi M, le Maire A, Bey G, Phillips K, Boulahtouf A, Perdu E, Zalko D, Bourguet W, Balaguer P. Environmental Health Perspectives. 2011; 119(9):1227. [PubMed: 21561829]
78. Riu A, le Maire A, Grimaldi M, Audebert M, Hillenweck A, Bourguet W, Balaguer P, Zalko D. Toxicological Sciences. 2011; 122(2):372. [PubMed: 21622942]
79. Inbar Y, Schneidman-Duhovny D, Dror O, Nussinov R, Wolfson H. Proc of RECOMB 3692 of Lecture Notes in Computer Science. 2007:423.
80. Molecular Operating Environment. 1010 Sherbooke St West, Suite #910, Montreal, QC, Canada: H3A 2R7: Chemical Computing Group Inc; 2012. 2012.10
81. Trott O, Olson AJ. Journal of Computational Chemistry. 2010; 31(2):455. [PubMed: 19499576]
82. Lewis SN, Brannan L, Guri AJ, Lu P, Hontecillas R, Bassaganya-Riera J, Bevan DR. PLoS One. 2011; 6(8):e24031. [PubMed: 21904603]
83. Schwartz, RL.; Phoenix, T.; Foy, BD. Learning Perl. 5th. O'Reilly Media, Inc.; 2008.
84. Kellenberger E, Rodrigo J, Muller P, Rognan D. Proteins. 2004; 57(255-242)
85. Triballeau N, Acher F, Brabet I, Pin JP, Bertrand HO. Journal of Medicinal Chemistry. 2005; 48(7):2534. [PubMed: 15801843]
86. Irwin JJ, Shoichet BK. Journal of Chemical Information and Modeling. 2005; 45:177. [PubMed: 15667143]
87. Sushko I, Salmina E, Potemkin V, Poda G, Tetko I. Journal of Chemical Information and Modeling. 2012; 52(8):2310. [PubMed: 22876798]
88. Kazius J, McGuire R, Bursi R. J Med Chem. 2005; 48(1):312. [PubMed: 15634026]
89. Oprea T. Journal of Computer-Aided Molecular Design. 2000; 14(3):251. [PubMed: 10756480]
90. Tanrikulu Y, Rau O, Schwarz O, Proschak E, Siems K, Muller-Kuhrt L, Schubert-Zsilavecz, Schneider G. ChemBIoChem. 2009; 10:75. [PubMed: 19067454]
91. Willson TM, Brown PJ, Sternbach DD, Henke BR. Journal of Medicinal Chemistry. 2000; 43(4): 527. [PubMed: 10691680]
92. Young PW, Buckle DR, Cantello BCC, Chapman H, Clapham JC, Coyle PJ, Haigh D, Hindley RM, Holder JC, Kallender H, Latter AJ, Lawrie KWM, Mossakowska D, Murphy GJ, Cox LR, Smith SA. Journal of Pharmacology and Experimental Therapeutics. 1998; 284(2):751. [PubMed: 9454824]
93. Kitchen DB, Decornez H, Furr JR, Bajorath J. Nature Reviews Drug Discovery. 2004; 3:935.
94. Lehmann J, Lenhard J, Oliver B, Ringold G, Kliewer S. Journal of Biological Chemistry. 1997; 272(6):3406. [PubMed: 9013583]
95. Wiggins J, Rajapakse R. Expert Opinion on Drug Metabolism and Toxicology. 2009; 5(10):1279. [PubMed: 19743890]
96. Lewis JD, Lichtenstein GR, Deren JJ, Sands BE, Hanauer SB, Katz JA, Lashner B, Present DH, Chuai S, Ellenburg JH, Nessel L, Wu GD. Gastroenterology. 2008; 134:688. [PubMed: 18325386]
97. Patil SA, Moss AC. Expert Review of Gastroenterology and Hepatology. 2008; 2(2):177. [PubMed: 19072352]

98. Guasch L, Sala E, Castell-Auví A, Cedó L, Liedl KR, Wolber G, Muehlbacher M, Mulero M, Pinent M, Ardévol A, Valls C, Pujadas G, Garcia-Vallvé S. PLoS One. 2012; 7(11):e50816. [PubMed: 23226391]

99. Paliwal S, Yadav D, Yadav R, Paliwal S. Medicinal Chemistry Research. 2011; 20:656.

100. Petersen R, Christensen K, Assimopoulou A, Fretté X, Papageorgiou V, Kristiansen K, Kouskoumvekaki I. Journal of Computer-Aided Molecular Design. 2011; 25(2):107. [PubMed: 21069556]

101. Sohn, Ys; Lee, Y.; Park, C.; Hwang, S.; Kim, S.; Baek, A.; Son, M.; Suh, JK.; Kim, HH.; Lee, KW. Bulletin of the Korean Chemical Society. 2011; 32(1):201.

