

# **Refinement of the Docking Component of Virtual Screening for PPAR $\gamma$ Therapeutics Using Pharmacophore Analysis and Molecular Dynamics**

Stephanie Nicole Lewis

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State  
University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy  
In  
Genetics, Bioinformatics, and Computational Biology

David R. Bevan, Chair  
Josep Bassaganya-Riera  
Jill Sible  
Liqing Zhang

June 11, 2013  
Blacksburg, VA

Keywords: Virtual screening, PPAR $\gamma$ , Drug discovery, Molecular docking,  
Pharmacophore screening, Molecular dynamics

Copyright 2013, Stephanie Nicole Lewis

# Refinement of the Docking Component of Virtual Screening for PPAR $\gamma$ Therapeutics Using Pharmacophore Analysis and Molecular Dynamics

Stephanie Nicole Lewis

## ABSTRACT

Exploration of peroxisome proliferator-activated receptor-gamma (PPAR $\gamma$ ) as a drug target holds applications for treating a wide variety of chronic inflammation-related diseases. Type 2 diabetes (T2D), which is a metabolic disease influenced by chronic inflammation, is quickly reaching epidemic proportions. Although some treatments are available to control T2D, more efficacious compounds with fewer side effects are in great demand. Drugs targeting PPAR $\gamma$  typically are compounds that function as agonists toward this receptor, which means they bind to and activate the protein. Identifying compounds that bind to PPAR $\gamma$  (i.e. binders) using computational docking methods has proven difficult given the large binding cavity of the protein, which yields a large target area and variations in ligand positions within the binding site. We applied a combined computational and experimental concept for characterizing PPAR $\gamma$  and identifying binders. The goal was to establish a time- and cost-effective way to screen a large, diverse compound database potentially containing natural and synthetic compounds for PPAR $\gamma$  agonists that are more efficacious and safer than currently available T2D treatments. The computational molecular modeling methods used include molecular docking, molecular dynamics, steered molecular dynamics, and structure- and ligand-based pharmacophore modeling. Potential binders identified in the computational component funnel into wet-lab experiments to confirm binding, assess activation, and test preclinical efficacy in a mouse model for T2D and other chronic inflammation diseases. The initial process used provided  $\alpha$ -eleostearic acid as a compound that ameliorates inflammatory bowel disease in a pre-clinical trial. Incorporating pharmacophore analyses and binding interaction information improved the method for use with a diverse ligand database of thousands of compounds. The adjusted methods showed enrichment for full agonist binder identification. Identifying lead compounds using our method would be an efficient means of addressing the need for alternative T2D treatments.

## DEDICATION

I would like to dedicate this dissertation to all the people in my life, both family and friends, who have done everything humanly possible to help me achieve my goals. Among those are most notably my mother, Stephanie R. Lewis, my grandmother, Bessie White, my husband, Steve Huff, our son, Ayden James, and my brothers, Darin and Daniel Lewis. Thank you for being my daily cheerleaders, entertainers, and taking pride in everything about me including my shortcomings. To my best friend and soul sister, Adria Markowski, thank you for taking the PhD journey with me, and showing me that greatness can be both dreamed and achieved. Thank you to Dennie Munson for being my mom away from home and my champion in all things academic. Thank you to my advisor, David Bevan, for seeing greatness in me, and allowing a level of freedom and independence that seems far too uncommon, but highly necessary, in graduate level education. Your confidence in me has truly helped my development as a leader, mentor, and scientist. Thank you to my amazing colleagues in the Bevan lab for being exceptional sounding boards for ideas, both great and not so great. Thank you to my graduate committee for constantly challenging my thought process and conclusions, and the PREP/IMSD faculty for facilitating professional guidance. The conversations with each of you have shaped my development as a researcher and guided me toward high-impact scientific thought. Thank you to all the students with whom I have worked while at Virginia Tech. It is through you that I see great potential and promise in the scientific endeavors to come. Lastly, I dedicate this to those who did not believe I could rise to this level of success. Through their lack of faith for various reasons at various times, I have been motivated to exceptional belief in my abilities and myself.

## ACKNOWLEDGEMENTS

I would like to acknowledge and thank the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health (NIH) for funding through the Ruth L. Kirschstein National Research Service Awards for Individual Predoctoral Fellowships to Promote Diversity in Health-Related Research (1F31DK091186-01A1).

I would like to acknowledge and thank the VT-PREP and IMSD programs and the faculty of the program for their financial and developmental support. These programs are funded through NIH: Biomedical and Behavioral Sciences Research Training Grant R25 GM072767 (Virginia Tech Initiative to Maximize Student Diversity) and Grant R25 GM066534 (Virginia Tech Post Baccalaureate Research and Education Program).

Additional sources responsible for funding of the experimental component were Grant 5R01AT4308 from the National Center for Complementary and Alternative Medicine and The National Institute for Allergy and Infectious Diseases Contract HHSN272201000056C.

## Table of Contents

1	Introduction.....	1
1.1	Chronic inflammation .....	1
1.1.1	Type 2 diabetes .....	1
1.1.2	Inflammatory bowel disease .....	2
1.2	PPAR $\gamma$ .....	3
1.3	Motivation for dissertation project.....	3
1.4	Organization of the dissertation .....	4
2	Virtual screening as a technique for PPAR modulator discovery.....	5
2.1	Abstract .....	6
2.2	Introduction .....	7
2.3	Characteristics of PPAR $\gamma$ and the activation process.....	7
2.4	Agonists and the ligand-binding domain of PPAR $\gamma$ .....	9
2.5	Docking .....	12
2.6	Virtual screening .....	13
2.7	Limitations of virtual screening .....	14
2.8	Docking and virtual screening successes .....	15
2.9	Relevance to PPAR $\gamma$ agonist discovery .....	15
2.10	Future directions .....	16
2.11	Acknowledgements .....	17
3	Dietary $\alpha$ -eleostearic acid ameliorates experimental inflammatory bowel disease in mice by activating peroxisome proliferator-activated receptor- $\gamma$ .....	18
3.1	Abstract .....	19
3.2	Introduction .....	20
3.3	Methods.....	21
3.3.1	Docking procedure.....	21
3.3.2	Structural model selection: Re-docking component.....	22
3.3.3	Structural model selection: Cross-docking component .....	22
3.3.4	Small-scale in-house ligand library construction.....	23
3.3.5	Docking analysis for re-docking and cross-docking.....	23
3.3.6	Docking analysis for small-scale VS .....	24
3.3.7	Ligand binding assay .....	24
3.3.8	Transfection of RAW 264.7 cells .....	25
3.3.9	Animal procedures.....	25
3.3.10	Histopathology .....	25
3.3.11	Immunophenotyping.....	25
3.3.12	Quantitative real-time RT-PCR .....	26
3.3.13	Statistical analysis .....	26
3.4	Results .....	26
3.4.1	Selection of structural model: Re-docking component.....	26
3.4.2	Selection of structural model: Cross-docking component .....	27
3.4.3	Conjugated trienes showed association with PPAR $\gamma$ in silico.....	31

3.4.4	ESA bound to and modulated PPAR $\gamma$ in vitro.....	32
3.4.5	ESA ameliorated clinical signs of IBD.....	33
3.4.6	Immunophenotypes for harvested tissues.....	35
3.4.7	Histological trends mimicked clinical activity.....	35
3.4.8	Gene expression suggested PPAR $\gamma$ -dependent and -independent mechanisms.....	36
3.5	Discussion.....	37
3.6	Acknowledgements.....	41
3.7	Supporting Information.....	42
4	Improved success of PPAR $\gamma$ docking using pharmacophore modeling.....	70
4.1	Abstract.....	71
4.2	Introduction.....	72
4.3	Methods.....	73
4.3.1	Catalog of PPAR $\gamma$ 3D structures.....	73
4.3.2	Pharmacophore modeling.....	74
4.3.3	Characterization of crystal structure binding cavities.....	77
4.3.4	Identification of ligand-based pharmacophores.....	78
4.3.5	Computational docking.....	79
4.3.6	Validation of docking protocol.....	80
4.3.7	Statistical analysis.....	81
4.3.8	Virtual screen process flow.....	81
4.4	Results.....	82
4.4.1	Shared features were identified for each ligand type.....	82
4.4.2	Binding cavity differences validated the need for docking with multiple structure models.....	84
4.4.3	Ligand binding reproducibility of the PPAR $\gamma$ structure models.....	86
4.4.4	Predictability of binding with PPAR $\gamma$ structures.....	87
4.4.5	Inclusion of the decoys to assess false positive prediction.....	89
4.4.6	Screening of known drugs validated screening and scoring methods.....	91
4.5	Discussion.....	93
4.6	Acknowledgements.....	97
4.7	Supporting Information.....	98
5	Effects of activation state and agonism on the molecular dynamics of PPAR $\gamma$ .....	108
5.1	Abstract.....	109
5.2	Introduction.....	110
5.3	Methods.....	113
5.3.1	Structure preparation for molecular dynamics simulations.....	113
5.3.2	System equilibration.....	114
5.3.3	Production simulation and analysis.....	114
5.3.4	Principal components analysis.....	115
5.3.5	Steered molecular dynamics.....	115
5.4	Results.....	117
5.4.1	RMSD analysis: System stabilization.....	117
5.4.2	RMSD analysis: Conformation clustering.....	118
5.4.3	Loop regions showed noticeable fluctuations from initial conformations.....	120
5.4.4	Principal components analysis.....	121

5.4.5	Presence of interactions over time .....	124
5.4.6	Steered MD: Changes in force indicated differences in interactions.....	125
5.4.7	Persistent forces indicated key residues for ligand binding.....	126
5.5	Discussion .....	131
5.6	Acknowledgements .....	132
6	Conclusions.....	133
7	Bibliography .....	137
8	Appendices.....	154
8.1	Appendix A: Executive summary .....	154
8.2	Appendix B: PPAR $\gamma$ PDB table.....	156
8.3	Appendix C: Virtual screening process.....	187
8.3.1	Steps for process .....	187
8.3.2	Scripts for performing virtual screening.....	188
8.3.3	Step 1: Prepare the ligand files for docking (prepare_ligand.sh).....	188
8.3.4	Step 2: command line items for dividing ligand files into separate directories for VS on supercomputer. ....	189
8.3.5	Step 3: Large-scale screening script for running serial dockings on a supercomputer (lsvs_mpi.sh). ....	189
8.3.6	Step 4: Post processing of the docked poses (post-process.sh). ....	190
8.3.7	vina_pdbqt2pdb_multi.pl.....	192
8.3.8	vina_parse_E.pl.....	195
8.3.9	all_energy_pose_multi.pl.....	196
8.3.10	calc_all_dist_auto.pl .....	198
8.3.11	weighted_score_key_intxns.pl.....	204
8.4	Appendix D: Molecular dynamics input files .....	209
8.4.1	Energy minimization (EM).....	209
8.4.2	Isochoric-isothermal ensemble (NVT) .....	209
8.4.3	Isothermal-isobaric ensemble (NPT) .....	210
8.4.4	Production MD.....	211
8.4.5	SMD .....	212
8.4.6	Steps for preparing SMD protein and ligand files .....	213

## List of Figures

<b>Figure 2.1</b> Rosiglitazone bound to LBD of PPAR $\gamma$ . Helices are labeled with H, followed by a number. Key residues involved in hydrogen bonding are labeled. Blue dashed lines represent bonding interactions between the hydrogen atoms of the residue and the oxygen atoms of the ligand. (PDB ID 3DZY) (56). .....	10
<b>Figure 2.2</b> Rosiglitazone bound to LBD of PPAR $\gamma$ . Helices are labeled with H, followed by a number. Some of the key residues involved in hydrophobic interactions are labeled. (PDB ID 3DZY) (56). .....	10
<b>Figure 3.1</b> Predicted docked conformations for $\alpha$ -eleostearic (purple), punicic (cyan), calendic (orange), jacaric (green), and catalpic (gold) acids relative to the rosiglitazone-occupied portion of the binding cavity (mesh surface) in the rigid PPAR $\gamma$ structure model. Key residues with which hydrogen bonding occurs are labeled. Atom-specific coloring: red = oxygen; gray = carbon; blue = nitrogen. Table 3.4 contains distance measurements for each docked pose.....	32
<b>Figure 3.2</b> Ligand-binding (A) and reporter assay (B) results for ESA bound to PPAR $\gamma$ with rosiglitazone (Ros) as a positive control. (A) Ligand binding was assessed as a measure of mean polarization for the displaced Fluormone <sup>TM</sup> molecule versus increasing concentrations of either ligand. (B) Reporter activity was measured as relative luciferase activity for various concentrations of ESA versus 1 $\mu$ M Ros. Error bars represent standard deviation, while asterisks (*) indicate significance ( $p \leq 0.05$ ) between the data sets.....	33
<b>Figure 3.3</b> Effect of ESA on disease activity scores for PPAR $\gamma$ -expressing (A) and PPAR $\gamma$ -null (B) mice with experimental IBD. PPAR $\gamma$ -null refers to lack of functional PPAR $\gamma$ product in colon epithelial and immune cells only. Data points represent averaged disease scores for each group with error bars representing standard deviation. Asterisk (*) indicates significance ( $p \leq 0.05$ ). .....	34
<b>Figure 3.4</b> Effect of ESA on immune cell subsets of PPAR $\gamma$ -expression and PPAR $\gamma$ -null mice with experimental IBD. Tissues examined included blood (A and D) and spleen (B, C, and E). Values represent least square means for percentage of gated cells with error bars to indicate standard error. Letters indicate significance ( $p \leq 0.05$ ) where a shared letter indicates groups which are not statistically significantly different. ....	35
<b>Figure 3.5</b> Effect of ESA on histopathological lesions in colons from PPAR $\gamma$ -expressing and PPAR $\gamma$ -null mice with experimental IBD. Epithelial erosion (Erosion) (A), immune cell infiltration (Infiltration) (B), and mucosal thickness (Thickness) (C) were assessed and averaged for all the DSS-treated group of samples. Data are presented as mean score with error bars to indicate standard deviation. Letters indicate significance ( $p \leq 0.05$ ) where a shared letter indicates groups which are not statistically significantly different. ....	36
<b>Figure 3.6</b> Effect of ESA on colonic concentrations of IL-6 (A), VCAM-1 (B), and ICAM-1 (C) in PPAR $\gamma$ -expressing and PPAR $\gamma$ -null mice with experimental IBD. The mean ratio of expression for each protein relative to constitutively expressed $\beta$ -actin is shown with error bars to indicate standard deviation. Letters indicate significance ( $p \leq 0.05$ ) where a shared letter indicates groups which are not statistically significantly different.....	36



**Figure 3.7** Visual assessments of molecular surface differences that result in unsuccessful docking of specific ligand types to the selected PPAR $\gamma$  structure model. Farglitazar is represented in both panels with atom-specific coloring. (A) 1ZGY and 1FM9 surface representations are green mesh and solid gray, respectively. The three poses predicted for farglitazar relative to 1ZGY are shown in magenta, cyan, and yellow. (B) Side chain rotamers for F282 and F363 are responsible for the differences in cavity surface at the rear of the cavity. Surface colors for 1ZGY and 1FM9 are the same as in (A). Atom-specific coloring: gray/black = carbon, blue = nitrogen, red = oxygen, white = hydrogen, and yellow = sulfur. .... 40

**Figure 3.S1** Colored ribbon representation of PPAR $\gamma$  showing three layers of helical “sandwich”, and co-crystallized rosiglitazone (PDB ID 1FM6 (45)). Helices for each layer are colored, with helix H12, which sits at the rear of the binding cavity (AF-2 region), colored in red. Rosiglitazone is colored in green, with oxygen, nitrogen, and sulfur atoms colored red, blue, and yellow, respectively. The insert (upper right) shows a close-up view of the molecular surface of the binding cavity. The thiazolidinedione head group of rosiglitazone sits at the rear of the binding cavity where it can interact with S289, H323, H449, and Y473 in order to change the conformation of the AF-2 region and activate the protein. .... 42

**Scheme 4.1** Process flow diagram for combined pharmacophore screening 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. .... 74

**Figure 4.1** Composite of PharmaGist feature groups overlaid onto representative ligands from each ligand type. The 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 above. 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. .... 83

**Figure 4.2** 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 tested within the binding cavity of the crystal structures 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 the grid established within the binding cavity 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. .... 85

<b>Figure 4.3</b> Distribution of percent similarity values for the 33 envelopes. Percent similarity reflects the proportion of common points between each pair of envelopes (N = 1089 pairs). .....	85
<b>Figure 4.4</b> ROC space for 33 PDB structures used for docking. PDB IDs for the four structures with the highest TPR are labeled. Scales for axes were reduced to show data detail. The black dashed line indicates the division between the states (better prediction versus worse prediction). .....	91
<b>Figure 4.5</b> 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.....	95
<b>Figure 5.1</b> KEGG pathway for PPARs (213, 214).....	111
<b>Figure 5.2</b> A 3D representation of the key residues responsible for ligand-independent and ligand-dependent stabilization of the AF-2 region. Rosiglitazone is shown in surface and stick representation within the binding cavity. Ligand-independent residues are colored magenta. Ligand-dependent residues are colored cyan. H12 is colored green. .	112
<b>Figure 5.3</b> Ribbon representations of the five states for MD simulations. The apo structures, derived from structure 1PRG, did not have a ligand within the binding site of the crystal structure. The ligand in the bound systems (2PRG) was rosiglitazone, which is shown as sticks and an atom-colored surface. Helix 12 is shown in green for all structures to illustrate the placement of the helix in the active and inactive states. The unbound active (UA) structure is the same as the bound active (BA) structure but with the ligand removed. The area for the ligand is shown as a dot surface to indicate the ligand was removed from that position. Atom colors: carbon = gray, oxygen = red, nitrogen = blue, sulfur = yellow. ....	113
<b>Figure 5.4</b> Ribbon representations of the SMD systems. Ligands within the 3D structure are shown in stick and surface representation. Each ligand is represented in 2D under each 3D structure. Atom colors: carbon = gray, oxygen = red, nitrogen = blue, sulfur = yellow.....	116
<b>Figure 5.5</b> Illustration of the origin of pulling for SMD simulations. The areas of the PPAR $\gamma$ structure through which each ligand was pulled are shown in orange. Panel B, which is rotated approximately 90 degrees, contains an arrow and spring that loosely illustrates the direction of pulling for the ligands. ....	117

<b>Figure 5.6</b> RMSD plots over 100 ns for three replicates of each system. Running averages were calculated with a block length of 1000 points. The panels indicate RMSD for the (A) AA, (B) UA, (C) BA, (D) AI, and (E) BI systems. ....	118
<b>Figure 5.7</b> Representative structures for each system taken from RMSD cluster one. The structures are the median member of the first cluster for the representative replicate. The active and inactive systems are compared separately. (a) Systems AA1 (gray), UA1 (green), and BA2 (coral) are superimposed. (b) Systems AI1 (blue) and BI1 (magenta) are superimposed. The H2'-H3 loop is highlighted in a darker version of the previously mentioned colors. ....	119
<b>Figure 5.8</b> Backbone RMSF for three replicates of each system. Panels show RMSF for the (A) AA, (B) UA, (C) BA, (D) AI, and (E) BI systems. A 3D representation of the PPAR $\gamma$ structures is shown in panel F as a reference for the RMSF peaks. Peaks on one RMSF curve are numbered to show where highlighted regions of the protein are located for all the curves. ....	121
<b>Figure 5.9</b> Clustering of conformations within the first two principal components for each system. Graphs reflect all the recorded conformations sampled for the last 50 ns of three replicates. Frequency with which each conformation occurred is shown with the heat map scale where red is most frequent and blue is least frequent. Transparent white lines indicate the origin axes and establish quadrants for conformational sampling relative to activation and bound states. The top half of each graph corresponds to inactive motion, while the bottom half corresponds to active motion. The left half indicates unbound states, while the right half indicates bound states. Panels: (A) AA, (B) UA, (C) BA, (D) AI, and (E) BI systems. ....	122
<b>Figure 5.10</b> Segments of covariance graphs that correspond to H3 (plotted on Y-axis and shown in red and orange in panel F) relative to the rest of the PPAR $\gamma$ structure (X-axis). The rectangles indicate areas on the plots showing differences in correlated (red) and anti-correlated (blue) motion. The colors for the rectangles correspond to the colored regions in panel F: green = H2'-H3 loop (atoms 175-210), orange = H3 (atoms 211-288), purple = s3 to the middle of H7 (atoms 415-486), and blue = H11 (atoms 709-747). Graphs indicate motion for each atom of the backbone. The systems represented by each panels are (a) AA1, (b) AI1, (c) UA1, (d) BA2, and (e) BI1. ....	123
<b>Figure 5.11</b> Number of contacts within 1 nm of each pair of residues listed. Dotted lines indicate ligand-independent residue pairs, while solid lines indicate ligand-dependent residue pairs. Panels show the number of contacts for the (A) AA1, (B) AI1, (C) UA1, (D) BA2, and (E) BI1 replicates. ....	125
<b>Figure 5.12</b> Calculated forces over time for (A) rosiglitazone, (B) R-enantiomer, and (C) S-enantiomer as each was pulled from the binding cavity of respective PPAR $\gamma$ structures. ....	126
<b>Figure 5.13</b> Average total energies measured for interactions between the listed residues and ligands. The electrostatic (black bars) component and van der Waals (VDW; white bars) component of the energy values are shown for residues within 1.0 nm of a ligand as it was pulled from the binding cavity. Panels correspond to the pulling simulations for (a) rosiglitazone, (b) the R-enantiomer, and (c) the S-enantiomer. Error bars indicate standard deviation for each mean. The top 20 energy values are shown for each ligand. Residues with no energy values for a particular system indicate those for which the energies were not in the top 20 values. ....	127

<b>Figure 5.14</b> Average interaction energy fluctuations over time for five residues relative to rosiglitazone (A, D, G, J, M), the R-enantiomer (B, E, H, K, N), and the S-enantiomer (C, F, I, L, O). .....	128
<b>Figure 5.15</b> Common and unique residues with large energy values over the course of the pulling simulations. Residues unique to each structure are highlighted in red while residues shared between two structures are shown in either blur or green to indicate which structures possessed shared residues. Circles indicate location of the highlighted residues. Panels correspond to the pulling simulations for (a) rosiglitazone, (b) the R-enantiomer, and (c) the S-enantiomer. ....	129
<b>Figure 5.16</b> Average energies over time for three residues unique to rosiglitazone binding. ....	129
<b>Figure 5.17</b> Average energies over time for two residues unique to enantiomer binding. ....	130

## List of Tables

<b>Table 2.1</b> PDB IDs of published crystal and NMR structures for PPAR $\gamma$ with various ligands bound. Resolution values are in Angstroms (Å). The “Reference number” columns list references for each PDB ID. All PDB IDs list Homo sapiens as the protein source for the PPAR $\gamma$ chain. Access date: 23 March 2009. (An updated list can be found in Appendix A.) .....	11
<b>Table 2.2</b> List of some commonly used molecular dynamics and docking software packages with developer URL. This is not intended to be a comprehensive list of all available dynamics and docking programs available. Available programs are typically free to download for academic use, but some require the purchase of a license for use. .	13
<b>Table 3.1</b> Average RMSD and free energy of binding (kcal/mol) for re-docking of rosiglitazone (N = 5). .....	27
<b>Table 3.2</b> Average RMSD and free energy of binding from cross-docking for various ligands relative to each listed PDB ID (top row) (N = 3). .....	29
<b>Table 3.3</b> Full names and structures for compounds listed by ligand ID in Table 3.2. Ligand IDs from respective PDB files were used. Ligand structures can be found in Table 3.S1. ....	30
<b>Table 3.4</b> Distance measurements (in Angstroms [Å]) for docked conjugated triene poses displayed in Figure 3.1. Distances were measured between carboxylic oxygen atoms of fatty acids and listed atoms for each residue. Free energy of binding is measured in kilocalories per mole of ligand (kcal/mol). No value is listed for rosiglitazone as this refers to the crystal conformation (denoted "N/A") Residues are labeled as the amino acid designation plus the atom name (e.g., S289.OG refers to the oxygen atom in the gamma position on serine 289). .....	34
<b>Formulas 3.S1</b> .....	43
<b>Table 3.S1</b> List of ligands used for virtual screening .....	44
<b>Table 3.S2</b> List of atoms for key residues common to selected rosiglitazone crystal structures used to assess potential interactions between docked poses and the protein structure model. ....	53
<b>Table 3.S3</b> List of atoms for key residues common to selected fatty acid-bound crystal structures used to assess potential interactions between docked poses and the protein structure model. ....	54
<b>Table 3.S4</b> List of atoms for key residues common to rosiglitazone- and fatty acid-containing PDB structures used to assess potential interactions between docked poses and the protein structure model. ....	55
<b>Table 3.S5</b> Predicted hydrophobic and hydrogen bond interactions for ligands in cross-docking test set relative to a reference list of interactions common to rosiglitazone and selected fatty acids. Poses were taken from docking of each ligand into each of the three listed PPAR $\gamma$ PDB files (top row). Ligand IDs refer to compounds listed in Table 3.3. .	56
<b>Table 3.S6</b> Presence or absence of potential hydrogen bond interactions between indicated residues of selected protein structure models and replicate poses of ligands listed by ID (Lig ID). A single “x” indicates one potential interaction for the listed residue was found for the specified ligand, whereas more than one “x” indicates more than one interaction (e.g., “xx” indicates two interactions found). (N = 3) .....	58

<b>Table 3.S7</b> Predicted hydrophobic and hydrogen bond interactions for ligands in small-scale screening test set relative to a reference list of interactions common to rosiglitazone and selected fatty acids (Table 3.S4). Poses were taken from docking of each ligand into each of the three listed PPAR $\gamma$ PDB files (top row). Predicted free energy of binding is listed as kcal/mol.....	60
<b>Table 3.S8</b> Presence or absence of potential hydrogen bond interactions between indicated residues of selected protein structure models (top row) and ligand poses. A single “x” indicates one potential interaction for the listed residue was found for the specified ligand, whereas more than one “x” indicates more than one interaction (e.g., “xx” indicates two interactions found). .....	65
<b>Table 3.S9</b> Predicted free energy of binding and interaction counts for conjugated trienes. Docking was performed using AD4 with three top-binding replicates for each ligand (150 total conformations). The highest energy conformation with the highest number of hydrogen bonds was used for analysis in Table 3.4.....	69
<b>Table 4.1</b> PDB structures used for the structure-based pharmacophore analysis and virtual screening. Ligand IDs and names are listed for the co-crystallized compounds, followed by the ligand type, structure resolution, and the citation for the paper associated with the structure.....	75
<b>Table 4.2</b> List of ligand IDs for agonists and antagonists used in the ligand-based pharmacophore analysis. The PDB ID for the crystal structure containing the ligand and the ligand type are listed for each ligand ID. Citations not included in Table 4.1 are included for the additional structures mentioned in this table. ....	78
<b>Table 4.3</b> 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 PDB ID) as the pivot molecule. ....	82
<b>Table 4.4</b> Groupings for structure models based on envelope similarity. PDB and ligand IDs match those used in Table 4.1. ....	86
<b>Table 4.5</b> RMSD values for re-docking of native ligands into corresponding crystal structure models. Results are listed by PDB ID and RMSD values are measured in Angstroms (Å). ....	87
<b>Table 4.6</b> True positive rates (TPRs) for each PDB structure in the training phase. 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.....	88
<b>Table 4.7</b> Counts of ligands that matched each ligand-based pharmacophore model from three lists of compounds: known actives from PDB crystal structures, known actives from the DUD database, and decoy compounds from the DUD database. The pharmacophore models used are listed alphabetically in the top row with the ligand type indicated in the following row.....	90
<b>Table 4.8</b> Results from the docking of the MSUSDrugs compounds into 1ZGY for which docked poses were deemed successful. All compounds possessed interactions with at least the 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 MOE and the ToxAlerts server. The characters “+” and “-“ indicate if the compound matched or did not match, respectively, the 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. .... 92

**Table 4.S1** RMSD values for all tested PDB ligands relative to each of the 33 structure models tested during the training phase. PDB (PDBID) and ligand (LigID) identifiers are listed for each ligand. RMSD values were calculated with the crystal structure-derived pose of the respective ligand..... 98

**Table 4.S2** Scoring table for MSUSDrugs compounds docked into the 1K74 structure model. Only the top ten results that satisfied the sorting criteria are shown. Toxicity was evaluated by two different programs with an “X” to indicate the compound was reported as toxic. Ligand pharmacophore model screening was applied to assess potential activity. The marker “+” indicates the compound matched the model, while “-“ indicates a non-match. “Bind Free E” refers to the free energy of binding prediction calculated by AutoDock Vina. Inhibition constants (Ki) were calculated based on the predicted free energy values. Interactions are listed relative to residues within the binding cavity that are known to interaction with the positive control rosiglitazone. Predicted interactions were measured with cutoffs and represented as “x” for hydrogen bond, “o” for hydrophobic, “/” for the threshold value of 4.0 Å, and “-“ indicates no distances within the threshold were measured. .... 105

**Table 4.S3** Scoring table for MSUSDrugs compounds docked into the 2I4J structure model. Only the top ten results that satisfied the sorting criteria are shown. Toxicity was evaluated by two different programs with an “X” to indicate the compound was reported as toxic. Ligand pharmacophore model screening was applied to assess potential activity. The marker “+” indicates the compound matched the model, while “-“ indicates a non-match. “Bind Free E” refers to the free energy of binding prediction calculated by AutoDock Vina. Inhibition constants (Ki) were calculated based on the predicted free energy values. Interactions are listed relative to residues within the binding cavity that are known to interaction with the positive control rosiglitazone. Predicted interactions were measured with cutoffs and represented as “x” for hydrogen bond, “o” for hydrophobic, “/” for the threshold value of 4.0 Å, and “-“ indicates no distances within the threshold were measured. .... 106

**Table 4.S4** Scoring table for MSUSDrugs compounds docked into the 2I4Z structure model. Only the top ten results that satisfied the sorting criteria are shown. Toxicity was evaluated by two different programs with an “X” to indicate the compound was reported as toxic. Ligand pharmacophore model screening was applied to assess potential activity. The marker “+” indicates the compound matched the model, while “-“ indicates a non-match. “Bind Free E” refers to the free energy of binding prediction calculated by AutoDock Vina. Inhibition constants (Ki) were calculated based on the predicted free energy values. Interactions are listed relative to residues within the binding cavity that are known to interaction with the positive control rosiglitazone. Predicted interactions were measured with cutoffs and represented as “x” for hydrogen bond, “o” for hydrophobic, “/” for the threshold value of 4.0 Å, and “-“ indicates no distances within the threshold were measured. .... 107

<b>Table 5.1</b> Systems established for MD simulations. Each system was assigned a two-letter abbreviation as defined in Figure 5.3, and each replicate was given a number (1 through 3). All systems contained five sodium ions that contributed to the total number of atoms. ....	114
<b>Table 5.2</b> Systems established for SMD simulations. Each system was assigned a two-letter abbreviation as defined in Figure 5.4, and each replicate was given a number (1 through 3). All systems included five sodium ions.....	117
<b>Table 5.3</b> RMSD clustering using a cutoff of 0.1 nm. Percentages for the top five clusters only are shown. The final row yields the total percentage of conformations contained in the top five clusters. ....	118
<b>Table 5.4</b> RMSD values in nanometers (nm) for the center conformation of cluster one for each 2PRG-derived system. Only systems with the same number of atoms were compared.....	119
<b>Table 5.5</b> RMSD values in nanometers (nm) for the middle member of cluster one for each 1PRG-derived system. Only systems with the same number of atoms were compared.....	120
<b>Table 5.6</b> Percentage of motion possessed by each of the top five eigenvalues for the last 50 ns of each trajectory. The index value in the first column refers to the component of motion. The final row is the total percent of motion for the five components of each replicate.....	122
<b>Table 5.7</b> List of residues with which interactions were purposed to be necessary for ligand binding. “Yes” and “No” indicate energy terms that either persisted or quickly dissipated, respectively. “Weak” indicates energy terms with less pronounced energy curves but potential interactions. ....	130



## List of Abbreviations

Abbreviation	Description
13-HODE	13-(S)-hydroxyoctadecadienoic
9-HODE	9-(S)-hydroxyoctadecadienoic
AD4	AutoDock 4
ADT	AutoDock Tools
AF-2	activation function-2
ANOVA	analysis of variance
AP-1	activator protein-1
aP2	adipocyte protein 2
C/EBP $\alpha$	CCAAT-enhancer-binding protein-alpha
C/EBP $\beta$	CCAAT-enhancer-binding protein-beta
C/EBP $\delta$	CCAAT-enhancer-binding protein-delta
CAM	complementary and alternative medicine
CD	Crohn's disease
CDC	Centers for Disease Control and Prevention
CLA	conjugated linoleic acid
COM	center of mass
DAI	disease activity index
DBD	DNA-binding domain
DMEM	Dulbecco's Modified Eagle Medium
DSS	dextran sodium sulfate
DUD	Directory of Useful Decoys
ERK1/2	extracellular signal-regulated kinase-1/2
ER $\alpha$	estrogen receptor-alpha
ESA	$\alpha$ -eleostearic acid
FDA	U.S. Food and Drug Administration
FN	false negative
FP	false positive
FPR	false positive rate
GLUT-4	glucose transporter-4
HNF4 $\alpha$	hepatocyte nuclear factor-4 $\alpha$
HTS	high-throughput screening
IACUC	institutional animal care and use committee
IBD	inflammatory bowel disease
ICAM-1	intercellular adhesion molecule-1
ID	identification number
IEC	intestinal epithelial cells
IL-6	interleukin-6
LANCL2	lanthionine synthetase component C-like protein 2

Abbreviation	Description
LBD	ligand-binding domain
LGA	Lamarckian genetic algorithm
LXXLL	nuclear receptor box
MCP-1	monocyte chemotactic protein-1
MD	molecular dynamics
MLN	mesenteric lymph nodes
MMTV-Cre-	mouse mammary tumor virus, Cre recombinase
MOE	Molecular Operating Environment
NF- $\kappa$ B	nuclear factor kappa-light-chain-enhancer of activated B cells
NHR	nuclear hormone receptor
NMR	Nuclear magnetic resonance
NPT	isothermal-isobaric
NVT	isochoric-isothermal
PDB	protein data bank
PEPCK	phosphoenolpyruvate carboxykinase
PME	Particle-mesh Ewald
PPAR $\alpha$	peroxisome proliferator-activated receptor alpha
PPAR $\gamma$	peroxisome proliferator-activated receptor gamma
PPAR $\gamma$ -flfl	PPAR $\gamma$ -floxed
PPAR $\delta$	peroxisome proliferator-activated receptor delta
PPRE	PPAR response element
PXR	pregnane X receptor
QSAR	quantitative structure-activity relationship
RA	retinoic acid
RCSB	Research Collaboratory for Structural Bioinformatics
RMSD	root mean-squared deviation
RMSF	root mean-squared fluctuation
ROC	receiver operating characteristic
RXR $\alpha$	retinoid X receptor alpha
SAS	Statistical Analysis Software
SMD	steered molecular dynamics
SRC-1	steroid receptor co-activating factor-1
STAT	signal transducer and activator of transcription
T2D	type 2 diabetes mellitus
TN	true negative
TNF- $\alpha$	tumor necrosis factor-alpha
TP	true positive
TPR	true positive rate
TZD	thiazolidinedione

Abbreviation	Description
UC	ulcerative colitis
UCSF	University of California, San Francisco
VCAM-1	vascular cell adhesion molecule-1
VDW	van der Waals
Vina	AutoDock Vina
VS	virtual screening

## Attributions

All documents included as the chapters in this dissertation were written by the candidate. Dr. Bevan as primary research advisor and Dr. Bassaganya-Riera as co-adviser edited and provided suggestions for improving all manuscripts for publication. Chapters 1 and 6 of this dissertation were written by the candidate in order to tie the intermediate chapters together. Appendices were compiled by the candidate and provide summaries of other papers written with author contributions and scripts written by the candidate for obtaining data presented in the dissertation. The specifics of author contributions are listed below by chapter. Each author's initials are used to specify which author contributed each part.

Author names used are as follows:

Stephanie N. Lewis (SNL)

Amir J. Guri (AJG)

Josep Bassaganya-Riera (JBR)

Pinyi Lu (PL)

David R. Bevan (DRB)

Zulma Garcia (ZG)

Raquel Hontecillas (RH)

Justin A. Lemkul (JAL)

Lera Brannan (LB)

William J. Allen (WJA)

### Chapter 2

SNL wrote this paper with input on content and corrections prior to publication from JBR and DRB. All authors collectively conceived the initial project idea and direction. SNL was responsible for submission of the completed manuscript and correspondence with the journal editor during the publication process.

### Chapter 3

SNL wrote the majority of this paper with input from each listed author. JBR, RH, and DRB performed major editorial support and aided in experimental design. SNL and LB performed the computational work, while AJG and PL conducted the experimental work. Analysis of results and derivation of conclusions was conducted by SNL, AJG, RH, JBR, and DRB. SNL was responsible for submission of the completed manuscript and correspondence with the journal editor during the publication process.

### Chapter 4

SNL wrote the majority of this paper with input from DRB on experimental design. SNL performed the majority of the data collection. LB performed the initial ligand-based pharmacophore analysis that contributed to the methods developed for this chapter. LB is acknowledged at the end of the chapter for her contribution. ZG performed initial structure-based pharmacophore screenings and provided some of the written material for the related sections. SNL conducted the derivation of conclusions. JBR and DRB provided editorial support.

### Chapter 5

SNL wrote the manuscript with editorial support from DRB. SNL performed all of the data collection and analyses. DRB, JAL, and WJA provided input on experimental design, methods, and analyses. JAL and WJA, though not listed as authors, are acknowledged for the guidance provided. SNL conducted the derivation of conclusions.

# **1 Introduction**

## **1.1 Chronic inflammation**

Symptoms associated with many diseases can be related in some way to a chronic inflammation state (1). Type 2 diabetes mellitus (T2D), obesity, inflammatory bowel disease (IBD) and the associated manifestations Ulcerative Colitis and Crohn's Disease, atherosclerosis, and colorectal cancer are all examples of diseases that are marked by chronic inflammation in various tissues throughout the body (1, 2). Inflammation in the abdominal area, specifically adipose deposits in the abdomen, is associated with T2D and obesity (3-5). Inflammation in the intestines is associated with IBD and colorectal cancer (6-9). Inflammation within the circulatory system is associated with atherosclerosis (10). Inflammation in each of these tissues can trigger different gene regulatory events and subsequently lead to different phenotypes. Despite the differences in the onset of these diseases, they all share the potential for regulation of the inflammatory processes by targeting a single regulatory protein that mediates processes contributing to each disease (11).

T2D and IBD were the two diseases of interest in this dissertation given the close metabolic links between them. In recent years, diagnoses for T2D have approached epidemic proportions. Estimates in 2007 suggested 23.5 million individuals over the age of 20 have been diagnosed with T2D, with an average annual increase of about 1.6 million individuals (12). In 2010, the estimate increased to 25.8 million individuals (13). The statistics on individuals with IBD are not as clear as those with T2D because there is no established set of symptoms associated with the disease (14, 15). Often the symptoms expressed, severity, and genetic influences can vary from patient to patient (16). There is an increase, however, in the number of individuals that seek medical treatment for symptoms that can be associated with the disease (14). Additionally, IBD is often misdiagnosed as other related diseases (17), or goes relatively unnoticed by affected individuals with less severe symptoms (16). Both diseases can affect individuals in any age range, with some ethnic predisposition for each. According to the American Diabetes Association and the Centers for Disease Control (CDC), T2D is often seen in individuals of African, Native American, Hispanic, and Asian descent (13). The CDC has suggested that IBD is prevalent in Caucasian and Ashkenazic Jewish individuals (16).

### **1.1.1 Type 2 diabetes**

The incidence of T2D has increased significantly over the last decade. According to the CDC, the number of individuals diagnosed with T2D will continue to increase despite the existence of effective therapeutics (13). The major symptom seen in T2D patients is insulin resistance, which is a result of an inflammation-related stress response in overwhelmed adipocytes (18). In the lean, healthy state, normal function includes triglyceride and glucose intake from diet. The triglycerides are shuttled to the adipose tissue where they can be broken down into fatty acids within the adipocytes. These fatty acids are transported to peripheral tissues like skeletal muscle, where they are used in

ATP production. Glucose is also taken into skeletal muscle via insulin-regulated glucose transporters for use in ATP production. With increased caloric intake, we see an increase in the blood levels of triglycerides and glucose that leads to the overweight, and eventually the obese, state (18). For most individuals, the body can compensate for the increase in calories by increasing fat storage in the adipose tissue. Alternatively, an individual that performs frequent, high-impact physical activity can “burn off” the excess calories through an increased need for ATP (19). Ultimately, the high levels of fatty acids in the blood lead to problems in biological function (18).

If the increase in caloric intake persists, the pro-inflammatory state can be reached. Here, the adipocytes are overwhelmed by the high levels of triglycerides. The hypertrophic adipocytes trigger a stress response in which the adipocytes release monocyte chemoattractant protein-1 (MCP-1). This protein promotes monocyte infiltration into the adipose tissue where these cells differentiate into macrophages. The macrophages secrete tumor necrosis factor-alpha (TNF- $\alpha$ ), which is a pro-inflammatory cytokine that inhibits triglyceride intake in the adipose tissue. As a result, obese individuals exhibit increased levels of triglycerides in the blood. Adipocytes that have not undergone apoptosis due to the hypertrophic state can still release fatty acids into the blood. At this point, an individual would exhibit high levels of triglycerides and fatty acids in the blood. The body will proceed to either find alternative storage or excrete as much of the fats as possible. Often the fat is stored as ectopic lipid droplets in the muscle and liver. These droplets impede mitochondrial function, which reduces ATP production. This reduction triggers a signal in the muscle to stop the influx of glucose from the blood. The subsequent high blood glucose levels signal increased insulin production, to which insulin receptors on the muscle will not respond in the persisting overwhelmed state caused by the high triglyceride and fatty acid levels. This state is termed insulin insensitivity or insulin resistance depending on the severity of the condition. The term stems from the lack of insulin-triggered glucose uptake by the transporters on peripheral tissues. (18)

The above pathogenesis translates to obesity, weight gain, and insulin insensitivity in T2D. All of these symptoms can be linked to high-fat diets, lack of physical activity, and chronic inflammation in the adipose tissue. One common and long-used treatment for T2D is prescription of compounds in the thiazolidinedione (TZD) compound family. These compounds are known to counteract insulin insensitivity, lower blood glucose levels, and suppress inflammation. TZDs were discovered as an effective T2D treatment before the protein target was identified as PPAR $\gamma$  in 1994 (20).

### **1.1.2 Inflammatory bowel disease**

The specific causes for the inflammation response that leads to inflammatory bowel disease (IBD) are unclear. What is known is that a response to commensal bacteria in the colon results in chronic inflammation that damages the colonic epithelium. Additional details about IBD are provided in Chapter 3 as the introduction to the content therein.

## 1.2 PPAR $\gamma$

Originally labeled as an orphan receptor within the nuclear hormone receptor (NHR) superfamily of proteins (21), PPAR $\gamma$  has become the focus of a significant amount of research over the last 20 years. The popularity of this protein as a target of interest stems from the many processes regulated by PPAR $\gamma$  and the other PPAR subtypes,  $-\alpha$  and  $-\beta/\delta$ . PPARs as a group of transcription factors are responsible for regulating inflammation, glucose and lipid homeostasis, cell differentiation and proliferation, and vascular function and integrity (22).

The first PPAR subtype was identified in 1990 as a protein that induced proliferation of peroxisomes, which are eukaryotic organelles involved in catabolism of long-chain fatty acids (2, 23). This first subtype was classified as PPAR $\alpha$ . Subsequent studies suggested the protein regulates peroxisome proliferation rather than directly inducing the process, which is where the role as a transcriptional regulator became more clear (24). PPAR- $\beta/\delta$  was identified next (25), followed by PPAR $\gamma$  (26, 27). PPAR $\gamma$  was determined to be the predominant regulator of adipogenesis and the target for anti-diabetic TZDs. As studies with PPARs have progressed, PPARs have been established as important for transcriptional regulation of metabolic, inflammatory, and vascular functions.

The PPAR $\gamma$  gene gives rise to several splice variations, with the predominate translatable isoforms being  $\gamma 1$  and  $\gamma 2$  (28, 29). PPAR $\gamma 2$  is predominately found in adipose tissue, while PPAR $\gamma 1$  is ubiquitously expressed throughout the body in small concentrations (28). Additional information about PPAR $\gamma$  can be found in the introduction sections of subsequent chapters.

## 1.3 Motivation for dissertation project

As previously mentioned, TZDs have been identified as an effective T2D treatment. The problem with TZDs as a therapeutic is the high incidence of serious side effects that occur in T2D patients taking these medications. Side effects for rosiglitazone and troglitazone include fluid retention, weight gain, hepatotoxicity, an increased risk of heart attack, and death (30, 31). Most recently, pioglitazone use has been linked to an increased risk for bladder cancer (32, 33). The Food and Drug Administration (FDA) has already removed troglitazone drugs from the market (34). In 2011, the FDA placed restrictions on Avandia® (GlaxoSmithKline), which contains rosiglitazone as the active ingredient (35). Only in extreme cases where no other drugs or treatments have worked can this drug be prescribed.

The removal of one drug and restriction of another has fueled the debate regarding the appropriateness of PPAR $\gamma$  as a target (36-38). It is important to note that the processes regulated by this protein play a significant role in the onset of disease, so restoring some level of homeostasis by targeting PPAR $\gamma$  is necessary to combat disease. Therefore, the treatment options are more the issue than the target, and an urgent need exists for therapeutic alternatives.

Lipids and fatty acids are endogenous PPAR $\gamma$  ligands, and they are the primary mediators of gene expression or repression in metabolic processes. It has been established that prostanoids like prostaglandin, hydrated fatty acids, nitrated fatty acids, oxidized fatty acids, and lysophosphatidic acids are PPAR $\gamma$  agonists (39-41). Other natural compounds have been experimentally identified to activate PPAR $\gamma$  including conjugated linoleic acid (CLA), which is found in ruminant fats (42). CLA, though effective, is found in very low concentrations in ruminant fats, hence simple dietary consumption is not enough to see benefits (42). The low concentration in fats also hinders any isolation processes. Ideally, it should be possible to find compounds that are abundant in nature, exhibit PPAR $\gamma$  agonism, and ameliorate T2D at a feasible, low-risk dosage. A compound similar to endogenous ligands would hopefully regulate gene expression in the same way and also exclude other non-relevant proteins in order to treat the disease with minimal side effects. It is also possible for currently available drugs to be repurposed for use in treating T2D. Established drugs have the benefit of years of clinical trials and real-world testing by consumers. Finding an established therapeutic with minimal side effects in treating other diseases would reduce the therapeutic development time-period in which compounds are tested to establish safe dosing and disease outcomes. This reduces the development time “from bench-top to bedside”, and gets beneficial drugs into the hands of patients more quickly, which would reduce the healthcare burden of T2D, IBD, and potentially other chronic inflammation-related diseases in a more timely fashion.

#### **1.4 Organization of the dissertation**

The dissertation has been compiled as a collection of published peer-reviewed papers and manuscripts prepared for publication submission. Chapter 1 serves as a general introduction to PPAR $\gamma$  as a protein target of interest and the research goal of establishing computational docking into and dynamic analysis of PPAR $\gamma$ . Chapter 2 is a published review article that will serve as a literature review. Chapter 3 is a published article detailing a proof of concept small-scale virtual screen with experimental validation and pre-clinical trial. The original citations for Chapters 2 and 3 are provided on the title pages. Minor editorial corrections were made for inclusion in the dissertation. Chapter 4 outlines a large-scale attempt at screening for PPAR $\gamma$  modulators with the inclusion of pharmacophore modeling. Chapter 5 describes analyses of PPAR $\gamma$  structure using molecular dynamics techniques. Chapter 6 contains concluding remarks that tie all of the chapters together and address the goals mentioned in Chapter 1. Chapter 6 is followed by the bibliography of references cited. The dissertation ends with appendices that contain executive summaries of other publications completed during the PhD process, a table of available PPAR $\gamma$  crystal structures, and scripts used in the docking and dynamics processes.



## **2 Virtual screening as a technique for PPAR modulator discovery**

Stephanie N. Lewis<sup>1,2</sup>, Josep Bassaganya-Riera<sup>3</sup>, and David R. Bevan<sup>2</sup>

<sup>1</sup>Genetics, Bioinformatics, and Computational Biology Program, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

<sup>2</sup>Department of Biochemistry (0308), Virginia Polytechnic Institute and State University, 201 Engel Hall, Blacksburg, VA 24061.

<sup>3</sup>Nutritional Immunology and Molecular Nutrition. Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Washington Street (0477), Blacksburg, VA 24061

Copyright © 2010 Stephanie N. Lewis et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: SN Lewis, J Bassaganya-Riera, and DR Bevan. (2010) Virtual Screening as a Technique for PPAR Modulator Discovery. PPAR Research. DOI: 10.1155/2010/861238.

## 2.1 Abstract

Virtual screening (VS) is a discovery technique to identify novel compounds with therapeutic and preventative efficacy against disease. Our current focus is on the *in silico* screening and discovery of novel peroxisome proliferator-activated receptor-gamma (PPAR $\gamma$ ) agonists. It is well recognized that PPAR $\gamma$  agonists have therapeutic applications as insulin sensitizers in type 2 diabetes or as anti-inflammatories. VS is a cost- and time-effective means for identifying small molecules that have therapeutic potential. Our long-term goal is to devise computational approaches for testing the PPAR $\gamma$ -binding activity of extensive naturally occurring compound libraries prior to testing agonist activity using ligand-binding and reporter assays. This review summarizes the high potential for obtaining further fundamental understanding of PPAR $\gamma$  biology and development of novel therapies for treating chronic inflammatory diseases. This goal can be achieved through evolution and implementation of computational screening processes for immunotherapeutics in conjunction with experimental methods for calibration and validation of results.

## 2.2 Introduction

Transdisciplinary research has become a common means of addressing the most pressing societal problems. Past discoveries of scientific hallmarks have favored exploring the depths of established ideas across scientific disciplines to better understand biological systems and processes. This is possible because the wealth of scientific knowledge has only scratched the surface of how biological systems work, and often exploring the unknown intricacies of biological networks requires knowledge of more than one scientific realm.

The vast amount of information readily available to the scientific community presents a valuable and perpetually renewing resource. However, this overabundance also poses a problem. There is simply too much information within too many areas of science for one person with expertise in a single field to rapidly make novel advances. Take for example the question of what factors determine whether an individual suffers from a particular disease. When designing a treatment, one can look at the symptoms, the cause of the symptoms, genetic differences between healthy and afflicted individuals, genetic differences between individuals with the same disease but slightly different symptoms, methods for treating the symptoms, methods for controlling or correcting the disease, and methods for screening for the disease. This list includes, but is not limited to, disciplines such as genetics, bioinformatics, biochemistry, pharmacology, and medicine, and it is the combination of all these disciplines that facilitates the development of effective preventive and therapeutic approaches.

In a more general sense, there is also an increasing need for integrating computational and experimental approaches. Computers have become a large and vital part of scientific exploration and serve to simplify and expedite processes that could take months to years for an individual to complete. First, computers allow for organization of scientific knowledge. Second, they allow for sharing of ideas and discoveries in an effective and timely fashion. Third, computers allow individuals to better analyze experimental results and develop more efficacious test methods. The fourth and ultimate benefit of computer technology to science is improved efficiency due to a reduced necessity for time, money, and resources.

Peroxisome proliferator-activated receptor (PPAR) research is one of many areas that may benefit from advances in computational biology and other transdisciplinary approaches. Mixtures of computational and experimental studies have given insight into characteristics of PPARs, particularly PPAR- $\gamma$  (PPAR $\gamma$ ) and its modulators, as well as the role of these proteins in treating type 2 diabetes (T2D), gastrointestinal diseases, and genetic disorders associated with glucose homeostasis and lipid uptake.

## 2.3 Characteristics of PPAR $\gamma$ and the activation process

PPAR $\gamma$  is one of three known PPAR isoforms ( $\alpha$ ,  $\delta$ , and  $\gamma$ ). PPARs belong to the nuclear hormone receptor superfamily and have been found to regulate inflammation, immunity, and metabolism (3, 11). Members of this superfamily are structurally and functionally conserved transcription factors that regulate both target gene expression and repression after ligand binding occurs (43). A diverse set of natural and synthetic molecules is

classified as ligands that can induce activation and expression of PPARs. These ligands include nutrients, non-nutrient endogenous ligands, and drugs such as thiazolidinediones (TZDs) and fibrates (3, 11, 44). Known endogenous and dietary agonists include conjugated linoleic acid (CLA), 9-(S)-hydroxyoctadecadienoic (9-HODE), 13-(S)-hydroxyoctadecadienoic (13-HODE) acid, and 15-deoxy- $\Delta^{12,14}$ -prostaglandin J2 (15d-PGJ2) (11, 41).

A great deal of literature focuses on increasing insulin sensitivity by controlling PPAR $\gamma$  interactions and altering gene expression of various transcription factors. PPAR $\gamma$  is a component of an extensive group of controls for adipogenesis and glucose homeostasis, and both of these processes directly affect obesity and T2D (45). PPAR $\gamma$  is located in high concentrations in adipocytes, and has also been found in significant amounts in the retina, cells of the immune system, and colonic epithelial cells (11, 46). Functionally, PPAR $\gamma$  down-regulates the expression of pro-inflammatory cytokines by antagonizing the activities of transcription factors such as AP-1 and NF- $\kappa$ B, and favoring the nucleocytoplasmic shuttling of the activated p65 subunit of NF- $\kappa$ B (3). As a consequence of the important roles PPARs play in controlling metabolic homeostasis and inflammatory processes, they are all well recognized as molecular targets for drugs against metabolic diseases, such as T2D (47-49), and treatment of immunoinflammatory disorders.

Structurally, PPAR $\gamma$  is composed of a DNA-binding domain (DBD), a hinge region, and a ligand-binding domain (LBD). The first step in PPAR $\gamma$  activation is disassociation of co-repressors after binding of retinoic acid (RA) to a single retinoid X receptor (RXR) subunit. This step is an essential part of numerous endocrine system pathways (45). The ligand-bound RXR then associates with ligand-bound PPAR $\gamma$ . To become fully active, the PPAR $\gamma$ -RXR $\alpha$  heterodimer requires association of co-activator molecules (45). Agonist binding to PPAR $\gamma$  regulates activity by causing conformational changes to the LBD, which is composed of approximately 250 amino acids near the C-terminal end of the protein (50). Mediation of activity is a direct result of changes to the transcription activation function-2 (AF-2) domain (45, 51). These changes vary depending on the type of ligand that binds to the LBD. Changes to AF-2 allow for co-activator recruitment, followed by transcriptional activation.

Co-activator recruitment is based on a LXXLL binding motif (nuclear receptor box) found on both PPAR $\gamma$  and co-activators like steroid receptor co-activating factor-1 (SRC-1) that associate for transcription induction after the conformational change of the AF-2 region (2, 43, 52). The DNA binding domains of PPAR $\gamma$ -RXR $\alpha$  interact with PPAR response elements (PPRE) found within the genome (53). Such elements include 5' regions for aP2 and PEPCK genes as part of adipogenesis, which suggests PPAR $\gamma$  plays a major role in fat cell-specific gene function (53). Though PPAR $\gamma$  is typically known to interact with DNA, it can also interact directly with other proteins to induce activity. For example, as preadipocytes differentiate, expression of C/EBP $\beta$  and C/EBP $\delta$  directly activate PPAR $\gamma$  and C/EBP $\alpha$ , which promote further differentiation and full insulin sensitivity (53). Alternatively, binding by specific ligands can induce activity as well. The use of TZDs in the treatment of T2D improves insulin resistance by increasing GLUT-4 levels and decreasing the levels of cytokines that induce insulin resistance, such as TNF- $\alpha$  and IL-6 (53) by antagonizing the activity of pro-inflammatory transcription

factors (3). Therefore, it is important to note that understanding the interactions involved in co-activator recruitment is crucial for predicting activity after ligand binding, and ultimately treatment of insulin insensitivity and inflammation.

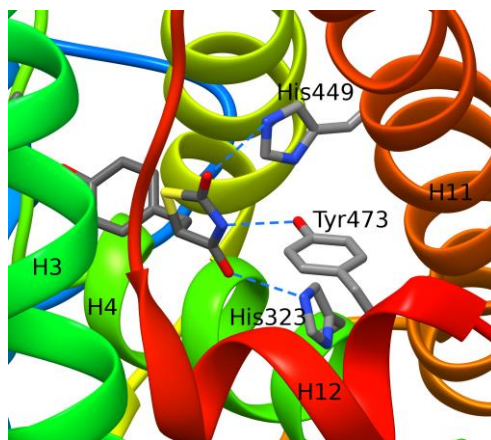
## **2.4 Agonists and the ligand-binding domain of PPAR $\gamma$**

Fatty acids and lipid metabolites have been found to be endogenous ligands for PPAR $\gamma$ . A recent study by Waku et al. (39) gives insight into how these ligands bind covalently to Cys285, thereby modifying PPAR $\gamma$  conformations. In particular, these covalent modifications induce rearrangement of the side-chain network around the created covalent bond in order to generate different transcriptional strengths. This attenuation of strength is specific to the ligand type and conformation. Waku et al. also mention that Ile267 and Phe287 are two key residues repositioned by covalent binding of fatty acids (39). It is also important to note that for some fatty acids, formation of a complex containing two fatty acid units is necessary for binding within the LBD of PPAR $\gamma$  (41).

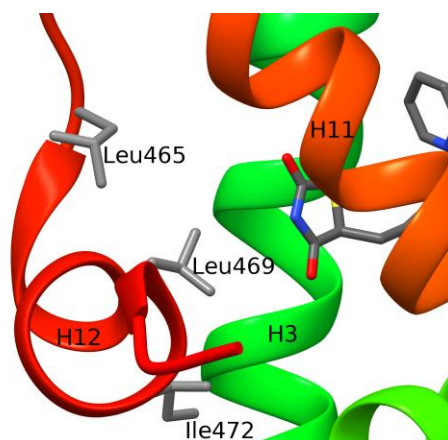
Synthetic ligands that can interact with PPAR $\gamma$  can be divided into at least three classes: full agonists, partial agonists, and antagonists. Full agonists bind and alter the conformation of the AF-2 domain allowing co-activators to bind for activation of genes in both adipogenic and insulin sensitivity processes. Partial agonist binding leads to a change that allows for recruitment of co-activators responsible for insulin sensitivity without affecting adipogenesis. Antagonists show high affinity, but do not activate PPAR $\gamma$ , suggesting the conformational change to AF-2 is either not enough to allow co-activator association or is similar to that of the inactive conformation. A study conducted by Kallenberger and colleagues showed that the dynamics of the AF-2 region plays a major role in the genetic regulation capabilities of PPAR $\gamma$ . Binding of a ligand reduces AF-2 mobility and allows for regulation of gene expression. Furthermore, the AF-2 region of PPAR $\gamma$  can undergo natural mutations, which result in severe insulin insensitivity and cause noticeable changes in dynamics of that AF-2 region (51).

PPAR $\gamma$  agonists typically possess a small polar region and a hydrophobic region that form hydrogen bonds and hydrophobic interactions, respectively, within the LBD. Hydrogen bonding typically occurs between His323, Tyr473, and His449 of the PPAR $\gamma$  LBD and carbonyl oxygen atoms of the ligand (Figure 2.1) (45, 52, 54). Hydrogen bonding of the ligand to Tyr473 is key to the stabilization of the AF-2 region (52, 55). The hydrophobic moiety interacts with other residues in the cavity, such as Leu465, Leu469, and Ile472, establishing hydrophobic interactions to stabilize the domain (Figure 2.2) (45, 52, 54).

In the case of partial agonists, key interactions are different, which result in lesser degrees of AF-2 stabilization and differential stabilization of distinct regions of the LBD (41). Either of these events leads to activation as a result of a shift of the ligand polar group away from the hydrogen-bonding residues. This shift can prevent hydrogen bonding or lead to a different hydrogen-bonding network. Changes in the hydrophobic interactions between the ligand and residues within the LBD also exist. The combination of these events results in conformational changes different enough from those caused by full agonist binding to elicit only partial activation and recruitment of different co-activators (41, 54).



**Figure 2.1** Rosiglitazone bound to LBD of PPAR $\gamma$ . Helices are labeled with H, followed by a number. Key residues involved in hydrogen bonding are labeled. Blue dashed lines represent bonding interactions between the hydrogen atoms of the residue and the oxygen atoms of the ligand. (PDB ID 3DZY) (56).



**Figure 2.2** Rosiglitazone bound to LBD of PPAR $\gamma$ . Helices are labeled with H, followed by a number. Some of the key residues involved in hydrophobic interactions are labeled. (PDB ID 3DZY) (56).

Antagonists for PPAR $\gamma$  have not received the same amount of research interest as the full and partial agonists. Therefore, little information is available on the binding of this type of ligand to the  $\gamma$  isoform. Antagonists for PPAR $\alpha$ , however, have provided insight into how ligands of this class might interact with PPAR $\gamma$  due to the conservation of the mode of co-repressor binding. Typically, co-repressors bind to PPAR $\alpha$  in the absence of ligand. The complex is then stabilized by antagonists, which disrupt any potential interactions with co-activators, and thereby prevent the initiation of transcription (57).

The LBD of PPAR $\gamma$  is a large, T-shaped cavity (54) with a volume of approximately  $1440\text{\AA}^3$  (45, 54), which can easily accommodate many different ligands due to the dynamics of the ligand-binding pocket (58). It is important to note that the type of ligand determines which co-activator associates with the PPAR $\gamma$ -RXR $\alpha$  heterodimer. The co-activator then determines the target gene for regulation and the direction of regulation (up or down). Thus, knowing the final conformation of the LBD that is necessary to elicit a specific activity is crucial for therapeutic development (43).

Until recently, available crystal structures for PPAR $\gamma$  generally were composed solely of

the PPAR $\gamma$  LBD with a ligand bound, a RXR $\alpha$  LBD heterodimerized to PPAR $\gamma$ , and a short segment of a co-activator protein. Chandra and colleagues have published three new crystal structures (3DZU, 3DZY, and 3E00 (56)) for PPAR $\gamma$  composed of the DBD, the hinge region, and the LBD with ligand bound (56). These structures are in complex with RXR $\alpha$ , polypeptides that mimic the LXXLL motif for co-activator binding, and a short DNA segment representative of a PPRE. Observations related to heterodimerization of PPAR $\gamma$  and RXR $\alpha$ , as well as activation of response elements are reported in this study. The LBD and DBD of PPAR $\gamma$  are positioned closely together, which aids in coupling of the PPAR $\gamma$  LBD to the relatively wide space between the LBD and DBD of RXR $\alpha$  (56). The study also discusses the polarity of the PPAR $\gamma$ -RXR $\alpha$  heterodimer, which is determined by the (C)-terminal extension of PPAR $\gamma$  and the DBD interactions of the two subunits. Table 2.1 contains a list of all currently available structures for PPAR $\gamma$ , which can be found in the RCSB PDB online database (59, 60).

**Table 2.1** PDB IDs of published crystal and NMR structures for PPAR $\gamma$  with various ligands bound. Resolution values are in Angstroms (Å). The “Reference number” columns list references for each PDB ID. All PDB IDs list Homo sapiens as the protein source for the PPAR $\gamma$  chain. Access date: 23 March 2009. (An updated list can be found in Appendix A.)

PDB ID	Res. <sup>1</sup> (Å)	Release Date	Reference Number	PDB ID	Res. (Å)	Release Date	Reference Number
1FM6	2.1	2/16/2001	(45)	2Q6S	2.4	10/23/2007	(55)
1FM9	2.1	2/16/2001	(45)	2Q8S	2.3	10/14/2008	(61)
1I7I	2.3	3/9/2002	(62)	2QMV	n/a (NMR)	9/2/2008	(63)
1K74	2.3	12/5/2001	(64)	2VSR	2	8/19/2008	(41)
1KNU	2.5	12/19/2002	(65)	2VST	2.3	8/19/2008	(41)
1NYX	2.7	7/15/2003	(66)	2VV0	2.5	8/19/2008	(41)
1PRG	2.2	1/13/2001	(52)	2VV1	2.2	8/19/2008	(41)
1RDT	2.4	11/9/2004	(67)	2VV2	2.8	8/19/2008	(41)
1WM0	2.9	9/7/2004	(68)	2VV3	2.8	8/19/2008	(41)
1ZEO	2.5	4/25/2006	(69)	2VV4	2.3	8/19/2008	(41)
1ZGY	1.8	7/26/2005	(70)	2ZK0	2.4	2/24/2009	(39)
2ATH	2.3	8/25/2006	(71)	2ZK1	2.6	2/24/2009	(39)
2F4B	2.1	2/14/2006	(72)	2ZK2	2.3	2/24/2009	(39)
2FVJ	2	5/16/2006	(73)	2ZK3	2.6	2/24/2009	(39)
2G0G	2.5	5/16/2006	(74)	2ZK4	2.6	2/24/2009	(39)
2G0H	2.3	5/16/2006	(74)	2ZK5	2.5	2/24/2009	(39)
2GTK	2.1	9/26/2006	(75)	2ZK6	2.4	2/24/2009	(39)
2HFP	2	9/19/2006	(76)	3B3K	2.6	10/28/2008	(77)
2HWQ	2	8/7/2007	(78)	3BC5	2.3	11/18/2008	(79)
2HWR	2.3	8/7/2007	(78)	3CDP	2.8	1/13/2009	(80)
2I4J	2.1	4/17/2007	(54)	3CDS	2.7	12/30/2008	(77)
2I4P	2.1	4/17/2007	(54)	3CS8	2.3	6/3/2008	(81)
2I4Z	2.3	4/17/2007	(54)	3CWD	2.4	7/8/2008	(82)
2OM9	2.8	4/24/2007	(83)	3D24	2.1	6/10/2008	(82)

Table 2.1 continued.

2P4Y	2.3	1/8/2008	(84)	3D6D	2.4	12/30/2008	(77)
2POB	2.3	3/18/2008	(85)	3DZU	3.2	10/28/2008	(56)
2PRG	2.3	7/19/1999	(52)	3DZY	3.1	10/28/2008	(56)
2Q59	2.2	10/23/2007	(55)	3E00	3.1	10/28/2008	(56)
2Q5P	2.3	10/23/2007	(55)	3ET0	2.4	2/17/2009	(86)
2Q5S	2	10/23/2007	(55)	3ET3	2	2/17/2009	(86)
2Q61	2.2	10/23/2007	(55)	3PRG	2.9	8/30/1999	(43)
2Q6R	2.4	10/23/2007	(55)	4PRG	2.9	5/27/1999	(87)

1 <sup>1</sup>Res. = Resolution

## 2.5 Docking

Docking can be defined as predicting both ligand conformation and orientation within a targeted binding site (88). Experimentally derived crystal and NMR protein structures are used as the basis for docking, and the physics involved is based on what is known about atomic and molecular interactions, as well as laws of thermodynamics. All docking methods must include sampling ligand conformations, generating poses of the ligand within the receptor binding site, and scoring the poses.

Before beginning a docking study, one must select from three conformational searching methods: systematic, random, and simulation. The systematic method explores the degrees of freedom possessed by the torsional bonds of a molecule. To achieve this goal, the ligand parts are introduced incrementally in order to obtain an energetically favorable conformation. Random searching, as the name implies, is based on generating random torsional variations of an initial conformation to test against the target. Simulation methods utilize molecular dynamics and energy minimization, and serve best when coupled with one or both of the above searching methods (88).

A large number of docking and dynamics software packages and online servers exist (Table 2.2), many of which are freely available for academic research. The variations in calculation methods and results make each program slightly different. Therefore, the researcher must pick which docking programs are ideal for his or her study. Studies have been performed to assess which programs are ideal for specific screening approaches or particular protein families. For instance, Kellenberger and colleagues published a comparative evaluation of eight widely used docking programs for screening accuracy in 2004 (89). Of the eight docking programs tested, GLIDE, GOLD, and SURFLEX provided the best docking and ranking accuracy within a 2.0 Å cutoff for root-mean squared deviation (RMSD), whereas QXP showed promising docking accuracy but reduced ranking performance. For ranking, FlexX outperformed QXP with percent scoring errors of 15% and 55%, respectively. Efficacy in screening of a compound database was found with SURFLEX, with 8 hits for ligands that bind to a difficult target out of 50 total compounds. GLIDE, GOLD, and FlexX were deemed good programs for virtual screening with hit values of 5, 4, and 4, respectively. Regarding docking times, FRED, which did not perform as well with scoring and docking accuracy, took the least amount of time to perform docking calculations of the eight programs tested, followed by DOCK and FlexX. No single program was deemed the best docking software, but the study demonstrated that the characteristics of the ligand and the target have a significant



effect on the efficiency of the docking program used (89).

**Table 2.2** List of some commonly used molecular dynamics and docking software packages with developer URL. This is not intended to be a comprehensive list of all available dynamics and docking programs available. Available programs are typically free to download for academic use, but some require the purchase of a license for use.

Dynamics	
Program	Developer Site
Amber	<a href="http://ambermd.org/">http://ambermd.org/</a>
AMMP	<a href="http://www.cs.gsu.edu/~cscrwh/progs/progs.html">http://www.cs.gsu.edu/~cscrwh/progs/progs.html</a>
BALLview	<a href="http://www.ballview.org/">http://www.ballview.org/</a>
CHARMM	<a href="http://www.charmm.org/">http://www.charmm.org/</a>
GROMACS	<a href="http://www.gromacs.org">http://www.gromacs.org</a>
LOOS Library	<a href="http://membrane.urmc.rochester.edu/Software">http://membrane.urmc.rochester.edu/Software</a>
YASARA	<a href="http://www.yasara.org/">http://www.yasara.org/</a>
YUP	<a href="http://rumour.biology.gatech.edu/YammpWeb/">http://rumour.biology.gatech.edu/YammpWeb/</a>
ZMM	<a href="http://www.zmmsoft.com/">http://www.zmmsoft.com/</a>
Docking	
Program	Developer Site
AutoDock	<a href="http://autodock.scripps.edu/">http://autodock.scripps.edu/</a>
FlexX	<a href="http://www.biosolveit.de/FlexX/">http://www.biosolveit.de/FlexX/</a>
FRED	<a href="http://www.eyesopen.com/">http://www.eyesopen.com/</a>
GLIDE	<a href="http://www.schrodinger.com/">http://www.schrodinger.com/</a>
GOLD	<a href="http://www.ccdc.cam.ac.uk/">http://www.ccdc.cam.ac.uk/</a>
Sculptor	<a href="http://sculptor.biomachina.org/">http://sculptor.biomachina.org/</a>
SLIDE	<a href="http://www.bch.msu.edu/~kuhn/projects">http://www.bch.msu.edu/~kuhn/projects</a>
SURFLEX	<a href="http://www.tripos.com/">http://www.tripos.com/</a>
UCSF DOCK	<a href="http://dock.compbio.ucsf.edu/">http://dock.compbio.ucsf.edu/</a>

## 2.6 Virtual screening

Because the process of finding a novel compound showing bioactivity can be time-consuming and expensive, structure-based drug design has been established as a vital first step to therapeutic development (90). Screening for ligand conformations can be performed using a ligand-based or a structure-based approach (91, 92). Ligand-based design uses known active and inactive compounds to generate a pharmacophore (92), which is often used in conjunction with quantitative structure-activity relationship (QSAR) analysis to determine ligand-protein interactions. Receptor-based design requires the availability of the receptor structure, which is used to examine the interactions that occur with any members of a large database of ligands (91). Computational screening of large databases of molecules against the three-dimensional structure of a protein has the potential to provide rapid and accurate prediction of the binding modes and affinities of

possible hits for lead optimization. One can pre-screen a database of thousands of compounds and narrow the field of ligands to two or three potential hits in a significantly reduced amount of time compared to laboratory experimental methods. This smaller group would increase the efficiency of experimental assays and new agonist discovery. VS, which incorporates high-throughput docking techniques, is a means to explore the LBD of a protein and make predictions about ligand binding. This technique categorizes ligands that bind to the protein of interest and allows predictions to be made about activation or inhibition of the protein.

Development begins with creating an algorithm that can be followed to set up the testing, followed by running the testing, and finally analyzing the results. Schneider and Böhm define these three issues that must be addressed when performing iterative structure generation respectively as, the construction problem, the docking problem, and the scoring problem (93). Deciding which protein crystal structure to use for all ligands, establishing a set of test parameters, and deciding which ligands to include in the test library make up the first part of the process. Typically, a crystal structure with the highest resolution and fewest missing atoms and residues will be selected. Setting parameters involves re-docking of published structures to reproduce experimentally observed docking conformations (88). The compound database, which can contain numbers of compounds in the thousands (94), should contain small molecules that, based on known chemical interactions between residues of the binding site and known ligands, will bind in varying degrees to the protein of interest and potentially yield the desired effect (e.g. conformational change and activation or inhibition of protein). Protein flexibility is also an important and necessary part of predicting orientations and interactions for many protein families (90), and therefore time should be taken to consider how to incorporate receptor flexibility as well as the binding site microenvironment (i.e. water and/or ions in the binding site). Once the conditions for docking have been established, docking, which is the second step, is relatively straightforward.

The final step, analysis, can often be the most daunting due to the variety of ways output can be interpreted and analyzed. The type of program used to perform simulations has a significant effect on analysis methods because of the information returned. Some programs may be better suited for calculating inhibition constants and free energy of binding estimations than others, whereas still other programs may provide more variables for consideration. There are many different approaches to analyzing results that one can take for scoring the results of a docking study, and these approaches involve examination of interactions on either a fragment or atomic level (93).

## **2.7 Limitations of virtual screening**

Because VS is not, as of yet, a stand-alone process, ligand-binding and reporter assays are essential processes for validating *in silico* results. Docking predicts ligands that may elicit the desired activity, and assays further refine the group of viable candidates to a select group of hits, which, at a specific concentration, will activate the protein of interest.

Further research into lesser-understood biochemical processes is necessary to improve upon the reliability of VS as a stand-alone process. These processes include protein flexibility and induced-fit adaptations, the role of water in solvation, desolvation and

ligand binding, and the involvement of electrostatics (88, 95). Though these unknowns can prove to be problematic when looking at a single computational method, combining strategies is a way to improve upon successful hit rates. Overall, VS saves time and resources when searching libraries of compounds to narrow candidates down to a handful of potential hits that can then be tested experimentally. There is also potential to find hits that may not have been discovered using experimental processes alone.

Another factor that can limit VS productivity is the amount of information available when building a compound library. An information-rich environment is available when considering natural compounds for the treatment and prevention of diseases. Natural plant extracts typically contain a vast number of components that one would need to sift through in order to find the one compound or multiple synergistic compounds that elicit a desired mechanistic affect (i.e. activation of PPARs). VS would prove useful after fractionation of natural extracts and chemical elucidation of key peaks to aid in identifying which compounds within a library are the bioactive compounds. This has the potential to minimize the need for serial HTS when testing for a lead candidate. It is important to note that fractionation is not a necessary step for VS, but can be useful for guiding database building when examining natural extracts for bioactivity.

## **2.8 Docking and virtual screening successes**

Despite the present inability of VS to replace HTS, the two can be complementary approaches to candidate pharmaceutical and nutraceutical searching because of the potential for one method to find activators or inhibitors for which the other method does not show results (95). Klebe (95) mentions in a review of VS strategies a comparison study performed by two groups searching for *Escherichia coli* dihydrofolate reductase inhibitors from a database of approximately 50,000 compounds. The VS portion of the study revealed a number of compounds previously unknown as inhibitors due to insufficient concentrations of the compounds being used during experimental testing (95-98). Klebe also provides a list of targets that have previously been addressed by virtual screening. These targets include nuclear receptors such as retinoic acid receptor and thyroid hormone receptor (95).

## **2.9 Relevance to PPAR $\gamma$ agonist discovery**

Docking techniques would prove useful in the development of new PPAR-based therapeutics, including *in silico* screening of synthetic agonists and natural compounds from plant extracts (i.e. botanicals), all of which have shown promise in the treatment and prevention of immunoinflammatory diseases through PPAR $\gamma$  agonism. Docking and simulation techniques provide a means to pre-screen for and enrich compounds with PPAR $\gamma$  agonism and thereby increase the efficiency of HTS. Docking also allows for structure-based searches for analogues and derivatives of known agonists.

To date, there have been several studies utilizing standard docking methods (74, 99-101) and VS methods (74, 97, 102-105). Most of these studies focus on derivatives or analogues of a particular compound showing high affinity for PPAR $\gamma$ . Studies of this nature can serve two purposes: identify hits for therapeutic development and provide insight into ligand-protein interactions and ligand selectivity.

Xu and colleagues (101) published a study in 2003 in which docking methods were used to look at interactions between PPAR $\gamma$  and eighteen known synthetic and natural agonists in order to determine the pharmacophore of PPAR $\gamma$  agonists. The group determined that PPAR $\gamma$  agonists must have a polar head group and hydrophobic tail in order to form necessary hydrogen bonding and hydrophobic contacts with the LBD, respectively (101).

In another study, Lu and colleagues (74) conducted a structure-based VS search for PPAR $\gamma$  partial agonists as candidates for treatment of T2D with fewer side effects than full agonists. The search revealed a class of ligands that could then be used to test against PPAR $\gamma$ . Two compounds of the class were identified as partial agonists with selectivity among the three PPAR subtypes, and would serve as candidates for further testing. Using VS, they were able to suggest determinants in ligand specificity. The computational results were coupled with x-ray crystallography and assessment of *in vitro* and *in vivo* protein activity (74).

A study regarding natural products identified as PPAR $\gamma$  agonists conducted by Salam and colleagues (103) also utilized structure-based VS to identify 29 potential agonists for experimental testing. Of those compounds, 6 were found to induce PPAR $\gamma$  transcriptional activity *in vitro*. The study also provided insight into the mechanism underlying the flavonoid-induced conformational change and activation of PPAR $\gamma$  (103).

## 2.10 Future directions

Naturally occurring compounds with preventive or therapeutic activity (nutraceuticals) represent a widely used Complementary and Alternative Medicine (CAM) modality and are an alternative to pharmaceuticals (i.e. TZDs) for treating various chronic diseases such as T2D. These natural compounds can modulate gene expression (106) and are typically safer than synthetic counterparts. In the case of T2D, nutraceuticals have the potential to decrease the risk of myocardial infarction, weight gain, and edema associated with current synthetic PPAR $\gamma$  agonist treatments (49, 107). Unfortunately, finding a compound that elicits a desired activity is not always easy because isolating a single compound from a bioactive extract is time consuming and expensive (108) and the mechanism by which the compound works is often unknown (109). VS, in combination with conventional experimental methods, has the potential to put the discovery of bioactive botanical constituents in a better competitive position with mainstream pharmaceutical research by reducing time and costs. For instance, in a study published by Rollinger and colleagues (110), a chemical feature-based pharmacophore modeling VS technique, in combination with ethnopharmacology, was utilized to identify inhibitors for cyclooxygenase (COX) I and II. Of the thousands of compounds listed in the four databases used (WDI, NCI, NPD, and DIOS), the success rate of finding known inhibitors within these three-dimensional databases was enhanced through the use of VS techniques (110).

A preliminary comparison of a small group of PDB PPAR $\gamma$  structures shows an overall conservation of backbone conformation across the available structures. This is relevant to the selection of a single macromolecule for large-scale automated testing. These findings suggest it is possible to select one macromolecule for all ligand types with a limited degree of error. It is important to note that though there is a relative consensus position for all key residues, some variation in the positions of key residues due to ligand

interactions are present. Therefore, this issue must be considered and several structures must be examined when deciding on a single macromolecule crystal structure for VS.

Another computational method that may prove useful is molecular dynamics (MD), which involves the use of computational chemistry to predict the dynamics of complex molecular systems and the macroscopic properties of those systems based on detailed atomic knowledge (111). Implementing MD would prove useful for examination of conformational changes and molecular interactions, which would allow for expansion upon what is known about how PPARs interact with ligands and other macromolecules.

To discover potential nutraceutical/CAM hits, further assessment of PPAR $\gamma$  and ligand characteristics is necessary to determine the best screening approach and which scoring functions compare for analysis. If the components of an extract are known or if one can speculate as to which compounds are present, a database of chemically related compounds could be created to test against PPAR $\gamma$ , and a smaller hit group can be identified for experimentation. Another necessary element is collecting experimentally proven properties for comparison to computationally derived data. Future work could also encompass finding co-agonists and pan-agonists for PPAR subtypes.

## **2.11 Acknowledgements**

Funding for this work was partially provided by the NIH Biomedical and Behavioral Sciences Research Training Grant R25 GM072767 (Virginia Tech Initiative to Maximize Student Diversity) and Grant 5R01AT4308 from the National Center for Complementary and Alternative Medicine.

### **3 Dietary $\alpha$ -eleostearic acid ameliorates experimental inflammatory bowel disease in mice by activating peroxisome proliferator-activated receptor- $\gamma$**

Stephanie N. Lewis<sup>1,2,3,4</sup>, Lera Brannan<sup>2</sup>, Amir J. Guri<sup>3</sup>, Pinyi Lu<sup>1,3,4</sup>, Raquel Hontecillas<sup>3,4</sup>, Josep Bassaganya-Riera<sup>1,3,4</sup>, and David R. Bevan<sup>1,2,4</sup>

<sup>1</sup>Genetics, Bioinformatics, and Computational Biology Program, Virginia Tech, Blacksburg, VA, USA

<sup>2</sup>Department of Biochemistry, Virginia Tech, Blacksburg, VA, USA

<sup>3</sup>Nutritional Immunology and Molecular Medicine Laboratory, Virginia Bioinformatics Institute, Blacksburg, VA, USA

<sup>4</sup>Center for Modeling Immunology to Enteric Pathogens, Virginia Bioinformatics Institute, Blacksburg, VA, USA

Copyright: © 2011 Lewis et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: SN Lewis, L Brannan, AJ Guri, P Lu, R Hontecillas, J Bassaganya-Riera, and DR Bevan. (2011) Dietary  $\alpha$ -Eleostearic Acid Ameliorates Experimental Inflammatory Bowel Disease in Mice by Activating Peroxisome Proliferator-Activated Receptor- $\gamma$ . PLoS ONE. 6(8): e24031. doi:10.1371/journal.pone.0024031.

### 3.1 Abstract

**Background:** Treatments for inflammatory bowel disease (IBD) are modestly effective and associated with side effects from prolonged use. As there is no known cure for IBD, alternative therapeutic options are needed. Peroxisome proliferator-activated receptor-gamma (PPAR $\gamma$ ) has been identified as a potential target for novel therapeutics against IBD. For this project, compounds were screened to identify naturally occurring PPAR $\gamma$  agonists as a means to identify novel anti-inflammatory therapeutics for experimental assessment of efficacy.

**Methodology/Principal Findings:** Here we provide complementary computational and experimental methods to efficiently screen for PPAR $\gamma$  agonists and demonstrate amelioration of experimental IBD in mice, respectively. Computational docking as part of virtual screening (VS) was used to test binding between a total of eighty-one compounds and PPAR $\gamma$ . The test compounds included known agonists, known inactive compounds, derivatives and stereoisomers of known agonists with unknown activity, and conjugated trienes. The compound identified through VS as possessing the most favorable docked pose was used as the test compound for experimental work. With our combined methods, we have identified  $\alpha$ -eleostearic acid (ESA) as a natural PPAR $\gamma$  agonist. Results of ligand-binding assays complemented the screening prediction. In addition, ESA decreased macrophage infiltration and significantly impeded the progression of IBD-related phenotypes through both PPAR $\gamma$ -dependent and -independent mechanisms in mice with experimental IBD.

**Conclusions/Significance:** This study serves as the first significant step toward a large-scale VS protocol for natural PPAR $\gamma$  agonist screening that includes a massively diverse ligand library and structures that represent multiple known target pharmacophores.

## 3.2 Introduction

Inflammatory bowel disease (IBD) is a chronic and recurring inflammatory disease with two clinical manifestations: ulcerative colitis (UC) and Crohn's disease (CD). UC and CD affect over 4 million individuals in the United States and accrue a significant portion of the estimated \$1.7 billion in health care costs for prevalent gastrointestinal diseases (CDC2007). While the etiopathogenesis of IBD remains unclear, it has been suggested that chronic mucosal inflammation characteristic of IBD is associated with a disruption in immune homeostasis (112). As such, treatments for IBD should correct this immune dysregulation in order to prevent or reduce gut mucosal damage.

There is no cure for IBD, but treatments are available to combat the associated symptoms. One such treatment, 5-aminosalicylic acid, targets the nuclear hormone receptor peroxisome proliferator-activated receptor-gamma (PPAR $\gamma$ ), which is highly expressed in the colonic epithelial and immune cells (22, 113-117). PPAR $\gamma$  and PPAR $\delta$  serve as targets for the treatment of inflammatory and immune-mediated diseases because of the role they play in maintaining homeostasis and suppressing inflammation (1, 112, 117, 118). PPAR $\gamma$  in particular is known to play a role in transcriptional regulation of anti-inflammatory processes via co-activator recruitment (1, 2, 116). Ligand-induced activation of PPAR $\gamma$  can antagonize the activity of pro-inflammatory transcription factors such as nuclear factor kappa-light-chain-enhancer of activated B cells (NF- $\kappa$ B), signal transducer and activator of transcription (STAT), and activator protein (AP)-1 (119). Other IBD treatments currently available include infliximab, which is an anti-tumor necrosis factor-alpha (TNF- $\alpha$ ) antibody (120, 121), and corticosteroids, which systemically suppress immunity (122). These medications are modestly successful for the long-term management of IBD but are associated with significant side effects, including increased risk of infection and cancer (123, 124). Interestingly, the insulin-sensitizing PPAR $\gamma$  agonists used for treating type 2 diabetes, such as rosiglitazone and pioglitazone, have proven useful at ameliorating IBD effects in humans with UC (30). However, rosiglitazone, and other PPAR $\gamma$  agonists of the thiazolidinediones (TZD) class of anti-diabetic drugs, are unlikely to be adopted by gastroenterologists for the treatment of IBD due to associated side effects (30) including hepatotoxicity, weight gain, fluid retention leading to edema, and congestive heart failure (49). In this regard, the U.S. Food and Drug Administration (FDA) restricted the use of rosiglitazone in 2010 due to its side effects, whereas the European Medicines Agency completely banned its use in the European market. Natural therapeutics, such as fatty acids that induce PPAR $\gamma$  activation, might be a safer alternative to current treatments and TZDs.

Our group has conducted several preclinical animal model studies to suggest that supplementation of diet with fatty acids, such as conjugated linoleic acid (CLA) (42, 118) or agonistic botanicals, is effective at ameliorating colonic inflammation in mouse and pig models of IBD through a PPAR $\gamma$ -dependent mechanism (42, 118, 125, 126). In an effort to expedite the drug and natural product therapeutic discovery process, virtual screening (VS) can complement traditional experimental methods for identification of novel PPAR $\gamma$  agonists. VS represents a cost- and time-efficient means of screening thousands of compounds within thematic libraries that justify further experimental assessment (95). We are undertaking VS to identify novel PPAR $\gamma$  agonists within a



collective of large compound databases. As a feasibility test, we screened a small group of known and proposed agonists, with the inclusion of known negative controls. The focus of this small-scale screen was to test our PPAR $\gamma$  structural model, and assess binding of natural compounds, with significant emphasis on conjugated trienes.

Conjugated trienes were selected due in part to their structural similarity to CLA. In addition, conjugated trienes exhibit effectiveness at ameliorating chronic inflammation (5, 127). One such compound,  $\alpha$ -eleostearic acid (ESA; 9Z11E13E-18:3), has been found at concentrations of 60-80% in tung and bitter melon seed oils (128). ESA has been shown to suppress tumor angiogenesis (129) and MCF-7 breast cancer cell proliferation via PPAR $\gamma$  activation (130), induce apoptosis via lipid peroxidation (131), and induce autophagy-dependent cell death through AKT/mTOR and ERK1/2 signal targeting (132). Evidence also indicates that punicic acid plays a significant role in increasing lipid peroxidation (133) and inhibiting TNF- $\alpha$ -induced neutrophil hyperactivation to protect against experimentally induced colon inflammation in rats (7). Our group has found that punicic acid ameliorates type 2 diabetes-induced inflammation by activating PPAR $\gamma$  and PPAR $\alpha$ , and repressing TNF- $\alpha$  expression in white adipose tissue and liver (127) and increases peripheral insulin sensitivity (134) without causing any adverse side effects (135). We have also demonstrated that punicic acid prevents experimental IBD through PPAR $\gamma$ - and PPAR $\delta$ -dependent mechanisms (136). Catalpic acid improves abdominal fat deposition, improves glucose homeostasis and up-regulates PPAR $\alpha$  expression in adipose tissue of mice (5). Though these plant-derived conjugated trienes suggest anti-inflammatory efficacy in various disease models, it has been suggested that ESA induces a greater degree of antioxidant activity than punicic acid in mice (137). Punicic acid ameliorates both diabetes (136) and gut inflammation (127) without causing side effects (135), whereas ESA elicits mainly anti-inflammatory and anti-carcinogenic effects (129-132). A goal of this study was to test the effectiveness of ESA in an experimental IBD model. Additionally, small-scale VS was conducted to test the predictability of our VS protocol for identifying PPAR $\gamma$  full agonists in the hopes of finding natural therapeutics and/or prophylactics for treating IBD and other chronic inflammation-related diseases. The computational portion of our study revealed information complementary to the predictions of our *in vitro* analysis, pre-clinical efficacy, and mechanistic testing in mice.

### **3.3 Methods**

#### **3.3.1 Docking procedure**

AutoDock 4.0 (138) (AD4) was used for structural model testing, while AutoDock Vina (139) (Vina) was used for screening a subset of our in-house ligand database against the selected structural models of PPAR $\gamma$ . AutoDock Tools 1.5.2 (ADT) was used to build the appropriate charged protein and ligand files for docking. Default values for the Lamarckian Genetic Algorithm (LGA) were used for docking with AD4, with the exception of the maximum number of energy evaluations, which was reduced to 250,000. Adjusting this number reduced the screening time without significantly affecting pose prediction. Five iterations of AD4 with 50 poses generated per iteration were conducted for the re-docking step totaling 250 poses per protein structure model. Vina was used for cross-docking and to run the small-scale screening. Three Vina iterations were conducted

for each ligand in the cross-docking step, while a single run was conducted for the small-scale screening. As a means to further sample conjugated triene geometry, three AD4 iterations of 50 poses each were run for each compound, which was a total of 150 poses per conjugated triene for each selected protein structure model. Scripts available through the AD4 development site (<http://autodock.scripps.edu/>) were modified and used to automate the screening process. Modifications to the scripts included exchanging the AD4 executable for the Vina executable and all subsequent necessary changes for Vina functionality.

### **3.3.2 Structural model selection: Re-docking component**

Five structures with co-crystallized rosiglitazone were downloaded from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (59, 60) (<http://www.pdb.org>). The selected structure IDs were 1FM6 (45), 1ZGY (140), 2PRG (52), 3CS8 (81), and 3DZY (56). These structures were evaluated to identify a PPAR $\gamma$  structural model that would be appropriate for docking in a full agonist-like pose. Completeness of structure, crystal resolution, and re-docking ability were the factors considered. Re-docking refers to the ability of a docking program to reproduce the co-crystallized binding geometry and orientation of the associated ligand given a rigid macromolecule state. The PDB structures were superimposed and rosiglitazone was isolated from each protein structural model with the UCSF Chimera software package (141).

Re-docking was conducted with both native and non-native initial rosiglitazone conformations. Native refers to use of coordinates for the co-crystallized ligand structure of the respective protein structure model, whereas non-native refers to use of initial coordinates not found in the original PDB file. For the native test, each isolated rosiglitazone was re-docked into its respective protein structure (e.g., five protein models each with a different rosiglitazone coordinate files). For the non-native test, a single rosiglitazone structure was randomly selected for re-docking into all five structure models. Ligand flexibility and random initial geometry for the ligand reduced possible bias associated with use of a native ligand for one test structure, which was non-native for the other four. A comparison of results for the native and non-native ligand re-docking suggested the randomized initial conformation for rosiglitazone does not affect pose prediction as the predicted poses for both test sets were similar (data not shown). The non-native procedure involved docking of a single ligand structure to the protein structures, which is similar to what would be used for large-scale screening. Therefore, data from the non-native re-docking was analyzed and provided here. Both the superimposed positioning and the use of a single rosiglitazone model established a relatively controlled test set: overlaid coordinate space for the test structures, which translated to similar grid areas, with a single ligand coordinate file for testing.