102. Bassaganya-Riera J, Hontecillas R. Clinical Nutrition. 2006; 25:454. [PubMed: 16698153]

103. Lakatos P. World Journl of Gastroenterology. 2006; 12(38):6102.

104. Azoulay L, Yin H, Filion KB, Assayag J, Majdan A, Pollak MN, Suissa S. British Medical Journal. 2012; 344:e3645. [PubMed: 22653981]

105. Lewis JD, Ferrara A, Peng T, Hedderson M, Bilker WB, Quesenberry CPJ, Vaughn DJ, Nessel L, Selby J, Strom BL. Diabetes Care. 2011; 34(4):916. [PubMed: 21447663]

106. Cohen J. Diabetologia. 2006; 49(6):1454. [PubMed: 16601971]

107. Henney JE. JAMA: Journal of the American Medical Association. 2000; 283(17):2228.

108. Nissen SE, Wolski K. Archives of Internal Medicine. 2010; 170(14):1191. [PubMed: 20656674]

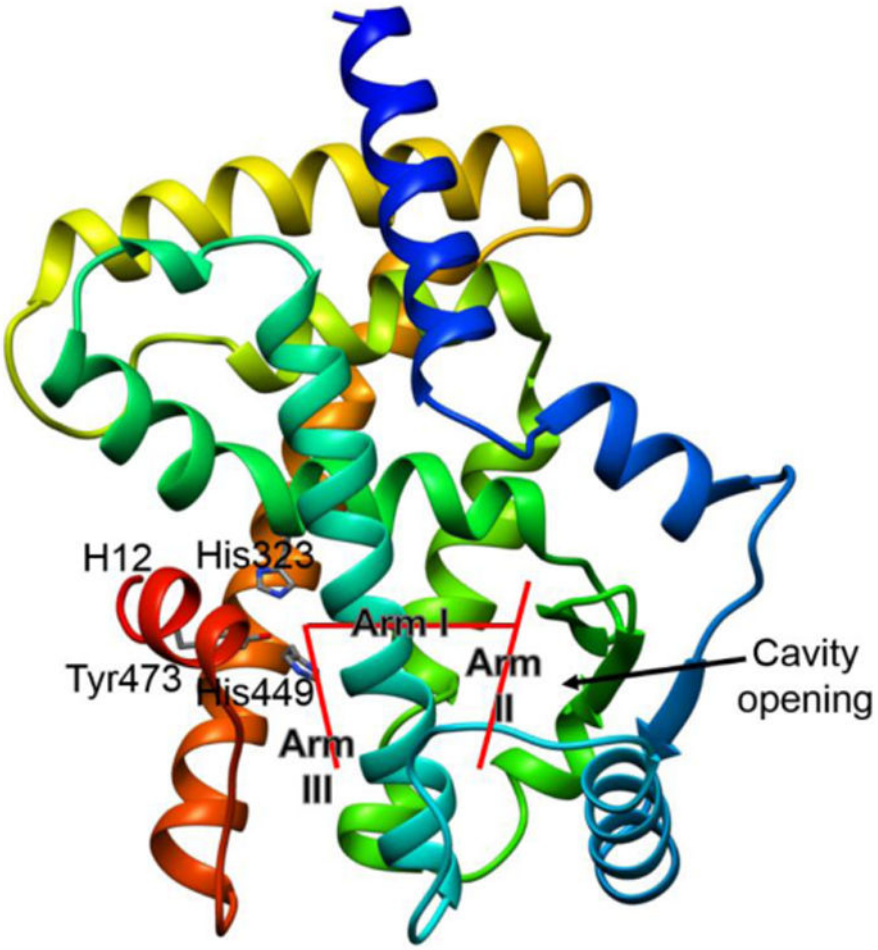


Fig. 1.

Rainbow-colored ribbon representation of PPARγ showing structural features of the binding cavity. The arms of the binding cavity, helix H12, and three hydrogen-bonding residues (His323, His449, and Tyr473) are labeled. USCF Chimera was used to create this image and Microsoft Office PowerPoint was used to label the image

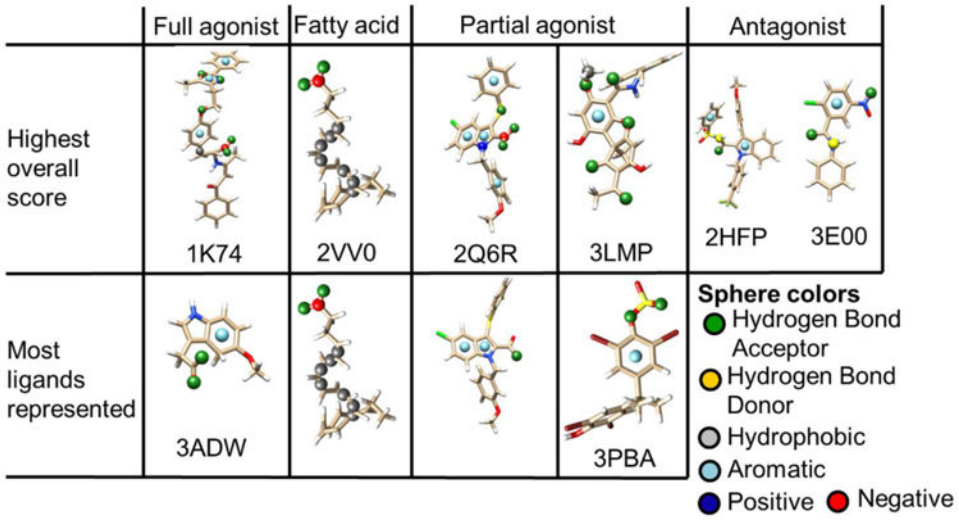


Fig. 2.