### **3.3.3 Structural model selection: Cross-docking component**

Co-crystallized ligands from various PDB files were used for cross-docking to test predictability for other known agonists. Cross-docking refers to docking different ligand structures isolated from multiple PDB structures of the same protein to a single selected model structure. Ligands from 1FM9 (45), 2F4B (72), 2HWQ (78), 2I4J (54), 2I4P (54),

2VSR (41), 2VST (41), 3ET3 (86), 2VV0 (41), 2VV1 (41), 2ZK1 (39), and 2ZK2 (39) were included in the library for this purpose (Table 3.3).

### **3.3.4 Small-scale in-house ligand library construction**

Our small-scale ligand library included the rosiglitazone structure from re-docking, several of the cross-docking ligands, known PPAR $\gamma$  agonists, and known inactive compounds. Inclusion of the ligands from the re-docking and cross-docking steps served as controls for successful and unsuccessful docking. A search of published literature was conducted to find both naturally and synthetically derived compounds shown experimentally to either activate or not activate PPAR $\gamma$  (103, 142-144). Structural models for non-crystallized ligands were downloaded from the UCSF ZINC database online (<http://zinc.docking.org/>). Any structures not available through ZINC were built using the Dundee PRODRG2 server (145) (<http://davapc1.bioch.dundee.ac.uk/prodrg/>). Structures built with PRODRG2 were examined to ensure conservation of stereochemistry. Charges for all of the ligands in the database and the protein were generated using ADT. Eighty-one compounds total were tested in this study. A complete list of ligands included in the test library can be found in Table 3.S1.

### **3.3.5 Docking analysis for re-docking and cross-docking**

The most energetically favorable pose for each ligand of the re-docking (25 lowest energy poses) and cross-docking (108 lowest energy poses) steps were used for analysis. Reference poses for root mean-squared deviation (RMSD) calculations were taken from crystal structure complexes for each ligand. These protein-ligand complex structures were superimposed onto the test structures to obtain a common coordinate space prior to the RMSD calculation. For re-docking, RMSD values are exact given each PPAR $\gamma$ -rosiglitazone complex was used as the reference for the respective results. However, the reported RMSD values for cross-docking were relative rather than absolute given the co-crystallized reference ligand coordinates are not relative to the protein structure models used for testing. The idea of relative RMSD stems from differences in side chain rotamers between the crystal structures. Side chain position is governed, in part, by ligand binding, which meant differences could be seen in binding cavity residue positions when the rosiglitazone-bound test structures were compared to each additional PPAR $\gamma$  structure model. These differences, which affect intramolecular interactions, resulted in minor deviations of the backbone on some regions for the superimposed structures relative to the test structure. This could mean the position of each co-crystallized reference ligand relative to the test structures was shifted slightly as well. However, there were areas of the backbone that superimposed without noticeable deviations. As the deviations between backbone positions were not consistent, adjusting for any rotamer-induced shifts in co-crystallized ligand coordinates was not feasible. Therefore, RMSD values for docked poses for each ligand were deemed “relative” as an acknowledgement of these minor variations in coordinates. An average RMSD, population standard deviation, and variance were calculated for each ligand (See Formulas document of Supporting Information). Re-docking and cross-docking results for each ligand relative to each test protein structure were deemed successful if the RMSD was less than 2.0 Å (89).

Docking success versus failure for re-docking and cross-docking was assessed qualitatively as well. Docked poses for rosiglitazone on the surface of the protein or near the opening of the binding cavity were deemed unsuccessful. Poses for which the molecule was not properly oriented, such as the imidazole ring of rosiglitazone positioned near the cavity opening rather than near the rear of the pocket, were deemed unsuccessful as well given such orientations would not match the co-crystallized coordinates. Similar conditions relative to each cross-docking ligand were also identified and assessed.

### **3.3.6 Docking analysis for small-scale VS**

To prepare for analysis of the small-scale VS results, interactions from various crystal structures were identified and cataloged. Reported crystal structure interactions for the five rosiglitazone-containing structures from the re-docking step and six fatty acid-containing structures from the cross-docking step were compiled using RCSB Ligand Explorer (146). Residue atoms common to more than one interaction list for a specific ligand type were pooled and used as a reference list for analysis after docking. As such, there were two master interaction lists: rosiglitazone-like interactions (Table 3.S2) and fatty acid-like interactions (Table 3.S3). Common interactions between the two lists were also noted (Table 3.S4).

Perl (147) scripts to automate pose distance measurement calculations and pose interaction predictions were also composed and used. The most energetically favorable docked pose for each ligand relative to the macromolecule were pooled for analysis. Only the potential for a ligand to fall into the full agonist category of ligands was assessed in depth for this study. Full agonism has been suggested to require interactions with Ser289, His323, His449, and Tyr473, which are residues positioned in the portion of the binding cavity proximal to the activation function-two (AF-2) region (Figure 3.S1). Interactions in this region govern AF-2 conformational changes necessary for PPAR $\gamma$  activation. Distance measurements between the top docked poses (77 lowest energy poses) were calculated and used to predict interactions. Interactions similar to those seen in the pooled crystal structure data were deemed “successful”. Potential hydrogen bonds were assessed based on distances between the donor/acceptor heavy atoms of the test ligand pose and four key residues. Lengths measuring less than 3.3 Å were considered potential hydrogen bond interactions (103, 146). Potential hydrophobic interactions were set to a distance threshold of 3.9 Å between carbon atoms (146). Predicted interactions for each ligand were counted and a screen for the presence of hydrogen bond interactions with the key residues listed above was conducted to determine docking success.

### **3.3.7 Ligand binding assay**

ESA was introduced at various concentrations (0.001-10 $\mu$ M) to solution containing PPAR $\gamma$  protein complexed with a fluorophore-bound compound (Fluormone<sup>TM</sup>, Invitrogen). This mixture was allowed to incubate for 20 hours. The ability of the test compound, which here was ESA, to displace Fluormone<sup>TM</sup> was calculated as mean polarization, where a decrease in polarization corresponded to an increase in ligand binding activity as previously described (148).

### **3.3.8 Transfection of RAW 264.7 cells**

RAW 264.7 mouse macrophage precursor cells (ATCC, Manassas, VA) were grown in 24-well plates in DMEM high glucose medium (Invitrogen, Carlsbad, CA) containing 10% fetal bovine serum until 60-70% confluence. Transfected cells were treated with varying concentrations of ESA (0, 1, 5, and 10  $\mu\text{M}$ ; Sigma) or rosiglitazone (1  $\mu\text{M}$ ; Cayman Chemicals, Ann Arbor, MI) for 24 hours. Other details of the protocol were as previously described (148, 149). Relative luciferase activity was calculated as a ratio between beginning and ending chemiluminescence values for a 10-second time period.

### **3.3.9 Animal procedures**

The protocol for animal care and genotyping of the mice was described previously (118). An ESA-supplemented diet was tested against a control (AIN-93G-based) diet in a dextran sodium sulfate (DSS)-induced IBD mouse model. Sixty mice were divided according to diet (ESA versus control), genotype (PPAR $\gamma$  flfl; MMTV-Cre-/PPAR $\gamma$ -floxed versus epithelial cell- and immune cell-specific PPAR $\gamma$  flfl; MMTV-Cre+/PPAR $\gamma$ -null), and DSS-challenge. Ten mice (5 for each genotype) from the control diet group and 9 mice (4 PPAR $\gamma$ -floxed and 5 PPAR $\gamma$ -null) from the ESA diet group were not given DSS-treated water as a control for the disease state. Drinking water with 2.5% DSS was administered to the test mice for a period of seven days. Body weights and disease activity index (DAI) values were recorded each day of the seven-day DSS treatment period. Procedures for assigning DAI values have been previously described (118). Mice were euthanized on day seven of the DSS challenge by CO<sub>2</sub> asphyxiation followed by secondary thoracotomy. Blood was withdrawn from the heart, after which spleen, mesenteric lymph nodes (MLNs), and colonic samples were examined for gross pathological lesions and isolated from each mouse. Organs were examined to assign scores based on size and macroscopic inflammatory lesions (0-3). Spleen and MLN were crushed to produce single-cell suspensions for flow cytometry, while colon samples were used for mRNA isolation and histological examination. This study was approved by the Virginia Tech Institutional Animal Care and Use Committee (IACUC) on May 15, 2008 under animal welfare assurance number A3208-01.

### **3.3.10 Histopathology**

Experimental design for histopathology was previously described (118, 125). Epithelial erosion, mucosal thickness, and immune cell infiltration were each assessed and scored (0-4) for colon cross-sectional samples stained with hematoxylin and eosin from each mouse.

### **3.3.11 Immunophenotyping**

Whole blood and MLN cells were seeded onto 96-well plates and treated with fluorochrome-conjugated antibodies. Monocyte/macrophage subsets were assessed using anti-F4/80-PE-Cy5 (5 mg/mL, eBioscience) and anti-CD11b-Alexa Fluor 700 (2 mg/mL, eBioscience). The lymphocyte subset was assessed with anti-CD4-Alexa Fluor 700 (2 mg/mL; BD Pharmingen), anti-CD8-PerCp-Cy5.5 (2 mg/mL, eBioscience), anti-CD3 PE-Cy5 (2 mg/mL; BD Pharmingen), anti-FoxP3-PE (2 mg/mL, eBioscience), and anti-IL10-

PE as previously described (4). Flow results were computed with a BD LSR II flow cytometer and data analysis was performed with the FACS Diva software package (BD).

### **3.3.12 Quantitative real-time RT-PCR**

Total RNA was isolated from colonic tissue using procedures previously described (125). PCR was performed on complementary DNA (cDNA) using Taq DNA polymerase (Invitrogen, Carlsbad, CA) and previously described methods and conditions (118, 125). cDNA concentrations for genes of interest were examined by quantitative real-time PCR using an iCycler IQ System and the iQ SYBR green supermix (Bio-Rad). A standard curve was generated for each gene using methods previously described (125). In addition, a melting curve analysis was performed for each product using previously described methods (125) in order to determine the number of products synthesized while excluding non-specific products and primer dimers. Real-time PCR was used to quantify the starting amount of nucleic acid of each unknown cDNA sample. Primer sequences, the length of the PCR product, and gene accession numbers have been outlined previously (125, 149). Primers used for this study were the forward and reverse cohorts of VCAM-1, ICAM-1, IL-6, and  $\beta$ -actin (125).

### **3.3.13 Statistical analysis**

Data were analyzed as a completely randomized design with statistical significance assessed using the analysis of variance (ANOVA) method. The general linear model procedure of the Statistical Analysis Software (SAS) package (SAS Institute Inc., Cary, NC) was run for weight, DAI, flow cytometry data, and histopathology scores to determine variance across and significance between treatment groups. Statistical significance was assessed based on a probability value ( $p$ ) less than or equal to 0.05. Significant models were further assessed using the Fisher's Protected Least Significant Difference multiple comparison method.

## **3.4 Results**

### **3.4.1 Selection of structural model: Re-docking component**

Structures with co-crystallized rosiglitazone (example given in Figure 3.S1) were used for re-docking because rosiglitazone was the positive control in the experimental studies, it is a known PPAR $\gamma$  agonist, and the purpose of this docking feasibility test was to find compounds that mimic rosiglitazone-induced activation. The top scoring pose from each of the five 50-pose replicates was selected for further analysis. This selection method was applied for each of the five starting structures, giving a total of 25 poses for comparison.

The RMSD and free energy of binding were averaged for the five poses for each protein structure model (Table 3.1). Additionally, the population-based standard deviation and variance were calculated. The average pose RMSD values for three structures, 1FM6, 1ZGY, and 2PRG, were within 2.0 Å of the crystal structure position. Of these three, 1ZGY possessed the highest standard deviation and variance values, which suggested that some poses with low and high RMSD values should be present. Examination of the poses for all five structures revealed that the lowest RMSD value (0.99 Å) for all rosiglitazone

poses was in the 1ZGY pose group as was the pose with the highest RMSD value (3.05 Å). Thus, we favored the 1ZGY structure for further docking studies because this structure enabled docking at the known rosiglitazone binding position as well as docking at other energetically favorable positions within the binding site, suggesting that it might accommodate ligands of diverse structure. To further confirm this selection, cross-docking with known ligands from other PDB structures was conducted.

**Table 3.1** Average RMSD and free energy of binding (kcal/mol) for re-docking of rosiglitazone (N = 5).

PDB ID	Resolution (Å)	RMSD			kcal/mol		
		Mean	Standard Deviation	Variance	Mean	Standard Deviation	Variance
1FM6	2.1	1.76	0.561	0.314	-7.58	0.487	0.237
1ZGY	1.8	1.91	0.925	0.856	-7.19	0.247	0.061
2PRG	2.3	1.84	0.357	0.128	-7.66	0.228	0.052
3CS8	2.3	2.81	0.101	0.010	-6.63	0.184	0.034
3DZY	3.1	2.82	0.183	0.034	-7.06	0.133	0.018

### 3.4.2 Selection of structural model: Cross-docking component

1ZGY, 1FM6, and 2PRG were included in the cross-docking testing as each showed successful re-docking and contained ligand-binding domains without missing loops or sequence segments. Structures 3CS8 and 3DZY were missing the H2'-H3 loop and did not result in accurate pose prediction for rosiglitazone. Rosiglitazone poses for 3CS8 and 3DZY occupied the portion of the binding cavity opening in which the H2'-H3 loop would normally sit (data not shown). This loop proved necessary for successful agonist docking given the poor success rate of re-docking in the absence of this region.

Vina was used for cross-docking instead of AD4 as the former was more time-efficient for the number of ligands used and the number of replicates to be carried out. It has also been reported that Vina better predicts poses for ligands with higher numbers of torsions (150), which was the case for some of the ligands used in cross-docking. Replicates were conducted with Vina for two reasons: to determine if replicates would be necessary in a larger-scale study, and to aid in the protein structure model selection process. Three replicate screens were run and each lowest-energy pose was analyzed (3 protein models x 3 replicates x 12 ligands = 108 lowest energy poses). Analysis of the cross-docking results included a comparison of RMSD values, free energy of binding, and number and identity of known interactions between each ligand and PPAR $\gamma$  based on the crystal structures of the complexes. Results from comparison of RMSD values and free energy of binding are listed in Table 3.2, with full ligand names listed in Table 3.3. To simplify the process of cross-docking of several ligands to multiple receptor structures, the initial crystal protein-ligand complexes were superimposed prior to docking. This practice allowed for RMSD values to be easily calculated between the docked ligand poses and crystal reference poses as the structures shared coordinate space.

The results relative to each of the test structure models were not completely consistent across all the models. The lowest overall average RMSD was seen with 1ZGY for the (2S)-ureidofibrate-like derivative. This ligand did not dock as well into 1FM6 and 2PRG. A similar comparative docking pattern was seen for 4-HDHA. Only one ligand, PTG taken from PDB ID 2ZK1 (PTG-1), docked within the 2.0 Å threshold across the three structural models. It should be noted here that the PTG structure taken from PDB ID 2ZK2 possessed different charges than the same compound from 2ZK1. The difference in charge is most likely due to the difference in crystallization states. 2ZK2 had glutathione covalently bound to PTG-1 as part of crystallization, whereas 2ZK1 did not. The glutathione-PTG-1 compound would therefore have more atoms over which charges would be distributed.

The RMSD, standard deviation, and variance values for farglitazar, 9-HODE, indeglitazar, and PTG-1 showed the most consistency across the three proteins, with PTG-1 showing favorable average RMSD values and negligible variance for each protein structure. For PTG-1, this suggested the ligand docked similarly to all three protein structures. When the replicate poses for the four compounds were assessed visually, the deviations for the 9-HODE poses were due in large part to variation in the placement of the hydrophobic tail portion, the PTG-1 poses docked more similarly to 9-HODE than the PTG-1 reference structure, and the indeglitazar poses occupied the middle portion of the binding cavity rather than the rear activation site. The placement of the indeglitazar and PTG-1 poses appeared to be due to the shape of the binding cavity at the rear of the pocket, which was mentioned previously to be the issue with farglitazar. This hindrance was seen to a lesser degree with PTG-1 as there is sufficient space to allow interactions despite lack of exact congruence to the co-crystallized reference. Indeglitzar and farglitazar poses were consistently unsuccessful due to the binding cavity restriction, whereas PTG-1 occupied a fatty acid-like orientation given the similarity of this compound to the types of ligands that can appropriately fill the allotted molecular space.

All of the poses had negative calculated free energy of binding values given the ligand structures and charge environment of the binding cavity. These values were energetically feasible, but were not an indication of the most favorable conformation for ligands that did not agree with the reference structure geometry. Therefore, RMSD and free energy of binding measurements were not enough to determine successful cross-docking for PPAR $\gamma$ . A visual assessment of poses suggested rosiglitazone and fatty acid compounds dock the best into the selected models. As such, interactions from crystal structures containing these compounds were used to generate a list of favorable interactions that might indicate successful docking. The residues considered are listed in Table 3.S4.



**Table 3.2** Average RMSD and free energy of binding from cross-docking for various ligands relative to each listed PDB ID (top row) (N = 3).

	1FM6			1ZGY			2PRG		
	RMSD (Å)								
PDB Ligand ID	Mean	SD <sup>1</sup>	Var <sup>2</sup>	Mean	SD	Var	Mean	SD	Var
243	2.82	0.199	0.040	2.82	0.014	0.000	2.60	0.040	0.002
570	3.19	0.000	0.000	3.08	0.007	0.000	3.13	0.050	0.003
4HD	1.81	0.365	0.134	1.40	0.018	0.000	2.19	0.236	0.056
9HO	1.73	0.184	0.034	1.85	0.270	0.073	1.70	0.162	0.026
DRH	2.74	0.030	0.001	1.55	0.209	0.044	2.17	0.251	0.063
DRJ	1.63	0.807	0.652	1.72	0.417	0.174	2.03	0.175	0.031
DRY	3.23	0.002	0.000	2.26	0.024	0.001	1.89	0.019	0.000
EHA	2.47	0.524	0.275	2.45	0.386	0.149	1.89	0.007	0.000
ET1	2.83	0.001	0.000	2.68	0.003	0.000	2.72	0.001	0.000
HXA	2.49	0.616	0.380	1.99	0.171	0.029	1.85	0.009	0.000
PTG-1	1.78	0.000	0.000	1.78	0.005	0.000	1.65	0.019	0.000
PTG-2	2.68	0.023	0.001	2.53	0.244	0.059	2.53	0.091	0.008
	Free energy of binding (kcal/mol)								
PDB Ligand ID	Mean	SD	Var	Mean	SD	Var	Mean	SD	Var
243	-6.87	0.094	0.009	-6.57	0.047	0.002	-6.47	0.309	0.096
570	-10.43	0.047	0.002	-11.00	0.000	0.000	-10.50	0.082	0.007
4HD	-7.00	0.163	0.027	-7.53	0.170	0.029	-6.97	0.047	0.002
9HO	-6.40	0.082	0.007	-6.70	0.082	0.007	-6.37	0.094	0.009
DRH	-8.23	0.047	0.002	-8.83	0.125	0.016	-8.17	0.047	0.002
DRJ	-8.63	0.094	0.009	-8.80	0.294	0.087	-8.37	0.094	0.009
DRY	-10.03	0.047	0.002	-10.13	0.047	0.002	-10.37	0.047	0.002
EHA	-10.10	0.082	0.007	-10.10	0.082	0.007	-10.60	0.000	0.000
ET1	-8.10	0.000	0.000	-8.50	0.000	0.000	-8.50	0.000	0.000
HXA	-7.00	0.327	0.107	-7.90	0.082	0.007	-7.33	0.047	0.002
PTG-1	-7.00	0.000	0.000	-7.23	0.047	0.002	-7.20	0.082	0.007
PTG-2	-7.17	0.170	0.029	-7.10	0.082	0.007	-7.47	0.094	0.009

<sup>1</sup>SD = Standard Deviation

<sup>2</sup>Var = Variance

**Table 3.3** Full names and structures for compounds listed by ligand ID in Table 3.2. Ligand IDs from respective PDB files were used. Ligand structures can be found in Table 3.S1.

<b>PDB Ligand ID</b>	<b>PDB ID</b>	<b>Reference</b>	<b>Ligand Name<sup>1</sup></b>
243	2VST	(41)	13-hydroxyoctadecadienoic acid (13-HODE)
570	1FM9	(45)	GI262570 (Farglitazar)
4HD	2VV1	(41)	(4S,5E,7Z,10Z,13Z,16Z,19Z)-4-hydroxydocosa-5,7,10,13,16,19-hexaenoic acid (4-HDHA)
9HO	2VSR	(41)	9-hydroxyoctadecadienoic acid (9-HODE)
DRH	2I4P	(54)	(2S)-2-[4-[2-(1,3-benzoxazol-2-yl-heptyl-amino)ethyl]phenoxy]-2-methyl-butanoic acid ((2S)-ureidofibrate-like derivative)
DRJ	2I4J	(54)	(2R)-2-[4-[2-(1,3-benzoxazol-2-yl-heptyl-amino)ethyl]phenoxy]-2-methyl-butanoic acid ((2R)-ureidofibrate-like derivative)
DRY	2HWQ	(78)	[(1-{3-[(6-benzoyl-1-propyl-2-naphthyl)oxy]propyl}-1H-indol-5-yl)oxy]acetic acid (5-substituted indoleoxyacetic acid analogue)
EHA	2F4B	(72)	(5-{3-[(6-benzoyl-1-propyl-2-naphthyl)oxy]propoxy}-1H-indol-1-yl)acetic acid (Indol-1-yl acetic acid)
ET1	3ET3	(86)	3-[5-methoxy-1-(4-methoxyphenyl)sulfonyl-indol-3-yl] propanoic acid (indeglitazar)
HXA	2VV0	(41)	Docosa-4,7,10,13,16,19-hexaenoic acid
PTG	2ZK1 2ZK2	(39)	15-deoxy-delta(12,14)-prostaglandin J2 (PTG)

<sup>1</sup>Abbreviations for ligands mentioned in the text are in parentheses following the full name of the compound.

Inclusion of the interaction criteria improved the target structure model selection process. Based on the crystal structure interactions common to rosiglitazone and known fatty acid agonists, the number of possible interactions (Table 3.S5) and instances of key residue hydrogen bonding (Table 3.S6) were counted for all the poses. Both sets of data suggested that 1ZGY was the most appropriate model relative to 1FM6 and 2PRG for the purposes of this study. Poses docked into the 1ZGY model all showed at least one key interaction, whereas the other two models returned poses for some ligands that did not exhibit any known interactions. Additionally, fatty acid and fatty acid-derivatives returned the most favorable poses of all the cross-docking ligand types. If interaction analysis is included in the selection process, we see 1ZGY as the predominate candidate

for the target structure model in a screen involving rosiglitazone-like and fatty acid compounds.

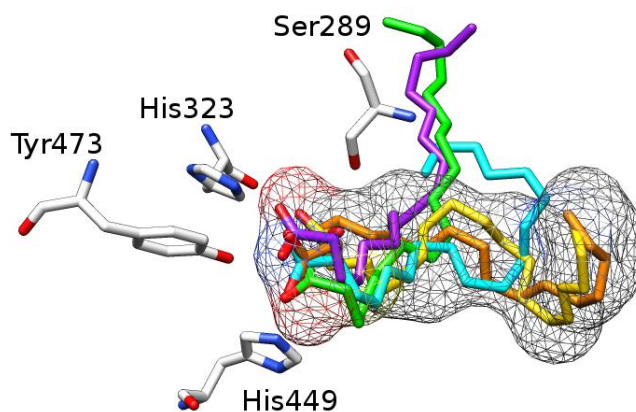
### 3.4.3 Conjugated trienes showed association with PPAR $\gamma$ in silico

For the small-scale screen, a library of seventy-seven compounds was selected. These compounds included known active and inactive compounds, with alternate stereochemistry for some structures. This test set allowed for screening of active versus inactive, rosiglitazone-like versus non-TZDs, and molecularly simple versus complex compounds. The interaction data (Tables 3.S7 and 3.S8) reinforced the assumption that the selected target structure model could accommodate rosiglitazone-like and fatty acid compounds. The cross-docking ligands included in the screen docked similarly to what was seen with the cross-docking test. Most of the rosiglitazone-like compounds studied by Markt et al. (144) showed successful docking. These compounds were Chemical Abstracts Service (CAS)# 264908-13-6, CAS# 651724-09-3, CAS# 853652-40-1, BRL48482, BVT13, CLX-M1, KRP297, and NNC61-4424 (Table 3.S1). Isomers of these compounds with differences in stereochemistry were used as well. Some of these structures did not dock as well, which was expected given it has been suggested from crystal structure studies that chirality can affect agonist activity (54). We also saw lack of favorable docking for bulkier compounds, which contain multiple ring and aromatic components, and compounds with multiple hydroxyl groups. These ligands included phenolic extracts taken from *Glycyrrhiza glabra* roots isolated by Kuroda et al. (143),  $\alpha$ -santonin-derived compounds identified by Tanrikulu et al. (142), and flavonoids screened by Salam et al. (103) (Table 3.S1). The compounds from Kuroda et al. (143) and Tanrikulu et al. (142) compounds were numbered according to extraction fraction and deviation from the original  $\alpha$ -santonin scaffold, respectively. The Kuroda et al. subset included compounds that induced low level activation. The Tanrikulu et al. subset contained one highly active compound (Tanrikulu\_1), one moderately active compound (Tanrikulu\_2), and six inactive compounds (Tanrikulu\_3 through Tanrikulu\_8). The selected Salam et al. compounds were apigenin, biochanin-A, chrysin, dihydroquercetin, genistein, hesperidin, psi ( $\psi$ )-baptigenin, and vitexin. The unsuccessful docking of known active compounds in these groups indicated the receptor structure was not appropriate for docking of these molecule types.

All of the conjugated trienes docked successfully but with similar geometry and energy scores, so a more detailed test for these compounds was conducted to see if a predominant ligand could be identified. AD4 was used to dock jacaric, catalpic, calendic, eleostearic, and punicic acids into the selected structural model, 1ZGY. Three iterations of 50 poses each were run and the lowest energy pose for each run for each fatty acid was selected and compared (15 lowest energy poses). The numbers of potential hydrogen bonds and hydrophobic interactions for each pose were calculated (Table 3.S8). The lowest energy pose with the most potential hydrogen bond interactions was selected for each triene and used for analysis. As there are no crystal structures available with any of these compounds co-crystallized, interactions from PDB structures with fatty acids bound were used to generate an interaction reference list (Table 3.S3). The four key residues that formed hydrogen bonds with rosiglitazone also formed hydrogen bonds with these fatty acids. Therefore, poses that possessed these interactions were deemed successful

agonists. Unsuccessful poses were those lacking the agonist interactions and poses with the reactive polar group pointed away from the activation site.

All the conjugated trienes showed favorable docked poses and exhibited interactions with residues associated with PPAR $\gamma$  activation (Table 3.4). The triene poses occupied a space similar to that seen with rosiglitazone (Figure 3.1), and exhibited interactions with key residues. Of all the replicate poses for triene docking, the ESA replicates consistently exhibited the most negative free energy of binding (Table 3.S9). Hydrogen bond interactions with only two of the four key residues were seen; however, it is not clear if interactions with all four residues are absolutely necessary for activation, or if a reduced number of interactions can still induce activation. It is feasible that a reduced number of specific interactions may contribute to the specificity seen with ligand-induced co-activator recruitment for PPARs. A comparison of distance measurements for the interactions showed two Tyr473-involved interactions for ESA, puniceic acid, and jacaric acid. Given the distance measurements, it was proposed that the acid head group straddles Tyr473, with one oxygen atom closer to one histidine side chain than the other. This was confirmed when the poses were visually assessed. The number of hydrophobic interactions was more consistent for the ESA poses compared to puniceic and jacaric acids. As previously mentioned, it is known that puniceic acid binds to PPAR $\gamma$  and modulates its activity, while ESA possesses greater antioxidant effects. Given the combination of what was known experimentally about the compounds and the predicted free energy of binding and interactions, ESA was selected as a candidate for validation using a ligand-binding assay and further experimental testing *in vivo*.

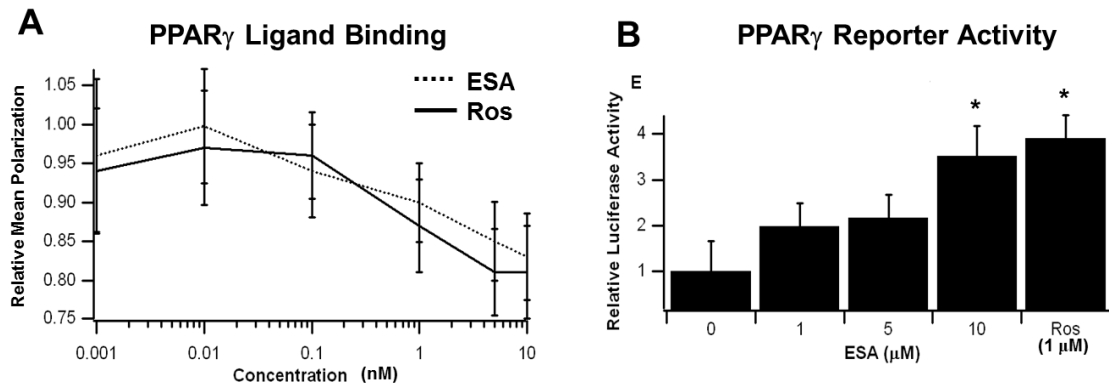


**Figure 3.1** Predicted docked conformations for  $\alpha$ -eleostearic (purple), puniceic (cyan), calendic (orange), jacaric (green), and catalpic (gold) acids relative to the rosiglitazone-occupied portion of the binding cavity (mesh surface) in the rigid PPAR $\gamma$  structure model. Key residues with which hydrogen bonding occurs are labeled. Atom-specific coloring: red = oxygen; gray = carbon; blue = nitrogen. Table 3.4 contains distance measurements for each docked pose.

#### 3.4.4 ESA bound to and modulated PPAR $\gamma$ *in vitro*

The results of our molecular docking efforts and various published studies (7, 127, 129-133, 135, 136) indicated that conjugated trienes, specifically ESA, may bind to and modulate PPAR $\gamma$  activity. Ligand-binding and reporter activity assays were conducted to test this assumption. A cell-free ligand-binding assay was implemented to determine if ESA associated with PPAR $\gamma$  *in vitro* and possessed a similar depolarization pattern to

rosiglitazone. The results suggested the depolarization pattern for ESA was similar to that seen with the rosiglitazone positive control with no significant difference between the two curves (Figure 3.2A).

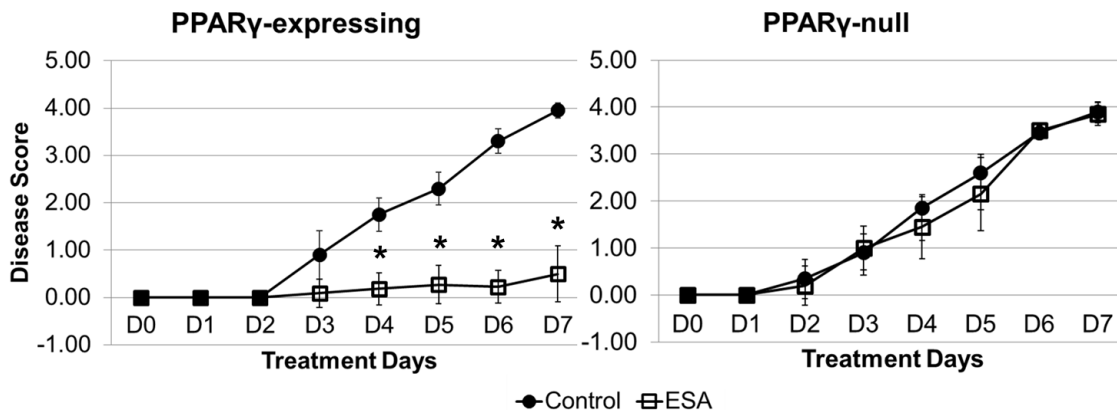


**Figure 3.2** Ligand-binding (A) and reporter assay (B) results for ESA bound to PPAR $\gamma$  with rosiglitazone (Ros) as a positive control. (A) Ligand binding was assessed as a measure of mean polarization for the displaced Fluormone<sup>TM</sup> molecule versus increasing concentrations of either ligand. (B) Reporter activity was measured as relative luciferase activity for various concentrations of ESA versus 1 $\mu$ M Ros. Error bars represent standard deviation, while asterisks (\*) indicate significance ( $p \leq 0.05$ ) between the data sets.

An assessment of PPAR $\gamma$  activity modulation was conducted using RAW 264.7 cells and varying ESA concentrations (0-10  $\mu$ M). Relative luciferase activity was measured to determine ligand-induced activation. The reporter assay suggested ESA does modulate PPAR $\gamma$  activity, but at a concentration 10-fold higher than the rosiglitazone control (Figure 3.2B), suggesting that there may be a difference in either potency or uptake by the cells between both compounds.

### 3.4.5 ESA ameliorated clinical signs of IBD

Under our DSS-induced IBD model, ESA significantly ameliorated IBD in mice with the wild phenotype (i.e., PPAR $\gamma$ -floxed). This observation was based on the significant difference between DAI for the last four days of the seven-day challenge (Figure 3.3). IBD-related disease phenotypes were milder in the ESA-fed PPAR $\gamma$ -expressing group of mice compared to the ESA-fed cell-specific PPAR $\gamma$ -null mice. The control groups (no ESA) for both genotypes showed no improvement in IBD phenotypes over the seven-day time course. Therefore, ESA was effective in ameliorating disease-associated phenotypes in mice with DSS colitis through a PPAR $\gamma$ -dependent mechanism.



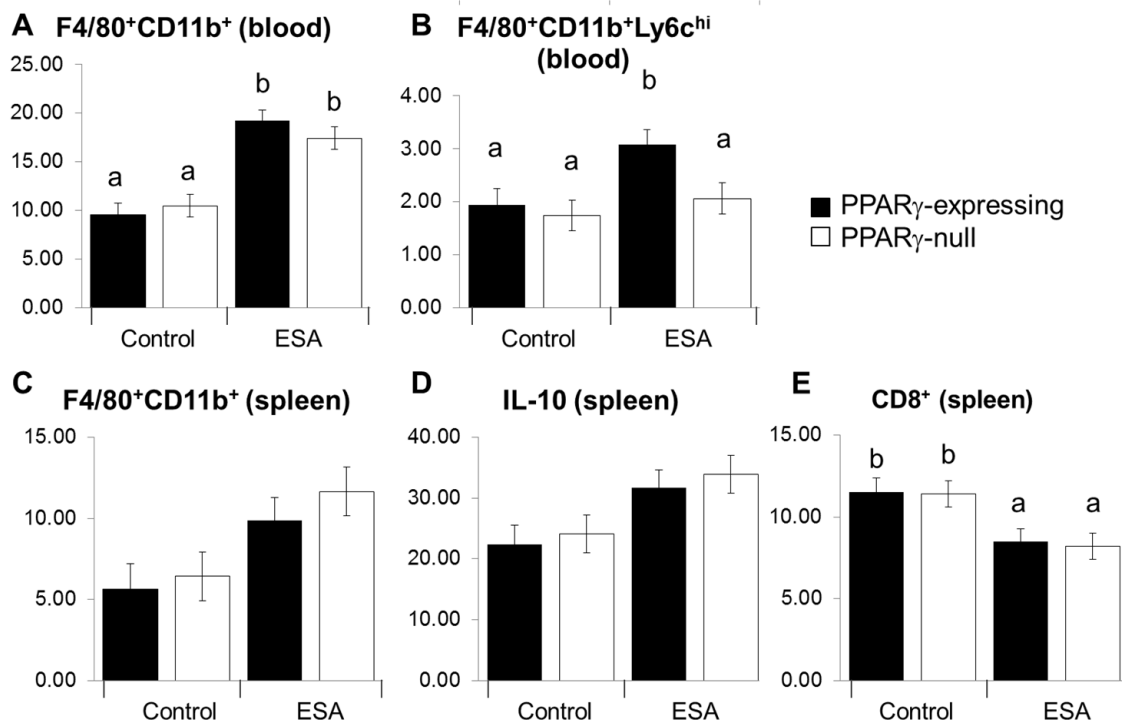
**Figure 3.3** Effect of ESA on disease activity scores for PPAR $\gamma$ -expressing (A) and PPAR $\gamma$ -null (B) mice with experimental IBD. PPAR $\gamma$ -null refers to lack of functional PPAR $\gamma$  product in colon epithelial and immune cells only. Data points represent averaged disease scores for each group with error bars representing standard deviation. Asterisk (\*) indicates significance ( $p \leq 0.05$ ).

**Table 3.4** Distance measurements (in Angstroms [Å]) for docked conjugated triene poses displayed in Figure 3.1. Distances were measured between carboxylic oxygen atoms of fatty acids and listed atoms for each residue. Free energy of binding is measured in kilocalories per mole of ligand (kcal/mol). No value is listed for rosiglitazone as this refers to the crystal conformation (denoted "N/A") Residues are labeled as the amino acid designation plus the atom name (e.g., S289.OG refers to the oxygen atom in the gamma position on serine 289).

Ligand	Color	Residue	Distance (Å)	kcal/mol
eleostearic acid	purple	H323.NE2	3.16	-5.6
		Y473.OH	3.01	
		Y473.OH	3.27	
punicic acid	cyan	H449.NE2	2.84	-4.28
		Y473.OH	3.03	
		Y473.OH	3.07	
calendic acid	orange	H449.NE2	2.81	-4.47
		Y473.OH	3.10	
catalpic acid	gold	S289.OG	3.05	-4.48
		H323.NE2	3.03	
		Y473.OH	3.26	
jacaric acid	green	H449.NE2	2.84	-4.5
		Y473.OH	3.16	
		Y473.OH	3.10	
rosiglitazone	gray mesh	S289.OG	3.02	N/A
		H323.NE2	2.83	
		H449.NE2	3.02	
		Y473.OH	2.85	

### 3.4.6 Immunophenotypes for harvested tissues

Changes in immune cell subsets due to DSS-induced colitis were assessed in the harvested tissues to investigate the modulation of inflammation by ESA (Figure 3.4). Flow cytometry was used to characterize the phenotype of macrophages and T cell subsets. DSS augmented the percentages of monocytes or macrophages in the blood and spleen (Figure 3.4A and 3.4C). A significant increase in blood monocytes was found in ESA-treated mice. The PPAR $\gamma$ -expressing mice on the ESA diet exhibited a higher percentage of monocytes expressing lymphocyte antigen 6 complex-high (Ly6C<sup>hi</sup>), which was not seen in the PPAR $\gamma$ -null group (Figure 3.4B) indicating a PPAR $\gamma$  dependency of this effect. Higher levels of IL-10 were observed in the spleen of the ESA-fed mice for both genotypes although these numerical differences were not statistically significant between the two diets for the PPAR $\gamma$ -expressing genotype (Figure 3.4D). Lastly, we found a numerical decrease in CD8<sup>+</sup> T-cells in the ESA diet group (Figure 3.4E), where the change was PPAR $\gamma$ -independent.

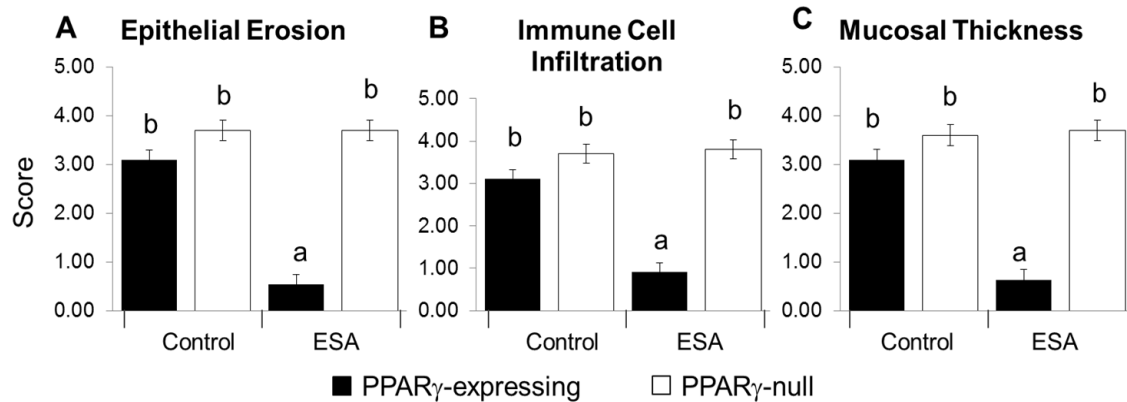


**Figure 3.4** Effect of ESA on immune cell subsets of PPAR $\gamma$ -expression and PPAR $\gamma$ -null mice with experimental IBD. Tissues examined included blood (A and D) and spleen (B, C, and E). Values represent least square means for percentage of gated cells with error bars to indicate standard error. Letters indicate significance ( $p \leq 0.05$ ) where a shared letter indicates groups which are not statistically significantly different.

### 3.4.7 Histological trends mimicked clinical activity

There was a significant decrease in epithelial erosion (Figure 3.5A), mucosal thickness (Figure 3.5B), and immune cell infiltration (Figure 3.5C) in the ESA-fed PPAR $\gamma$ -expressing mice but not in ESA-fed PPAR $\gamma$ -null mice. This suggested amelioration of experimental IBD phenotypes by ESA is PPAR $\gamma$ -dependent. This agreed with the DAI

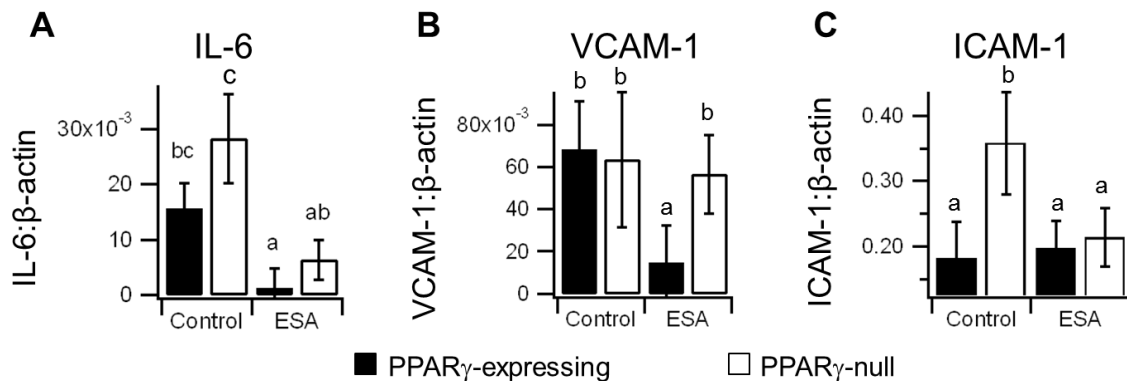
data and further indicated an ESA-associated PPAR $\gamma$ -dependent improvement in IBD phenotypes.



**Figure 3.5** Effect of ESA on histopathological lesions in colons from PPAR $\gamma$ -expressing and PPAR $\gamma$ -null mice with experimental IBD. Epithelial erosion (Erosion) (A), immune cell infiltration (Infiltration) (B), and mucosal thickness (Thickness) (C) were assessed and averaged for all the DSS-treated group of samples. Data are presented as mean score with error bars to indicate standard deviation. Letters indicate significance ( $p \leq 0.05$ ) where a shared letter indicates groups which are not statistically significantly different.

### 3.4.8 Gene expression suggested PPAR $\gamma$ -dependent and -independent mechanisms

There was a marked decrease in IL-6 and VCAM-1 mRNA expression between the control- and ESA-fed PPAR $\gamma$ -expressing groups (Figure 3.6A and 3.6B). The IL-6 decrease appeared to be PPAR $\gamma$ -independent, while the VCAM-1 decrease was PPAR $\gamma$ -dependent. We also found a decrease in ICAM-1 expression between the control and ESA diet groups, but this decrease also occurred in the PPAR $\gamma$ -null mice suggesting ESA can induce ICAM-1 regulation in a PPAR $\gamma$ -independent manner (Figure 3.6C).



**Figure 3.6** Effect of ESA on colonic concentrations of IL-6 (A), VCAM-1 (B), and ICAM-1 (C) in PPAR $\gamma$ -expressing and PPAR $\gamma$ -null mice with experimental IBD. The mean ratio of expression for each protein relative to constitutively expressed  $\beta$ -actin is shown with error bars to indicate standard deviation. Letters indicate significance ( $p \leq 0.05$ ) where a shared letter indicates groups which are not statistically significantly different.



### 3.5 Discussion

The VS model protein structure and parameters used in this study allowed for prediction of docking conformations for rosiglitazone-like and fatty acid compounds. The re-docking results for rosiglitazone, cross-docking results for PTG-1 and 9-HODE, and the conjugated triene docking all suggested 1ZGY is appropriate for screening fatty acids and TZD-like compounds. Potential for docking of fatty acid derivative partial agonists, like (2S)-ureidofibrate-like derivative, was also seen, but not fully assessed for this study as full agonism was the binding type of interest. Thus, we have successfully established a VS parameter set appropriate for a large-scale PPAR $\gamma$  full agonist search amongst fatty acids and fatty acid derivatives.

Information regarding interactions known to occur with PPAR $\gamma$  agonists is a suitable means to identify docking success. However, the success rate may be improved by incorporating even more criteria. Such criteria include a more extensive list of key interactions and/or establishment of distinct lists to specify interactions characteristic of each ligand category (e.g., full agonist, partial agonist, and antagonist). Based on the number of interactions and presence of interactions with key residues, we were able to determine which ligand types do and do not fit our selected target structure model. Combining this with RMSD data allowed us to see which types of ligands dock away from the binding cavity given the molecular environment of the selected target structure model. This information regarding ligands that would be excluded in a screen for compounds that interact similarly to what is seen with rosiglitazone can be used to identify one or more additional target structure models to incorporate into a large-scale screen. RMSD data, however, would not be available from a screen of unknowns, and conclusions would therefore have to be drawn from the interaction and free energy data.

Due to the high degree of precision observed with the cross-docking ligands, it was determined that a single pose for each ligand would be sufficient for the initial analysis step in a large-scale screen. Replicates were necessary for the pre-screening analysis in which parameters and structure models were tested for predictability. Replicates are useful in docking studies to ensure any conclusions are based on consistent interactions. However, running replicates for a library numbering in the thousands is computationally time-consuming and less than practical given replicate poses may possess geometry that is exactly or close to the same. Rather than run replicates on the entire library of compounds, it would be feasible to run more detailed docking with compounds selected as successful binders of interest with the potential for experimental verification.

We observed a complementary relationship between the experimental ESA-IBD study and the computational screening results. In a recent review, we mentioned previous studies in which dual- or pan-agonistic effects have been associated with conjugated trienes (151). This information, coupled with other published studies regarding synthetic agonists and inactive compounds, provided a means to develop and test computational methods for identifying natural agonists. Our docking analysis suggested ESA possessed a more favorable binding energy compared to the other conjugated trienes. Though comparative relationships have not been established between ESA and all the tested trienes, we do know that ESA possesses greater antioxidant effects than puniceic acid in mice (152). It is plausible that the differences in efficacy between the compounds is

interaction-related, which may result in conformational changes that attenuate co-activator recruitment and subsequent transcriptional regulation. The interaction aspect may have been picked up by our study, but the dynamic significance was not. This second aspect would require further computational testing to see if differences in protein stability and conformation can be detected between the protein-ligand complexes.

The ligand binding and reporter assays verified that ESA binds to and modulates PPAR $\gamma$ . Our docking study suggested fewer interactions occurred in the PPAR $\gamma$ -ESA complex compared to PPAR $\gamma$ -rosiglitazone. It is possible that the absence of interactions with Ser289 and His449 could result in a different level of ligand-induced activity attenuation or the interactions with His323 and Tyr473 may be more important for fatty acid-induced agonism. Given the different levels of agonism, which is ligand-dependent, it is plausible that the specificity toward anti-inflammatory mechanisms observed as PPAR $\gamma$ -dependent in the pre-clinical trial were influenced by some difference in agonism specific to ESA. This notion is further supported by the absence of rosiglitazone-associated phenotypes seen in studies published by other groups (9, 152). Both the Shah et al. and Ramakers et al. studies involved testing rosiglitazone against DSS-induced colitis in mice (9, 152). Ramakers et al. showed weight gain in mice treated with rosiglitazone prior to DSS challenge, followed by significantly greater weight loss compared to the control after DSS challenge (9). Increases in the severity of colitis-specific colon phenotypes were also seen, but with a decrease in inflammation (9). The Shah et al. study indicated a PPAR $\gamma$ -dependent rosiglitazone-induced decrease in macrophage recruitment, but showed no other significant changes to the levels of other cytokines (152).

We have shown that the immune modulatory actions of ESA may be both PPAR $\gamma$ -dependent and PPAR $\gamma$ -independent in mice with experimental IBD, although its effects on disease activity and colonic lesions are dependent on expression of PPAR $\gamma$  by immune and epithelial cells. It is known that PPAR $\gamma$  is highly expressed in immune cells, intestinal epithelial cells (IECs), and adipocytes, with lower expression levels throughout various tissues of the body. Recently, our group published work in which the severity of IBD was tested in a mouse model for IEC-specific PPAR $\gamma$  deletion in a C57BL/6 background (8). It was determined that the absence of PPAR $\gamma$  from IECs resulted in significantly worse disease scores, greater loss of body weight, and increased inflammation in the colon, spleen, and MLN compared to mice expressing PPAR $\gamma$  (8). Further, it was concluded that the presence of PPAR $\gamma$  in IEC contributes to anti-inflammatory effects, regulation of immune cell distribution, and gene expression regulation necessary to counteract IBD symptoms (8).

Additionally, there are studies in which PPAR $\gamma$  expression and the effect of ESA on disease pathogenesis have been evaluated in breast cancer cell lines (130, 153), pre-adipocytes (154), and colon cancer cell lines (40). In all cases the fatty acid was capable of significantly ameliorating the disease via PPAR $\gamma$ -dependent responses such as induced apoptosis of cancer cells (40, 130, 153) and reduced lipid storage during differentiation (154). Other conjugated trienes, such as punicic acid and catalpic acid (5, 127) have shown reduced inflammation responses in cancer, cardiovascular disease (155), and obesity (5, 127, 155). All of these studies are strong examples of how PPAR $\gamma$  mediates

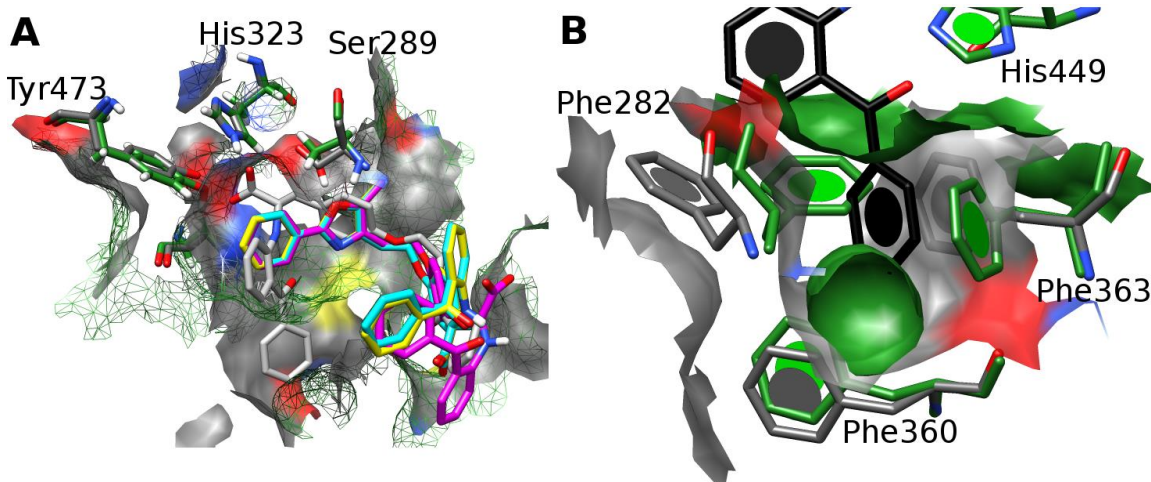
inflammatory, metabolic, proliferation, signal transduction, and cellular motility processes (8) in various cell types.

It is possible that the presence of other nuclear receptors in the cells play a role in ESA-mediated effects. PPAR $\delta$  in the colon may play a role in ESA-mediated IBD amelioration given the possibility of dual-agonist and pan-agonist modulation seen with PPARs, and the ability of all three PPARs to accommodate fatty acids. Further computational and experimental tests would be necessary to determine whether ESA mediates both PPAR $\gamma$  and PPAR $\delta$  transcriptional regulation, which has been previously described for CLA (118). The anti-inflammatory responses induced by ESA, which appeared to be PPAR $\gamma$ -independent, might also be attributed to other unforeseen targets in the system. For instance, we previously described the potential of PPAR $\gamma$  agonists to bind to lanthionine synthetase component C-like protein 2 (LANCL2) (148). Such an association is one proposed molecular mechanism of regulating disease-related inflammatory effects in a PPAR $\gamma$ -independent manner.

Beyond what is seen in IBD, it has been shown that ESA binds to and activates estrogen receptors in breast cancer cell lines (156). It is also known that hepatocyte nuclear factor-4 $\alpha$  (HNF4 $\alpha$ ), which is essential for maintaining lipid homeostasis via gene regulation and regulating hepatocyte differentiation, is activated by fatty acids (157). It has been suggested that PPAR $\alpha$  ligands can interfere with HNF4 $\alpha$  activity (158), but the mechanism by which this occurs is not fully understood. As conjugated trienes like punicic acid activate PPAR $\alpha$  in adipocytes (127), and PPAR $\alpha$  and fatty acids are present in liver tissue also, it seems feasible that conjugated trienes could come in contact with and bind HNF4 $\alpha$  as well. To our knowledge such a study involving HNF4 $\alpha$  and ESA or any other conjugated trienes has not been conducted.

The ability of the binding cavity to accommodate many different ligand types represents a major technical obstacle when performing computational docking into PPAR $\gamma$  as a therapeutic target. The issue stems from the dynamic nature of the binding cavity and changes in protein conformation necessary to accommodate different agonists. This dynamic nature is not possible with rigid macromolecule docking techniques, and incorporation of flexibility can be difficult given the number of residues that can possess variable positions and the number of possible rotamers for each residue. The rigidity of crystal structures combined with the variability of residue side chain positions proved an issue for docking non-native ligands to the selected structure model. For example, the docked poses for farglitazar across the three protein structure models examined in the cross-docking step reflected a lack of appropriate molecular volume at the rear of the binding pocket to accommodate the benzyl ketone group on the ligand (Figure 3.7A). When the three structure models were compared to the 1FM9 crystal structure in which farglitazar was co-crystallized, the space necessary to accommodate the benzyl ketone group of farglitazar was missing given the differences in the side chain positions for Phe282 and Phe363 (Figure 3.7B). These residues do not pose an issue for rosiglitazone docking, but occupied the portion of the cavity in which farglitazar should have docked, which prevented successful cross-docking of this compound to the selected structure models. As such, selection of a single model to appropriately accommodate a narrow range of ligands and selection of several models to use with a diverse ligand library are two avenues toward identifying PPAR $\gamma$  agonists *in silico*. The first technique is used

widely, but the second is not as common due to the amount of time necessary to properly identify target structure models. Given the molecular exclusion of the more hydrophobic compounds in our small-scale screen, the second technique would be ideal for dealing with a diverse library, such as the one we have constructed. Therefore, further testing with additional protein structure models capable of accommodating bulkier and more hydrophobic compounds would be necessary.



**Figure 3.7** Visual assessments of molecular surface differences that result in unsuccessful docking of specific ligand types to the selected PPAR $\gamma$  structure model. Farglitazar is represented in both panels with atom-specific coloring. (A) 1ZGY and 1FM9 surface representations are green mesh and solid gray, respectively. The three poses predicted for farglitazar relative to 1ZGY are shown in magenta, cyan, and yellow. (B) Side chain rotamers for F282 and F363 are responsible for the differences in cavity surface at the rear of the cavity. Surface colors for 1ZGY and 1FM9 are the same as in (A). Atom-specific coloring: gray/black = carbon, blue = nitrogen, red = oxygen, white = hydrogen, and yellow = sulfur.

An additional technique for improving predictability is molecular dynamics simulation and analysis, which is also extremely time consuming and can prove problematic since parameters for ligands must be developed. Conformational sampling of the PPAR $\gamma$  binding cavity via MD is one means of gleaning useful information in a relatively short amount of time. This technique would provide information about predominant conformations adopted by PPAR $\gamma$  that would aid in the selection of multiple target structure models for docking, and can be easily verified by the large number of available crystal structures.

PPAR $\gamma$  has proven a difficult protein to explore as a drug target given dynamic and specificity issues. The large binding cavity and ability of the protein to accommodate a wide range of compounds presents an issue for rigid docking screening. The ability of the protein to bind compounds of different compound families requires a degree of ligand diversity that is often not employed in conventional VS studies. As a means to improve our method, we are currently testing additional PPAR $\gamma$  crystal structures as docking targets. As a consequence of this study, we have established a need for at least one additional target structure model that can accommodate bulkier compounds. An analysis of MD simulations for unbound active, bound active, and unbound inactive forms of PPAR $\gamma$  are ongoing. These simulations, combined with further analysis of available crystal structure models, will allow us to develop additional target structure models. Incorporating conformational variability by screening against multiple protein

conformations of the same protein should improve our screening process. We propose matching ligand and protein pharmacophores prior to screening to reduce the incidence of screening ligands against a protein structure into which the ligands cannot fit or where the charge environment is inappropriate.

The diversity of our compound database is being expanded as well, and will include an extensive list of known PPAR $\gamma$  agonists, decoy compounds that mimic known agonist structure but are inactive toward PPAR $\gamma$ , drugs currently available for treatment of other diseases, and extracts tested experimentally for PPAR $\gamma$  modulation. Such a library would improve enrichment, which is part of the separation of binders from non-binders. Further, inclusion of a weighting system based on the occurrence of known interactions would improve the separation process. With a diverse library in which available therapeutics are included, it may be possible to identify lesser known drug interactions with PPAR $\gamma$  linked to side effects seen with patients taking medications for cancer and neurological diseases. Given the success of our current study and the pending improvements to our method for testing of diverse ligand types, we are making progress toward an extensive and highly effective means to computationally identify feasible PPAR $\gamma$ -targeted drug candidates. Ideally, the established methods could be applied to the other PPARs, other nuclear hormone receptors, and alternate protein family targets where similar considerations must be made.

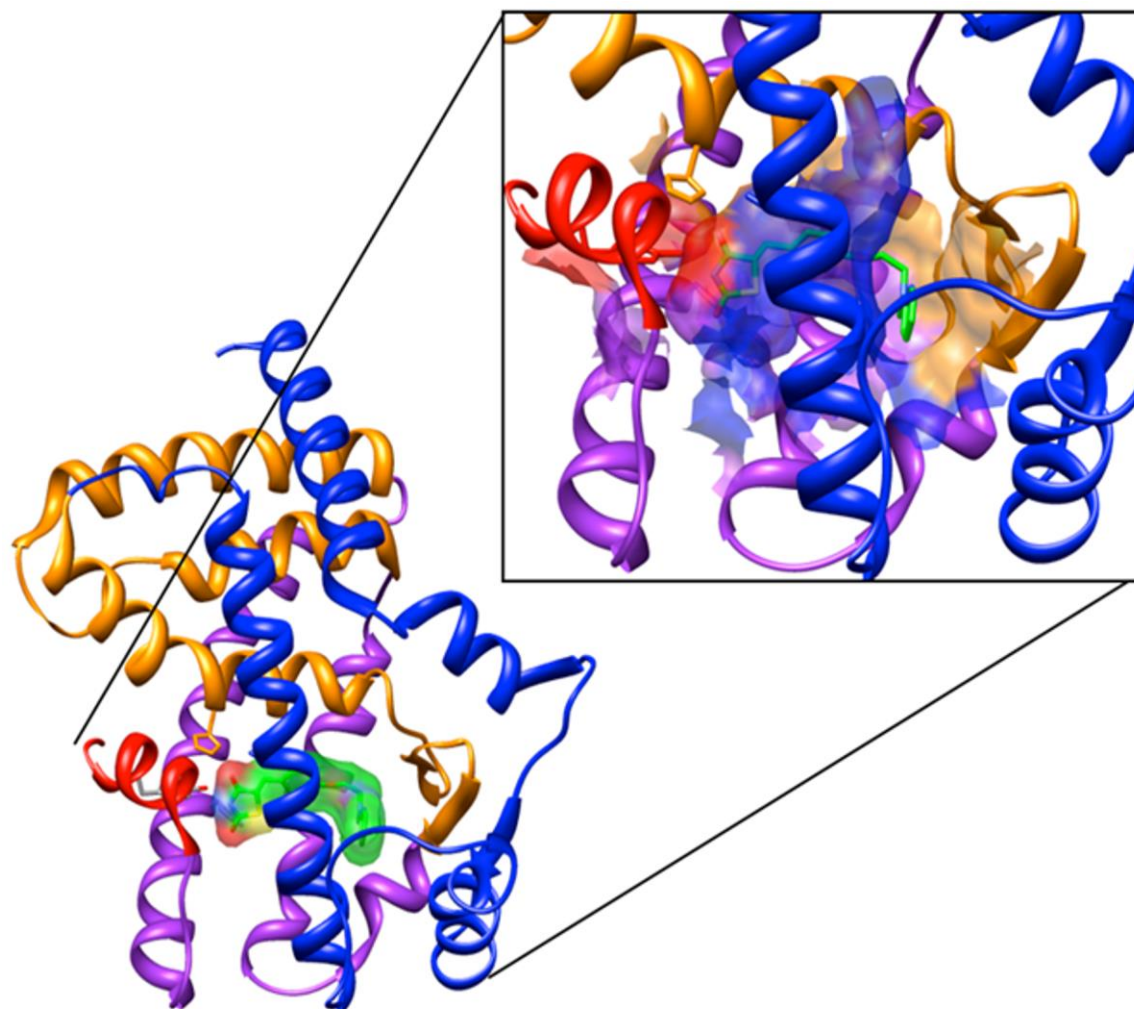
This study exemplifies how experimental methods can be used to complement and verify computational predictions. We have demonstrated that it is possible to predict ligand association given information known about the binding cavity of the target. We have also established a means to reduce the need for researcher intervention in assessing successful binding by incorporating a search for key interactions. More specifically, we have successfully established a protocol for screening fatty acid compounds against PPAR $\gamma$  for agonism, and were able to predict that ESA and other conjugated trienes would bind to and activate PPAR $\gamma$  using molecular docking. These predictions have been verified through *in vitro* assays both here and in our previous work (127, 136). *In vivo* efficacy was assessed as well to determine if disease-associated benefits could be seen given the activation of PPAR $\gamma$  by ESA. In this regard, ESA did induce both PPAR $\gamma$ -dependent and -independent responses that ameliorated disease activity and intestinal lesions in IBD. The scope of this work implies the techniques described here can aid in streamlining drug discovery and development techniques as the technology develops.

### **3.6 Acknowledgements**

Funding: NIH Biomedical and Behavioral Sciences Research Training Grant R25 GM072767 (Virginia Tech Initiative to Maximize Student Diversity). National Center for Complementary and Alternative Medicine Grant 5R01AT4308. The Genetics, Bioinformatics, and Computational Biology PhD program, Virginia Tech, Blacksburg, VA 24601.

Additional funding agencies: National Institute for Allergy and Infectious Diseases Contract HHSN272201000056C. National Institute of Diabetes and Digestive and Kidney Disease of the National Institutes of Health Grant 1F31DK091186.

### 3.7 Supporting Information



**Figure 3.S1** Colored ribbon representation of PPAR $\gamma$  showing three layers of helical “sandwich”, and co-crystallized rosiglitazone (PDB ID 1FM6 (45)). Helices for each layer are colored, with helix H12, which sits at the rear of the binding cavity (AF-2 region), colored in red. Rosiglitazone is colored in green, with oxygen, nitrogen, and sulfur atoms colored red, blue, and yellow, respectively. The insert (upper right) shows a close-up view of the molecular surface of the binding cavity. The thiazolidinedione head group of rosiglitazone sits at the rear of the binding cavity where it can interact with S289, H323, H449, and Y473 in order to change the conformation of the AF-2 region and activate the protein.

### Formulas 3.S1

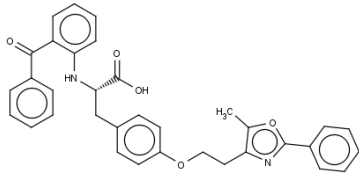
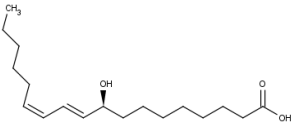
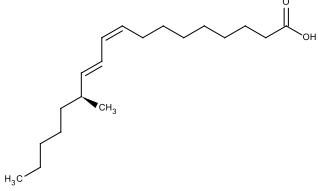
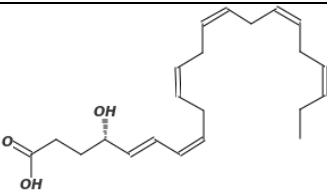
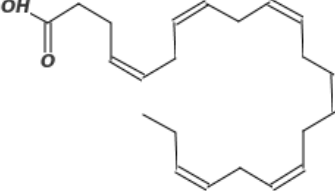
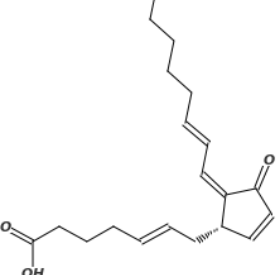
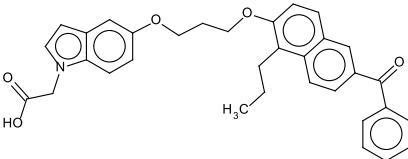
Standard Deviation ( $\sigma$ ) refers to the degree to which the measured values deviate from the mean.

$\sigma = \sqrt{\sum_{k=1}^n (x_k - \mu)^2}$ , where  $x$  is a single value,  $k$  is the index for the value,  $\mu$  is the population mean, and  $n$  is the number of values. Variance defines how far the measured values are from each other and is defined as  $\sigma^2$ .

Root Mean-Squared Deviation (RMSD) is a measure of the change in geometry and orientation of a pose from that of the reference control structure.

$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n (v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2}$ , where  $n$  number of equivalent pairs of atoms, and  $v$  and  $w$  are the coordinate sets for the pose and reference structures.

**Table 3.S1** List of ligands used for virtual screening.

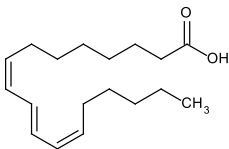
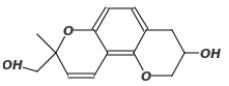
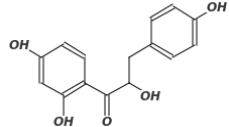
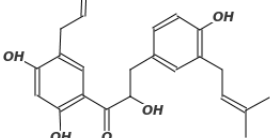
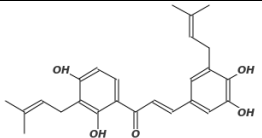
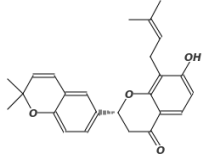
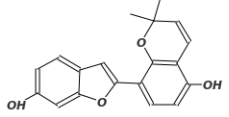
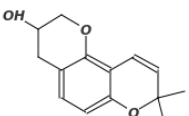
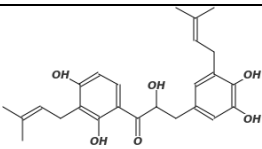
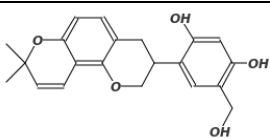
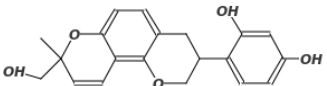
Ligand	Activity	Reference	Structure
GI262570 (Farglitazar)	active	(45)	
9-hydroxyoctadecadienoic acid	active	(41)	
13-hydroxyoctadecadienoic acid	active	(41)	
(4S,5E,7Z,10Z,13Z,16Z,19Z)-4-hydroxydocosa-5,7,10,13,16,19-hexaenoic acid	active	(41)	
Docosa-4,7,10,13,16,19-hexaenoic acid	active	(41)	
15-deoxy-delta(12,14)-prostaglandin J2	active	(39)	
(5-{3-[(6-benzoyl-1-propyl-2-naphthyl)oxy]propoxy}-1H-indol-1-yl)acetic acid	active	(72)	



**Table 3.S1** continued

Ligand	Activity	Reference	Structure
[(1-{3-[(6-benzoyl-1-propyl-2-naphthyl)oxy]propyl}-1H-indol-5-yl)oxy]acetic acid	active	(78)	
(2R)-2-(4-{2-[1,3-benzoxazol-2-yl(heptyl)amino]ethyl}phenoxy)-2-methylbutanoic acid	active	(54)	
(2S)-2-(4-{2-[1,3-benzoxazol-2-yl(heptyl)amino]ethyl}phenoxy)-2-methylbutanoic acid	active (partial agonist)	(54)	
3-[5-methoxy-1-(4-methoxyphenyl)sulfonyl-indol-3-yl]propanoic acid	active	(86)	
(9Z,11E)-octadeca-9,11-dienoic acid (CLA)	active		
Calendic Acid	active		
Catalpic Acid	active		

**Table 3.S1** continued

Ligand	Activity	Reference	Structure
Jacaric Acid	active		
Kuroda_No10	inactive	(143)	
Kuroda_No15	inactive	(143)	
Kuroda_No16	active	(143)	
Kuroda_No2	active	(143)	
Kuroda_No34	active	(143)	
Kuroda_No38	active	(143)	
Kuroda_No39	active	(143)	
Kuroda_No3	active	(143)	
Kuroda_No5	active	(143)	
Kuroda_No6	active	(143)	

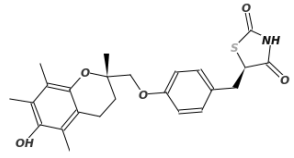
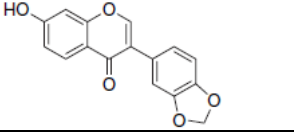
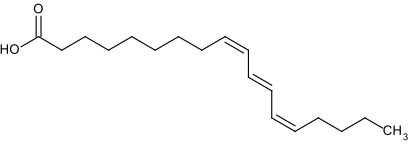
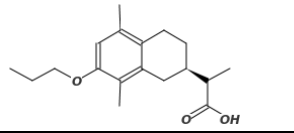
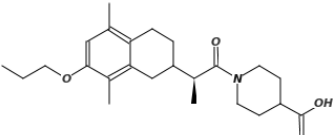
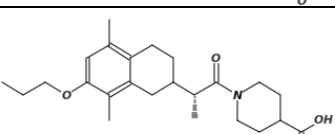
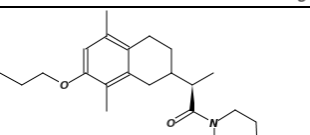
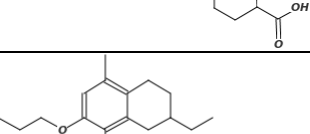
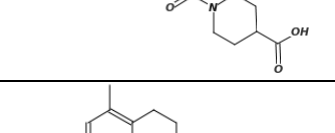
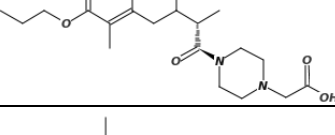
**Table 3.S1** continued

Ligand	Activity	Reference	Structure
Markt_264908-13-6_1	active	(144)	
Markt_264908-13-6_2	active	(144)	
Markt_264908-13-6_3	active	(144)	
Markt_264908-13-6_4	active	(144)	
Markt_651724-09-3_1	active	(144)	
Markt_651724-09-3_2	active	(144)	
Markt_853652-40-1_1	active	(144)	
Markt_853652-40-1_2	active	(144)	
Markt_BRL48482_1	active	(144)	
Markt_BRL48482_2	active	(144)	

**Table 3.S1** continued

Ligand	Activity	Reference	Structure
Markt_BVT13	active	(144)	
Markt_CLX-M1_1	active	(144)	
Markt_CLX-M1_2	active	(144)	
Markt_KRP297_1	active	(144)	
Markt_KRP297_2	active	(144)	
Markt_NNC61-4424_1	active	(144)	
Markt_NNC61-4424_2	active	(144)	
Tesaglitazar	active	(144)	
Troglitazone_1	active	(144)	
Troglitazone_2	active	(144)	
Troglitazone_3	active	(144)	

**Table 3.S1** continued

Ligand	Activity	Reference	Structure
Troglitazone_4	active	(144)	
$\psi$ -baptigenin	active	(103)	
Punicic Acid	active		
Tanrikulu1	active	(142)	
Tanrikulu2_1	active	(142)	
Tanrikulu2_2	active	(142)	
Tanrikulu2_3	active	(142)	
Tanrikulu2_4	active	(142)	
Tanrikulu3_1	inactive	(142)	
Tanrikulu3_2	inactive	(142)	

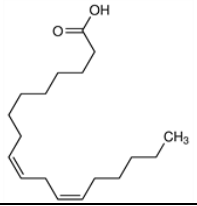
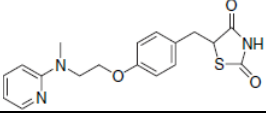
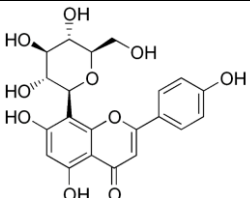
**Table 3.S1** continued

Ligand	Activity	Reference	Structure
Tanrikulu3_3	inactive	(142)	
Tanrikulu3_4	inactive	(142)	
Tanrikulu4	inactive	(142)	
Tanrikulu5	inactive	(142)	
Tanrikulu6	inactive	(142)	
Tanrikulu7_1	inactive	(142)	
Tanrikulu7_2	inactive	(142)	
Tanrikulu7_3	inactive	(142)	
Tanrikulu7_4	inactive	(142)	
Tanrikulu8_1	inactive	(142)	

**Table 3.S1** continued

Ligand	Activity	Reference	Structure
Tanrikulu8_2	inactive	(142)	
Tanrikulu8_3	inactive	(142)	
Tanrikulu8_4	inactive	(142)	
$\alpha$ -Eleostearic Acid	active		
Apigenin	active	(103)	
$\beta$ -Eleostearic Acid	low active		
Biochanin-A	low active	(103)	
Chrysin	low active	(103)	
Dihydroquercetin	inactive	(103)	
Genistein	low active	(103)	
Hesperidin	low active	(103)	

**Table 3.S1** continued

Ligand	Activity	Reference	Structure
Omega-3 conjugated linoleic acid	inactive		
Rosiglitazone	active		
Vitexin	inactive	(103)	



**Table 3.S2** List of atoms for key residues common to selected rosiglitazone crystal structures used to assess potential interactions between docked poses and the protein structure model.

<b>Residue Name</b>	<b>Residue Number</b>	<b>Atom Name<sup>1</sup></b>	<b>Residue Name</b>	<b>Residue Number</b>	<b>Atom Name</b>	<b>Residue Name</b>	<b>Residue Number</b>	<b>Atom Name</b>
ILE	281	CG2	HIS	323	NE2	LYS	367	CD
GLY	284	C	ILE	326	CG2	LYS	367	CE
GLY	284	CA	LEU	330	CD1	HIS	449	CE1
CYS	285	CB	LEU	330	CD2	HIS	449	NE2
GLN	286	CG	ILE	341	CG2	LEU	469	CD1
SER	289	CB	MET	348	CE	TYR	473	OH*
SER	289	OG	MET	364	CE			

<sup>1</sup>Atom names are represented by two or three characters, where the first is the atom type, the second is the position relative to the residue structure, and the number of the atom is more than one is as the designated position (e.g., CG2 corresponds to the second carbon atom in the gamma position).

\*Hydroxyl oxygen.

**Table 3.S3** List of atoms for key residues common to selected fatty acid-bound crystal structures used to assess potential interactions between docked poses and the protein structure model.

Residue Name	Residue Number	Atom Name <sup>1</sup>	Residue Name	Residue Number	Atom Name	Residue Name	Residue Number	Atom Name
PHE	226	CD1	HIS	323	NE2	SER	342	CA
PRO	227	CD	ILE	326	CG2	SER	342	N
LEU	228	CB	ILE	326	O	GLU	343	CA
ILE	281	CG2	TYR	327	CE2	MET	348	CE
PHE	282	CD1	MET	329	CB	LEU	353	CD2
PHE	282	CE1	MET	329	CE	PHE	363	CB
PHE	282	CZ	LEU	330	CA	PHE	363	CD2
CYS	285	CB	LEU	330	CD1	PHE	363	CE2
ARG	288	CB	LEU	330	CD2	PHE	363	CZ
ARG	288	CD	LEU	333	CD1	MET	364	CE
SER	288	CG	VAL	339	CG1	MET	364	CG
ARG	288	CZ	VAL	339	CG2	LYS	367	CE
ARG	288	NE	ILE	341	C	HIS	449	CE1
ARG	288	NH1	ILE	341	CA	HIS	449	NE2
SER	289	CB	ILE	341	CB	TYR	473	CZ
SER	289	OG	ILE	341	CG2	TYR	473	OH*

<sup>1</sup>Atom names are represented by two or three characters, where the first is the atom type, the second is the position relative to the residue structure, and the number of the atom is more than one is as the designated position (e.g., CG2 corresponds to the second carbon atom in the gamma position).

\*Hydroxyl oxygen.

**Table 3.S4** List of atoms for key residues common to rosiglitazone- and fatty acid-containing PDB structures used to assess potential interactions between docked poses and the protein structure model.

<b>Residue Name</b>	<b>Residue Number</b>	<b>Atom Name<sup>1</sup></b>	<b>Residue Name</b>	<b>Residue Number</b>	<b>Atom Name</b>
ILE	281	CG2	ILE	341	CG2
SER	289	CB	MET	348	CE
SER	289	OG	MET	364	CE
HIS	323	NE2	LYS	367	CE
ILE	326	CG2	HIS	449	CE1
LEU	330	CD1	HIS	449	NE2
LEU	330	CD2	TYR	473	OH*

<sup>1</sup>Atom names are represented by two or three characters, where the first is the atom type, the second is the position relative to the residue structure, and the number of the atom is more than one is as the designated position (e.g., CG2 corresponds to the second carbon atom in the gamma position).

\*Hydroxyl oxygen.

**Table 3.S5** Predicted hydrophobic and hydrogen bond interactions for ligands in cross-docking test set relative to a reference list of interactions common to rosiglitazone and selected fatty acids. Poses were taken from docking of each ligand into each of the three listed PPAR $\gamma$  PDB files (top row). Ligand IDs refer to compounds listed in Table 3.3.

Ligand ID	1FM6		1ZGY		2PRG	
	Hydrophobic	Hydrogen bond	Hydrophobic	Hydrogen bond	Hydrophobic	Hydrogen bond
243	8	2	10	1	6	4
	6	0	5	4	8	4
	3	1	11	2	9	5
4HD	9	3	8	5	12	2
	14	1	10	5	13	0
	15	3	9	3	16	0
570	11	0	12	1	5	0
	9	0	11	1	6	0
	9	0	7	1	5	0
9HO	9	4	9	3	9	4
	8	5	15	6	10	4
	7	5	10	6	5	4
DRH	7	0	14	6	5	4
	9	0	12	3	7	0
	9	0	13	3	8	4
DRJ	9	0	12	4	12	5
	13	4	13	4	10	5
	13	4	9	4	12	4
DRY	12	0	7	5	11	3
	9	0	6	3	9	4
	13	0	8	3	8	3
EHA	10	0	6	2	10	3
	10	2	6	3	9	4
	9	2	14	0	11	4
ET1	5	0	8	1	6	1
	5	0	8	1	7	1
	4	0	8	1	7	1
HXA	14	3	10	5	13	4
	6	0	10	3	10	2
	7	0	8	3	14	2
PTG1	16	5	13	6	15	5
	16	5	13	6	14	2
	16	5	11	6	15	4

**Table 3.S5** continued

<b>Ligand ID</b>	<b>1FM6</b>		<b>1ZGY</b>		<b>2PRG</b>	
	Hydrophobic	Hydrogen bond	Hydrophobic	Hydrogen bond	Hydrophobic	Hydrogen bond
PTG2	6	2	11	1	16	4
	5	2	10	1	16	5
	9	5	11	1	11	3

**Table 3.S6** Presence or absence of potential hydrogen bond interactions between indicated residues of selected protein structure models and replicate poses of ligands listed by ID (Lig ID). A single “x” indicates one potential interaction for the listed residue was found for the specified ligand, whereas more than one “x” indicates more than one interaction (e.g., “xx” indicates two interactions found). (N = 3)

Lig ID	1FM6				1ZGY				2PRG			
	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH
243		x		x		x			x	x	x	x
						x	xx	x	x	x	x	x
				x		x		x	x	x	x	xx
4HD		xx		x	x	x	x	xx	x		x	
		x			x	x	x	xx				
	x		x	x	x		x	x				
570							x					
							x					
							x					
9HO	x		x	xx		x	x	x	x	x	x	x
	x	x	x	xx	x	x	xx	xx	x	x	x	x
	x	x	x	xx	x	x	xx	xx	x	x	x	x
DRH					x	x	xx	xx	x	x	x	x
						xx		x				
						xx		x	x	x	x	x
DRJ					x	x	x	x	x	xx		xx
	x	x	x	x	x	x	x	x	x	xx		xx
	x	x	x	x	x	x	x	x	x	x	x	x

**Table 3.S6** continued

<b>Lig ID</b>	<b>1FM6</b>				<b>1ZGY</b>				<b>2PRG</b>			
	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH
DRY					X	X	X	XX	X	X		X
						X	X	X	X	X	X	X
						X	X	X	X	X		X
EHA	X			X		X	X	X	X	X		XX
		X		X					X	X		XX
ET1							X				X	
							X				X	
HXA		X	X	X	X	X	X	XX		XX		XX
					X		X	X	X		X	
					X		X	X	X		X	
PTG1	X	X	X	XX	X	X	XX	XX	X	X	X	XX
	X	X	X	XX	X	X	XX	XX	X		X	
	X	X	X	XX	X	X	XX	XX	X	X	X	X
PTG2		X	X				X		X	X	X	X
		X	X				X		X	X	X	XX
	X	X	X	XX			X		X	X	X	

**Table 3.S7** Predicted hydrophobic and hydrogen bond interactions for ligands in small-scale screening test set relative to a reference list of interactions common to rosiglitazone and selected fatty acids (Table 3.S4). Poses were taken from docking of each ligand into each of the three listed PPAR $\gamma$  PDB files (top row). Predicted free energy of binding is listed as kcal/mol.

Ligand	1FM6			1ZGY			2PRG		
	kcal/mol	Hydrophobic	Hydrogen Bonds	kcal/mol	Hydrophobic	Hydrogen Bonds	kcal/mol	Hydrophobic	Hydrogen Bonds
Farglitazar	-10.4	10	0	-11.2	10	0	-10.5	5	0
Indol-1-yl acetic acid	-10.1	8	2	-10.4	8	2	-10.6	9	4
5-substituted indoleoxyacetic acid analogue	-10.1	9	0	-10.4	9	0	-10.4	10	3
(2R)-ureidofibrate-like	-8.7	14	4	-9.0	14	4	-8.3	10	5
(2S)-ureidofibrate-like	-8.2	10	0	-8.9	10	0	-8.3	6	4
9-HODE	-6.5	9	5	-6.6	9	5	-6.6	8	4
13-HODE	-7.0	6	2	-6.6	6	2	-6.3	7	4
Indeglitazar	-8.1	4	0	-8.5	4	0	-8.5	6	1
$\alpha$ -Eleostearic Acid	-6.4	8	5	-6.5	8	5	-6.5	6	4
apigenin	-8.1	0	0	-8.1	0	0	-7.8	11	2
$\beta$ -Eleostearic Acid	-6.1	10	5	-6.2	10	5	-6.0	13	5
BiochaninA	-8.0	0	0	-7.7	0	0	-7.8	2	0
Calendic Acid	-6.1	10	5	-5.8	10	5	-5.7	11	4
Catalpic Acid	-5.9	8	4	-5.9	8	4	-6.0	8	2
chrysin	-7.8	6	0	-8.4	6	0	-7.8	0	0
CLA (18C:c9,t11)	-6.0	9	1	-6.4	9	1	-6.3	8	4
Dihydroquercetin	-8.5	3	0	-8.0	3	0	-7.7	6	2



**Table 3.S7** continued

<b>Ligand</b>	<b>1FM6</b>			<b>1ZGY</b>			<b>2PRG</b>		
	kcal/mol	Hydrophobic	Hydrogen Bonds	kcal/mol	Hydrophobic	Hydrogen Bonds	kcal/mol	Hydrophobic	Hydrogen Bonds
Genistein	-8.1	0	0	-7.5	0	0	-8.0	4	0
Hesperidin	-9.2	6	0	-10.0	6	0	-9.2	9	0
Jacaric Acid	-5.9	13	5	-6.2	13	5	-5.8	7	4
Kuroda_No10	-7.1	9	0	-7.3	9	0	-7.2	8	0
Kuroda_No15	-7.3	2	0	-7.5	2	0	-7.5	10	3
Kuroda_No16	-7.7	0	0	-8.7	0	0	-8.9	6	0
Kuroda_No2	-8.4	6	1	-8.8	6	1	-9.2	0	0
Kuroda_No3	-8.7	3	0	-9.2	3	0	-9.0	6	0
Kuroda_No34	-9.8	4	0	-9.9	4	0	-9.2	4	0
Kuroda_No38	-8.9	0	0	-9.4	0	0	-9.5	9	1
Kuroda_No39	-8.4	12	0	-8.1	12	0	-8.6	10	3
Kuroda_No5	-7.5	6	0	-9.2	6	0	-8.7	8	0
Kuroda_No6	-7.5	8	0	-9.0	8	0	-8.1	11	0
Markt_264908-13-6_1	-9.9	8	3	-10.5	8	3	-10.1	11	2
Markt_264908-13-6_2	-10.4	6	4	-9.6	6	4	-10.3	5	4
Markt_264908-13-6_3	-10.1	14	3	-10.0	14	3	-9.8	13	0
Markt_264908-13-6_4	-9.5	2	0	-8.8	2	0	-9.9	4	2
Markt_651724-09-3_1	-8.3	5	0	-8.7	5	0	-8.4	7	0
Markt_651724-09-3_2	-8.6	5	0	-8.8	5	0	-8.7	14	2

**Table 3.S7** continued

<b>Ligand</b>	<b>1FM6</b>			<b>1ZGY</b>			<b>2PRG</b>		
	kcal/mol	Hydrophobic	Hydrogen Bonds	kcal/mol	Hydrophobic	Hydrogen Bonds	kcal/mol	Hydrophobic	Hydrogen Bonds
Markt_853652-40-1_1	-10.2	9	2	-10.3	9	2	-10.7	9	4
Markt_853652-40-1_2	-10.2	14	5	-10.6	14	5	-10.7	11	4
Markt_BRL48482_1	-8.9	10	1	-9.3	10	1	-8.9	7	2
Markt_BRL48482_2	-9.1	7	4	-9.3	7	4	-9.5	12	4
Markt_BVT13	-8.1	4	0	-8.6	4	0	-8.2	8	0
Markt_CLX-M1_1	-9.9	10	4	-10.2	10	4	-9.9	13	4
Markt_CLX-M1_2	-8.4	6	0	-8.6	6	0	-8.7	3	0
Markt_KRP297_1	-9.2	9	0	-9.3	9	0	-9.2	7	2
Markt_KRP297_2	-9.4	6	1	-9.3	6	1	-9.3	12	4
Markt_NNC61-4424_1	-9.5	14	5	-9.4	14	5	-9.8	6	4
Markt_NNC61-4424_2	-9.2	10	5	-8.2	10	5	-9.3	12	2
Markt_tesaglitazar	-8.7	4	5	-8.6	4	5	-8.6	6	4
Markt_troglitazone_1	-10.2	3	4	-10.4	3	4	-10.5	8	4

**Table 3.S7** continued

<b>Ligand</b>	<b>1FM6</b>			<b>1ZGY</b>			<b>2PRG</b>		
	kcal/mol	Hydrophobic	Hydrogen Bonds	kcal/mol	Hydrophobic	Hydrogen Bonds	kcal/mol	Hydrophobic	Hydrogen Bonds
Markt_troglitazone_2	-9.9	4	4	-10.1	4	4	-10.6	10	4
Markt_troglitazone_3	-9.5	5	0	-9.9	5	0	-9.9	5	2
Markt_troglitazone_4	-9.2	5	0	-9.7	5	0	-9.7	5	2
Omega-3	-6.6	11	5	-6.5	11	5	-6.7	9	4
Ψ-Baptigenin	-8.1	4	0	-8.2	4	0	-8.5	8	2
Punicic Acid	-6.3	10	4	-5.7	10	4	-6.2	11	3
Rosiglitazone	-8.7	6	4	-8.5	6	4	-8.7	9	4
Tanrikulu1	-7.6	9	0	-7.5	9	0	-7.3	7	0
Tanrikulu2_1	-9.3	4	0	-8.7	4	0	-8.8	3	0
Tanrikulu2_2	-8.2	13	0	-9.0	13	0	-8.7	8	0
Tanrikulu2_3	-9.6	10	1	-9.4	10	1	-8.8	3	0
Tanrikulu2_4	-8.8	8	0	-9.1	8	0	-8.8	4	0
Tanrikulu3_1	-9.2	7	0	-8.8	7	0	-8.8	6	0
Tanrikulu3_2	-9.0	10	0	-9.0	10	0	-8.7	10	0
Tanrikulu3_3	-8.8	6	1	-8.5	6	1	-8.7	8	1
Tanrikulu3_4	-9.1	6	0	-8.5	6	0	-8.5	10	1
Tanrikulu4	-8.4	4	0	-8.8	4	0	-9.4	8	0
Tanrikulu5	-8.1	4	0	-9.2	4	0	-9.4	8	1
Tanrikulu6	-8.7	3	0	-8.4	3	0	-8.8	4	0
Tanrikulu7_1	-9.6	7	0	-9.5	7	0	-9.1	4	0
Tanrikulu7_2	-9.1	6	0	-9.0	6	0	-9.1	10	4
Tanrikulu7_3	-9.1	7	0	-9.4	7	0	-9.1	3	0

**Table 3.S7** continued

<b>Ligand</b>	<b>1FM6</b>			<b>1ZGY</b>			<b>2PRG</b>		
	kcal/mol	Hydrophobic	Hydrogen Bonds	kcal/mol	Hydrophobic	Hydrogen Bonds	kcal/mol	Hydrophobic	Hydrogen Bonds
Tanrikulu7_4	-9.2	8	0	-9.6	8	0	-9.1	5	0
Tanrikulu8_1	-9.5	9	1	-9.8	9	1	-9.2	7	0
Tanrikulu8_2	-9.3	9	0	-9.4	9	0	-9.5	11	4
Tanrikulu8_3	-9.6	4	0	-9.4	4	0	-9.2	12	3
Tanrikulu8_4	-9.5	8	0	-9.8	8	0	-9.4	12	5
Vitexin	-9.1	4	0	-11.2	4	0	-8.6	7	0

**Table 3.S8** Presence or absence of potential hydrogen bond interactions between indicated residues of selected protein structure models (top row) and ligand poses. A single “x” indicates one potential interaction for the listed residue was found for the specified ligand, whereas more than one “x” indicates more than one interaction (e.g., “xx” indicates two interactions found).