Composite of PharmaGist feature groups overlaid onto representative ligands from each ligand type. The protein data bank (PDB) ID in which each ligand was found is listed in the corresponding cell. Cells without a PDB ID are the same ligand as the one listed in the top cell for a given column. Sphere colors for the feature group are as indicated in the legend. Each ligand is shown in stick representation with atom-specific coloring: carbon (tan), oxygen (red), nitrogen (blue), hydrogen (white), and sulfur (yellow). The ligand for 3PBA contained bromine, which is shown in dark red. UCSF Chimera was used to generate the 3D images of the ligands, and Microsoft Office PowerPoint was used to combine the figure components.

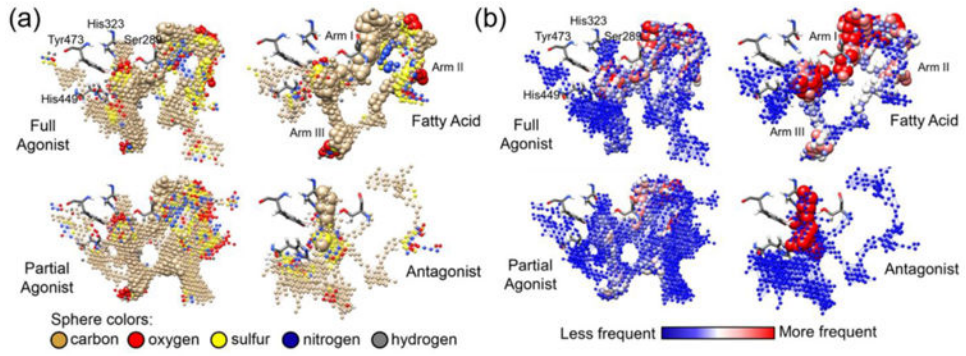


Fig. 3.

Compiled envelopes for each activity class shown as (a) an atom-specific visual histogram and (b) grid point-specific heat map-like representation. The four hydrogen-bonding residues and the arm designation for the binding cavity are shown for each panel on the top two envelopes. The intersection of arms I and II is the entrance to the cavity, while the intersection of arms I and III is the far end of the cavity near helix 12. (a) Each atom type is shown with atom-specific coloring indicated by the legend. Sphere size increases with frequency at which the indicated atom type appeared within the most energetically favorable envelopes for all structures within the activity class. (b) The spheres represent the grid points of all the crystal structures tested. Sphere size increases with frequency at which the grid point appeared (atom type neglected) within the most energetically favorable envelopes for all structures within the activity category. Sphere color also indicates frequency of point appearance as the color progresses from blue to red, which is indicated by the color scale under the figure. UCSF Chimera was used to generate the envelope images, and Microsoft PowerPoint was used to label the images.

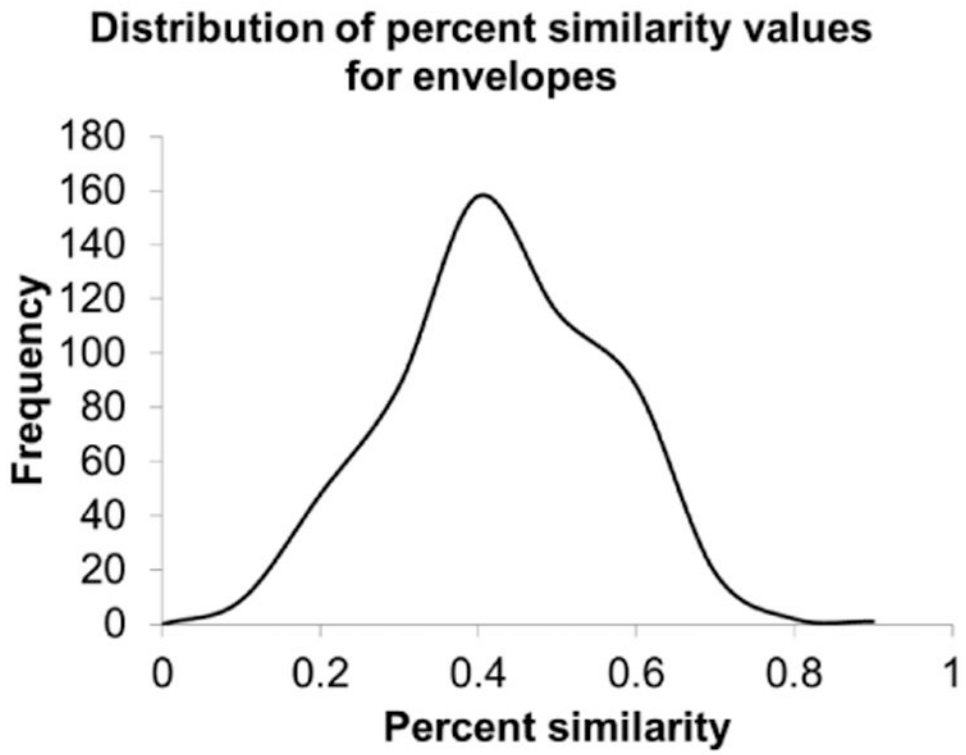


Fig. 4.