Ligand	1FM6				1ZGY				2PRG			
	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH
Farglitazar												
9-HODE	x	x	x	xx	x	x	x	xx	x	x	x	x
13-HODE		x		x		x		x	x	x	x	x
Indol-1-yl acetic acid		x		x		x		x	x	x	x	x
5-substituted indoleoxyacetic acid analogue									x	x	x	
(2R)- ureidofibrate-like derivative	x	x	x	x	x	x	x	x	x	xx		xx
((2S)-ureidofibrate-like derivative									x	x	x	x
Indeglitazar											x	
CLA (18C:c9,t11)			x				x		x	x	x	x
Calendic Acid	x	x	x	xx	x	x	x	xx	x	x	x	x
Catalpic Acid	x	x	x	x	x	x	x	x	x		x	
Jacaric Acid	x	x	x	xx	x	x	x	xx	x	x	x	x
Kuroda_No10												
Kuroda_No15									x		x	x
Kuroda_No16												

**Table 3.S8** continued

<b>Ligand</b>	<b>1FM6</b>				<b>1ZGY</b>				<b>2PRG</b>			
	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH
Kuroda_No2			x				x					
Kuroda_No34												
Kuroda_No38									x			
Kuroda_No39									x		x	x
Kuroda_No3												
Kuroda_No5												
Kuroda_No6												
Markt_264908-13-6_1	x		x	x	x		x	x		x		x
Markt_264908-13-6_2	x		x	XX	x		x	XX	x	x	x	x
Markt_264908-13-6_3	x		x	x	x		x	x				
Markt_264908-13-6_4										x		x
Markt_651724-09-3_1												
Markt_651724-09-3_2										x		x
Markt_853652-40-1_1		x		x		x		x	x	x	x	x
Markt_853652-40-1_2	x	x	x	XX	x	x	x	XX	x	x	x	x
Markt_BRL48482_1			x				x			x		x
Markt_BRL48482_2	x		x	XX	x		x	XX	x	x	x	x
Markt_BVT13												
Markt_CLX-M1_1	x	x	x	x	x	x	x	x	x	x	x	x
Markt_CLX-M1_2												
Markt_KRP297_1										x		x

**Table 3.S8** continued

<b>Ligand</b>	<b>1FM6</b>				<b>1ZGY</b>				<b>2PRG</b>			
	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH
Markt_KRP297_2			x				x		x	x	x	x
Markt_NNC61-4424_1	x	x	x	xx	x	x	x	xx	x	x	x	x
Markt_NNC61-4424_2	x	x	x	xx	x	x	x	xx		xx		
Tesaglitazar	x	xx	x	x	x	xx	x	x	x	x	x	x
Troglitazone_1	x		x	xx	x		x	xx	x	x	x	x
Troglitazone_2	x		x	xx	x		x	xx	x	x	x	x
Troglitazone_3										x		x
Troglitazone_4										x		x
ψ-baptigenin										x		x
Punicic Acid	x	x	x	x	x	x	x	x	x		x	x
Tanrikulu1												
Tanrikulu2_1												
Tanrikulu2_2												
Tanrikulu2_3			x				x					
Tanrikulu2_4												
Tanrikulu3_1												
Tanrikulu3_2												
Tanrikulu3_3			x				x				x	
Tanrikulu3_4										x		
Tanrikulu4												
Tanrikulu5										x		

**Table 3.S8** continued

<b>Ligand</b>	<b>1FM6</b>				<b>1ZGY</b>				<b>2PRG</b>			
	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH	HIS 323 NE2	HIS 449 NE2	SER 289 OG	TYR 473 OH
Tanrikulu6												
Tanrikulu7_1												
Tanrikulu7_2									x	x	x	x
Tanrikulu7_3												
Tanrikulu7_4												
Tanrikulu8_1		x				x						
Tanrikulu8_2										x	xx	x
Tanrikulu8_3									x	x	x	
Tanrikulu8_4									xx	x	x	x
$\alpha$ -Eleostearic Acid	x	x	x	xx	x	x	x	xx	x	x	x	x
Apigenin										x		x
$\beta$ -Eleostearic Acid	x	x	x	xx	x	x	x	xx	x	x	x	xx
Biochanin-A												
Chrysin												
Dihydroquercetin										x	x	
Genistein												
Hesperidin												
Omega-3 conjugated linoleic acid	x	x	x	xx	x	x	x	xx	x	x	x	x
Rosiglitazone	x	x	x	x	x	x	x	x	x	x	x	x
Vitexin												



**Table 3.S9** Predicted free energy of binding and interaction counts for conjugated trienes. Docking was performed using AD4 with three top-binding replicates for each ligand (150 total conformations). The highest energy conformation with the highest number of hydrogen bonds was used for analysis in Table 3.4.

<b>Ligand</b>	<b>kcal/mol</b>	<b>Hydrogen bond</b>	<b>Hydrophobic</b>
$\alpha$ -Eleostearic	-5.75	2	15
	-5.73	1	15
	-5.6	3	16
Calendic	-4.49	0	29
	-3.95	1	13
	-4.47	2	7
Catalpic	-4.72	1	13
	-4.48	3	9
	-4.31	1	23
Jacaric	-4.81	2	11
	-4.97	1	9
	-4.5	3	16
Punicic	-4.3	2	16
	-4.28	3	15
	-3.78	1	10

## **4 Improved success of PPAR $\gamma$ docking using pharmacophore modeling**

Stephanie N. Lewis<sup>1,2,3</sup>, Zulma Garcia<sup>2</sup>, Raquel Hontecillas<sup>1,3</sup>, Josep Bassaganya-Riera<sup>1,3,4</sup>, and David R. Bevan<sup>1,2</sup>

<sup>1</sup>Genetics, Bioinformatics, and Computational Biology Program, Virginia Tech, Blacksburg, VA, USA

<sup>2</sup>Department of Biochemistry, Virginia Tech, Blacksburg, VA, USA

<sup>3</sup>Nutritional Immunology and Molecular Medicine Laboratory, Center for Modeling Immunity to Enteric Pathogens, Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA USA

<sup>4</sup>Department of Biomedical Sciences and Pathobiology, Virginia-Maryland College of Veterinary Medicine, Virginia Tech, Blacksburg, VA USA

## 4.1 Abstract

Peroxisome proliferator-activated receptor-gamma (PPAR $\gamma$ ) falls within a class of nuclear hormone receptors responsible for regulating various metabolic and immune processes. The PPAR family of receptors possesses a relatively large binding cavity that imparts a level of promiscuity not common to other nuclear receptors. This feature increases the challenge of using computational methods to identify PPAR binders, which are compounds that dock favorably into a structural model of the protein. Utilizing both ligand- and structure-based pharmacophore methods, we sought to improve the prediction of potential agonists by grouping ligands according to pharmacophore features. These pharmacophore feature groups were then matched to multiple PPAR $\gamma$  crystal structures to account for differences in the structural features of the binding cavity that could favor binding of particular types of ligands. For 22 of the 33 receptor structures evaluated we saw an increase in true positive rate (TPR) when screening was restricted to compounds sharing features seen with rosiglitazone. The TPR improvement varied with each receptor model in that some structures showed positive docking of specific classes of ligands while others suggested accommodation of multiple ligands with differing activity type. In the context of diverse compound database screening against several models of PPAR $\gamma$ , pharmacophore pre-screening reduced the number of compounds that needed to be screened while retaining the list of true positives identified with a given receptor structure model. Structure models were selected based on differences in binding cavity environment, and each model was paired with one or more ligand pharmacophore models. A large-scale screening using a marketed drug database verified predictability of the selected structure models. This study highlights the steps necessary to improve screening for PPAR $\gamma$  ligands using multiple structure models and ligand-based pharmacophore data.

## 4.2 Introduction

Peroxisome proliferator-activated receptor-gamma (PPAR $\gamma$ ) is a ligand-activated transcription factor within the nuclear hormone receptor (NHR) superfamily (2). It is one of three PPAR subtypes:  $-\alpha$ ,  $-\beta/\delta$ , and  $-\gamma$  (45), and it is involved in transcriptional regulation of numerous biological processes including glucose and lipid homeostasis (2, 3, 11, 45), adipocyte differentiation (2, 159), cell proliferation, and inflammation regulation (1, 3, 11, 112, 160). PPAR $\gamma$  is highly expressed in various cells that make up the immune system (161), as well as other cell types throughout the body (4, 22, 162, 163). Given the regulatory roles of PPAR $\gamma$ , this protein holds potential as a target for the treatment of metabolic and chronic inflammatory diseases, such as metabolic syndrome and inflammatory bowel disease.

In general, compounds that bind to PPAR $\gamma$  can be either agonists or antagonists (41, 51, 54, 57). Agonists can be 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 compounds (45, 54). Agonist binding induces a conformational change in the activation function-2 (AF-2) domain that governs co-repressor release and co-activator recruitment (2, 43, 51, 52, 163). 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 PPAR $\gamma$ -agonist-co-activator complex, which is heterodimerized with retinoic acid-bound retinoid X receptor-alpha, can associate with PPAR response elements (PPRE) on DNA to induce transcription (45, 53, 163). Antagonist binding differs in that binding does not induce co-repressor release. Instead the inactive conformation is stabilized by binding of the antagonist (57). Typically, PPAR $\gamma$  ligands can be described as having a three-module structure consisting of a reactive polar head group, followed by a linker, and an effector segment (61, 164). The polar head group tends to interact with residues near helix (H) 12, while the effector segment interacts with residues near H3 and the opening to the binding cavity.

Some reports of experimental application of high-throughput screening (HTS) to identify PPAR $\gamma$  modulators have appeared (165, 166), though none of the HTS methods used have been widely adopted. Moreover, experimental methods can be costly if one wants to screen a diverse collection of large numbers of compounds, and would require a significant time commitment. Virtual screening is a computational technique for streamlining the drug discovery and development process (95). 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 predicted to be unfavorable or nonexistent.

Large-scale virtual screening studies with PPAR $\gamma$  structures often result in generally disappointing outcomes (167-171). 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 identify features within the binding cavity that would favor interactions between the target receptor and the atoms of potential ligands (90, 91). 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 (92).

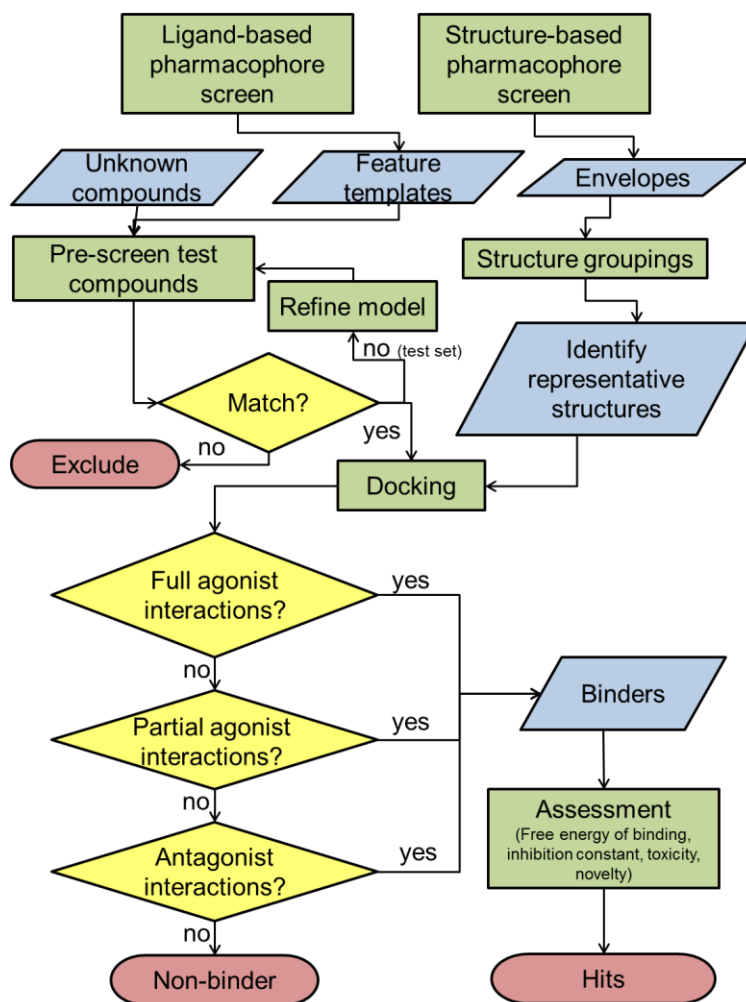
These two pharmacophore modeling methods independently provide useful structural information for establishing criteria for identifying binders in a search for potential agonists. Studies using one method or the other have shown identification of agonists for therapeutic development that satisfied specific pharmacophore criteria using screening methods performed on compound datasets of varying size (172-177). Pharmacophore modeling has been applied to virtual screening of ligands for various protein targets including the other PPAR subtypes  $\alpha$  and  $\beta/\delta$  (178). A combination of the two pharmacophore methods may narrow the search to facilitate compound screening with respect to PPAR $\gamma$  (169-171). Through application of this concept, pharmacophore modeling can be an important tool in the discovery and design of novel drugs for PPAR $\gamma$  (179).

In this study, we set out to identify criteria that could be applied in a virtual screen of a diverse ligand database using multiple structure models for PPAR $\gamma$ . The goal was to reduce the amount of docking necessary to identify binders while improving the frequency with which true potential binders were identified. Using pharmacophores as filters (180), 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 work out a scoring system where pharmacophores and potential interactions are considered when proposing binders for validation (Scheme 4.1). A test of the methods with a dataset of compounds with unknown PPAR $\gamma$  binding ability indicated that identification of compounds that should bind to PPAR $\gamma$  is possible given a diverse compound database, our combined theoretical approach, and the selected criteria for identifying binders.

## **4.3 Methods**

### **4.3.1 Catalog of PPAR $\gamma$ 3D structures**

PPAR $\gamma$  structures from the RCSB Protein Data Bank (PDB) (59, 60) were cataloged. As of January 2012, 97 structures were available, of which most were complexed 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 4.1 and 4.2). The activity classes explored in this study were full agonist, fatty acid, 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 PPAR $\gamma$  and are structurally different from synthetic agonists.



**Scheme 4.1** Process flow diagram for combined pharmacophore screening 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.

### 4.3.2 Pharmacophore modeling

A subset of structures and associated ligands was used for the pharmacophore searching process. A total of 33 crystal structures containing ligands with clearly defined activity types were used for the structure-based pharmacophore analysis (Table 4.1). The selected structures covered the four activity types: 11 full agonist-, 5 fatty acid- 15 partial agonist-, and 2 antagonist-containing structures.

For the ligand-based pharmacophore modeling, the ligands from the selected PDB structures were separated into groups based on activity. Additional ligands were included in the full and partial agonist groups for the ligand-based analysis to expand the number of ligand features that were considered. The final counts for each activity type were 22 full agonists, 5 fatty acids, 21 partial agonists, and 2 antagonists (Table 4.2).

**Table 4.1** PDB structures used for the structure-based pharmacophore analysis and virtual screening. Ligand IDs and names are listed for the co-crystallized compounds, followed by the ligand type, structure resolution, and the citation for the paper associated with the structure.

PDB ID	Ligand ID	Ligand Name	Ligand Type	Res. (Å)	Citation
1FM6	BRL	2,4-THIAZOLIDINEDIONE, 5-[[4-[2-(METHYL-2-PYRIDINYLAMINO)ETHOXY]PHENYL]METHYL]-(9CL)	full	2.10	(45)
1FM9	570	2-(2-BENZOYL-PHENYLAMINO)-3-{4-[2-(5-METHYL-2-PHENYL-OXAZOL-4-YL)-ETHOXY]-PHENYL}-PROPIONIC ACID	full	2.10	(45)
1K74	544	2-(1-METHYL-3-OXO-3-PHENYL-PROPYLAMINO)-3-{4-[2-(5-METHYL-2-PHENYL-OXAZOL-4-YL)-ETHOXY]-PHENYL}-PROPIONIC ACID	full	2.30	(64)
1RDT	570	2-(2-BENZOYL-PHENYLAMINO)-3-{4-[2-(5-METHYL-2-PHENYL-OXAZOL-4-YL)-ETHOXY]-PHENYL}-PROPIONIC ACID	full	2.40	(67)
1WM0	PLB	2-[(2,4-DICHLOROBENZOYL)AMINO]-5-(PYRIMIDIN-2-YLOXY)BENZOIC ACID	partial	2.90	(68)
1ZGY	BRL	2,4-THIAZOLIDINEDIONE, 5-[[4-[2-(METHYL-2-PYRIDINYLAMINO)ETHOXY]PHENYL]METHYL]-(9CL)	full	1.80	(70)
2FVJ	RO0	1-(3,4-DIMETHOXYBENZYL)-6,7-DIMETHOXY-4-[[4-(2-METHOXYPHENYL)PIPERIDIN-1-YL]METHYL]ISOQUINOLINE	partial	1.99	(73)
2G0G	SP0	3-FLUORO-N-[1-(4-FLUOROPHENYL)-3-(2-THIENYL)-1H-PYRAZOL-5-YL]BENZENESULFONAMIDE	partial	2.54	(74)
2G0H	SP3	N-[1-(4-FLUOROPHENYL)-3-(2-THIENYL)-1H-PYRAZOL-5-YL]-3,5-BIS(TRIFLUOROMETHYL)BENZENESULFONAMIDE	partial	2.30	(74)
2HFP	NSI	3-(4-METHOXYPHENYL)-N-(PHENYLSULFONYL)-1-[3-(TRIFLUOROMETHYL)BENZYL]-1H-INDOLE-2-CARBOXAMIDE	antagonist	2.00	(76)
2I4J	DRJ	(2R)-2-(4-{2-[1,3-BENZOXAZOL-2-YL(HEPTYL)AMINO]ETHYL}PHENOXY)-2-METHYLBUTANOIC ACID	full	2.10	(54)
2I4P	DRH	(2S)-2-(4-{2-[1,3-BENZOXAZOL-2-YL(HEPTYL)AMINO]ETHYL}PHENOXY)-2-METHYLBUTANOIC ACID	partial	2.10	(54)
2I4Z	DRH	(2S)-2-(4-{2-[1,3-BENZOXAZOL-2-YL(HEPTYL)AMINO]ETHYL}PHENOXY)-2-METHYLBUTANOIC ACID	partial	2.25	(54)

Table 4.1 continued.

PDB ID	Ligand ID	Ligand Name	Ligand Type	Res. (Å)	Citation
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	2.80	(83)
2P4Y	C03	(2R)-2-(4-CHLORO-3-{{3-(6-METHOXY-1,2-BENZISOXAZOL-3-YL)-2-METHYL-6-(TRIFLUOROMETHOXY)-1H-INDOL-1-YL}METHYL}PHENOXY)PROPANOIC ACID	partial	2.25	(84)
2POB	GW4	N-[(2S)-2-[(2-BENZOYLPHENYL)AMINO]-3-{4-[2-(5-METHYL-2-PHENYL-1,3-OXAZOL-4-YL)ETHOXY]PHENYL}PROPYL]ACETAMIDE	full	2.30	(85)
2PRG	BRL	2,4-THIAZOLIDINEDIONE, 5-[[4-[2-(METHYL-2-PYRIDINYLAMINO)ETHOXY]PHENYL]METHYL]-(9CL)	full	2.30	(52)
2Q5P	241	(2S)-2-(3-{{1-(4-METHOXYBENZOYL)-2-METHYL-5-(TRIFLUOROMETHOXY)-1H-INDOL-3-YL}METHYL}PHENOXY)PROPANOIC ACID	partial	2.30	(55)
2Q5S	NZA	5-CHLORO-1-(4-CHLOROBENZYL)-3-(PHENYLTHIO)-1H-INDOLE-2-CARBOXYLIC ACID	partial	2.05	(55)
2Q6R	SF2	5-CHLORO-1-(3-METHOXYBENZYL)-3-(PHENYLTHIO)-1H-INDOLE-2-CARBOXYLIC ACID	partial	2.41	(55)
2Q6S	PLB	2-[(2,4-DICHLOROBENZOYL)AMINO]-5-(PYRIMIDIN-2-YLOXY)BENZOIC ACID	partial	2.40	(55)
2VSR	9HO	(9S,10E,12Z)-9-HYDROXYOCTADECAN-10,12-DIENOIC ACID	full/ fatty acid	2.05	(41)
2VST	243	(9Z,11E,13S)-13-HYDROXYOCTADECAN-9,11-DIENOIC ACID	full/ fatty acid	2.35	(41)
2VV0	HXA	DOCOSA-4,7,10,13,16,19-HEXAENOIC ACID	full/ fatty acid	2.55	(41)
2VV1	4HD	(4S,5E,7Z,10Z,13Z,16Z,19Z)-4-HYDROXYDOCOSA-5,7,10,13,16,19-HEXAENOIC ACID	full/ fatty acid	2.20	(41)
2VV2	5HE	(5R,6E,8Z,11Z,14Z,17Z)-5-HYDROXYICOSA-6,8,11,14,17-PENTAENOIC ACID	full/ fatty acid	2.75	(41)
3B3K	LRG	(2S)-2-(BIPHENYL-4-YLOXY)-3-PHENYLPROPANOIC ACID	full	2.60	(77)
3CDP	YRG	(2S)-2-(4-CHLOROPHENOXY)-3-PHENYLPROPANOIC ACID	partial	2.80	(80)



Table 4.1 continued.

PDB ID	Ligand ID	Ligand Name	Ligand Type	Res. (Å)	Citation
3CS8	BRL	2,4-THIAZOLIDINEDIONE, 5-[[4-[2-(METHYL-2-PYRIDINYLAMINO)ETHOXY]PHENYL]METHYL]-(9CL)	full	2.30	(81)
3DZU	PLB	2-[(2,4-DICHLOROBENZOYL)AMINO]-5-(PYRIMIDIN-2-YLOXY)BENZOIC ACID	partial	3.20	(56)
3DZY	BRL	2,4-THIAZOLIDINEDIONE, 5-[[4-[2-(METHYL-2-PYRIDINYLAMINO)ETHOXY]PHENYL]METHYL]-(9CL)	full	3.10	(56)
3E00	GW9	2-CHLORO-5-NITRO-N-PHENYLBENZAMIDE	antagonist	3.10	(56)
4PRG	072	(+/-)(2S,5S)-3-(4-(4-CARBOXYPHENYL)BUTYL)-2-HEPTYL-4-OXO-5-THIAZOLIDINE	partial	2.90	(87)

### 4.3.3 Characterization of crystal structure binding cavities

Generation of structure-based pharmacophores involved application of AutoLigand, which uses a grid-based representation of binding potentials to identify sites of favorable interactions in a protein binding site (181). This approach calculates atom-specific grids within the binding cavity for the selected PDB crystal structures with the resulting affinity potential grid being called an envelope (181). The atom types considered for the envelopes were carbon, oxygen, nitrogen, hydrogen, and sulfur. Prior to generating the envelopes, UCSF Chimera (141) was used to overlay the structures and save protein coordinates to the same coordinate space. The 1FM6 PDB structure (45) was used as the reference for overlaying the structures. 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. A tally of identical points was calculated for the favorable envelopes within each activity class. A representative envelope for each activity type was created using the tallies for the grid points. UCSF Chimera (141) was used to visualize the representative envelopes, and the tallies were modifiers for grid point representation. Two visualization techniques were applied: (1) sphere radius was adjusted by using the tallies as radius multipliers (181), and (2) sphere color was changed as a gradient to measure frequency of point appearance for all envelopes in a class independent of atom type. The first representation is a visual histogram of atom distribution, while the second is a heat-map-like view of point frequency. Larger spheres represent regions where grid points were common between the envelopes of an activity class, while smaller spheres represented more unique regions of atom placement for the compared envelopes. Binding cavity arms were designated as arms I through III for visualization and evaluation of the binding cavity. Names were based on crystal structure data and naming within the associated publications (144, 182). 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 next grouped based on envelope similarity. Activity class was not considered for grouping, which allowed for comparison of crystal structure binding pockets across activity classes.

**Table 4.2** List of ligand IDs for agonists and antagonists used in the ligand-based pharmacophore analysis. The PDB ID for the crystal structure containing the ligand and the ligand type are listed for each ligand ID. Citations not included in Table 4.1 are included for the additional structures mentioned in this table.

PDB ID	Ligand ID	Ligand Type	PDB ID	Ligand ID	Ligand Type
1FM6 <sup>a</sup>	BRL	Full agonist	2ZK1 (39)	PTG	Full agonist
1FM9	570	Full agonist	2ZK6 (39)	C08	Partial agonist
1K74	544	Full agonist	2ZNO (183)	S44	Full agonist
1RDT	571	Full agonist	3ADW (184)	MYI	Full agonist
1WM0	PLB	Partial agonist	3ADX (184)	IMN	Full agonist
2ATH (71)	3EA	Full agonist	3AN4 (185)	M7R	Full agonist
2FVJ	RO0	Partial agonist	3B0Q (186)	MC5	Full agonist
2G0G	SP0	Partial agonist	3B1M (187)	KRC	Partial agonist
2G0H	SP3	Partial agonist	3B3K	LRG	Full agonist
2HFP	NSI	Antagonist	3CDP	YRG	Partial agonist
2I4J	DRJ	Full agonist	3CDS (77)	GRR	Full agonist
2I4P	DRH	Partial agonist	3E00	GW9	Antagonist
2P4Y	C03	Partial agonist	3ET0 (86)	ET0	Full agonist
2POB	GW4	Full agonist	3FUR (188)	Z12	Partial agonist
2Q59 (55)	240	Full agonist	3GBK (189)	2PQ	Full agonist
2Q5P	241	Partial agonist	3H0A (190)	D30	Partial agonist
2Q5S <sup>a</sup>	NZA	Partial agonist	3H00 (190)	DKD	Full agonist
2Q61 (55)	SF1	Partial agonist	3HOD (191)	ZZH	Full agonist
2Q6R	SF2	Partial agonist	3K8S (192)	Z27	Partial agonist
2VSR	9HO	Fatty acid	3LMP (193)	CEK	Partial agonist
2VST	243	Fatty acid	3NOA (28)	5BC	Full agonist
2VV0	HXA	Fatty acid	3OSI (194)	XDH	Partial agonist
2VV1	4HD	Fatty acid	3OSW (194)	XDI	Partial agonist
2VV2	5HE	Fatty acid	3PBA (195)	ZXG	Partial agonist
2XKW (196)	P1B	Full agonist			

<sup>a</sup>Ligands included in grouping with antagonists to fulfill PharmaGist server requirement of at least three compounds.

#### 4.3.4 Identification of ligand-based pharmacophores

The ligand-based pharmacophores were determined using PharmaGist (179, 197), which is a freely available web server for ligand feature prediction. 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). These values are different from the default settings, which weighed aromatics the heaviest (3.0), followed by hydrogen bonding groups (1.5) and charged groups (1.0). The weights were changed from the default values because hydrogen bonding is necessary for full agonism in PPAR $\gamma$ . Files containing coordinates for groups of ligands based on activity class were uploaded to PharmaGist. The server uses each uploaded ligand as a pivot molecule onto which all the other compounds are overlaid for feature comparison and scoring. 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. With this approach one molecule was used as the pivot molecule onto which the other was overlaid to determine feature similarity.

PharmaGist calculates a feature group score based on the number of features present on each ligand, the features shared between overlaid ligands, and the weight assigned to each feature type that was common for a given number of ligands. The feature groups then were ranked and assessed by these scores. The highest scoring feature group and the group representing the most ligands were evaluated for each activity class (full agonist, fatty acid agonist, partial agonist, and antagonist). 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.

3D pharmacophore models were generated for the lists of PharmaGist features with the Molecular Operating Environment (MOE) software suite (198) using the pivot molecules as the representative ligands for each activity class. These models facilitated quick screening of ligand databases for ligands that matched the pharmacophores. The 3D pharmacophore models were used in MOE to cluster compounds in the databases of PPAR $\gamma$  ligands from the PDB and from the database of PPAR $\gamma$  known and decoy compounds obtained from the Directory of Useful Decoys (DUD) (169). A test of the pharmacophore models against the known PPAR $\gamma$  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 TZD compounds, respectively. An additional full agonist model based on rosiglitazone, and a fatty acid model based on 9-hydroxyoctadeca-10,12-dienoic acid (9HO; PDB ID 2VSR) were included. These models recovered the missing ligands. As some compounds did match the initial models, they were maintained in the ligand pharmacophore model set. The final number of pharmacophore models used was eight.

#### **4.3.5 Computational docking**

The training portion of virtual screening was conducted with a total of 3308 ligands: 96 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 PPAR $\gamma$ . The compiled ligand database included duplicates of known agonists; the DUD active dataset included compounds from the PDB, and the PDB list contained multiples of some ligands. This level of redundancy provided checks for reproducibility and consistency in pose prediction for known agonists. Docking was performed against the 33 superimposed PPAR $\gamma$  structure models for which binding cavities were characterized using AutoLigand. Identical grid boxes were established for docking into each structure model. The grid box size was set to 30 Å by 40 Å by

40 Å with 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<sup>3</sup>.

AutoDock Vina (139) (Vina) was used for docking against the selected PDB structure models. Initially, re-docking was done by docking the ligand from each structure back into its cognate receptor (199). Cross-docking was then done to see if the structures could accommodate other known ligands, which involved docking known ligands from multiple PDB structures into each selected protein structure model (199).

Perl (147) scripts written in-house were used for post-processing of the docked results. This analysis included retrieving lowest energy poses and the associated energy scores, measuring distances between interacting atoms, and formatting of output for ease of reading and assessment. Nine separate poses were predicted for each ligand. Only the most energetically favorable pose for each ligand was used for analysis. Reference poses for root-mean-squared deviation (RMSD) calculations were those from the crystal structure complexes, therefore RMSD values were only calculated for the PDB-associated ligands. The terms “successful pose” and “success” were applied to docked poses with an RMSD value less than or equal to 2.0 Å (89).

Interactions between protein and ligand atoms were based on distances between heavy atoms. Distances less than or equal to 3.3 Å indicated hydrogen bonds (103, 146), while distances less than or equal to 3.9 Å indicated hydrophobic interactions (146). Predicted interactions were compared to two crystal structure-derived interaction lists that were derived from five PPAR $\gamma$ -rosiglitazone PDB structures. One list contained eight residues common to the five PDB structures that showed interactions with rosiglitazone. The second list was composed of the four residues (Ser289, His323, His449, and Tyr473) known to form hydrogen bonds with multiple full agonists. Ligand Explorer (146) was used to identify key interactions in the original PDB structures. Here the term “success” was used to describe poses for which hydrogen-bond interactions were calculated with the four critical residues.

#### **4.3.6 Validation of docking protocol**

A total of 2319 compounds from the MicroSource U.S. Drugs (MSUSDrugs) dataset of the ZINC online chemical catalog (200) were used for the validation set. These compounds are all commercially available drugs. As there are drugs currently available that activate PPAR $\gamma$ , it was expected that known agonists like rosiglitazone would be in this test set. No pre-screening was performed, however, in order to maintain a level of blind testing. Thus, this database was classified as unknowns because the binding activity was not known prior to screening. The MSUSDrugs compounds were docked into four PPAR $\gamma$  structure models selected after binding cavity analysis to identify representative models ideal for screening. The docking procedure used for the training step (described above) was also used for the validation step.

As a validation of the structure model selection and pharmacophore-based ligand screening process, additional filters were applied to the MSUSDrugs docking results in order to rank and identify 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 predict potential activity types for each ligand. Compounds were further evaluated for toxicity based on presence of reactive or unstable molecular features and existing therapeutic patents that include targeting of

PPAR $\gamma$ . Toxicity was assessed using MOE (198) and the ToxAlerts server (201). MOE references two publications in which potentially toxic and reactive molecular groups were reported (202, 203). ToxAlerts included these two publications and 23 others dating from 1976 to 2013.

### 4.3.7 Statistical analysis

Various statistical methods were used to test significance within the data. 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, which was 300. A distribution of the values was obtained by generating a histogram. Screening statistics included indicators of success and failure at accurately predicting active compounds while excluding decoys. The primary values considered were specificity and sensitivity. Specificity is a measure of the proportion of negatives accurately identified as such. Ideal specificity values should be greater than 0.95 or 95 percent. Specificity was used to calculate the false positive rate (FPR), which is the frequency with which non-binders are inaccurately identified as binders. Ideal FPRs would approach zero. Sensitivity, which is also called the true positive rate (TPR), is the probability that a randomly selected item within the test set was a binder accurately predicted as such. Sensitivity improves as the value approaches 1.00 or 100 percent. Specificity and sensitivity are calculated given true positive, false positives, true negative, and true positives.

Distinctions between true positives (TP; binders identified as binders) and false negatives (FN; binders inaccurately predicted to be non-binders) were determined based on the RMSD values for docking the PDB ligands, and 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 predicted to be binders) and true negatives (TN; non-binders identified as non-binders) were assessed using interaction predictions for the DUD decoy compound docked poses. FPs would possess interactions with the key residues, while TNs would not. TPRs and FPRs were used to assess the receiver operating characteristic (ROC) space for the structures. The ROC method is a statistic that indicates the ability of a test to separate data into two populations (204). Evaluation of the ROC space allowed for identification of structures that more accurately predicted binders while excluding non-binders.

### 4.3.8 Virtual screen process flow

As illustrated in Scheme 4.1, the process flow starts with pharmacophore screening with both ligand-based and structure-based methods. The data obtained in these steps established features necessary for binder identification. The ligand-based leg served two purposes: to establish ligand features common to known agonists and to screen compounds with unknown binding character to propose activity type. In the case of known agonists, the pharmacophore models were refined to satisfy most if not all of the known dataset. For the screening of unknowns, compounds that did not match the models were excluded, while those that did match continued on to the docking step. In this study, the known and unknown compounds that did not match the ligand-based pharmacophore models were retained to test validity of the process flow. The structure-based leg of the process served to establish representative structure models to use for docking. Docking

results were then scrutinized for potential interactions with the structure models. Only full agonist criteria were used here, but criteria for identifying partial agonists and antagonists could be included to detect binders that fit those activity types. Currently, the interactions necessary for partial agonist and antagonist binding that are distinct from full agonist interactions are not clear. Therefore, these activity classes were excluded for the test of the process flow. Any compounds not matching the binder identification criteria at this step were deemed non-binders, while those that did match were classified as binders. The binders underwent additional assessment for predicted free energy of binding and inhibition constant, potential toxicity, and novelty of the compound in the context of PPAR $\gamma$  binding. Compounds that satisfied this assessment were classified as hits, which are compounds that could be developed as a lead for experimental validation.

## 4.4 Results

### 4.4.1 Shared features were identified for each ligand type

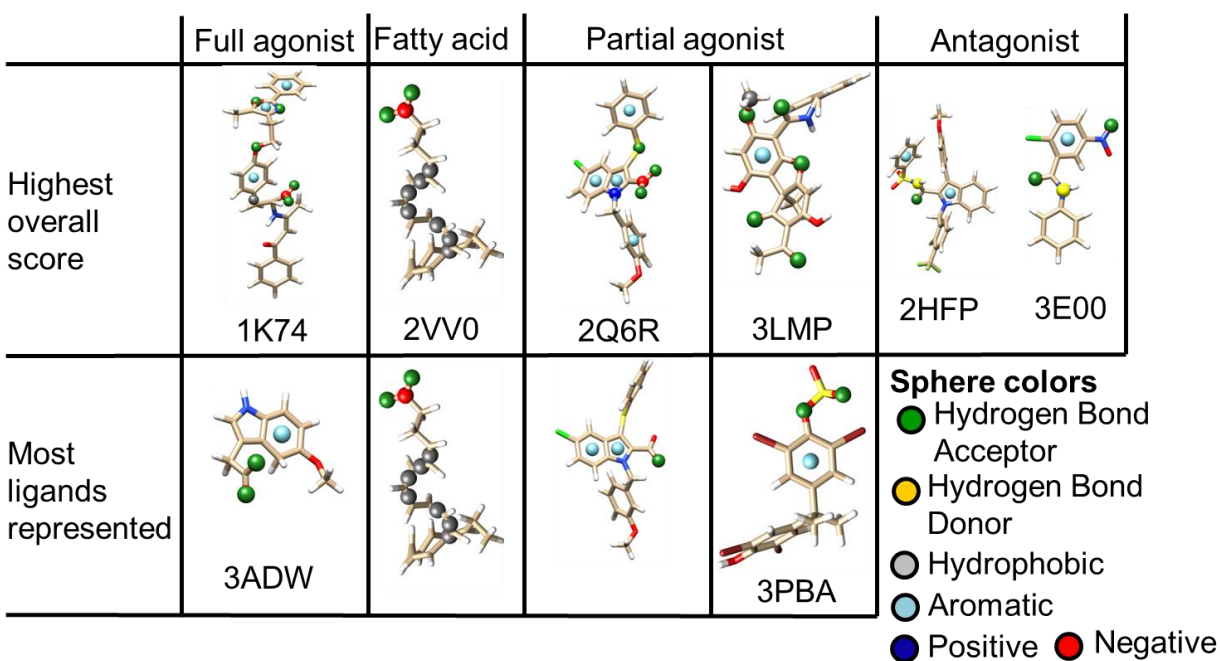
The PharmaGist online server reported several lists of features shared by compounds within each activity class. The highest overall scoring feature set and the set representative of the most ligands were assessed (Table 4.3 and Figure 4.1). Of the full agonists, three molecules aligned to give the highest overall feature score of 48.5 with ten features common to these three molecules. Within the partial agonist class, three compounds made up the highest overall feature score group with a score of 39.2 and nine shared features. The three ligands in this case were molecularly similar and were excluded in a second run of the PharmaGist server to determine if an additional predominant feature set could be found for the partial agonists. Doing so returned a set of three molecules scoring 39.9 with seven shared features. This second group of features was noticeably different from the initial partial agonist feature set, and it shared features seen with the full agonists. The highest scoring cluster for the fatty acids showed ten features common to the five ligands with a score of 27.2. 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.

**Table 4.3** 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 PDB ID) as the pivot molecule.

	Full agonist	Fatty acid	Partial agonist (22 ligands)	Partial agonist (19 ligands)	Antagonist
Highest overall score	48.5	27.2	39.2	39.9	NSI (2HFP): 8.0 GW9 (3E00): 10.0
Most ligands represented	33.6	27.2	20.0	28.9	

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 PPAR $\gamma$ . The three features for the full agonists and second partial agonist set were two hydrogen bond

acceptor groups and an aromatic moiety. Interestingly, the placement of these groups relative to the binding cavity was different, however, 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. For the fatty acids, the similarity between the highest scoring cluster and the clustering representing most ligands was not surprising given this was the one ligand category where the ligand composition was highly similar for the compounds that were analyzed. 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 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 sets and the limited number of available structures, the feature sets for both ligands 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. The pivot molecule, which was a single ligand onto which each subsequent ligand was overlaid, was used as the representative ligand for generating the pharmacophore models for each feature set. Once the pharmacophore models were established, the next step was to evaluate the PPAR $\gamma$  crystal structures to assess binding cavity similarity and identify which ligand pharmacophores would be best suited for the different shaped cavities.



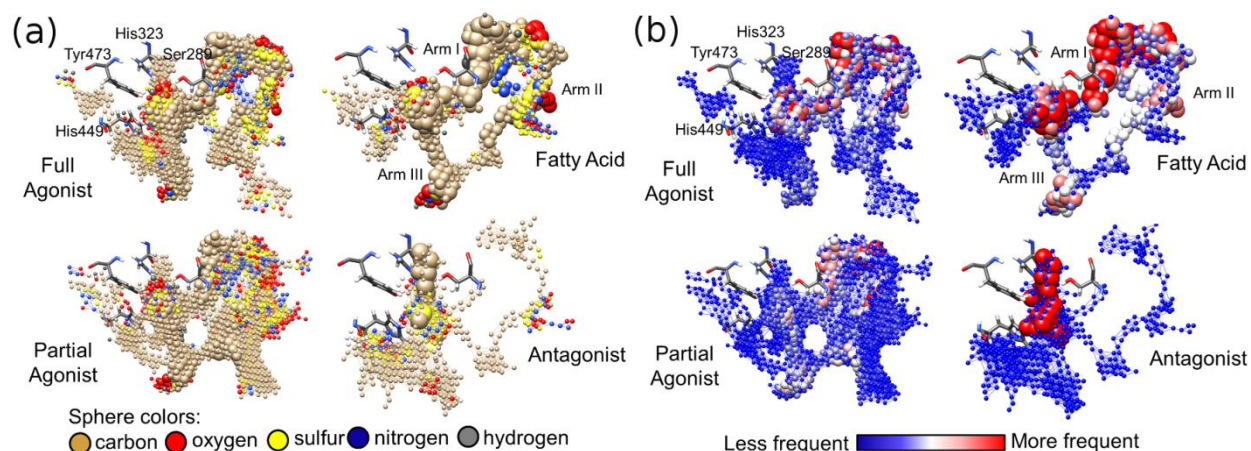
**Figure 4.1** Composite of PharmaGist feature groups overlaid onto representative ligands from each ligand type. The PDB ID in which each ligand was found is listed in the corresponding cell. Cells without a PDB ID are the same ligand as the one listed in the top cell for a given column. Sphere colors for the feature group are as indicated in the legend above. 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.

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

The envelopes generated using AutoLigand are comparable to structure-based pharmacophores used to predict potential ligand interactions given a protein structure. Envelopes were created for the 33 crystal structures listed in Table 4.1, and then combined based on the activity class of the ligand that was present in the crystal structure. Some similarities were observed in the envelopes across all activity classes (Figure 4.2). For instance, the atom types predicted to sit in various regions of the binding cavity were similar for each activity class (Figure 4.2A). Carbon atom points occupied most of the cavities. Oxygen atoms appeared most favorable near H12 and the opening of the binding cavity. A third cluster of oxygen atom points appeared at one end of arm III. The position of the more electronegative atom types suggested residues in these locations were either charged or polar. The frequency with which those atoms appeared for some regions was different, however, and presumably based on the type of activity described for the original crystal structure. The full agonist and fatty acid containing structures both showed higher quantities 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 for the composite envelope for the partial agonist-containing structures. The full agonist- and fatty acid-containing structures showed a definitive shape to the pocket in which these compounds fit, whereas the envelopes for the partial agonist-containing structures contained points spread over more of the pocket. The clustering or dispersion of spheres around the binding cavity should indicate conserved or promiscuous cavity shape, respectively. This clustering may also translate to conservation or promiscuity of activation patterns influenced by conformational changes upon ligand binding. It was observed for the antagonist group that points clustered near H12, which also occurred in the envelopes for the other activity types.

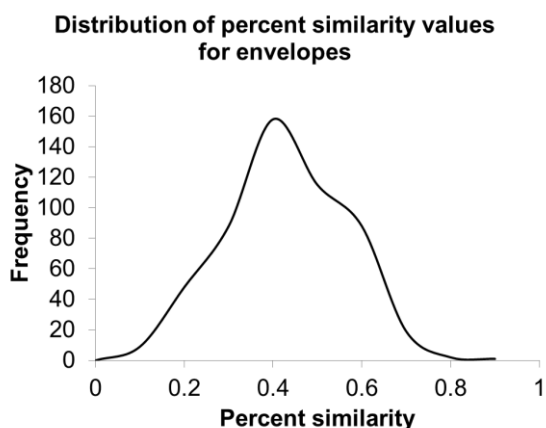
Counts of common grid points independent of atom type were also calculated and visualized in a heat-map-like representation for each activity class (Figure 4.2B). Such a representation indicated patterns in grid points favored for each activity type, which would reflect where ligand atoms should sit most favorably within the cavity for a given activity. Fatty acid-containing structures suggested favorable ligand positioning along the cavity entrance and arm I of the binding cavity. The full agonist-bound structures suggested atoms near H12 were prevalent, while the partial agonist-bound structures suggested atoms around H3 were preferred. The antagonist containing structures suggested atoms were favored around H12. This 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.





**Figure 4.2** 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 tested within the binding cavity of the crystal structures 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 the grid established within the binding cavity 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.

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 as reflected by envelope similarity. A distribution of these values showed a major peak around 0.4 or 40 percent similarity between pairs of envelopes (Figure 4.3). Using 0.4 as a cutoff value, an attempt was made to divide the models into groups with presumably similar binding cavity landscapes. The groupings were not clear with this cutoff and resulted in overlapping groups for all the models where structures appeared in multiple groups. Given all the envelopes were generated using crystal structures for PPAR $\gamma$  with bound ligands, the high degree of similarity with a low cutoff value was not surprising.



**Figure 4.3** Distribution of percent similarity values for the 33 envelopes. Percent similarity reflects the proportion of common points between each pair of envelopes (N = 1089 pairs).

The similarity cutoff for dividing the structure models into groups was increased to 0.6 or 60 percent envelope point similarity, which resulted in groupings that were more distinct. The distribution shows a second less pronounced peak around 0.6, which suggested this cutoff might be stricter. This cutoff retained similarities while allowing any differences between envelopes to be more easily seen. Five structure model groups were established (Table 4.4). Structures in Group 1 contained full agonist and fatty acids. Group 2 contained structure models with full and partial agonists in the original crystal structures. Group 3 structures contained fatty acids and partial agonists. Groups 4 and 5 contained partial agonist-bound structures alone. Group 6 was composed of the remaining crystal structures tested 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 assigned ligand pharmacophore models based on the types of ligands that should be best accommodated by the structure model (Table 4.4). The pairing of ligand-based and structure-based pharmacophore models presumes that only ligands that match the ligand pharmacophore will dock into any structure model within the group. 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 and was the next method to be applied in order to single out the representative structure models for virtual screening.

**Table 4.4** Groupings for structure models based on envelope similarity. PDB and ligand IDs match those used in Table 4.1.

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

#### 4.4.3 Ligand binding reproducibility of the PPAR $\gamma$ structure models

Re-docking and cross-docking were conducted to determine which of the PPAR $\gamma$  crystal structures would be suitable for virtual screening. With 2.0 Å as the threshold for RMSD values, re-docking results were favorable for 10 out of the 33 analyzed structures (Table 4.5). Successful docking results were found for structures with PDB codes 1WM0, 1ZGY, 2FVJ, 2G0G, 2G0H, 2POB, 2PRG, 2Q5P, 3DZU, and 4PRG, which are full or partial agonist-containing structures. During the cross-docking, all the structures showed successful docking of more than one non-native ligand despite the fact that a majority of the structures did not show successful re-docking. A table containing all the RMSD values for the re- and cross-docking poses is in the supplemental information (Supplemental Table 4.S1). The re- and cross-docking data provided an initial evaluation of how well the structure models could accurately identify binders. It was hypothesized that incorporating the pharmacophore screening would improve the predictability of structure models best suited for predicting ligands of the activity type represented by the

pharmacophore model, which would aid in representative model selection. This concept required additional statistical measures of docking potential.

**Table 4.5** RMSD values for re-docking of native ligands into corresponding crystal structure models. Results are listed by PDB ID and RMSD values are measured in Angstroms (Å).

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

#### 4.4.4 Predictability of binding with PPAR $\gamma$ structures

The information pertaining to successful or unsuccessful pose prediction was used to calculate the frequency at which a structure could accurately predict a binder pose. The initial calculation was conducted with just the PDB ligands so RMSD values could be used to assess true positive identification. The overall per-model TPRs were extremely low (2.08 to 18.75 percent success rates) for the 33 structure models (Table 4.6, original TPRs). The low rates may derive from the number and variety of ligands screened and rigid receptor structures during docking. Structures could be identified that served better for docking one type of ligand over the other types. For example, docking of full and dual agonists to 2PRG (PPAR $\gamma$ -rosiglitazone) was successful, and only one partial agonist docked well. Additionally, 2VV0 (PPAR $\gamma$ -HXA) was able to accommodate full and partial agonists though the native ligand is a fatty acid, which did not re-dock well. For some structures, preference of one activity type was evident such as 2I4P, in which a partial agonist was natively found and only partial agonists docked successfully to this structure. Antagonists in general did not dock well, but two structures, 1FM6 and 3DZY, returned favorable docking of the tested antagonists. These structures might prove useful in predicting antagonists in a docking study of compounds similar to the two antagonists screened. In this case, considerations would have to be set for the best means of distinguishing any potential antagonists from full agonists.

Pharmacophore filtering was applied to the docked poses using the rosiglitazone pharmacophore model for calculating adjusted TPRs. A subset of 22 out of the 33 structures showed some increase in TPR (Table 4.6, adjusted TPRs). Filtering reduced the total number of compounds for calculating the rate, which suggested appropriate pre-screening of test compounds 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.

**Table 4.6** True positive rates (TPRs) for each PDB structure in the training phase. 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				

It was possible that either the non-TZD accommodating structures were ideal for screening other 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. When compared to the grouped list of structure models based on envelope similarity, these structures were not in structure groups one and two, which were predicted to best fit the rosiglitazone model. In taking a closer look at the docking results, it was noticed that the compounds that docked well for the eleven structures were either exclusively partial agonists or compounds that matched multiple pharmacophore models (full agonists, partial agonists, fatty acids, and antagonists). Therefore, the structures with decreasing rates would not be useful for screening for full agonists. Some of these structures, however, may be better suited for partial agonist prediction.

Structures for which the rate increased too much would indicate either a structure most appropriate for full agonist docking or an improvement in rate due solely to the reduction of total actives in the TPR calculation. Two structures showed rates increasing by 100 percent. Of those, 2PRG showed favorable docking of multiple full, dual, and pan agonists, all of which matched the rosiglitazone pharmacophore model and therefore showed an increase in rate given the reduction in the total number of ligands considered for the calculation. The other structure, 2G0H, showed successful docking of partial agonists, one full and one pan agonist. This

structure would be less useful when screening for compounds of the full agonism activity type despite the increase in rate. Use of this model for docking might result in inaccurate non-binder identification given the greater success of partial agonist docking. The 2G0H structure showed a low number of successfully docked compounds and therefore would not be an ideal model despite the significant rate increase. Again, the increase in rate was due solely to the reduction of total actives.

The structures with a moderate increase in rate were those for which the ligands that docked well were partially those that matched the model and partially those that did not. For example, the rate for 1ZGY increased by 57.12 percent. The three compounds that did not match the rosiglitazone model were a fatty acid, partial agonist and full agonist with a stereoisomer that is a partial agonist. The fatty acid compound matched the fatty acid pharmacophores and no others. The other two ligands matched multiple pharmacophores but not rosiglitazone. This suggests that the structure model is ideal for screening compounds matching the selected pharmacophore while possessing enough variability in the binding cavity to screen for other compounds that potentially possess similar activity but may not completely match the selected pharmacophore. Additionally a structure that can accommodate compounds that do not match the pharmacophore models might prove useful for assess novel, uncharacterized ligands. 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 step was to assess the rate at which false positives were predicted and exclusion of true negative compounds.

#### **4.4.5 Inclusion of the decoys to assess false positive prediction**

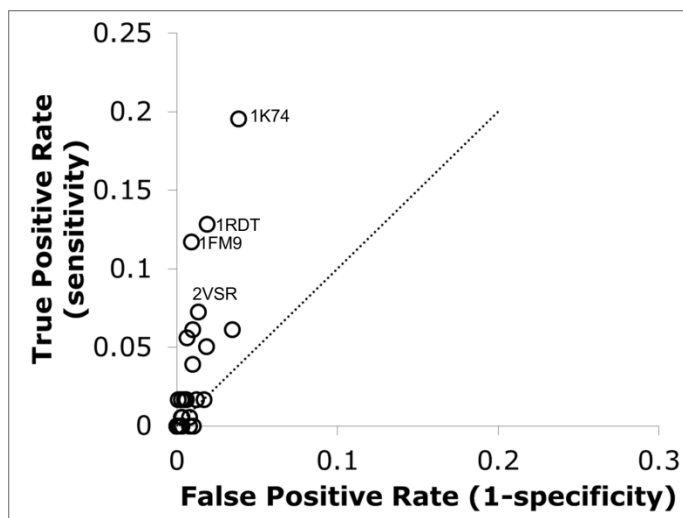
The ligand-based pharmacophore models generated from the PharmaGist feature sets were used to evaluate the training set of compounds to hypothesize structure model success. The predictability of each structure model either increased or decreased depending on the pharmacophore used for filtering in the previous test. For the eight pharmacophores used, all of the DUD known ligands matched at least one model, while the PDB and the DUD decoy lists contained some ligands that did not match any model (Table 4.7). The PDB ligand set possessed more ligands than the DUD active set. Including more ligands in the PharmaGist screening or adjusting the models based on structural information might recover the missing PDB compounds. For the immediate purposes of this study, the pharmacophores were sufficient for covering most of the training set and were not adjusted. Most of the training set matched the fatty acid and full agonist pharmacophores. This suggested the training set was best suited for assessing predictability of full agonist and fatty acid binding. For the remainder of the study, emphasis was placed on the full agonist and fatty acid agonist activity class to evaluate the success of the progress flow and identify representative model structures for virtual screening.

**Table 4.7** Counts of ligands that matched each ligand-based pharmacophore model from three lists of compounds: known actives from PDB crystal structures, known actives from the DUD database, and decoy compounds from the DUD database. The pharmacophore models used are listed alphabetically in the top row with the ligand type indicated in the following row.

	9HO	BRL	GW9	HXA	MYI	NSI	SF2	ZXG	Total ligands searched
	Fatty acid	Full	Antagonist	Fatty acid	Full	Antagonist	Partial	Partial	
PDB	53	49	5	82	65	16	24	51	99
DUD known	0	76	1	78	75	1	4	73	81
DUD decoy	0	1908	362	2564	1945	335	598	1910	2906

Inclusion of the decoy compounds allowed for assessment of false positive prediction. A single sensitivity and specificity value were calculated for each PDB structure used for the training phase given success or failure of docking determined by presence or absence, respectively, of the four hydrogen bond interactions. None of the structures did exceptionally well given the low TPRs. Despite the seemingly overall poor performance, some structures fared better than others for accurate prediction of true binders (Figure 4.4). 1K74 was the most successful at identifying true binders and neglecting true non-binders, which agrees with the RMSD-based TPR calculation. 1RDT was the next most successful, but did not exhibit as high a TPR for this step as it did in the RMSD-based TPR calculation. Both of these structures were in Group 2 of the grouped structure models, which indicated 1K74 would be better suited than 1RDT for full agonist and fatty acid agonist prediction. The representative structures for the other similarity-based structure groups was less clear given the relative low positive 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. The most successful structures 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. 1ZGY performed well in the TPR assessment analysis and possessed the same adjusted TPR as 1K74. This structure was also one of the most complete of the structures in Group 1 and was therefore selected as the representative structure for this group. 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 agonist binder prediction might be less successful with these structures. The 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.



**Figure 4.4** ROC space for 33 PDB structures used for docking. PDB IDs for the four structures with the highest TPR are labeled. Scales for axes were reduced to show data detail. The black dashed line indicates the division between the states (better prediction versus worse prediction).

#### 4.4.6 Screening of known drugs validated screening and scoring methods

As a means of validating the prediction methods, screening was performed with 2,319 established drug compounds. These compounds were docked into the four representative structure models and compared to the eight pharmacophore models. The interaction data and free energy of binding scores were used to rank each list, pharmacophore data was included for the models that were paired with each representative structure from the binding cavity comparison step. The top ten ranked binders for each structure model were evaluated totaling 40 compounds that were assessed for toxicity and novelty based on a patent and paper search for proposed binders. Data from the 1ZGY docking is shown here (Table 4.8). 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. Only the data for the first pose is shown in Table 4.8. The remaining data for the other three structures are in the supplemental information (Supplemental Tables 4.S2-4.S4). Some compounds appeared in more than one list suggesting some agreement in the docking results for all the structures. For instance, folic acid showed up as a hit in the 1ZGY, 1K74, and 2I4Z lists. The most negative free energy value was observed for the 1ZGY pose. Overall, the composition and order of hits was slightly different for all the lists. All of the hits matched at least one of the pharmacophore models proposed to fit the binding cavity except for some compounds docked into 2I4Z. Five of the ten hits in the 2I4Z list did not match either partial agonist model. It was proposed that the 2I4Z structure would be better suited for docking of partial agonists, but the full agonist interaction criteria was applied to identify binders. It is possible that the hits are an indication of additional pharmacophores to consider and/or a need for different interaction criteria to better identify binders with this structure model.

All of the compounds that appeared in the top-ranked list for 1ZGY either directly interact with PPAR $\gamma$  or function as a therapeutic in conjunction with other drugs for treating a PPAR-associated disease. Rosiglitazone and pioglitazone are well known PPAR $\gamma$  agonists. The presence of these compounds in the list is encouraging and immediately strengthens the predictability of the virtual screening process. Additionally, flufenamic acid is a non-steroidal anti-inflammatory drug that has been shown to activate PPAR $\gamma$  (205). The other four compounds

that appeared in patents were listed as compounds used in conjunction with PPAR $\gamma$  agonists. It was not clear in the patents if the compounds directly or indirectly activate PPAR $\gamma$ , but the addition of the compounds appeared to facilitate the activity of the agonists. Furthermore, the disease types listed in the patents included cancers, metabolic disorders, and inflammation-related diseases that have been linked to PPAR $\gamma$ -mediated regulatory processes. Only two of the compounds did not immediately return patent information for PPAR $\gamma$ -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 PPAR $\gamma$  and PPAR $\delta$  play a role (206). The literature suggests this compound directly binds to and activates PPAR $\gamma$  (207). The only unclear hit was ceftriaxone sodium trihydrate. No patents or published papers appeared to list the compound and PPAR $\gamma$ . This compound could be a promising hit, but the presence of toxic molecular groups could prove problematic for experiments beyond the *in vitro* stage.

**Table 4.8** Results from the docking of the MSUSDrugs compounds into 1ZGY for which docked poses were deemed successful. All compounds possessed interactions with at least the 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 MOE and the ToxAlerts server. The characters “+” and “-“ indicate if the compound matched or did not match, respectively, the 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 Alerts	Patent	M Y I	B R L	9 H O	H X A	Free Energy of Binding (kcal/mol)	Predicted Ki ( $\mu$ M)
18456289	Folic acid			EP1605950 A4	-	+	+	+	-9.8	0.066
01529323	methotrexate			EP1959950 A1	-	+	+	+	-9.6	0.092
01540998	pemetrexed disodium			US8362075 B2	-	+	+	+	-9.6	0.092
03956919	bezafibrate		yes	EP2089023 A2	-	-	+	+	-9.0	0.253
03952881	balsalazide disodium	yes	yes		-	-	+	+	-8.9	0.299
00968326	pioglitazone hydrochloride			10 different patents	-	+	-	-	-8.7	0.420
00086535	flufenamic acid		yes	EP2303252 A1	-	-	-	+	-8.5	0.588
35973845	ceftriaxone sodium trihydrate	yes	yes		-	-	-	+	-8.5	0.588
00968328	rosiglitazone		yes	US5002953 (expired 2012)	-	+	-	-	-8.4	0.696



## 4.5 Discussion

The prediction of PPAR $\gamma$  binders is a prominent area of research in recent years given therapeutic possibilities for numerous diseases including type 2 diabetes, inflammatory bowel disease, and colorectal cancer. Predicting binders encompasses use of various traditional bench-top methods like high-throughput screening and isothermal titration calorimetry, and newer computational methods like molecular docking and pharmacophore modeling. A combined approach of applied and theoretical methods is ideal for timely screening and identification of potential compounds for therapeutic development. Further development of virtual screening scoring and ranking is necessary to improve predictability 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 binder identification process. We have shown that adding pharmacophore and interaction filtering to the binder identification criteria improved predictability while reducing unnecessary screening of compounds that would not fit the models or should not bind given molecular composition.

In a 2004 review by Kitchen of docking and scoring methods, he mentioned considerations for improving the hit identification process (88). In the context of energy scoring, Kitchen suggested that scoring should derive from knowledge-based structural comparisons in which binding interactions are considered (88). 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 (88). Shape-dependent filters can be limiting as he states, but implementing 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.

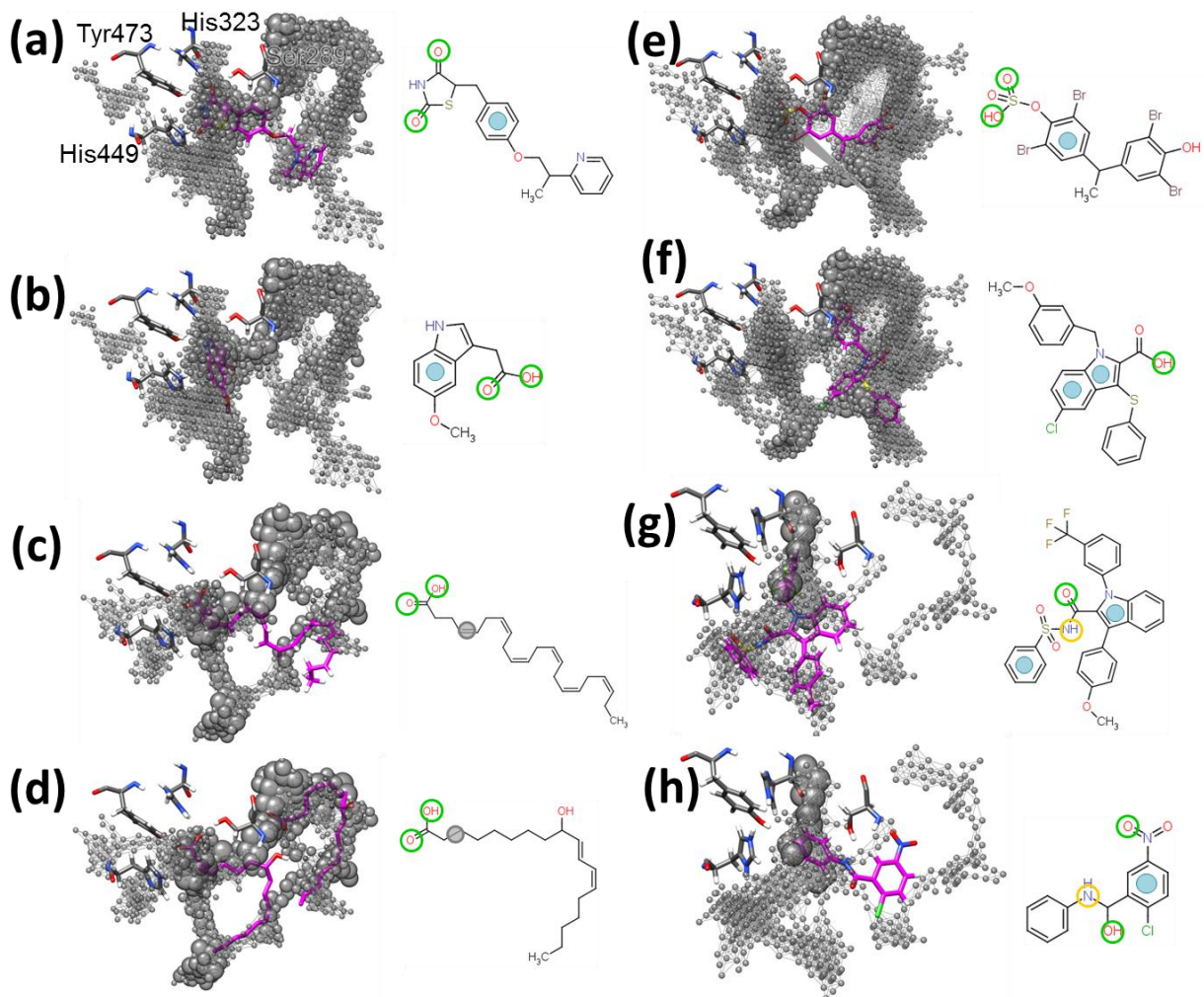
Few features were shared by all compounds of a single ligand type. Including more ligands might reduce common features while facilitating the need for multiple pharmacophores for a single ligand type. An examination of crystal structures suggests numerous interactions can arise with a given ligand and differences are observed between structures containing the same ligand. The promiscuity of PPAR $\gamma$  makes identifying all possible pharmacophores challenging but it has proven possible given what is known about binders.

Similarities among crystal structure binding cavities exist given all the structures considered for this study were of PPAR $\gamma$ , but ligand-induced differences were observed that required the use of multiple structure models for virtual screening. The subset of crystal structures used here to analyze binding cavity similarity was adequate to assess full agonist and fatty acid binding, but was not sufficient to assess partial agonists and antagonist binding. Partial agonist binding might be possible with additional analyses of key interactions and ligand pharmacophores. There are only three crystal structures for PPAR $\gamma$  that contain antagonists, two of which have the same ligand. More structures with bound antagonists are needed to better establish differences from

the other activity classes. Despite the lack of predictability of partial agonists and antagonists with this study, the information gathered for those activity types still serve as a comparison tool for identifying full agonists and a starting point for assessing partial agonist and antagonist binding.

An examination of the sample 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 4.5). 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 (41). 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 agrees with published crystal structure data suggesting that H3 is involved in partial agonism and is independent of H12 (54, 55). 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.

The differences seen in the pharmacophore models are beneficial to identifying activity class, and combining different models during assessment can strengthen any claims for a specific type. The full and partial agonist pharmacophore models suggest differences for the location of key features within the pocket necessary for binding. Combining the pharmacophore models would allow for distinguishing between full and partial agonists within arm I of the pocket. Compounds that satisfied the full agonist model and not the partial agonist models were most likely full agonists. Conversely, compounds that satisfy both partial agonist pharmacophores are more likely to be partial agonists versus compounds that satisfied one partial agonist and one full agonist model, or just the full agonist model.



**Figure 4.5** 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.

A limitation exists that pertains to the use of interactions as a filter for activity type. The limitation is lack of information about protein-ligand interactions that are truly indicative of a particular activity. Full agonist interaction information will only allow one to identify compounds that are potential full agonists. Within this activity type, there are variations that can further limit which agonists are identified 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 when the issue of trying to predict activity type is considered. Only full agonists can be identified with full agonist key interaction information. Therefore, the true limitation is how much information is known about the interactions that dictate the activity type. The full agonist key residues are well established; however, this is not the case for the other types of activity. A long list of potential interacting residues is known. 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. There may not be a clear distinction between subtype-specific full agonists, dual agonists, and pan agonists given all three categories should require association with the four hydrogen bonding residues. Nevertheless, there may be differences in which additional residues are necessary for other interactions that are shared between the PPAR subtypes. In this case, the residues would most likely share residue characteristics at key regions of the binding cavities but not necessarily the exact same residues. For all the additional agonism types, more analyses would be needed to identify specific residue lists for key interactions with each activity class.

Low success rates still exist for all the structures, but predictions were improved by filtering out unsuccessful and unnecessary docking. There is a need, however, 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.

Calculations could have been adjusted to account for the duplicate ligands. Adjustments would include removing duplicates, performing multiple RMSD calculations for the duplicate docked poses using the various reference structures, and post-processing evaluation of which duplicate poses were successful. It is the opinion of the authors that doing so may have introduced unintentional bias. The goal was an evaluation of predictability for multiple PDB structures with a level of reproducibility that contributes to enrichment of binder predictability. As the known ligand dataset contained duplicate copies of compounds, one would expect those compounds to dock the same because they should bind the same experimentally. As already suggested, the crystal structures indicated that variations do exist given a single compound. Ideally, the duplicates would be helpful to determine the predictability of all binding orientations for a given compound. It appeared that predictions of specific binding orientations were favored over others given the RMSD values for the duplicate compounds and the position of the docked poses within the binding cavity. Assessing the free energy of binding showed precision where RMSD values of the docked poses indicated success, and lack of precision in the energy values where RMSD indicated unsuccessful docking. This means variation in free energy of binding for duplicate

compounds suggested a lack of precision and reproducibility of favorable binding poses. Therefore, retaining or excluding structure models not suitable for docking required the inclusion of the duplicate compounds and strengthened the model selection process.

RMSD alone is not a useful metric for assessing structure predictability. Further, re- and cross-docking are not stringent enough to establish appropriate models for docking. RMSD values can be misleading, specifically in the instance where compounds contain ring structures that can orient in atom distinct but chemically indistinct positions (167). Brozell et al. provided an example of using intramolecular-distance based RMSD calculations to recover lost successful poses in the assessment of structure model docking success (167). This method does not help with assessment of docked poses for which no native crystal pose is known, as is the case when conducting virtual screening with a database of unknowns. Distances that indicated interactions retained a range of consistency regardless of which atoms of a ring structure oriented toward the residues of interest. Additionally, assessing presence of ligand features improved exclusion of non-binders.

Our analysis suggested pre-screening with ligand-based pharmacophore models and matching compounds with structures for which the binding cavity can accommodate key features is a useful method for reducing the computational time necessary to screen large, diverse compound databases for PPAR $\gamma$  binders. Furthermore, pharmacophore pre-screening may help with further classification of the type of agonism that a binder might possess. In such a case, the predicted interactions would be 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. Other factors to consider would include how well the features for the test compound align to the pharmacophore models and if compounds match more than one ligand model. With these considerations, it is feasible to significantly improve the predictability of potential therapeutics for difficult docking targets.

## **4.6 Acknowledgements**

The authors would like to thank Lera Brannan for her preliminary efforts with identifying ligand-based pharmacophore screening methods to use for this study.

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>

The authors would like to acknowledge the National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health (NIH) for funding this work (Grant number 1F31DK091186-01A1).

## 4.7 Supporting Information

**Table 4.S1** RMSD values for all tested PDB ligands relative to each of the 33 structure models tested during the training phase. PDB (PDBID) and ligand (LigID) identifiers are listed for each ligand. RMSD values were calculated with the crystal structure-derived pose of the respective ligand.

PDBID	LigID	1FM6	1FM9	1K74	1RDT	1WM0	1ZGY	2G0G	2G0H	2HFP	2P4Y	2POB	2Q5P	2Q5S	2VSR	3DZU	4PRG
1FM6	BRL	3.151	2.725	2.153	2.753	3.713	0.853	2.915	2.905	3.021	2.888	2.784	3.021	3.145	3.003	2.677	3.108
1FM9	570	3.249	3.163	1.618	1.666	3.489	3.238	3.207	3.302	2.177	3.262	1.389	3.021	3.099	3.008	3.07	2.99
1I7I	AZ2	1.462	1.52	1.483	1.682	3.002	2.924	3.545	2.456	2.674	2.809	2.845	3.052	2.596	2.578	3.088	2.834
1K74	544	3.189	1.513	2.969	1.848	2.911	3.079	3.121	3.219	3.216	3.254	1.432	2.918	3.017	3.215	2.816	2.92
1KNU	YPA	3.067	2.876	2.931	2.966	3.652	3.112	2.953	2.999	3.022	2.308	2.981	2.962	2.968	3.05	2.752	2.764
1NYX	DRF	3.302	2.955	2.858	2.777	2.808	3.219	2.714	3.212	3.414	2.92	2.144	3.129	2.364	2.867	2.785	2.85
1RDT	570	3.187	1.375	1.667	2.973	3.003	3.047	3.152	3.313	1.924	3.235	3.205	3.042	3.054	3.257	2.999	2.979
1WM0	PLB	2.233	3.086	2.918	2.117	0.954	1.957	2.732	3.164	2.189	2.82	1.423	2.924	2.871	2.844	1.102	2.849
1ZEO	C01	3.554	2.877	3.902	3.877	2.985	3.358	3.614	3.597	3.166	3.129	2.933	2.921	3.336	2.906	4.813	3.312
1ZGY	BRL	3.161	1.98	2.8	2.829	2.956	1.245	2.709	3.331	3.139	2.988	2.922	2.997	3.118	2.917	2.677	2.977
2ATH	3EA	3.191	2.795	2.687	2.693	2.696	2.943	2.899	3.223	3.053	2.775	3.121	2.902	2.772	3.077	2.777	2.922
2F4B	EHA	3.272	2.47	2.983	2.507	3.818	3.016	3.27	2.988	2.49	3.116	3.07	3.215	3.1	3.115	3.221	2.286
2FVJ	ROO	2.839	2.991	2.952	3.335	2.55	3.716	2.838	3.259	3.261	2.995	3.011	2.899	3.126	3.058	3.23	3.183
2G0G	SP0	3.572	3.436	3.224	2.769	3.24	3.454	0.919	0.958	2.729	3.166	2.746	2.899	3.239	3.079	3.251	3.19
2G0H	SP3	3.444	3.402	3.205	3.917	2.718	3.434	1.013	1.078	3.393	3.066	2.825	2.782	3.135	2.822	3.101	2.616
2GTK	208	2.895	2.704	2.714	2.685	3.707	2.825	2.644	3.193	2.907	2.851	2.709	1.642	2.368	1.526	2.349	2.967
2HFP	NSI	3.872	4.023	3.822	3.911	3.696	3.785	3.882	3.54	3.279	3.871	3.673	4.008	3.532	3.337	4.187	3.326
2HWQ	DRY	3.25	3.141	2.202	2.396	3.134	1.899	2.854	2.533	3.127	3.195	2.208	3.142	3.221	2.095	2.681	2.628
2HWR	DRD	3.201	1.737	3.008	2.232	3.806	3.077	3.209	2.558	3.073	3.071	3.131	3.046	3.251	3.161	2.676	2.37
2I4J	DRJ	3.365	3.309	2.997	4.155	3.137	1.47	3.073	2.866	2.674	2.787	2.197	2.616	2.681	3.264	2.879	3.394
2I4P	DRH	3.504	3.246	3.58	2.934	2.691	3.638	2.834	3.531	2.455	2.708	2.837	2.861	2.618	3.145	2.624	3.37
2I4Z	DRH	3.505	3.349	3.56	1.888	3.101	3.221	3.549	3.018	2.861	2.8	3.034	2.917	3.019	3.111	2.699	3.414
2OM9	AJA	2.59	2.654	2.659	4.474	2.823	2.792	2.605	2.66	3.143	3.159	2.524	3.148	1.927	0.744	2.824	3.113
2P4Y	C03	3.157	2.86	2.344	3.16	4.049	3.081	2.741	2.344	2.75	2.557	2.914	0.765	3.252	2.825	2.958	2.983
2POB	GW4	3.267	2.888	3.145	3.177	3.127	3.462	3.127	3.232	2.045	2.943	1.33	2.965	3.344	3.332	3.098	3.042
2PRG	BRL	3.1	3.367	2.793	2.767	3.012	3.194	2.798	3.259	3.154	2.923	2.899	2.486	3.056	2.76	2.465	2.9
2Q59	240	3.319	3.075	3.304	3.202	3.16	2.961	3.731	1.795	3.013	3.076	3.328	0.78	3.026	2.211	3.834	2.932
2Q5P	241	2.539	2.581	2.957	3.022	3.027	3.192	3.48	2.562	3.307	0.907	3.102	1.011	2.813	2.626	1.926	2.913

**Table 4.S1** continued.

PDBID	LigID	1FM6	1FM9	1K74	1RDT	1WM0	1ZGY	2G0G	2G0H	2HFP	2P4Y	2POB	2Q5P	2Q5S	2VSR	3DZU	4PRG
2Q5S	NZA	2.741	2.671	2.736	4.708	1.621	2.63	3.099	3.165	3.132	2.325	2.609	2.775	2.398	1.649	2.464	2.761
2Q61	SF1	2.739	2.872	2.694	2.504	1.648	2.903	3.053	2.982	3.345	2.355	2.524	2.699	0.887	1.669	2.418	1.434
2Q6R	SF2	2.741	2.96	2.215	2.6	1.673	2.802	3.195	3.25	3.187	2.288	2.789	2.773	2.779	2.613	2.366	1.538
2Q6S	PLB	1.909	3.071	2.763	1.278	0.801	2.932	2.956	2.596	2.248	2.795	1.796	2.956	2.915	2.96	1.225	2.867
2Q8S	L92	3.613	2.759	2.79	1.567	3.735	3.209	2.819	2.962	3.035	3.017	2.76	2.355	2.518	3.099	2.683	2.444
2VSR	9HO	3.58	3.426	3.51	3.155	2.768	2.812	2.463	2.244	3.901	2.764	3.167	2.821	2.593	2.336	2.619	3.661
2VST	243	2.689	3.028	2.912	3.035	3.772	3.085	2.438	2.629	3.438	2.626	3.062	2.432	2.253	2.844	2.818	2.908
2VV0	HXA	3.249	2.909	2.905	2.843	2.211	3.115	2.798	2.779	3.274	2.742	2.837	2.482	2.937	2.598	2.835	2.799
2VV1	4HD	3.21	2.657	2.722	2.537	2.946	3.345	2.705	3.029	3.2	3.134	2.675	2.492	2.758	2.568	3.272	2.64
2VV2	5HE	3.46	2.633	2.975	2.597	2.954	1.78	3.46	3.529	2.733	3.304	2.72	2.654	3.07	1.365	3.114	3.03
2VV3	4R8	3.208	2.903	2.421	2.935	3.022	3.183	3.25	3.358	2.71	3.068	4.235	2.373	3.122	1.871	3.181	2.722
2VV4	6OB	2.365	2.712	2.365	2.672	2.606	2.947	2.801	3.619	3.235	2.855	2.641	2.614	2.595	2.254	2.753	2.643
2VV4	6OC	3.357	2.769	2.765	2.751	2.935	3.342	2.903	2.66	2.936	3.162	2.804	2.989	2.94	2.839	2.889	3.349
2XKW	P1B	2.949	3.232	3.233	3.244	3.082	2.917	3.02	3.241	2.909	3.196	3.15	2.481	2.551	2.873	2.434	3.074
2XYJ	WLM	3.033	1.935	1.818	3.036	2.733	3.017	3.011	2.527	3.057	2.791	2.023	3.014	2.731	2.976	2.857	2.964
2XYW	08S	3.155	1.696	1.836	2.86	1.839	3.264	2.89	2.159	2.861	3.041	3.152	3.164	2.758	2.524	3.047	2.903
2XYX	Z00	3.088	2.764	2.721	2.771	2.66	3.126	2.876	3.129	3.179	2.743	3.035	2.687	3.077	2.625	3.34	2.985
2ZK1	PTG	3.755	2.761	2.606	2.627	3.733	3.602	2.809	2.713	2.79	2.883	2.721	2.869	3.068	2.449	3.062	3.478
2ZK2	PTG	3.677	2.678	2.71	2.712	2.397	2.939	2.566	2.593	3.108	2.866	2.449	2.572	2.344	2.478	2.934	3.499
2ZK3	OCX	3.687	2.883	2.227	2.894	3.005	3.464	3.046	3.125	2.897	2.917	3.05	2.773	2.921	2.126	3.183	3.054
2ZK4	OCR	2.725	2.696	2.57	2.973	2.776	2.72	2.739	2.715	3.074	2.696	3.003	2.859	2.724	2.766	2.618	2.709
2ZK5	NRO	2.44	2.902	2.661	2.937	2.053	2.99	2.916	2.791	2.676	3.005	3.098	2.75	2.827	2.908	2.522	2.951
2ZNO	S44	3.128	2.399	3.256	1.894	4.105	1.433	3.15	2.503	2.96	2.413	1.564	2.21	2.062	1.994	3.254	2.325
2ZVT	PTG	3.539	3.376	2.987	2.171	2.413	3.642	2.823	3.465	2.629	3.053	2.832	2.86	3.048	2.329	3.132	3.423
3ADS	IMN	3.268	3.659	2.209	3.616	3.217	3.384	3.894	3.794	3.506	3.331	3.242	3.078	3.275	3.183	3.631	3.429
3ADT	HID	1.036	2.304	2.355	2.394	3.139	2.039	2.106	2.077	2.928	3.448	3.594	2.99	2.84	2.686	2.893	3.582
3ADU	MYI	1.75	1.693	3.535	2.138	3.323	1.743	1.861	3.693	2.703	3.165	3.303	3.327	3.141	1.966	3.127	3.69
3ADV	SRO	3.27	2.75	2.696	2.674	2.819	2.414	3.084	2.779	3.127	2.343	2.855	2.664	2.706	2.172	2.416	3.473
3ADW	MYI	3.664	2.191	2.303	1.647	3.36	1.869	3.58	2.225	2.748	3.071	3.512	3.012	3.195	2.31	2.994	3.73
3ADW	OCR	3.118	3.367	3.316	2.659	2.879	3.201	3.207	2.144	3.416	2.284	3.022	2.284	2.872	2.495	3.456	3.184
3ADX	IMN	3.787	3.61	0.902	1.076	3.2	3.586	3.875	3.882	3.523	3.046	3.432	3.003	3.407	3.102	3.156	3.38
3ADX	NRO	2.539	2.646	2.872	2.762	1.753	3.231	2.963	2.985	3.385	3.03	3.009	2.713	2.471	2.799	2.772	2.863
3AN3	M7S	3.035	3.102	3.259	1.915	3.957	3.164	1.793	1.857	2.193	2.54	1.782	2.444	3.305	2.209	3.347	2.577

**Table 4.S1** continued.

PDBID	LigID	1FM6	1FM9	1K74	1RDT	1WM0	1ZGY	2G0G	2G0H	2HFP	2P4Y	2POB	2Q5P	2Q5S	2VSR	3DZU	4PRG
3AN4	M7R	3.192	3.048	3.274	2.63	3.858	3.142	1.885	3.206	3.141	2.666	1.826	3.15	2.751	2.341	2.432	2.641
3B0Q	MC5	2.889	3.4	1.084	2.665	2.754	3.426	2.582	3.358	3.206	2.827	2.932	2.939	3.038	3.171	2.515	2.876
3B0R	GW9	1.844	2.451	2.309	2.455	2.817	3.512	2.384	4.045	2.716	3.013	2.477	2.94	2.972	2.526	3.078	3.55
3B1M	KRC	3.229	2.906	3.607	2.447	1.767	1.488	2.822	4.751	2.36	1.829	2.621	2.952	2.179	1.614	2.871	2.123
3B3K	LRG	3.611	3.624	3.701	1.911	3.262	3.464	3.932	2.31	3.708	3.212	1.952	3.149	3.413	2.972	3.375	3.423
3BC5	ZAA	3.399	1.475	1.069	1.645	2.995	3.175	1.973	3.198	3.303	3.185	2.945	2.833	2.985	2.797	2.913	2.972
3CDP	YRG	3.632	1.824	1.724	3.386	3.123	3.232	2.042	2.324	2.678	3.537	1.714	3.329	3.456	2.885	2.605	3.561
3CDS	GRR	3.668	1.945	1.683	1.802	3.13	3.262	2.003	3.47	2.671	3.5	1.757	3.24	3.425	2.838	3.499	3.586
3CS8	BRL	3.475	2.735	2.167	2.792	2.39	1.344	2.615	2.917	3.132	2.86	2.817	2.993	3.164	2.877	2.952	3.064
3CWD	LNA	3.533	2.655	2.568	2.696	2.924	3.845	2.462	2.827	2.826	3.04	2.582	2.487	2.989	2.321	2.922	3.362
3CWD	LNB	3.485	2.764	2.684	2.811	2.815	3.957	2.56	2.942	2.915	3.037	2.676	2.328	3.002	2.185	3.029	3.314
3D6D	LRG	2.691	2.601	2.9	2.499	2.086	2.764	2.655	2.743	3.322	2.6	2.838	2.652	2.06	2.319	2.755	2.727
3DZU	PLB	2.741	1.306	2.725	3.132	1.111	2.961	2.514	3.194	2.235	2.795	1.804	2.972	2.921	2.994	1.095	2.885
3DZY	BRL	1.323	2.724	1.029	2.828	3.076	1.449	2.619	3.328	3.237	2.894	2.928	3.019	3.125	2.828	2.501	2.807
3E00	GW9	3.653	2.512	2.468	2.446	2.777	3.741	2.398	2.481	2.891	2.646	2.577	2.625	2.64	2.44	2.992	3.478
3ET0	ET0	3.894	2.202	2.272	1.703	3.291	3.31	2.331	3.774	3.182	3.456	3.322	3.46	3.15	2.611	2.979	3.641
3ET3	ET1	3.682	2.689	1.088	2.685	2.777	2.66	1.172	1.291	2.435	3.083	2.905	2.639	3.157	2.571	3.051	2.991
3FEJ	CTM	1.201	2.712	2.685	2.643	3.031	3.008	2.648	2.667	2.855	2.912	1.781	2.398	3.11	2.394	3.145	2.601
3FUR	Z12	2.832	2.653	2.653	2.728	2.646	3.003	2.849	3.324	3.055	2.929	2.641	2.553	2.847	2.995	2.989	2.838
3G8I	RO7	3.226	3.063	2.959	2.719	3.77	3.125	2.701	3.341	2.909	3.01	2.678	3.095	3.044	2.639	2.672	2.712
3G9E	RO7	2.866	2.893	2.71	2.921	2.448	3.139	2.682	4.175	2.553	2.877	2.801	3.053	2.233	2.446	2.614	2.716
3GBK	2PQ	3.28	3.177	2.831	3.106	3.883	3.144	2.877	3.227	3.111	3.115	3.068	2.879	3.283	3.398	2.76	2.545
3H0A	D30	2.982	2.47	2.364	2.552	2.169	2.03	3.137	3.249	3.187	2.5	3.11	2.173	3.125	1.569	2.678	1.872
3H00	DKD	3.476	2.684	2.71	2.733	3.147	3.57	2.654	3.102	3.722	3.478	2.779	3.32	3.439	2.817	3.67	3.244
3HOD	ZZH	3.655	2.662	1.779	1.816	3.447	3.322	2.674	3.308	2.238	3.531	1.93	3.39	3.469	3.001	3.575	3.442
3IA6	UNT	2.798	2.824	1.69	2.829	3.015	1.545	2.976	3.51	2.9	3.026	1.514	2.308	2.466	2.651	2.73	2.878
3K8S	Z27	2.78	3.069	1.41	1.882	2.891	3.064	2.328	2.234	2.989	2.709	2.824	2.833	2.986	2.739	2.817	2.966
3KMG	538	3.011	3.131	1.305	2.945	3.899	2.959	2.506	3.408	3.039	2.544	2.868	3.078	2.736	3.01	2.434	2.968
3LMP	CEK	2.582	2.023	3.634	2.436	3.724	2.26	1.744	4.777	1.501	1.67	2.344	2.903	1.865	1.829	3.218	2.58
3NOA	5BC	3.373	3.273	3.054	2.081	3.826	2.254	3.269	2.899	2.782	2.684	2.978	3.517	3.311	2.466	2.465	2.465
3OSI	XDH	3.243	2.929	2.403	1.296	2.302	3.286	2.598	2.599	3.217	1.187	2.314	2.241	2.3	2.332	2.418	3.073
3OSW	XDI	3.296	2.845	1.31	2.42	1.813	3.301	2.384	2.418	1.653	1.212	2.871	2.506	4.328	2.412	2.591	2.431
3PBA	ZXG	3.388	3.085	2.564	4.341	2.563	3.393	2.53	2.138	2.63	3.752	2.774	2.55	2.497	2.571	2.628	2.607



**Table 4.S1** continued.

PDBID	LigID	1FM6	1FM9	1K74	1RDT	1WM0	1ZGY	2G0G	2G0H	2HFP	2P4Y	2POB	2Q5P	2Q5S	2VSR	3DZU	4PRG
3SP6	IL2	3.325	3.146	3.121	1.714	2.364	3.055	1.659	3.111	2.558	2.504	2.261	2.858	2.879	2.408	3.046	2.887
4PRG	072	3.117	2.794	3.076	3.172	2.216	3.513	3.164	3.354	3.2	3.256	2.26	3.381	3.434	1.159	3.579	1.746

PDBID	LigID	2FVJ	2I4J	2I4P	2I4Z	2OM9	2PRG	2Q6R	2Q6S	2VST	2VV0	2VV1	2VV2	3B3K	3CDP	3CS8	3DZY	3E00
1FM6	BRL	2.83	3.21	3.417	2.8	3.094	2.196	3.018	2.899	2.859	2.945	2.807	2.219	3.071	3.477	2.957	3.101	3.354
1FM9	570	2.781	3.029	3.326	3.279	2.97	3.234	3.13	3.263	3.126	3.024	3.064	2.935	3.112	3.277	3.394	3.362	3.353
1I7I	AZ2	2.358	3.042	3.071	3.064	3.405	1.48	3.117	3.571	2.31	2.173	1.569	3.017	2.398	2.478	3.08	2.818	2.925
1K74	544	3.087	2.334	2.944	1.587	3.082	3.218	2.938	2.857	3.026	2.785	2.881	3.162	3.018	3.19	3.001	2.393	3.27
1KNU	YPA	2.304	3.212	2.499	3.174	3.05	1.523	2.746	2.487	2.432	3.043	2.888	2.956	2.719	2.779	2.966	2.208	3.128
1NYX	DRF	2.407	3.211	2.177	2.594	3.081	1.422	2.638	2.226	2.325	2.908	2.892	2.17	2.87	2.884	3.027	2.849	3.054
1RDT	570	3.13	2.935	3.021	2.724	3.039	3.217	3.046	3.52	2.853	1.88	3.069	2.814	3.093	3.261	3.385	2.393	3.27
1WM0	PLB	1.975	1.434	2.858	2.825	2.558	2.771	3.005	2.968	2.069	1.377	2.839	3.032	2.954	2.056	0.968	3.228	2.709
1ZEO	C01	2.959	3.373	3.581	3.445	3.533	3.009	3.082	2.966	2.829	2.904	2.912	3.485	2.825	2.947	3.4	3.477	3.34
1ZGY	BRL	2.957	3.376	3.321	2.311	3.317	2.221	2.988	2.885	2.822	2.953	2.833	2.914	2.912	3.102	3.001	3.034	3.345
2ATH	3EA	2.765	3.066	3.082	2.684	3.307	2.174	2.98	2.582	2.592	2.662	2.749	2.876	2.733	2.65	2.512	3.053	2.97
2F4B	EHA	3.064	2.21	2.525	3.224	3.005	1.931	3.214	4.16	3.105	3.301	2.98	3.009	3.106	3.032	2.312	3.131	2.639
2FVJ	RO0	0.557	3.275	2.886	2.89	3.037	2.554	3.102	4.189	3.025	3.366	3.074	3.02	3.026	2.593	3.382	3.488	1.452
2G0G	SP0	2.754	3.324	3.349	3.337	3.655	2.788	3.224	3.257	3.166	2.753	2.755	3.395	3.546	3.155	3.132	3.286	2.349
2G0H	SP3	2.891	3.278	2.75	2.752	3.491	3.367	3.256	3.597	2.948	2.755	2.862	3.338	3.028	2.992	3.325	3.32	3.004
2GTK	208	2.216	1.747	2.329	1.74	1.728	2.011	2.582	2.256	2.311	1.56	1.678	2.584	2.421	2.551	2.835	3.157	2.261
2HFP	NSI	3.975	3.884	3.896	3.86	3.289	3.326	3.587	3.336	3.449	4.055	3.391	3.304	3.911	3.394	3.811	3.89	3.017
2HWQ	DRY	2.88	2.268	2.605	3.146	2.476	3.295	3.238	3.039	2.847	3.24	3.183	3.167	2.959	2.766	3.368	3.139	3.239
2HWR	DRD	2.918	2.346	2.326	2.775	3.054	3.284	3.219	3.133	3.038	2.653	3.042	3.176	2.71	2.676	3.109	3.165	3.248
2I4J	DRJ	2.742	3.059	3.357	3.044	3.314	3.366	1.94	3.043	2.48	3.046	2.941	2.591	2.947	2.945	3.289	2.927	2.346
2I4P	DRH	2.729	3.072	2.825	3.021	3.218	2.526	2.903	3.012	2.326	3.028	2.051	2.873	2.882	2.999	3.296	3.561	2.553
2I4Z	DRH	2.773	3.458	3.391	3.041	3.295	2.916	2.943	3.736	2.425	3.053	2.666	2.888	2.859	2.985	3.311	2.941	2.636
2OM9	AJA	3.177	2.363	2.643	2.769	2.492	2.83	2.498	2.739	2.888	2.796	2.6	2.618	2.866	2.574	2.684	2.542	2.812
2P4Y	C03	1.053	3.009	2.977	3.021	3.106	2.396	2.965	4.31	0.986	1.127	3.026	3.076	2.799	1.109	2.852	3.013	2.625
2POB	GW4	3.087	2.781	3.006	3.262	2.978	3.204	3.328	3.934	2.819	2.086	3.372	2.856	3.18	3.275	3.364	3.431	3.323
2PRG	BRL	2.734	3.032	3.206	3.26	3.046	1.169	2.634	2.945	2.789	2.937	3.221	3.158	2.77	2.775	2.762	3.067	2.843
2Q59	240	1.601	2.888	3.261	2.89	3.077	2.881	3.252	2.673	0.969	0.891	2.242	2.3	1.079	3.023	3.001	3.278	1.207
2Q5P	241	1.142	3.015	0.834	3.163	2.835	2.332	2.565	3.006	1.075	0.769	3.016	2.449	3.123	1.167	2.65	3.174	2.585