Distribution of percent similarity values for the 33 consensus envelopes generated using protein data bank (PDB) crystal structures. Percent similarity reflects the proportion of common points between each pair of envelopes (N = 1089 pairs). Microsoft Excel was used to generate the graph

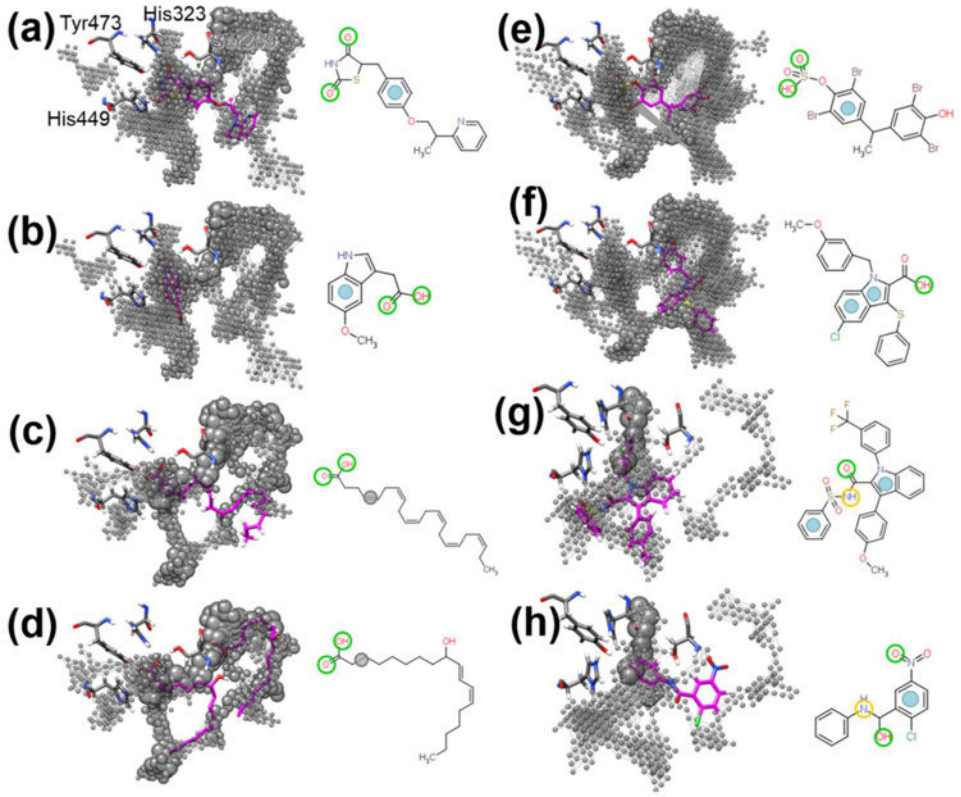
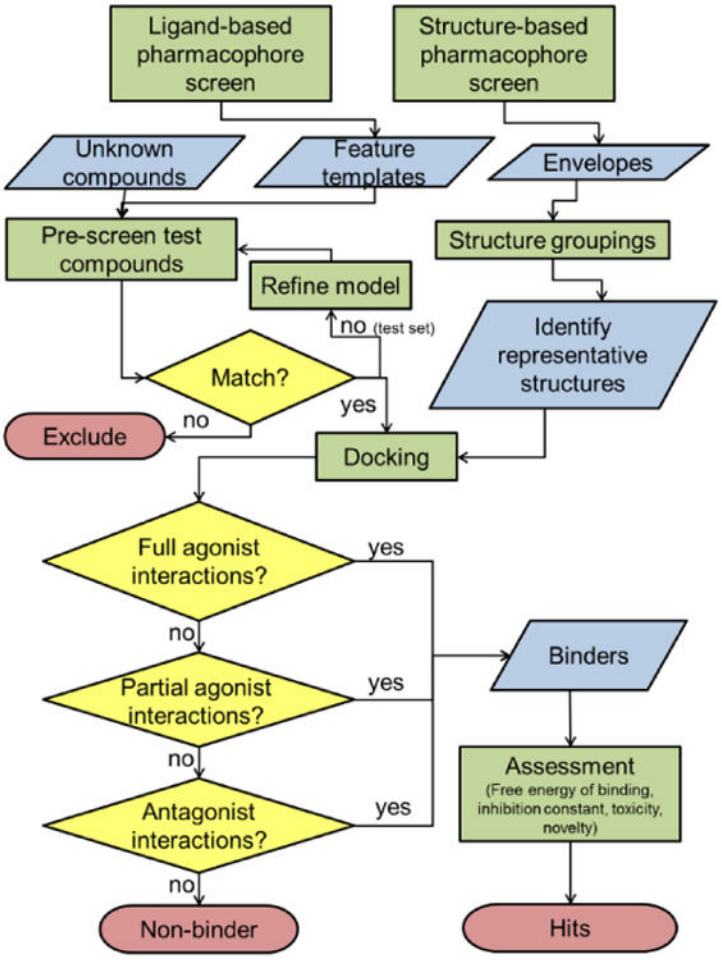


Fig. 5.

The ligands used for generating pharmacophore models in MOE relative to the ligand-specific envelopes. The full agonist envelope with (a) the 3ADW ligand MYI or (b) the 1FM6 ligand BRL. The fatty acid envelope with (c) the 2VV0 ligand HXA or (d) the 2VSR ligands 9HO. The partial agonist envelope with (e) the 3PBA ligand ZGX or (f) the 2Q6R ligand SF2. The antagonist envelope with (g) the 2HFP ligand NSI or (h) the 3E00 ligand GW9. 2D drawings of the ligands are shown to the right of the envelope for each panel with colored circles showing placement of the pharmacophore feature groups used for generating the models. Circle colors for the feature groups are cyan (aromatic), green (hydrogen bond acceptor), yellow (hydrogen bond donor), and gray (hydrophobic). All envelopes are shown as gray, transparent spheres relative to the crystal structure-derived pose for each ligand. Ligands in the 3D image are in stick representation with atom-specific coloring: carbon = magenta, oxygen = red, nitrogen = blue, sulfur = yellow, bromine = dark red, chlorine = green, hydrogen = white. The hydrogen-bonding residues are shown as sticks as well. Atom specific coloring was the same except the carbon atoms are gray. UCSF Chimera was used to generate the envelope images, MarvinSketch was used to make the two-dimensional ligand images, and Microsoft PowerPoint was used to compile and label the figure.



Scheme 1.

Process flow diagram for combined pharmacophore screening and docking methods to identify binders. The diagram includes nodes for assessing the test set to establish pharmacophores and nodes for incorporation of unknown compound data. Green boxes were used for processes, blue polygons represent data, yellow diamonds signify decision nodes, and red ovals indicate terminal nodes. Microsoft PowerPoint was used to create the flow chart.

Table 1

List of protein data bank (PDB) structures used for this study. Structures are listed by ligand type and include PDB identifiers (PDB ID), ligand identifier (Lig ID), ligand name, resolution (Res.) in Angstroms for structures used, citation for the PDB entry (Cit.), and indication of whether the receptor (Rec.), ligand (Lig.), or both were used in the study

PDB ID	Lig ID	Ligand Name	Ligand Type	Res. (Å)	Cit.	Rec.	Lig.
2HFP	NSI	3-(4-methoxyphenyl)-N-(phenylsulfonyl)-1-[3-(trifluoromethyl)benzyl]-1H-indole-2-carboxamide	antagonist	2.0	[45]	x	x
3E00	GW9	2-chloro-5-nitro-N-phenylbenzamide	antagonist	3.1	[46]	x	x
2VSR	9HO	(9S,10E,12Z)-9-hydroxyoctadeca-10,12-dienoic acid	fatty acid	2.05	[14]	x	x
2VST	243	(9Z,11E,13S)-13-hydroxyoctadeca-9,11-dienoic acid	fatty acid	2.35	[14]	x	x
2V_V0	HXA	docosa-4,7,10,13,16,19-hexaenoic acid	fatty acid	2.55	[14]	x	x
2V_V1	4HD	(4S,5E,7Z,10Z,13Z,16Z,19Z)-4-hydroxydocosa-5,7,10,13,16,19-hexaenoic acid	fatty acid	2.2	[14]	x	x
2V_V2	5HE	(5R,6E,8Z,11Z,14Z,17Z)-5-hydroxyicosa-6,8,11,14,17-pentaenoic acid	fatty acid	2.75	[14]	x	x
1FM6[1]	BRL	2,4-thiazolidinedione, 5-[[4-[2-(methyl-2-pyridinylamino) ethoxy] pheynyl] methyl]-(9CL)	full	2.1	[2]	x	x
1FM9	570	2-(2-benzoyl-phenylamino)-3-{4-[2-(5-methyl-2-phenyl-oxazol-4-yl)-ethoxy]-phenyl}-propionic acid	full	2.1	[2]	x	x
1I74	544	2-(1-methyl-3-oxo-3-phenyl-propylamino)-3-{4-[2-(5-methyl-2-phenyl-oxazol-4-yl)-ethoxy]-phenyl}-propionic acid	full	2.3	[47]	x	x
1FDT	570	2-(2-benzoyl-phenylamino)-3-{4-[2-(5-methyl-2-phenyl-oxazol-4-yl)-ethoxy]-phenyl}-propionic acid	full	2.4	[48]	x	
1ZEY	BRL	2,4-thiazolidinedione, 5-[[4-[2-(methyl-2-pyridinylamino) ethoxy] phenyl] methyl]-(9CL)	full	1.8	[49]	x	
2AT_H	3EA	2-{5-[3-(7-propyl-3-trifuoromethylbenzo[D]isoxaz ol-6-yloxy)propoxy]indol-1-yl}ethanoic acid	full		[50]		x
2ZNJ	DRJ	(2R)-2-(4-{2-[1,3-benzoxazol-2-yl(heptyl)amino]ethyl}phenox y)-2-methylbutanoic acid	full	2.1	[15]	x	x
2PQB	GW4	N-[(2S)-2-[(2-benzoylphenyl)amino]-3-{4-[2-(5-methyl-2-phenyl-1,3-oxazol-4-yl)ethoxy]phenyl}propyl]acet amide	full	2.3	[51]	x	x
2PRG	BRL	2,4-thiazolidinedione, 5-[[4-[2-(methyl-2-pyridinylamino) ethoxy] phenyl] methyl]-(9CL)	full	2.3	[16]	x	
2Q59	240	(2S)-2-(2-{[1-(4-methoxybenzol)-2-methyl-5-(trifluoromethoxy)-1H-indol-3-yl]methyl}phenoxy)propanoic acid	full		[52]		x
3B3K	LRG	(2S)-2-(biphenyl-4-yloxy)-3-phenylpropanoic acid	full	2.6	[53]	x	x
3CS8	BRL	2,4-thiazolidinedione, 5-[[4-[2-(methyl-2-pyridinylamino) ethoxy] phenyl] methyl]-(9CL)	full	2.3	[54]	x	
3DZY	BRL	2,4-thiazolidinedione, 5-[[4-[2-(methyl-2-pyridinylamino) ethoxy] phenyl] methyl]-(9CL)	full	3.1	[46]	x	
2XKW	P1B	(5R)-5-{4-[2-(5-ethylpyridin-2-yl)ethoxy]benzyl}-1,3-thiazolidine-2,4-dione	full		[55]		x
2ZK1	PTG	(5E,14E)-11-oxoprosta-5,9,12,14-tetraen-1-oic acid	full		[56]		x
2ZNO	S44	(2S)-2-(4-propoxy-3-{[({4-[(3S,5S,7S)-tricyclo[3.3.1.1∼3,7∼]dec-1-yl]phenyl}carbonyl)amino]me thyl}benzyl)butanoic acid	full		[57]		x
3ADW	MYI	(5-methoxy-1H-indol-3-yl)acetic acid	full		[58]		x
3ADX	IMN	indomethacin	full		[58]		x