**Table 4.S1** continued.

PDBID	LigID	2FVJ	2I4J	2I4P	2I4Z	20M9	2PRG	2Q6R	2Q6S	2VST	2VV0	2VV1	2VV2	3B3K	3CDP	3CS8	3DZY	3E00
2Q5S	NZA	2.786	1.623	1.678	2.742	2.719	2.81	2.397	2.707	2.358	2.645	1.602	2.489	2.23	2.554	2.563	2.375	2.505
2Q61	SF1	2.747	1.664	1.675	2.706	2.2	2.84	2.4	2.659	2.649	2.217	2.163	2.222	2.304	2.626	2.59	2.757	2.481
2Q6R	SF2	2.812	1.646	1.573	2.605	2.789	2.936	2.077	1.756	2.689	2.822	2.209	2.286	2.241	2.644	2.698	2.885	2.819
2Q6S	PLB	3.063	1.523	1.815	1.844	2.125	2.957	2.948	4.133	2.174	2.948	2.967	2.985	2.202	2.073	1.01	3.12	2.699
2Q8S	L92	2.877	2.985	3.054	1.751	3.094	2.594	3.125	2.662	2.665	2.997	2.771	2.424	2.726	2.948	2.753	2.981	3.025
2VSR	9HO	2.932	3.329	3.446	3.455	3.547	3.794	2.776	2.608	2.139	2.9	2.276	3.191	3.755	4.507	1.68	3.348	3.766
2VST	243	2.927	2.4	2.359	3.115	2.837	2.611	2.721	2.524	2.803	2.79	2.582	3.121	2.502	3.943	3.215	2.934	3.221
2VV0	HXA	2.743	3.041	3.233	2.947	3.09	3.151	2.714	2.211	2.756	2.196	2.923	1.738	2.757	2.461	2.245	3.021	2.852
2VV1	4HD	2.773	3.097	3.026	3.146	3.243	3.207	3.11	2.706	3.271	2.752	2.465	2.68	2.766	3.169	2.972	2.496	2.743
2VV2	5HE	2.673	3.179	2.87	3.122	2.991	2.004	2.989	2.708	3.054	2.426	2.733	2.941	2.589	2.737	3.108	2.907	2.514
2VV3	4R8	2.975	2.913	2.975	3.117	2.892	2.221	2.963	3.627	3.007	2.467	3.183	2.894	2.855	2.996	2.816	3.14	2.242
2VV4	6OB	2.853	3.165	2.982	2.967	3.031	3.445	1.386	4.017	2.84	2.881	3.625	2.121	2.764	2.667	2.82	2.582	2.763
2VV4	6OC	2.449	3.062	3.098	3.131	3.219	2.813	2.9	2.948	3.041	3.032	2.848	2.148	3.169	3.574	2.952	2.504	2.782
2XKW	P1B	3.194	2.88	2.866	2.832	2.441	2.946	2.41	2.503	3.037	3.291	2.616	2.862	1.858	2.033	1.514	3.12	3.175
2XYJ	WLM	2.451	2.982	3.119	3.111	3.172	2.872	2.72	3.783	2.675	3.13	2.054	2.961	3.047	3.082	2.943	2.826	2.848
2XYW	08S	2.925	2.809	3.02	1.916	3.054	3.031	2.602	2.715	2.591	2.021	2.868	2.099	3.157	3.112	3.307	2.833	2.622
2XYX	Z00	2.972	3.139	2.604	1.86	3.299	3.251	2.577	3.126	3.264	2.833	2.857	3.134	3.149	3.206	3.046	3.174	3.285
2ZK1	PTG	2.921	3.354	3.386	3.175	3.424	2.839	2.75	2.656	3.19	2.918	2.79	2.651	3.097	3.076	3.142	3.277	2.534
2ZK2	PTG	2.628	3.059	3.158	3.154	3.448	2.413	2.328	3.8	2.982	2.724	2.474	2.474	2.909	3.728	3.124	2.609	2.352
2ZK3	OCX	2.925	3.154	3.113	3.495	3.405	2.126	3.114	2.903	2.822	2.328	2.853	2.867	3.057	2.532	3.058	2.376	2.557
2ZK4	OCR	2.837	2.792	2.926	2.83	2.887	2.703	2.491	2.543	2.78	2.682	2.746	2.669	3.679	3.708	3.05	2.675	2.583
2ZK5	NRO	2.724	2.65	2.684	2.641	2.753	2.815	3.078	2.505	2.824	2.968	2.064	2.818	2.59	2.867	1.497	2.904	3.338
2ZNO	S44	2.9	3.043	3.202	2.009	2.036	1.33	2.107	4.202	3.16	2.192	2.118	1.495	3.181	3.164	2.935	3.409	2.826
2ZVT	PTG	2.361	2.963	3.082	3.105	3.233	2.779	3.352	2.384	3.092	2.823	2.828	2.843	3.129	3.068	2.97	2.988	3.214
3ADS	IMN	2.344	3.282	3.426	3.479	3.344	3.223	3.269	3.258	3.222	3.069	3.149	2.707	3.31	3.529	3.442	3.724	3.521
3ADT	HID	3.177	3.685	2.758	3.517	3.573	1.066	2.566	3.043	3.412	2.781	3.023	2.71	2.396	3.697	2.468	2.849	2.51
3ADU	MYI	2.082	3.56	3.617	3.298	3.688	1.792	2.859	3.288	3.683	3.386	3.252	2.941	2.068	3.68	2.215	3.599	2.011
3ADV	SRO	3.062	3.471	2.662	3.92	3.426	2.204	2.297	2.912	3.925	2.804	2.399	1.858	3.037	3.176	2.539	2.692	2.818
3ADW	MYI	3.396	3.455	3.402	3.492	3.576	2.251	2.855	3.32	3.58	3.395	3.182	2.688	2.16	3.591	2.133	3.227	2.298
3ADW	OCR	2.403	3.093	3.327	3.313	2.915	3.364	2.918	2.424	2.818	2.385	2.789	2.878	3.86	4.025	2.346	2.766	2.646
3ADX	IMN	2.22	3.399	3.452	3.396	3.423	3.351	2.93	3.19	3.356	2.801	2.906	3.374	3.007	2.879	3.386	3.353	2.236
3ADX	NRO	2.772	2.847	2.282	2.716	2.719	2.731	2.036	2.591	3.043	2.96	1.983	2.739	3.053	2.597	2.749	3.048	2.959
3AN3	M7S	3.101	3.248	3.26	2.258	2.358	2.453	2.507	3.82	2.69	1.607	2.431	3.16	3.158	3.149	2.896	3.165	3.132

**Table 4.S1** continued.

PDBID	LigID	2FVJ	2I4J	2I4P	2I4Z	20M9	2PRG	2Q6R	2Q6S	2VST	2VV0	2VV1	2VV2	3B3K	3CDP	3CS8	3DZY	3E00
3AN4	M7R	1.819	3.3	3.398	3.292	2.391	2.333	3.189	4.105	2.517	2.001	2.319	2.492	2.379	2.374	2.913	3.192	2.857
3B0Q	MC5	2.587	3.286	2.885	3.118	3.517	3.167	2.746	2.857	2.489	2.923	2.906	2.887	2.573	3.006	2.869	2.907	2.863
3B0R	GW9	2.062	3.447	3.454	3.544	3.404	2.953	3.132	3.016	2.942	2.879	2.242	2.489	2.961	2.94	3.145	2.47	2.544
3B1M	KRC	2.789	1.585	2.268	2.203	1.899	1.528	2.724	4.103	2.911	1.971	2.246	2.675	1.773	2.906	2.458	2.707	3.088
3B3K	LRG	2.083	3.452	3.33	3.137	3.556	3.221	3.249	3.295	3.391	3.104	3.167	3.079	3.351	3.436	3.261	3.432	1.831
3BC5	ZAA	2.361	3.022	2.997	3.123	2.919	2.873	3.178	3.764	3.085	2.856	1.945	2.777	2.844	2.974	3.161	3.03	2.432
3CDP	YRG	2.044	3.474	3.521	3.403	3.526	3.638	3.271	3.103	3.449	3.338	3.333	2.764	2.772	3.458	2.091	3.503	1.91
3CDS	GRR	2.027	3.458	3.539	3.437	3.544	3.32	3.312	3.115	3.506	2.957	1.766	2.775	2.755	3.438	3.023	3.45	1.931
3CS8	BRL	2.622	2.98	3.291	2.812	3.084	2.245	2.988	3.521	2.833	2.713	2.585	2.897	3.085	3.081	2.905	1.139	2.699
3CWD	LNA	2.571	3.174	3.221	3.28	3.535	2.404	2.369	2.834	3.027	2.383	2.376	2.791	3.118	3.259	2.742	3.08	3.13
3CWD	LNB	2.621	3.117	3.16	3.239	3.488	2.262	2.196	2.738	2.998	2.256	2.441	2.83	3.19	3.342	2.737	2.959	3.134
3D6D	LRG	2.555	2.723	2.776	4.287	2.869	2.95	2.723	2.015	2.759	2.496	2.608	2.551	2.526	2.708	2.775	2.55	2.976
3DZU	PLB	3.112	1.175	1.838	1.921	2.167	2.969	2.977	1.271	2.155	2.957	2.985	2.987	2.054	2.869	1.063	3.149	2.771
3DZY	BRL	2.715	2.479	2.857	3.325	3.2	2.268	2.998	2.86	2.85	2.922	2.811	2.477	2.944	3.109	2.96	3.08	3.093
3E00	GW9	2.314	2.926	2.948	2.94	3.42	2.546	2.739	2.716	2.899	2.787	2.552	2.555	2.949	2.571	2.871	1.564	2.405
3ET0	ET0	2.757	3.437	3.51	3.485	3.716	2.533	3.233	3.261	3.698	3.067	3.358	3.073	3.101	3.679	2.38	3.699	2.064
3ET3	ET1	2.497	3.432	3.476	3.407	3.356	2.726	3.047	2.771	3.061	2.683	2.702	2.644	3.063	2.952	3.305	3.456	3.185
3FEJ	CTM	2.726	3.04	2.544	2.481	2.499	1.61	2.222	3.826	2.768	2.928	2.973	2.445	2.837	2.786	2.756	1.251	2.962
3FUR	Z12	2.014	2.661	2.684	2.613	2.998	2.856	2.793	2.493	2.675	1.698	2.879	2.652	2.974	2.623	3.221	3.244	2.44
3G8I	RO7	2.767	2.307	3.203	2.716	2.908	2.606	2.932	2.997	2.901	2.906	3.101	2.451	2.432	2.712	2.906	3.752	3.035
3G9E	RO7	2.665	3.14	3.198	3.117	2.821	2.746	2.62	3.702	3.04	2.994	3.106	2.4	2.558	2.527	2.916	3.066	2.976
3GBK	2PQ	3.11	2.202	2.951	3.142	3.128	3.22	3.359	4.089	2.553	3.141	3.394	3.446	2.862	2.656	3.393	3.368	3.043
3H0A	D30	2.563	2.093	1.691	2.082	2.457	2.033	3.184	2.211	3.198	3.042	2.154	2.151	3.189	3.212	3.028	3.087	2.911
3H00	DKD	2.97	3.554	3.459	3.157	3.496	3.973	3.492	3.444	3.489	3.204	3.473	3.445	3.416	3.263	3.381	3.482	2.073
3HOD	ZZH	2.875	3.486	3.292	3.406	3.543	4.02	3.325	3.235	3.448	3.104	2.824	2.73	1.222	3.49	3.307	3.424	1.924
3IA6	UNT	2.884	2.325	2.411	1.758	2.813	2.627	3.179	2.633	2.688	3.019	2.328	2.305	2.895	2.997	2.875	1.102	3.121
3K8S	Z27	3.057	2.873	2.726	1.829	1.359	2.981	2.834	3.941	2.996	2.981	2.736	2.957	2.757	3.199	2.782	3.037	2.801
3KMG	538	2.372	1.537	2.283	2.931	1.564	2.979	2.947	3.761	2.298	1.455	2.792	3.026	2.721	2.812	3.191	3.024	2.569
3LMP	CEK	2.757	1.742	1.726	1.953	2.332	2.011	1.546	4.095	1.656	2.723	2.291	1.962	2.896	2.659	2.35	2.556	2.697
3NOA	5BC	3.162	2.262	2.402	2.662	2.527	3.184	2.697	3.18	2.627	3.121	2.405	3.098	3.105	2.599	2.586	3.688	2.974
3OSI	XDH	1.607	2.748	2.305	4.338	1.58	2.333	1.633	1.933	2.407	1.715	1.569	3.004	2.548	2.56	1.736	2.925	2.694
3OSW	XDI	2.571	2.419	1.328	2.381	3.105	3.213	2.266	1.959	4.42	2.334	1.313	2.414	2.19	2.565	2.391	3.067	2.366
3PBA	ZXG	0.899	2.607	2.635	2.595	2.595	2.594	2.976	3.794	2.419	2.574	2.582	3.125	2.561	2.36	3.21	3.298	2.582

**Table 4.S1** continued.

PDBID	LigID	2FVJ	2I4J	2I4P	2I4Z	20M9	2PRG	2Q6R	2Q6S	2VST	2VV0	2VV1	2VV2	3B3K	3CDP	3CS8	3DZY	3E00
3SP6	IL2	2.573	2.447	3.009	3.103	3.324	2.198	3.053	2.352	2.798	2.498	2.499	2.485	2.968	2.907	2.325	3.261	2.56
4PRG	072	2.554	3.245	1.018	2.649	3.323	3.293	3.145	2.372	3.442	3.301	2.703	3.314	3.16	2.264	3.238	2.839	3.208

**Table 4.S2** Scoring table for MSUSDrugs compounds docked into the 1K74 structure model. Only the top ten results that satisfied the sorting criteria are shown. Toxicity was evaluated by two different programs with an “X” to indicate the compound was reported as toxic. Ligand pharmacophore model screening was applied to assess potential activity. The marker “+” indicates the compound matched the model, while “-“ indicates a non-match. “Bind Free E” refers to the free energy of binding prediction calculated by AutoDock Vina. Inhibition constants (Ki) were calculated based on the predicted free energy values. Interactions are listed relative to residues within the binding cavity that are known to interaction with the positive control rosiglitazone. Predicted interactions were measured with cutoffs and represented as “x” for hydrogen bond, “o” for hydrophobic, “/” for the threshold value of 4.0 Å, and “-“ indicates no distances within the threshold were measured.

ZINC ID	Compound name	Tox. MOE	Tox Alerts	M Y I	B R L	Bind Free E (kcal/mol)	Ki (μM)	CYS .285. CB	HIS. 323. NE2	HIS. 449. NE2	ILE. 341. CG2	MET. 348. CE	MET. 364. CE	SER. 289. OG	TYR .473. OH
18456279	leucovorin calcium		X	-	+	-9.4	0.129	ooo// oo	x	x/	ooo	-	-	x	/x
06920406	methotrexate	X	X	-	+	-9.3	0.152	oooo o	x	xxx	/	-	oo	/	x/
18456289	folic acid		X	-	+	-9.3	0.152	oooo oo	x	x//	-	-	-	x	/x
01529323	methotrexate	X	X	-	+	-9	0.253	ooo/	xx	x	-	-	-	x	xx
18202555	leucovorin calcium		X	-	+	-9	0.253	o/o// /oo	x	x	o/o/	-	-	x	/x
03842753	trifluridine		X	-	+	-8.9	0.299	o	x	x	-	-	-	x	x
00156820	flumequine		X	-	+	-8.7	0.420	oooo	x	/	-	-	/	x	/x
03830400	cefamandole nafate		X	-	+	-8.7	0.420	oo//	x/	/	oo	o	-	xx	xx
03875368	nitrofurantoin	X	X	-	+	-8.4	0.696	/	x	x	-	-	-	x	/xxx
03927870	fludarabine phosphate	X	X	-	+	-8.4	0.696	o	x	x/	-	-	-	xx	x

**Table 4.S3** Scoring table for MSUSDrugs compounds docked into the 2I4J structure model. Only the top ten results that satisfied the sorting criteria are shown. Toxicity was evaluated by two different programs with an “X” to indicate the compound was reported as toxic. Ligand pharmacophore model screening was applied to assess potential activity. The marker “+” indicates the compound matched the model, while “-“ indicates a non-match. “Bind Free E” refers to the free energy of binding prediction calculated by AutoDock Vina. Inhibition constants (Ki) were calculated based on the predicted free energy values. Interactions are listed relative to residues within the binding cavity that are known to interaction with the positive control rosiglitazone. Predicted interactions were measured with cutoffs and represented as “x” for hydrogen bond, “o” for hydrophobic, “/” for the threshold value of 4.0 Å, and “-“ indicates no distances within the threshold were measured.

ZINC ID	Compound name	Tox MOE	Tox Alerts	9 H O	H X A	Bind Free E (kcal/mol)	Ki (μM)	CYS .285. CB	HIS. 323. NE2	HIS. 449. NE2	ILE. 341. CG2	MET. 348. CE	MET .364. CE	SER .289. OG	TYR .473. OH
03830429	cefoperazone		X	-	+	-10.9	0.010	o	/x/	x/	/oo	/	-	/x	x
30320646	nadide	X	X	+	+	-10.3	0.028	o	x	x	/	/	-	/xxx	x
12503222	famotidine		X	-	+	-6.9	8.755	/	x	x	-	-	-	xx	x
17860482	ascorbic acid			-	+	-5.6	78.550	/	/x	xx/	-	-	-	xx	x
01843030	allantoin		X	-	-	-5.4	110.090	-	x	x	-	-	-	/x/x	x
03831134	miglitol		X	-	+	-5.4	110.090	oo	x	/x	-	-	-	xx	/x
03860468	glucosamine hydrochloride		X	-	+	-5.4	110.090	-	x	x	-	-	-	x/	/
03860468	glucosamine hydrochloride		X	-	-	-5.4	110.090	-	x	x	-	-	-	x/	/
19364219	ethambutol hydrochloride		X	-	-	-5.4	110.090	ooo	x	/x/	-	-	-	xxx/	x
00113442	metronidazole	X		-	-	-5.2	154.294	oo	x	x	-	-	-	x	x

**Table 4.S4** Scoring table for MSUSDrugs compounds docked into the 2I4Z structure model. Only the top ten results that satisfied the sorting criteria are shown. Toxicity was evaluated by two different programs with an “X” to indicate the compound was reported as toxic. Ligand pharmacophore model screening was applied to assess potential activity. The marker “+” indicates the compound matched the model, while “-“ indicates a non-match. “Bind Free E” refers to the free energy of binding prediction calculated by AutoDock Vina. Inhibition constants (Ki) were calculated based on the predicted free energy values. Interactions are listed relative to residues within the binding cavity that are known to interaction with the positive control rosiglitazone. Predicted interactions were measured with cutoffs and represented as “x” for hydrogen bond, “o” for hydrophobic, “/” for the threshold value of 4.0 Å, and “-“ indicates no distances within the threshold were measured.

ZINC ID	Compound name	Tox MOE	Tox Alerts	S F 2	Z X G	Bind Free E (kcal/mol)	Ki (μM)	CYS .285. CB	HIS. 323. NE2	HIS. 449. NE2	ILE. 341. CG2	MET. 348. CE	MET. 364. CE	SER. 289. OG	TYR .473. OH
03830431	cefoperazone		X	-	+	-10.2	0.033	o	/x	x	o	-	-	x/	x
04474460	cholecalciferol			-	-	-9.6	0.092	oooo	x	x	oo	o	o	/	x
18456289	folic acid		X	-	+	-9.5	0.109	ooo	x	/x	o/	-	/o	xx	/
00538483	trazodone hydrochloride		X	-	-	-9.2	0.180	-	x	x	-	-	o	x	x
03978669	cholecalciferol			-	-	-9.2	0.180	oooo	x	x	o	-	o	//	x
03830398	cefamandole nafate		X	-	-	-9	0.253	/o	x	/x/	o	-	-	x	/x
03830490	cefuroxime axetil		X	-	-	-8.8	0.354	/o/o	x	//	-	-	o	x/xx	x
00002062	salsalate		X	-	+	-8.4	0.696	o/	x	x	-	-	-	/xxx/	/
03830487	cefuroxime sodium		X	-	+	-8.3	0.824	oo/	/	x/	//	-	-	xxx	x
05421253	sulfabenzamide	X	X	-	+	-8	1.367	o/o	/	x	-	-	oo	xx	/

## **5 Effects of activation state and agonism on the molecular dynamics of PPAR $\gamma$**

Stephanie N. Lewis<sup>1,2,3</sup>, Raquel Hontecillas<sup>1,3</sup> Josep Bassaganya-Riera<sup>1,3,4</sup>, and David R. Bevan<sup>1,2</sup>

<sup>1</sup>Genetics, Bioinformatics, and Computational Biology Program, Virginia Tech, Blacksburg, VA, USA

<sup>2</sup>Department of Biochemistry, Virginia Tech, Blacksburg, VA, USA

<sup>3</sup>Nutritional Immunology and Molecular Medicine Laboratory, Center for Modeling Immunity to Enteric Pathogens, Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA USA

<sup>4</sup>Department of Biomedical Sciences and Pathobiology, Virginia-Maryland College of Veterinary Medicine, Virginia Tech, Blacksburg, VA USA



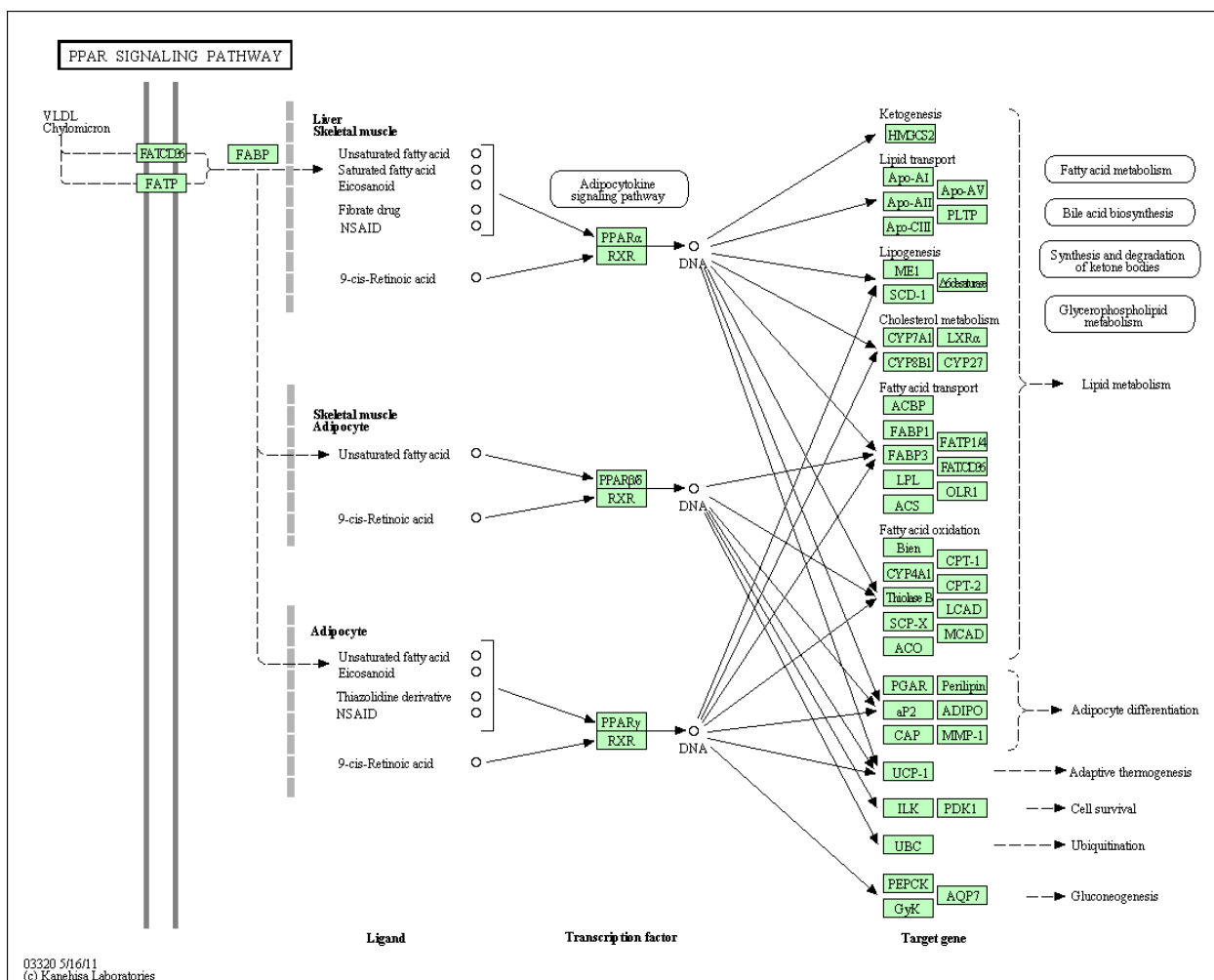
## 5.1 Abstract

Peroxisome proliferator-activated receptor-gamma (PPAR $\gamma$ ) is a nuclear receptor for which activation is mediated by ligand binding. The molecular composition of the ligand dictates conformational changes that lead to activation of the protein and co-activator recruitment for regulation of genes. Presently, little is known about the differences in dynamics between the active and inactive states, and the ligand bound and unbound states. In this study, we used molecular dynamics simulation to sample conformations for apo, active, inactive, unbound, and bound systems to evaluate differences in dynamics among systems. Differences in loop fluctuations appeared to dictate activation in both ligand-dependent and ligand-independent ways. The dynamics of the helix(H)-2'-H3, H2-strand1, and H9-H10 loops seemed to be affected most by the activation state. The presence of rosiglitazone appeared to have the greatest influence on protein dynamics. Additionally, steered molecular dynamics techniques were used to ascertain which amino acid residues contribute to binding of three different ligands. The interactions necessary for binding of full agonists and partial agonists varied according to key molecular features of each ligand. This study shows that unique residues necessary for binding of specific compounds can be identified. Such findings would improve drug discovery outcomes and facilitate prediction of downstream biological effects *in vivo*.

## 5.2 Introduction

Peroxisome proliferator-activated receptor-gamma (PPAR $\gamma$ ) is one of three subtypes responsible for metabolic, inflammatory, vascular, and cellular processes. PPARs are members of the nuclear hormone receptor (NHR) family of proteins that regulate expression of various genes. NHRs share structural similarity with sequence variations that govern ligand binding, co-regulator recruitment, and dimerization. The differences in sequence divide the NHRs into subfamilies, but the similarities contribute to dynamics shared by all NHRs. Teotico et al. proposed similarities in dynamics between pregnane X receptor (PXR), estrogen receptor- $\alpha$  (ER $\alpha$ ), and PPAR $\gamma$  using molecular dynamics (MD) simulations (208). The simulations suggested parallels in activation-function-2 (AF-2) domain dynamics for these receptors (208). AF-2 is the surface cleft onto which co-regulators associate through interactions with the LXXLL motif (209). This motif is common to NHRs and therefore presumes the correlated motion observed in the Teotico et al. study would be seen with other NHRs. As ligands are important for inducing conformational changes that prepare the AF-2 region for co-regulator recruitment, the effects mediated by ligand binding are of interest.

PPAR $\gamma$  was originally classified as an orphan receptor with unknown binding partners. The discovery that anti-diabetic thiazolidinediones (TZDs) target this protein influenced interest in and research of PPAR $\gamma$ . TZDs were classified as agonists because they activate rather than inhibit PPAR $\gamma$  and up-regulate biological processes mediated by the receptor, such as insulin sensitization and adipogenesis (2, 73, 210, 211). PPAR $\gamma$  agonism can be divided into four types: full, partial, dual, and pan. Full and partial agonists tend to be subtype-specific, while dual agonists bind to two PPAR subtypes, and pan agonists bind to all three. Dual and pan agonists can be of the full or partial activation class. The agonism type alters specificity for co-regulators, and subsequent recognition of certain PPAR response elements (PPREs) (211, 212) (Figure 5.1). Tissue type and the cell proteome also dictate specificity. PPAR $\gamma$  is predominantly expressed in adipose tissue (2). Full agonism in fat can turn on adipocyte differentiation, improve insulin sensitivity, and turn off inflammatory processes, whereas partial agonism might lead to expression of only insulin sensitizing genes. Additionally, studies have shown that different compounds that fall within the same activity class can regulate different sets of genes (212).

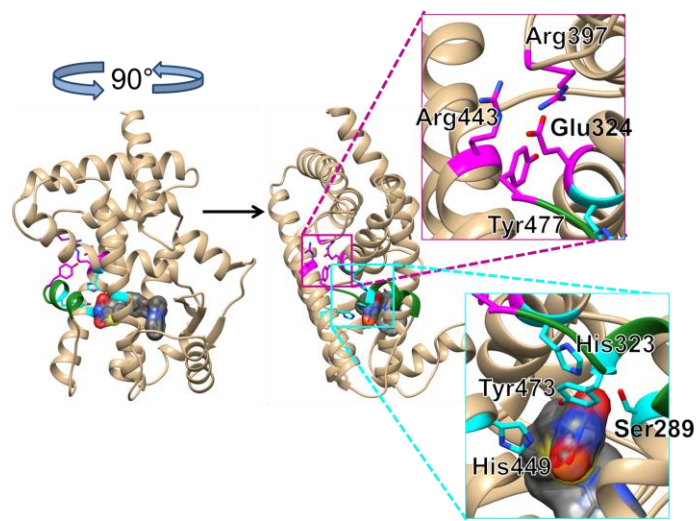


**Figure 5.1** KEGG pathway for PPARs (213, 214).

Over 100 crystal structures of PPAR $\gamma$ , usually in complex with a ligand, have been deposited in the RCSB Protein Data Bank (PDB) (59, 60). These static structures provide a wealth of knowledge regarding the interactions between ligands and PPAR $\gamma$ , but the dynamic differences that occur upon ligand binding are less clear. The literature suggests the active position of H12 is more stable than the inactive position (41, 52, 208, 215). However, PPAR $\gamma$  can oscillate between inactive and active conformations without a bound ligand. The binding of a ligand shifts conformational preference to the active state (41, 52, 208, 215). Full agonists can bind in H3-independent, H12-dependent manner while partial agonists bind in a H3-dependent, H12-independent mechanism. Each of these governs interactions with residues within the binding cavity to drive conformational changes.

Specific residues are involved in ligand-dependent and ligand-independent stabilization of tertiary structure (Figure 5.2). Residues Arg397, Arg443, Glu324, and Tyr477 are involved in ligand-independent stabilization of the AF-2 domain, which is composed of H12 and regions of H3, H3', and H4 (182). Residues Ser289, His323, His449, and Tyr473 are responsible for ligand-dependent stabilization of primarily H12, which directly influences the state of the AF-2 region (182). This is how differences in agonism

influence conformational variation of the AF-2 region and subsequent co-regulator recruitment.



**Figure 5.2** A 3D representation of the key residues responsible for ligand-independent and ligand-dependent stabilization of the AF-2 region. Rosiglitazone is shown in surface and stick representation within the binding cavity. Ligand-independent residues are colored magenta. Ligand-dependent residues are colored cyan. H12 is colored green.

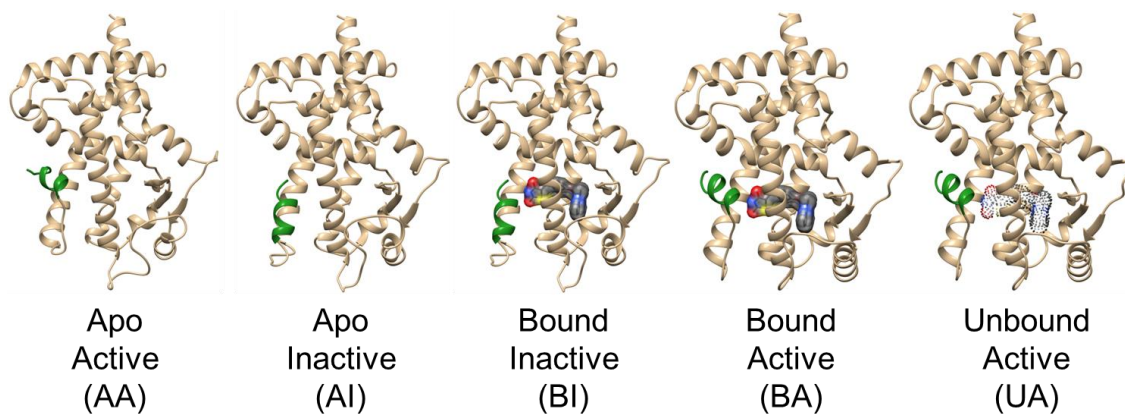
The combination of interactions that translate to AF-2 resurfacing can vary per agonism type, and even per ligand within an agonism type. Crystal structures provide some insight as to which interactions contribute to each activity type, but the large number of interactions recovered from crystal structure data and differences between those lists is problematic when attempting to distinguish between activity type. Further, differences in interaction information for the same ligand in multiple crystal structures suggests a shorter list of interactions are required for binding, while other interactions may be more a result of the binding event. For example, five rosiglitazone-containing structures each possess slightly different lists of interactions (45, 52, 56, 70, 81). A comparison of those lists narrows the list of common residues with which interactions may be necessary for binding to eight (Ile281, Gly284, Ser289, His323, Ile326, Ile341, His449, and Tyr473), but which are actually required for ligand binding and which are a consequence of binding is not clear.

The goal of this study was to apply MD simulation to examine the dynamics of apo, bound, unbound, active, and inactive PPAR $\gamma$  structures to ascertain differences between these states. This information, in conjunction with static structures of PPAR $\gamma$  in complex with ligands from experimental and computational studies, would indicate which regions of the protein are most influenced by the presence of a ligand in the binding cavity and the activation state of the protein. Given the similarity in tertiary structure between NHRs, this knowledge could contribute to agonist discovery for other orphan receptors. An additional goal was to ascertain key residues necessary for binding of different ligand types. Crystal structures provide some indication of the differences in interactions necessary for binding, but the specifics of which interactions are pivotal to each agonism type are less clear. Such knowledge would aid in classifying the activity of compounds that bind to PPAR $\gamma$ , thereby contributing to the drug development process and identification of novel therapeutics.

## 5.3 Methods

### 5.3.1 Structure preparation for molecular dynamics simulations

Two crystal structures from the RCSB PDB website (59, 60) were used for the simulations: 1PRG and 2PRG (52). The 1PRG structure is an apo, homodimer of the ligand-binding domain (LBD) of PPAR $\gamma$ , while 2PRG is a rosiglitazone-bound homodimer of the LBD complexed with a short peptide segment of steroid receptor coactivator-1 (SRC-1). As part of transcriptional regulation, PPAR $\gamma$  heterdimerizes with retinoid X receptor-alpha (RXR $\alpha$ ) (216). Therefore, the homodimer complex in the crystal structure is believed to be an artifact that arises from conservation of the dimerization interface for NHRs and exclusion of the RXR $\alpha$  protein from the crystallization process (216). Rosiglitazone is a known PPAR $\gamma$  agonist in the TZD compound group that has been used in many studies as a positive control (199, 211). Coordinates for crystal waters and the SRC-1 peptide were removed from the coordinate files. Simulations were done using a single subunit from the homodimers (Figure 5.3). Because 1PRG contains one subunit in the active conformation (chain A) and one subunit in the inactive conformation (chain B), based on the position of H12, each subunit was used in separate simulations for the apo-active (AA) and apo-inactive (AI) states. The 2PRG file contains two ligand-bound subunits, but chain B was missing atoms and hence excluded. The remaining chain from 2PRG (chain A) was used to represent the bound-active (BA) state, and removal of the ligand coordinates was used to create the unbound-active (UA) state. A bound-inactive (BI) state also was established by superimposing the AI and BA chains, and inserting the ligand coordinates for rosiglitazone from the BA chain into the AI chain. The placement of rosiglitazone within the binding cavity of the inactive form did not result in any unfavorable overlap of atoms. All of the above preparation steps were performed with UCSF Chimera (141).



**Figure 5.3** Ribbon representations of the five states for MD simulations. The apo structures, derived from structure 1PRG, did not have a ligand within the binding site of the crystal structure. The ligand in the bound systems (2PRG) was rosiglitazone, which is shown as sticks and an atom-colored surface. Helix 12 is shown in green for all structures to illustrate the placement of the helix in the active and inactive states. The unbound active (UA) structure is the same as the bound active (BA) structure but with the ligand removed. The area for the ligand is shown as a dot surface to indicate the ligand was removed from that position. Atom colors: carbon = gray, oxygen = red, nitrogen = blue, sulfur = yellow.

### 5.3.2 System equilibration

Simulations of the five systems were performed using GROMACS v4.5.4 (217) for system set-up and analysis. Each system was prepared in the same way, and three replicates were run for each system. A replicate was an independent simulation with a unique, randomly generated seed value for setting initial velocities. The protein or protein-ligand complexes were placed in a dodecahedral box filled with TIP3P water (24). A 1.0-nm distance was set between the solute and box edges to avoid periodic boundary artifacts. The AMBER03 force field (218, 219) was used to establish topologies for all the systems. Rosiglitazone parameters were built with Antechamber (220, 221). Five sodium ions were randomly substituted for water molecules within the box to balance the net charge of  $-5.0 e^-$  on the protein which established a net-neutral system. Table 5.1 lists the system size for each test system.

**Table 5.1** Systems established for MD simulations. Each system was assigned a two-letter abbreviation as defined in Figure 5.3, and each replicate was given a number (1 through 3). All systems contained five sodium ions that contributed to the total number of atoms.

Model	System	Number atoms	Simulation IDs	Waters	Total number of atoms
AA	Apo active	4,401	AA1, AA1, AA3	48,513	52,919
AI	Apo inactive	4,401	AI1, AI3, AI3	49,137	53,543
UA	Unbound active	4,422	UA1, UA2, UA3	39,123	43,550
BA	Bound active	4,466	BA1, BA3, BA3	39,072	43,543
BI	Bound inactive	4,445	BI1, BI2, BI3	47,376	51,826

Steepest descent energy minimization was performed using a step size of 0.01 nanometer (nm) with a maximum force cutoff of 10.0 kJ/mol/nm. The systems were then equilibrated in two phases during which the energy minimized systems were adjusted to biologically relevant conditions. To accomplish this, an isochoric-isothermal (NVT) ensemble was performed for 50,000 steps, followed by the isothermal-isobaric (NPT) ensemble for an additional 50,000 steps. In both ensembles, the time-step was set to 2 fs and the positions of the heavy atoms were restrained. Temperature for each system was increased to 300 K and regulated with velocity rescaling to dampen any significant temperature changes over the course of the simulation (222). The Parrinello-Rahman barostat was used to regulate pressure fluctuations (223, 224). The average pressure for the systems was 1 bar. Short-range cutoffs for neighbor searching atoms, the Particle-mesh Ewald (PME) summation cutoff for Coulombic interactions (225), and Lennard-Jones interactions were set to 1.0 nm.

### 5.3.3 Production simulation and analysis

After equilibration, unrestrained simulations were carried out for 100 ns per simulation. Root-mean-square deviation (RMSD) was calculated for each trajectory and the resulting curve was used as a qualitative evaluation of convergence. The 100 ns simulation time was deemed sufficient based on block averaging where the RMSD was averaged for successive windows of 10 ns from the end of the trajectory (i.e. last 10 ns, last 20 ns, last 30 ns, etc.). The resulting standard deviations were compared for all the blocks. Values less than two standard deviations (0.02 nm) for the last 50 ns were considered indicative

of convergence, and thus the last 50 ns of each trajectory was used for analyses. The Grace software package was used to generate 2D line art for the RMSD and RMSF data (226).

Analyses included root-mean-square fluctuation (RMSF) assessment and RMSD clustering. The cutoff for RMSD clustering was set to 0.2, 0.15, and 0.1 nm. Only clustering with the 0.1 nm cutoff resulted in more than one or two clusters of conformations. Conformations were clustered based on a neighbor searching method in which a cluster was established for a single conformation and all the conformations that satisfied the cutoff value were placed into that cluster. The remaining conformations were clustered similarly until all conformations were included in a cluster. The center, or median, conformation is the one with the most neighbors. The median conformation of the first cluster was used for analysis.

Distances between pairs of atoms over the course of the MD simulations also were analyzed as a way to compare conformational differences among states. A distance cutoff of 1.0 nm was assigned for assessing contacts between atom pairs.

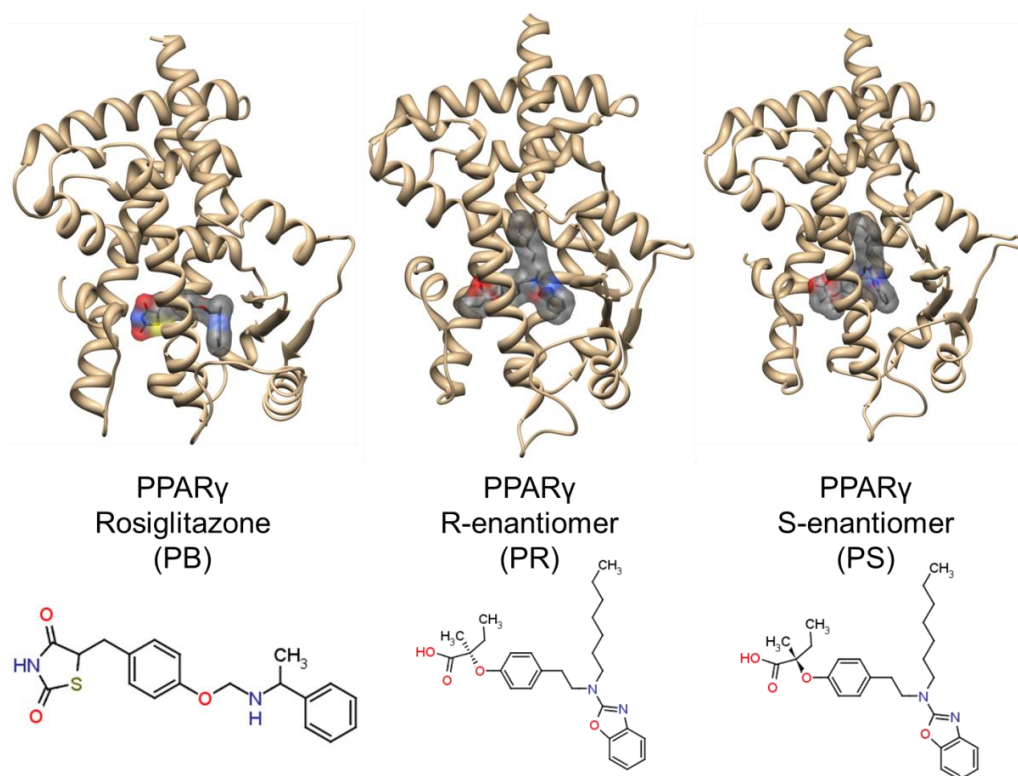
### 5.3.4 Principal components analysis

Atomic coordinates for 10,001 frames of each replicate were used to construct a covariance matrix. This matrix was used to calculate eigenvalues and eigenvectors, which correspond to the major magnitudes and directions of motion, respectively. The matrix allowed evaluation of correlated and anti-correlated motions that persisted during the trajectory. Additionally, eigenvalues indicated the percentage of motion detected by each component, while projections of the eigenvectors onto the trajectories illustrated the predominant motions for a given simulation. The first and second principal component projections of each simulation represented the major motions and were plotted against each other to identify clustering of sampled conformations.

### 5.3.5 Steered molecular dynamics

Steered molecular dynamics (SMD) is a computational method akin to atomic force microscopy in that forces are measured as one entity is separated from another under the influence of an external biasing force. In this case, forces were applied to the ligand to pull it from the binding cavity (227, 228). Force curves facilitate the identification of interactions that must be overcome in order for the ligand to exit the cavity. These interactions indicate which residues are potentially necessary for the binding process and stabilize the PPAR $\gamma$ -ligand complex in the active form. The assumption is made that pulling the ligand from the binding cavity should indicate the path of, and interactions necessary for, ligand binding (227). Three PDB structures were used for SMD: 2PRG (PPAR $\gamma$ -BRL; PB), 2I4J (54) (PPAR $\gamma$ -R-enantiomer; PR), and 2I4P (54) (PPAR $\gamma$ -S-enantiomer; PS) (Table 5.2 and Figure 5.4). As previously mentioned, 2PRG is a PPAR $\gamma$ -rosiglitazone complex, while 2I4J and 2I4P are complexes containing ligands that are stereoisomers. The 2I4J structure contains (2R)-2-[4-[2-(1,3-benzoxazol-2-yl-heptylamino)ethyl]phenoxy]-2-methyl-butanoic acid (R-enantiomer), which acts as a full agonist. The 2I4P structure contains (2S)-2-[4-[2-(1,3-benzoxazol-2-yl-heptylamino)ethyl]phenoxy]-2-methyl-butanoic acid (S-enantiomer), which acts as a partial agonist.





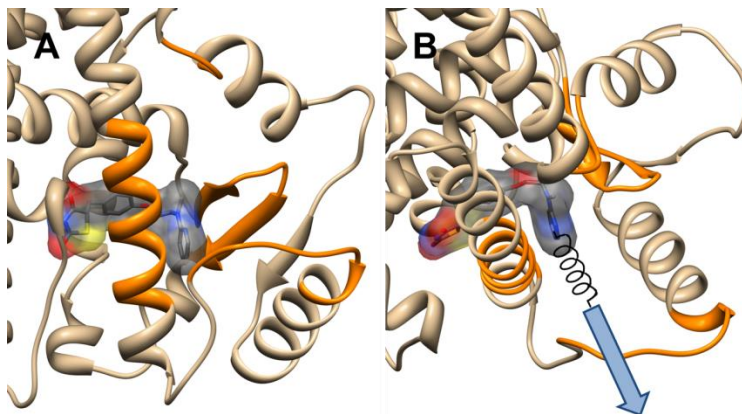
**Figure 5.4** Ribbon representations of the SMD systems. Ligands within the 3D structure are shown in stick and surface representation. Each ligand is represented in 2D under each 3D structure. Atom colors: carbon = gray, oxygen = red, nitrogen = blue, sulfur = yellow.

SMD was performed with GROMACS v4.0.7 (26). Complexes were inserted in a 12x12x15 nm rectangular box, which are considerably larger than those for the unrestrained simulations. These larger boxes allowed the ligand to exit the binding cavity without crossing a periodic boundary during pulling. A dummy atom of zero mass was attached by a virtual spring to the center of mass (COM) of the solvent-exposed portion of the ligand for pulling at a constant velocity. Only the solvent-exposed portion of the ligand was used because pulling from the COM of the whole ligand resulted in destabilization of the PPAR $\gamma$  structure. A pull vector was established for each complex that was positioned such that the ligand exited the cavity through an opening that was surrounded by H3, H5, the H2'-H3 loop and the s1-s2-s3  $\beta$ -sheet (Figure 5.5). The spring constant was set to 1000 kJ/mol·nm<sup>2</sup>, and the pull rate was set to 0.008 nm per second. These values were determined after several tests of different pull rates to find the ideal rate for the simulations without significant distortion of tertiary structure. Pulling simulations were run for 375,000 steps with a 2 fs time step to yield 750 ps of simulation time. This time frame was sufficient for the ligands to exit the binding cavity and be pulled far enough from the protein to ensure the force calculations dropped to near zero at the end of each simulation. The pulling was performed three times for each system. Temperature was set to 310 K and the Nose-Hoover thermostat was used to regulate temperature. The Parrinello-Rahman barostat was used to regulate pressure around 1 bar. Short-range cutoffs were set to 1.4 nm for the neighbor list, Coulomb, and van der Waals terms to account for residue rearrangement as the ligand was pulled from the cavity.



**Table 5.2** Systems established for SMD simulations. Each system was assigned a two-letter abbreviation as defined in Figure 5.4, and each replicate was given a number (1 through 3). All systems included five sodium ions.