J Comput Aided Mol Des. Author manuscript; available in PMC 2016 March 03.

PDB ID	Lig ID	Ligand Name	Ligand Type	Res. (Å)	Cit.	Rec.	Lig.
3AN4	M7R	(2R)-2-benzyl-3-(4-propoxy-3-{[({4-[(3S,5S,7S)-tricyclo[3.3.1.1∼3,7∼]dec-1-yl]phenyl}carbonyl)amino]me thyl}phenyl)propanoic acid	full		[59]		x
3B0Q	MC5	(5S)-5-({6-[(2-fluorobenzyl)oxy]naphthalen-2-yl} methyl)-1,3 -thiazolidine-2,4-dione	full		[60]		x
3CDS	GRR	(2S)-2-(4-ethylphenoxy)-3-phenylpropanoic acid	full		[53]		x
3ET0	ET0	3-(5-methoxy-1H-indol-3-yl)propanoic acid	full		[61]		x
3GBK	2PQ	2-[(1-{3-[4-(biphenyl-4-ylcarbonyl)-2-propylphenoxy]propyl}-1,2,3,4-tetrahydroquinolin-5-yl)oxy]-2-methylpropanoic acid	full		[62]		x
3HO0	DKD	(2S)-2-(4-phenethylphenoxy)-3-phenyl-propanic acid	full		[63]		x
3HOD	ZZH	(2S)-2-(4-benylphenoxy)-3-phenylpropanoic acid	full		[64]		x
3NOA	5BC	(5-{3-[4-(biphenyl-4-ylcarbonyl)-2-propylphenoxy]propoxy}-1H-indol-1-yl)acetic acid	full		[65]		x
1WM0	PLB	2-[(2,4-dichlorobenzoyl)amino]-5-(pyrimidin-2-yloxy)benzoic acid	partial	2.9	[66]	x	x
2FVJ	RO0	1-(3,4-dimethoxybenzyl)-6,7-dimethoxy-4-{[4-(2-methoxyphenyl)piperidin-1-yl]methyl}isoquinoline	partial	1.99	[67]	x	x
2G0G	SP0	3-fluoro-n-[1-(4-fluorophenyl)-3-(2-thienyl)-1h-pyrazol-5-yl]benzenesulfonamide	partial	2.54	[68]	x	x
2G0H	SP3	N-[1-(4-fluorophenyl)-3-(2-thienyl)-1h-pyrazol-5-yl]-3,5-bis(trifluoromethyl)benzenesu lfonamide	partial	2.3	[68]	x	x
2I4P	DRH	(2S)-2-(4-{2-[1,3-benzoxazol-2-yl(heptyl)amino]ethyl}phenox y)-2-methylbutanoic acid	partial	2.1	[15]	x	x
2I4Z	DRH	(2S)-2-(4-{2-[1,3-benzoxazol-2-yl(heptyl)amino]ethyl}phenox y)-2-methylbutanoic acid	partial	2.25	[15]	x	
2OM9	AJA	(6AR,10AR)-3-(1,1-dimethylheptyl)-1-hydroxy-6,6-dimethyl-6a,7,10,10a-tetrahydro-6h-benzo[C]chromene-9-carboxylic acid	partial	2.8	[69]	x	
2P4Y	C03	(2R)-2-(4-chloro-3-{[3-(6-methoxy-1,2-benzisoxazol-3-yl)-2-methyl-6-(trifluoromethoxy)-1h-indol-1-yl]methyl}sphenoxy)propanoi c acid	partial	2.25	[70]	x	x
2Q5P	241	(2S)-2-(3-{[1-(4-methoxybenzoyl)-2-methyl-5-(trifluoromethoxy)-1h-indol-3-yl]methyl}phenoxy)propanoic acid	partial	2.3	[52]	x	x
2Q5S[1]	NZA	5-chloro-1-(4-chlorobenzyl)-3-(phenylthio)-1h-indole-2-carboxylic acid	partial	2.05	[52]	x	x
2Q61	SF1	1-benzyl-5-chloro-3-(phenylthio)-1H-indole-2-carboxylic acid	partial				x
2Q6R	SF2	5-chloro-1-(3-methoxybenzyl)-3-(phenylthio)-1h-indole-2-carboxylic acid	partial	2.41	[52]	x	x
2Q6S	PLB	2-[(2,4-dichlorobenzoyl)amino]-5-(pyrimidin-2-yloxy)benzoic acid	partial	2.4	[52]	x	
3CDP	YRG	(2S)-2-(4-chlorophenoxy)-3-phenylpropanoic acid	partial	2.8	[71]	x	x
3DZU	PLB	2-[(2,4-dichlorobenzoyl)amino]-5-(pyrimidin-2-yloxy)benzoic acid	partial	3.2	[46]	x	
4PRG	072	(+/-)(2S,5S)-3-(4-(4-carboxyphenyl)butyl)-2-heptyl-4-oxo-5-thiazolidine	partial	2.9	[72]	x	x
2ZK6	C08	difluoro(5-{2-[(5-octyl-1H-pyrrol-2-yl-kappan)methylidene]-2H-pyrrol-5-yl-kappan}pentanoato)boron	partial		[56]		x
3B1M	KRC	(9AS)-8-acetyl-N-[(2-ethylnaphthalen-1-yl)methyl]-1,7-dihydroxy-3-methoxy-9A-methyl-9-oxo-9,9A-dihydrodibenzo[B,D]furan-4-carboxamide	partial		[73]		x
3FUR	Z12	2,4-dichloro-N-[3,5-dichloro-4-(quinolin-3-yloxy)phenyl]benzenesulfona mide	partial		[74]		x
3H0A	D30	[(4-{[2-(pent-2-yn-1-yloxy)-4-{[4-(trifluoromethyl)phenoxy]met hyl}phenyl]sulfanyl}-5,6,7,8-tetrahydronaphthalen-1-yl)oxy]acetic acid	partial		[63]		x
3K8S	Z27	2-chloro-N-{3-chloro-4-[(5-chloro-1,3-benzothiazol-2-yl)sulfanyl]phenyl}-4-(trifluoromethyl)benzenesulfo namide	partial		[75]		x

J Comput Aided Mol Des. Author manuscript; available in PMC 2016 May 01.

PDB ID	Lig ID	Ligand Name	Ligand Type	Res. (Å)	Cit.	Rec.	Lig.
3LMP	CEK	(9AS)-8-acetyl-1,7-dihydroxy-3-methoxy-9A-methyl-N-(1-naphthylmethyl)-9-oxo-9,9A-dihydrodibenzo[B,D]furan-4-carboxamide	partial		[76]		x
3OSI	XDH	4,4′-propane-2,2-diylbis(2,6-dichlorophenol)	partial		[77]		x
3OSW	XDI	4,4′-propane-2,2-diylbis(2,6-dibromophenol)	partial		[77]		x
3PBA	ZXG	2,6-dibromo-4-[2-(3,5-dibromo-4-hydroxyphenyl)propan-2-yl]phenyl hydrogen sulfate	partial		[78]		x

[1] Ligands included in grouping with antagonists to fulfill PharmaGist server requirement of at least three compounds.

Table 2

Feature scores for ligands submitted to the PharmaGist server. The highest overall scores and highest scores for the feature cluster with the most ligands are listed for the full agonist, fatty acid, and partial agonist categories. The pair-wise scores for the antagonists are included, which are relative to the indicated ligand (ligand ID followed by Protein data bank [PDB] ID) as the pivot molecule. The number of features reported for each group is reported in parentheses.

	Full agonist	Fatty acid	Partial agonist (21 ligands)	Partial agonist (19 ligands)	Antagonist
Highest overall score	48.5 (10)	27.2 (10)	39.2 (9)	39.9 (7)	NSI (2HFP): 8.0 (4) GW9 (3E00): 10.0 (4)
Most ligands represented	33.6 (3)	27.2 (10)	20.0 (3)	28.9 (3)	

Table 3

Groupings for structure models based on envelope similarity. Protein Data Bank (PDB) and ligand identifiers match those used in Table 1

Group #	PDB ID	Pharmacophore model ligand ID
1	1FM6, 1ZGY, 2PRG, 2VSR, 2VV2	BRL[1], MYI[1], 9HO[3], HXA[3]
2	1FM9, 1K74, 1RDT, 2I4Z, 2POB	BRL[1], MYI[1], SF2[2], ZXG[2]
3	2I4J, 2VSR, 2VST, 2VV0, 2VV1, 2VV2	SF2[2], ZXG[2], 9HO[3], HXA[3]
4	2I4Z, 2P4Y, 2Q5P	SF2[2], ZXG[2]
5	2Q5S, 2Q6R	SF2[2], ZXG[2]
6	1WM0, 2FVJ, 2G0G, 2G0H, 2HFP, 2I4P, 2OM9, 2Q6S, 3B3K, 3CDP, 3CS8, 3DZU, 3DZY, 3E00, 4PRG	

Activity type indicated by the following:

[1] full agonist,

[2] partial agonist,

[3] fatty acid

Table 4

Root mean-squared deviation (RMSD) values in Angstroms for re-docking of native ligands into corresponding crystal structure models. Results are listed by Protein Data Bank (PDB) ID. Representative structures were those from the envelope similarity groups that were appropriate for docking

PDB ID	RMSD	PDB ID	RMSD	PDB ID	RMSD	PDB ID	RMSD
1FM6	3.15	2P4Y	2.56	2I4J	3.06	2VV0	2.20
1FM9	3.16	2POB[1]	1.33	2I4P	2.83	2VV1	2.47
1K74	2.97	2Q5P[1]	1.01	2I4Z	3.04	2VV2	2.94
1RDT	2.97	2Q5S	2.40	20M9	2.49	3B3K	3.35
1WM0[1]	0.95	2VSR	2.34	2PRG[1]	1.17	3CDP	3.46
1ZGY[1]	1.25	3DZU[1]	1.10	2Q6R	2.08	3CS8	2.91
2G0G[1]	0.92	4PRG[1]	1.75	2Q6S	4.13	3DZY	3.08
2G0H[1]	1.08	2FVJ[1]	0.56	2VST	2.80	3E00	2.41
2HFP	3.28						

[1]Indicates structures that showed successful re-docking results (RMSD < 2.0 Å).

Table 5

True positive rates (TPRs) for each Protein Data Bank (PDB) structure in the training phase expressed as percentages. Successful poses were those that satisfied the RMSD cutoff of 2.0 Å. The original TPRs were calculated with the PDB ligands, and the adjusted TPRs were calculated using the subset of ligands that matched the rosiglitazone-based pharmacophore model. The percent change was calculated as the difference between the original and adjusted TPRs divided by the original value

PDB ID	Original TPR	Adjusted TPR	Percent change	PDB ID	Original TPR	Adjusted TPR	Percent change
1FM6	7.29	10.64	45.90	2Q5P	4.17	6.38	53.19
1FM9	12.50	17.02	36.17	2Q5S	3.13	2.13	-31.91
1K74	17.71	21.28	20.15	2Q6R	4.17	2.13	-48.94
1RDT	18.75	21.28	13.48	2Q6S	4.17	0.00	-100.00
1WM0	10.42	2.13	-79.57	2VSR	12.50	14.89	19.15
1ZGY	13.54	21.28	57.12	2VST	4.17	6.38	53.19
2OM9	5.21	4.26	-18.30	2VV0	11.46	12.77	11.41
2FVJ	8.33	8.51	2.13	2VV1	8.33	8.51	2.13
2G0G	9.38	17.02	81.56	2VV2	4.17	4.26	2.13
2G0H	5.21	10.64	104.26	3B3K	4.17	6.38	53.19
2HFP	3.13	4.26	36.17	3CDP	2.08	2.13	2.13
2I4J	10.42	6.38	-38.72	3CS8	7.29	2.13	-70.82
2I4P	10.42	6.38	-38.72	3DZU	4.17	0.00	-100.00
2I4Z	9.38	8.51	-9.22	3DZY	4.17	6.38	53.19
2P4Y	5.21	4.26	-18.30	3E00	6.25	10.64	70.21
2POB	15.63	21.28	36.17	4PRG	4.17	4.26	2.13
2PRG	10.42	21.28	104.26				

Table 7

Top-ranked results from the docking of the MicroSource U.S. Drugs compounds into 1ZGY for which docked poses were deemed most successful (e.g., most negative free energy of binding). All compounds possessed interactions with all four hydrogen-bonding residues. Compounds are listed by ZINC ID and compound name. The following two columns indicate if the compound was classified as toxic using the Molecular Operating Environment program and the ToxAlerts server. The characters “+” and “–” indicate if the compound matched or did not match, respectively, the ligand-based pharmacophore model listed (MYI, BRL, 9HO, or HXA). The remainder of the table lists the patent number in which the compound was referenced if one was found, the calculated free energy of binding score, and the calculated inhibition constant

ZINC ID	Compound Name	Tox. MOE	Tox Alert	Patent	MYI	BRL	9HO	HXA	Free Energy of Binding (kcal/mol)
18456289	Folic acid			EP1605950 A4	-	+	+	+	-9.8
01529323	methotrexate			EP1959950 A1	-	+	+	+	-9.6
01540998	pemetrexed disodium			US8362075 B2	-	+	+	+	-9.6
03956919	bezafibrate		yes	EP2089023 A2	-	-	+	+	-9
03952881	balsalazide disodium	yes	yes		-	-	+	+	-8.9
00968326	pioglitazone hydrochloride			10 different patents	-	+	-	-	-8.7
00086535	flufenamic acid		yes	EP2303252 A1	-	-	-	+	-8.5
35973845	Ceftriaxone sodium trihydrate	yes	yes		-	-	-	+	-8.5
00968328	rosiglitazone		yes	US5002953 (expired 2012)	-	+	-	-	-8.4