Model	System	Number atoms	Simulation IDs	Waters	Total number of atoms
PB	PPAR $\gamma$ -rosiglitazone	4,466	PB1, PB2, PB3	209,631	214,102
PR	PPAR $\gamma$ -R-enantiomer	4,469	PR1, PR2, PR3	209,760	214,234
PS	PPAR $\gamma$ -S-enantiomer	4,469	PS1, PS2, PS3	209,646	214,120



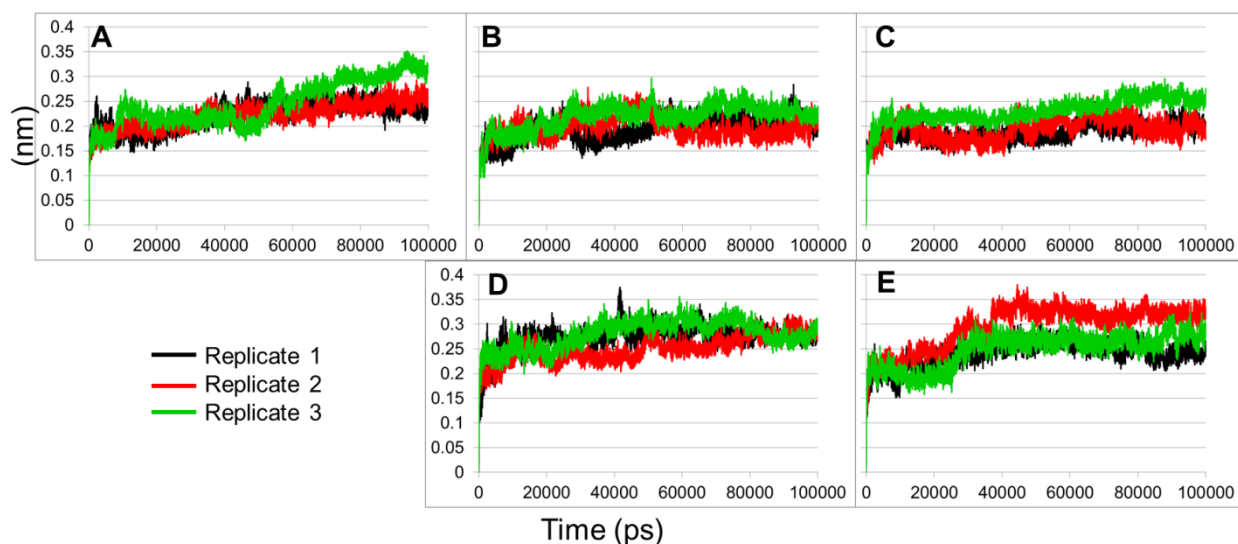
**Figure 5.5** Illustration of the origin of pulling for SMD simulations. The areas of the PPAR $\gamma$  structure through which each ligand was pulled are shown in orange. Panel B, which is rotated approximately 90 degrees, contains an arrow and spring that loosely illustrates the direction of pulling for the ligands.

Distances between atoms of the ligand and residues within the binding cavity were measured over the course of the pulling trajectory. Residues within 1.0 nm of the ligand as it exited the cavity were catalogued. Electrostatic and van der Waals energies between the ligand and the catalogued residues were calculated and averaged over the simulation time. Only the first twenty most negative energy values were considered for analysis.

## 5.4 Results

### 5.4.1 RMSD analysis: System stabilization

The RMSD analysis for all three replicates of each system indicated that the systems had converged within the criterion set for these simulations (Figure 5.6). That is, differences in block averages over the last 50 ns of each simulation fell within two standard deviations. Based on these results, the last 50 ns of each simulation was used for further analyses.



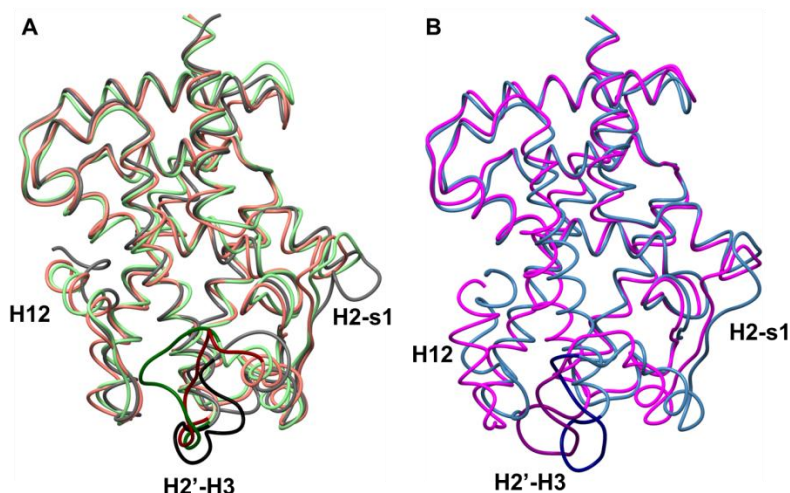
**Figure 5.6** RMSD plots over 100 ns for three replicates of each system. Running averages were calculated with a block length of 1000 points. The panels indicate RMSD for the (A) AA, (B) UA, (C) BA, (D) AI, and (E) BI systems.

### 5.4.2 RMSD analysis: Conformation clustering

Clustering of conformations from the last 50 ns of each simulation indicated minimal variation in the dynamics of individual replicates. The cutoff of 0.2 nm for distinct clusters produced one or two clusters for all replicates. A more stringent cutoff of 0.1 nm produced more clusters, and the first five clusters in most instances captured a large proportion of the sampled conformations (Table 5.3). The median conformation for the first cluster was taken as the representative coordinate set for each replicate. These frames were compared to identify structural differences among the systems.

**Table 5.3** RMSD clustering using a cutoff of 0.1 nm. Percentages for the top five clusters only are shown. The final row yields the total percentage of conformations contained in the top five clusters.

	AA			AI			UA			BA			BI		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	24.0	33.3	31.9	21.9	20.8	17.5	60.0	53.7	49.8	50.0	24.8	41.6	34.7	22.7	45.2
2	20.0	16.3	15.5	16.5	17.0	12.5	11.6	11.9	26.1	16.7	15.2	30.0	13.6	17.0	22.4
3	8.7	14.9	11.5	13.3	11.5	10.3	9.1	8.3	13.9	12.1	11.9	16.4	9.8	13.6	8.3
4	7.5	12.0	8.5	2.9	6.1	8.9	6.5	7.6	2.9	6.6	9.1	4.4	8.6	10.4	5.9
5	7.0	5.1	4.9	2.7	6.0	4.0	3.0	4.8	2.6	3.4	6.9	2.4	6.2	6.0	4.0
T	<b>67.2</b>	<b>81.6</b>	<b>72.3</b>	<b>57.3</b>	<b>61.4</b>	<b>53.2</b>	<b>90.2</b>	<b>86.3</b>	<b>95.3</b>	<b>88.8</b>	<b>67.9</b>	<b>94.8</b>	<b>72.9</b>	<b>69.7</b>	<b>85.8</b>



**Figure 5.7** Representative structures for each system taken from RMSD cluster one. The structures are the median member of the first cluster for the representative replicate. The active and inactive systems are compared separately. (a) Systems AA1 (gray), UA1 (green), and BA2 (coral) are superimposed. (b) Systems AI1 (blue) and BI1 (magenta) are superimposed. The H2'-H3 loop is highlighted in a darker version of the previously mentioned colors.

The representative structures from the first cluster for the UA and BA replicates were similar in conformation, which is indicated by RMSD values less than 0.2 nm between most structure pairs (Table 5.4). The UA systems appeared to predominately sample conformations similar to the BA systems given the small RMSD values. This was anticipated given the starting conformations for each set were the same with the exception of the presence or absence of the rosiglitazone. Visual inspection of the protein backbones for the first cluster representatives agreed with this (Figure 5.7A). The only noticeable difference between the UA and BA sets was the position of the H2'-H3 loop. This difference was most likely the result of the presence of the ligand in the BA set. Though an RMSD could not be reliably calculated between the AA set and the UA and BA sets due to the difference in the number of atoms, the cluster one structures were visually similar. An estimate of RMSD based on superposition of one structure over another suggested that the deviations from the AA representative structure to the UA and BA representative structures were both 0.11 nm. The only noticeable differences were in the H2'-H3 and H2-s1 loops.

**Table 5.4** RMSD values in nanometers (nm) for the center conformation of cluster one for each 2PRG-derived system. Only systems with the same number of atoms were compared.

	UA1	UA2	UA3	BA1	BA2
UA2	0.149				
UA3	0.174	0.171			
BA1	0.174	0.159	0.182		
BA2	0.180	0.162	0.203	0.153	
BA3	0.196	0.189	0.218	0.207	0.193

The cluster one structures for the AA set of simulations were similar to each other, but distinct from the AI and BI sets according to the larger RMSD values (Table 5.5). The most noticeable difference between the AA replicates and the other two is the location of

H12, which is in the inactive position for AI and BI (Figure 5.7B). Despite the addition of the rosiglitazone in set BI, H12 did not adopt the active position. This was most likely a result of the length of the simulation time. Noticeable differences in the location of H12 relative to H3 were observed for the AI and BI systems. The position of H12 was more consistent for the BI systems than for the AI set. It is understood that significant conformational rearrangement would be necessary for H12 to shift into the active position upon ligand binding. It stands to reason that the presence of rosiglitazone caused the difference in the position of H12 relative to H3 for the two sets. Next, the backbone fluctuations for each replicate were evaluated to determine specific regions of variations in protein dynamics.

**Table 5.5** RMSD values in nanometers (nm) for the middle member of cluster one for each 1PRG-derived system. Only systems with the same number of atoms were compared.

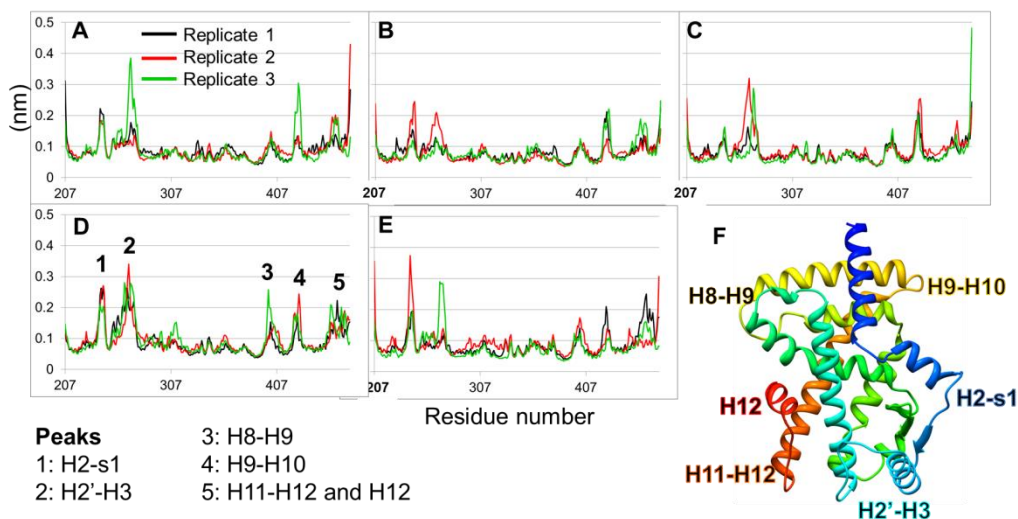
	AA1	AA2	AA3	AI1	AI2	AI3	BI1	BI2
AA2	0.190							
AA3	0.271	0.243						
AI1	0.269	0.267	0.301					
AI2	0.332	0.304	0.350	0.313				
AI3	0.305	0.272	0.245	0.256	0.297			
BI1	0.394	0.361	0.337	0.286	0.282	0.294		
BI2	0.429	0.399	0.444	0.389	0.268	0.366	0.292	
BI3	0.373	0.360	0.315	0.275	0.321	0.268	0.254	0.365

### 5.4.3 Loop regions showed noticeable fluctuations from initial conformations

Fluctuations in backbone atom positions over time revealed five regions where the structure deviated the most from the initial conformation in all the systems (Figure 5.8). These regions appeared in the H2-s1, H2'-H3, H8-H9, H9-H10, and H11-H12 loops, with some fluctuations occurring in H12. H12 showed less fluctuation in the simulations of the active systems than in those of the inactive systems. The H8-H9 peak was consistently low in the simulations of the active orientation, but appeared to be higher in the simulations of the inactive forms. The H8-H9 and H9-H10 loops showed opposing peaks for a portion of the simulations. If the H8-H9 loop showed a peak, then the H9-H10 region did not and vice versa. These two loops are positioned on opposite sides of the protein, and the observed fluctuations may be indicative of larger domain motions occurring on one side of the protein that influence fluctuations on the other side. The H2'-H3 and H9-H10 loops showed the highest degree of fluctuations in the most replicates. Though the H2'-H3 and H9-H10 loops lay at distal portions of the protein, it is possible that the motions of one region influenced the other as motion translates across the sandwiched helices. The H2-s1 loop showed fluctuations in all systems, but these fluctuations were less pronounced in the simulations of the active forms. This loop moved definitively less in the BA systems compared to the others. Overall, the state, either active or inactive, appeared to influence loop dynamics, but no ligand-based distinction could be made.

The minor dynamics inconsistencies for the loops in the replicates warranted further evaluation of the conformations sampled. It appeared that there may be some interplay in

loop dynamics, but how the combination of activation and bound states influenced these dynamics could not be determined. The nature of molecular dynamics presents a large amount of data, which can prove difficult to assess when trying to pull out subtle differences between the motions of key regions. Principal components analysis was the next step as it is a means by which the major motions for each replicate could be isolated and compared to aid in identifying system-dependent dynamic differences.



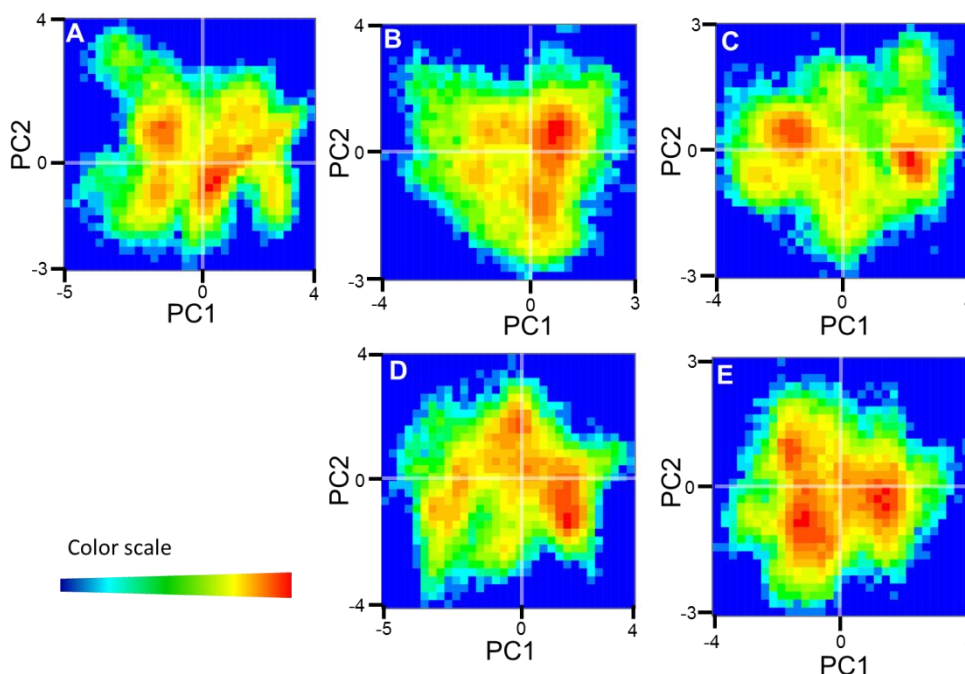
**Figure 5.8** Backbone RMSF for three replicates of each system. Panels show RMSF for the (A) AA, (B) UA, (C) BA, (D) AI, and (E) BI systems. A 3D representation of the PPAR $\gamma$  structures is shown in panel F as a reference for the RMSF peaks. Peaks on one RMSF curve are numbered to show where highlighted regions of the protein are located for all the curves.

#### 5.4.4 Principal components analysis

Principal components analysis was performed to examine global motion of PPAR $\gamma$ . The majority of the observed motions laid within the first and second principal components for each simulation (Table 5.6). Distinct clustering of the motions was observed for each system when the conformations sampled for the first two principal components were plotted against each other (Figure 5.9). At least two predominant clusters were identified for each system. Clusters were areas of several red pixels clustered together. Conformational sampling of all activation and bound states were seen for the various regions of the protein, but some systems indicated shifts toward sampling of specific states. The UA simulations showed conformations that shifted toward the bound state (Figure 5.9B), while the BA simulations showed conformations that shifted toward the active state (Figure 5.9C). The AI simulations indicated heavy sampling of the inactive state with a cluster of conformations occurring in the bound-active quadrant. The presence of the ligand in the BI simulations shifted the conformational sampling toward the unbound-active quadrant with less conformational sampling in the bound-inactive quadrant. What remained to be seen was which regions of the PPAR $\gamma$  structure contributed to favoring of one conformation versus another. This required an assessment of correlated and anti-correlated motion between different regions of the protein.

**Table 5.6** Percentage of motion possessed by each of the top five eigenvalues for the last 50 ns of each trajectory. The index value in the first column refers to the component of motion. The final row is the total percent of motion for the five components of each replicate.

	AA 1	AA 2	AA 3	UA 1	UA 2	UA 3	BA 1	BA 2	BA 3	AI1	AI2	AI3	BI1	BI2	BI3
1	15.0	26.2	30.6	18.7	17.0	17.8	16.4	20.1	29.0	25.3	28.8	22.4	19.4	23.5	18.4
2	12.4	11.0	11.9	10.4	9.1	14.9	7.6	14.1	8.4	11.2	12.9	15.2	10.1	11.9	9.0
3	8.1	7.0	7.2	8.4	6.9	5.3	6.8	9.4	6.6	6.2	5.8	5.9	8.3	8.1	8.8
4	5.6	4.4	5.1	6.9	4.5	4.9	5.8	5.6	4.0	5.6	4.3	5.2	4.7	3.7	4.9
5	4.8	4.1	3.8	3.6	3.8	3.9	4.7	3.6	3.4	3.7	3.3	4.0	3.9	3.6	4.4
	46.0	52.6	58.6	48.1	41.2	46.8	41.4	52.8	51.4	51.9	55.0	52.7	46.5	50.9	45.4



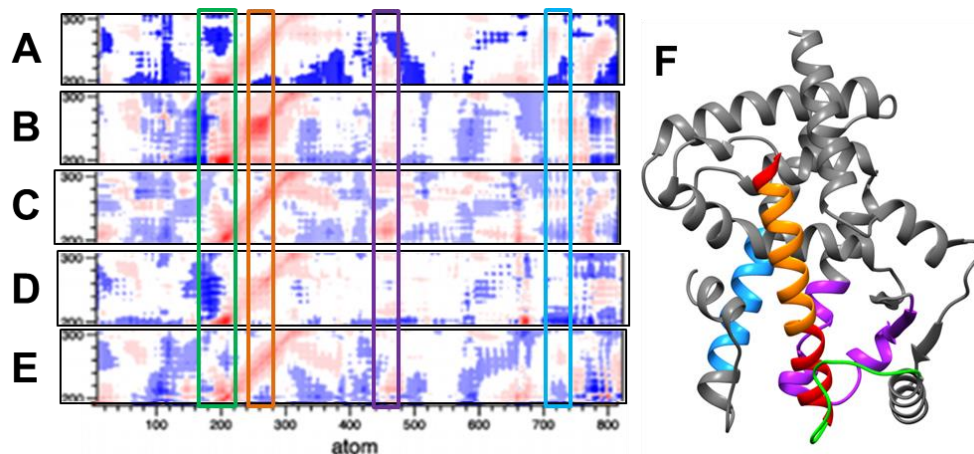
**Figure 5.9** Clustering of conformations within the first two principal components for each system. Graphs reflect all the recorded conformations sampled for the last 50 ns of three replicates. Frequency with which each conformation occurred is shown with the heat map scale where red is most frequent and blue is least frequent. Transparent white lines indicate the origin axes and establish quadrants for conformational sampling relative to activation and bound states. The top half of each graph corresponds to inactive motion, while the bottom half corresponds to active motion. The left half indicates unbound states, while the right half indicates bound states. Panels: (A) AA, (B) UA, (C) BA, (D) AI, and (E) BI systems.

An examination of the structure showed that H3 sits at the middle of the binding cavity and extends the length of the protein. Given the location of this region, all other portions of the protein should have some motion relative to it. Therefore, all regions of PPAR $\gamma$  were examined for motions relative to H3 (Figure 5.10). Here, only representative simulations are presented for ease of explanation. The major dynamics for all replicates of a given system were similar. The H2'-H3 loop (residues Lys265 to Glu276) exhibited correlated motion for the N-terminal half of H3 and anti-correlated motion for the rest in



the AA1 and BA2 system, but all correlated motion in the AI1 system (Figure 5.10, green rectangle). A shift from more correlated motion to more anti-correlated motion can be seen when the AI and BI segments were compared. The same could be seen from the UA1 state to the BA2 state. The area from s3 to the middle of H7 showed mostly independent motion, but appeared to move contrary to H3 in the AA1 system (Figure 5.10, purple rectangle). H11 showed an anti-correlated trend in motion with H3, but this motion was not as pronounced in the AI1 replicate. Overall, the AA1 and BA1 replicates showed more anti-correlated motion in the highlighted regions compared to the other systems. This data suggested that the H2'-H3 loop dynamics relative to H3 were dependent on the activation state and presence of a ligand. This loop moved with H3 in the AI system, but in a different direction relative to or independent of H3 in the other systems. Additionally, movement of H11 relative to H3 was anti-correlated in the active forms and mostly independent of H3 in the simulations of the inactive systems. Thus, it would appear H11 dynamics were more dependent on activation state than the presence of a ligand.

There are other regions that show greater correlated motion with H3. These include parts of H4 and H12. These segments make up the AF-2 domain that is the location for co-activator association. The movement of the regions around the binding cavity relative to H3 suggested conformational rearrangement given the activation state and presence of rosiglitazone. The presence of rosiglitazone influenced movement of the regions around the binding cavity either away from or against the motions of H3. Rosiglitazone is known to interact more with H12 than H3, and the simulations indicate that this interaction may be H3-independent. The portions of PPAR $\gamma$  away from the binding cavity show more correlated motion with H3, which may suggest concerted conformational rearrangement in this region to facilitate co-activator recruitment. Principal components analysis provided a global perspective of the major motions present in each system, but the existence of specific residue-residue interactions that govern activation was unclear. An analysis of contacts between residues that would designate interactions was needed.



**Figure 5.10** Segments of covariance graphs that correspond to H3 (plotted on Y-axis and shown in red and orange in panel F) relative to the rest of the PPAR $\gamma$  structure (X-axis). The rectangles indicate areas on the plots showing differences in correlated (red) and anti-correlated (blue) motion. The colors for the rectangles correspond to the colored regions in panel F: green = H2'-H3 loop (atoms 175-210), orange = H3 (atoms 211-288), purple = s3 to the middle of H7 (atoms 415-486), and blue = H11 (atoms 709-747). Graphs indicate motion for each atom of the backbone. The systems represented by each panels are (a) AA1, (b) AI1, (c) UA1, (d) BA2, and (e) BI1.

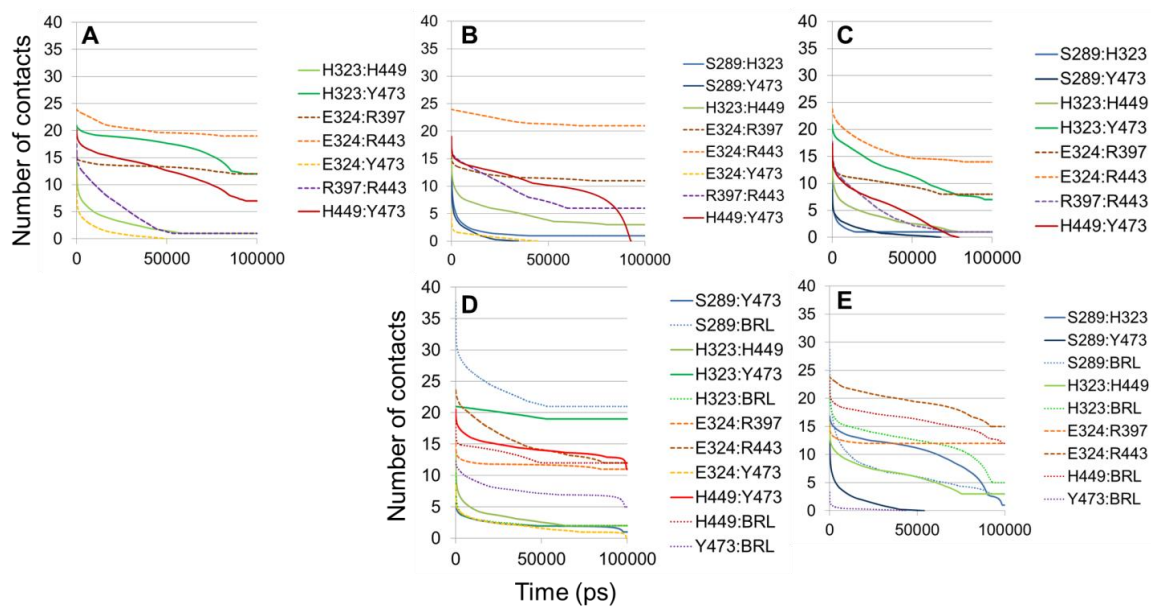
### 5.4.5 Presence of interactions over time

Crystal structure data indicate there are sets of interactions that should persist in the absence of ligand (ligand-independent) and interactions that should occur and persist in the presence of a ligand (ligand-dependent). It was hypothesized that combinations of these interactions should occur in the systems, and that the presence of certain interactions would be dictated by the activation state and presence of rosiglitazone. Ligand-independent interactions were present for all systems (Figure 5.11). In most cases interactions between Glu324 and Arg397 and between Glu324 and Arg443 persisted with minimal decrease in contacts over the course of the simulations. Less prominent ligand-independent interactions, such as Arg397-Arg443 and Glu324-Arg443, dropped off over time. The ligand-dependent interactions dropped off quickly for the AI systems (Figure 5.11B), with some persistence over time for the AA (Figure 5.11A) and UA systems (Figure 5.11C). Specifically, the H323-Y473 contacts were maintained over time for the AA and UA sets. This interaction appears to be dependent more on the activation state than the presence of a ligand.

The inclusion of rosiglitazone made a noticeable difference in the interactions present over the course of the simulations (Figures 5.11D and 5.11E). In both of the ligand-bound systems, additional ligand-dependent contacts were present and were maintained over the course of the trajectory. An assessment of interactions with rosiglitazone was included, which showed interactions between the ligand and key residues of the binding cavity. Thus, the simulations appropriately sampled ligand-bound conformations. Though H12 is in the inactive conformation for the BI set of simulations, ligand-dependent interactions did occur. It was previously mentioned that H12 moved away from the binding cavity in the BI trajectories, which is why an interaction between rosiglitazone and Tyr473 was not observed as it was in the BA simulations. It is also important to note that the overall number of contacts for the BI set was lower than those for the BA set. It is possible that there is an additional condition necessary for the ligand-dependent interactions to persist with stronger interactions in the BI simulations and for H12 to adopt the active position.

Overall, the most contacts that were consistent over the course of the simulation were seen in the BA set. This suggested the ligand-bound, active conformation possessed the most interactions that would contribute to stability of the active complex. The BI simulation followed next, suggesting the presence of the ligand contributed in some way to the stability of the protein despite the inactive conformation for H12. The ligand-dependent contacts did drop off over time, which indicated the position of H12, and thus the active conformation, is necessary to maintain the stability imparted by the ligand. The presence of some ligand-dependent interactions in the apo and unbound systems suggested that the structures were able to adopt some aspects of the active conformation without the ligand, which agrees with literature that suggests PPAR $\gamma$  can transition between the active and inactive states independent of a ligand (39, 43, 52).



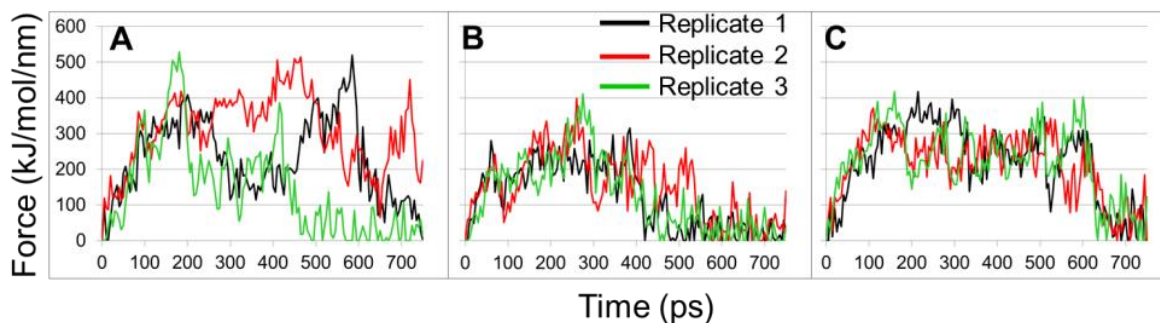


**Figure 5.11** Number of contacts within 1 nm of each pair of residues listed. Dotted lines indicate ligand-independent residue pairs, while solid lines indicate ligand-dependent residue pairs. Panels show the number of contacts for the (A) AA1, (B) AI1, (C) UA1, (D) BA2, and (E) BI1 replicates.

#### 5.4.6 Steered MD: Changes in force indicated differences in interactions

The MD analysis indicated that the dynamics of the loop regions were dependent to some extent on the activation state and the presence of a ligand. An assessment of surface dynamics provided hints to areas where protein-ligand interactions governed ligand binding, but a more detailed list of residues specific to binding was necessary to improve the ability to predict agonism. The pulling simulations provided a comparison of binding for three compounds that exhibit differences in activity due to differences in ligand molecular structure.

An examination of the force curves for the three pulling simulations for each system suggested differences in binding strength between the agonists (Figure 5.12). The force curves for rosiglitazone showed peaks that reached just over 500 kJ/mol/nm with definitive changes in force over time that suggested release of strong interactions (Figure 5.12A). Both enantiomer systems possessed force curves with broad major force peaks and valleys (Figure 5.12B and 5.12C). The presence and release of strong, ligand-dependent interactions was less clear in terms of the enantiomers and warranted a detailed residue-by-residue analysis of the energies that result from release of key binding interactions.

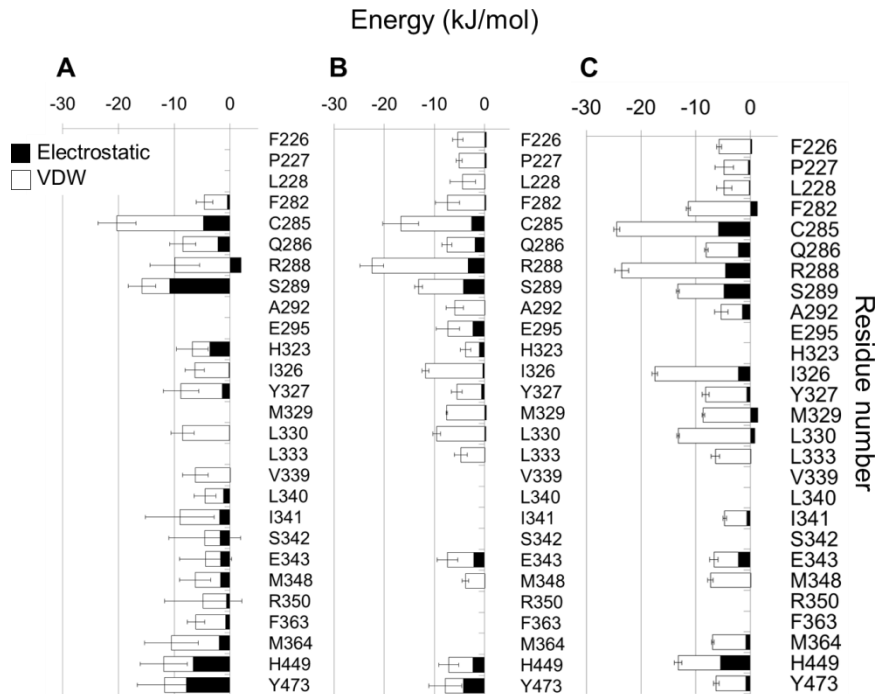


**Figure 5.12** Calculated forces over time for (A) rosiglitazone, (B) R-enantiomer, and (C) S-enantiomer as each was pulled from the binding cavity of respective PPAR $\gamma$  structures.

### 5.4.7 Persistent forces indicated key residues for ligand binding

The force curves indicated approximate time points for examining binding interactions. The average energy values for electrostatic and van der Waals terms over time indicated which residues interacted most strongly with each ligand (Figure 5.13). All three ligands showed large, negative energy values for Cys285, Arg288, and Ser289. Cys285 and Ser289, which are located on H3, occupy the middle of the binding cavity. H3 is positioned at the middle of the binding cavity, which means all three ligands sit close to, and potentially interact with, the residues on H3. The large energy values for all three compounds relative to Cys285 and Ser289 suggested these residues are involved in binding regardless of which ligand binds. Arg288 sits near the opening to the binding cavity. The van der Waals energy values for the enantiomer systems were higher than what was seen with rosiglitazone. This suggested that interactions with this residue were more important for the enantiomers than rosiglitazone.

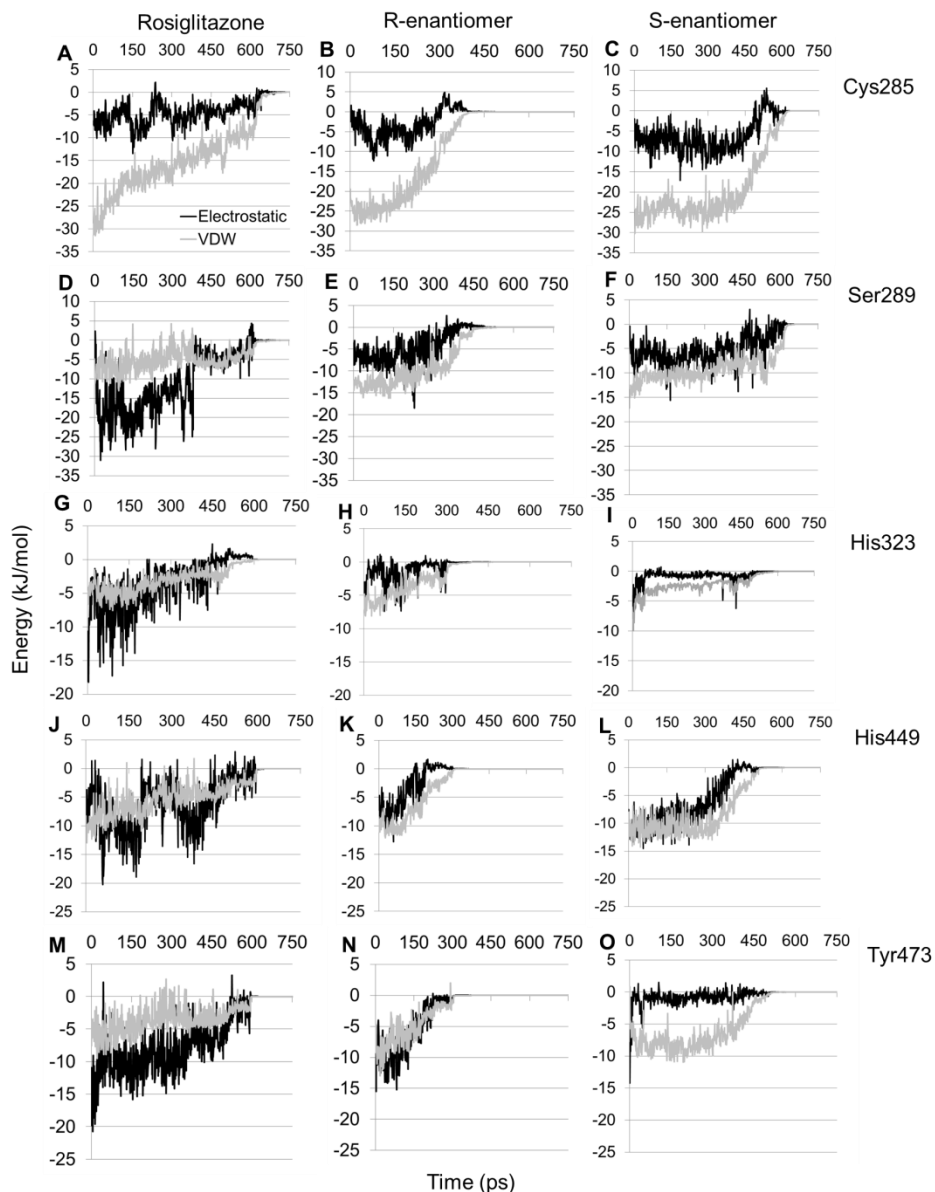
Interactions with the key hydrogen bond residues appeared to vary by ligand. The rosiglitazone simulations indicated highly negative electrostatic energy values for His449 and Tyr473, with some interactions between rosiglitazone and His323. The R-enantiomer registered noticeable energy values for these three residues, but the S-enantiomer pulling simulations did not suggest strong interactions with His323. Though both enantiomers showed energy values within the top 20 values reported for His449 and Tyr473, the partial agonist pulling suggested less energy was necessary to separate the ligand from Tyr473.



**Figure 5.13** Average total energies measured for interactions between the listed residues and ligands. The electrostatic (black bars) component and van der Waals (VDW; white bars) component of the energy values are shown for residues within 1.0 nm of a ligand as it was pulled from the binding cavity. Panels correspond to the pulling simulations for (a) rosiglitazone, (b) the R-enantiomer, and (c) the S-enantiomer. Error bars indicate standard deviation for each mean. The top 20 energy values are shown for each ligand. Residues with no energy values for a particular system indicate those for which the energies were not in the top 20 values.

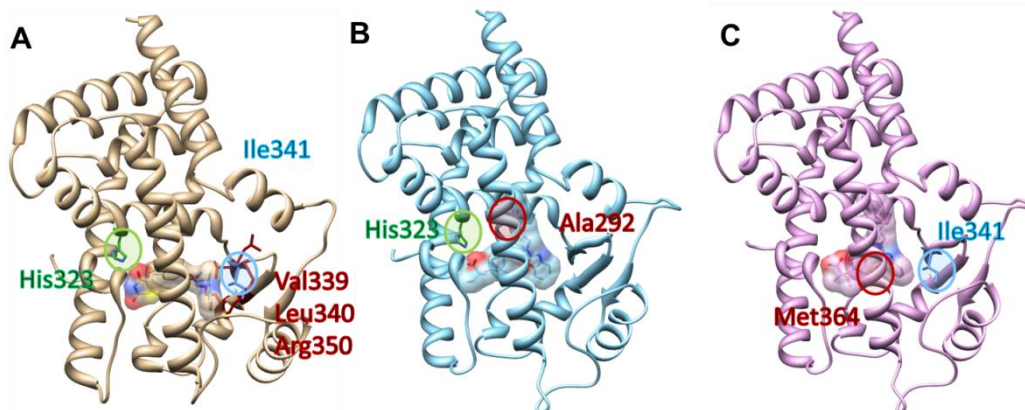
An examination of the average energy fluctuations over time for the residues with higher average energies indicated some residues showed stronger interactions with certain ligands than others (Figure 5.14). This analysis required examination of interactions with residues beyond the top 20 reported and made the details of which residues most likely contributed to holding the ligand within the binding cavity clearer. A drastic change in energy in a short time window indicated release of interactions between the residue and the ligand. All ligands showed persistent interactions with Cys285 (Figure 5.14A-5.14C) and His449 (Figure 5.14J-5.14L). A potentially stronger interaction was observed between rosiglitazone and Ser289 (Figure 5.14D) when compared to the enantiomers (Figure 5.14E and 5.14F). Energy values for rosiglitazone and His323 persisted over time (Figure 5.14G), which was not the case for the enantiomers. The energy terms for the R-enantiomer indicated the interaction between His323 and this ligand were quickly overcome during the pulling (Figure 5.14H). Interactions between His323 and the S-enantiomer were minimal given the small energies calculated with minimal fluctuations (Figure 5.14I). All ligands showed energy values for Tyr473, but the electrostatic component contributed more to the average total energy for rosiglitazone and the R-enantiomer (Figure 5.14M and 5.14N). The van der Waals portion of energy showed noticeable fluctuations for the S-enantiomer (Figure 5.14O). This difference suggested electrostatic interactions, most likely in the form of hydrogen bonds, are necessary for the full agonists to bind, but not the partial agonist. Taken together, this data indicated any interactions with Cys285, Ser289, His449, and Tyr473 do not necessarily depend on

activity type, but the strength and type of interaction does depend on activity type. Additionally, the His323-rosiglitazone interaction may be more specific to that ligand.



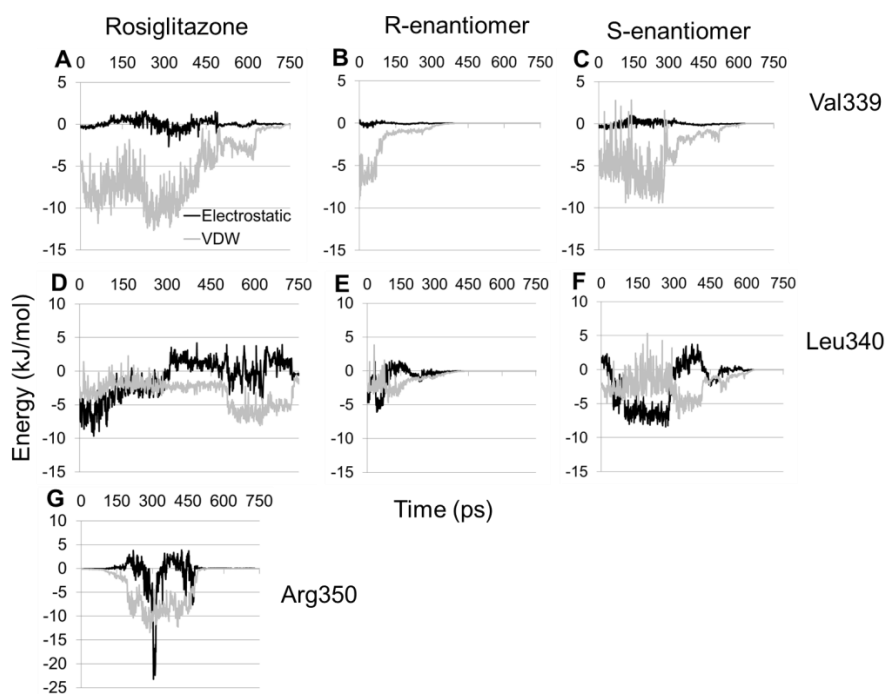
**Figure 5.14** Average interaction energy fluctuations over time for five residues relative to rosiglitazone (A, D, G, J, M), the R-enantiomer (B, E, H, K, N), and the S-enantiomer (C, F, I, L, O).

Most of the residues present in the energy terms lists were similar between the simulation sets, but some unique residues were observed for each list (Figure 5.15). More unique residue-ligand interactions were seen for the rosiglitazone pulling simulations (Figure 5.15A). Those residues were Val339, Leu340, and Arg350, which all sit near the opening of the binding cavity. The unique residues for the enantiomers were Ala292 on H3 for the full agonist R-enantiomer (Figure 5.15B) and Met364 on the H6'-H7 loop for the partial agonist S-enantiomer (Figure 5.15C). These residues sit on opposite sides of the binding cavity and may reflect interactions governed by the chiral differences for these ligands.



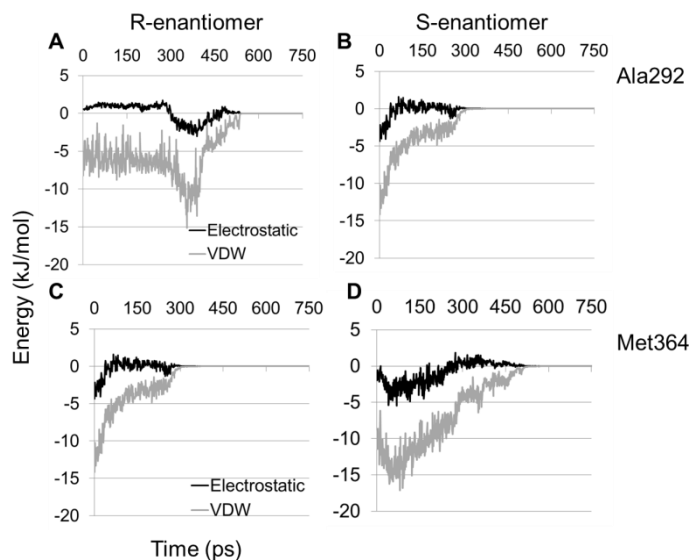
**Figure 5.15** Common and unique residues with large energy values over the course of the pulling simulations. Residues unique to each structure are highlighted in red while residues shared between two structures are shown in either blur or green to indicate which structures possessed shared residues. Circles indicate location of the highlighted residues. Panels correspond to the pulling simulations for (a) rosiglitazone, (b) the R-enantiomer, and (c) the S-enantiomer.

The residues unique to rosiglitazone did not show energy trends that indicated strong interactions for the enantiomers (Figure 5.16). Of the three, Arg350 was not within 1 nm of either enantiomer and therefore no energy values were calculated. The energy curves for Val339 and Leu340 showed persisting interactions with rosiglitazone (Figure 5.16A and 5.16D). This was not the case for the R-enantiomer (Figure 5.16B and 5.16E), but minor interactions may have been present for the S-enantiomer (Figure 5.16C and 5.16F). The energy curve for Arg350 appeared to reflect a strong, mostly Coulombic interaction with rosiglitazone.



**Figure 5.16** Average energies over time for three residues unique to rosiglitazone binding.

The residues unique to the enantiomers did not show energy trends for rosiglitazone (Figure 5.17) because these residues were not within 1 nm of this ligand during the pulling simulations. Ala292 appeared to have stronger interactions with the R-enantiomer than the S form (Figure 5.17A and 5.17B), while Met364 showed a more persistent interaction with the S-enantiomer than the R form (Figure 5.17C and 5.17D). The interaction data combined suggested similarity and differences in residues necessary for ligand binding. Table 5.7 indicates the residue associations deemed necessary for binding according to this data.



**Figure 5.17** Average energies over time for two residues unique to enantiomer binding.

**Table 5.7** List of residues with which interactions were purposed to be necessary for ligand binding. “Yes” and “No” indicate energy terms that either persisted or quickly dissipated, respectively. “Weak” indicates energy terms with less pronounced energy curves but potential interactions.

Residue	Rosiglitazone	R-enantiomer	S-enantiomer
Cys285	Yes	Yes	Yes
Arg288	No	Yes	Yes
Ser289	Yes	Weak	Weak
Ala292	No	Yes	No
His323	Yes	No	No
Val339	Yes	No	Weak
Leu340	Yes	Weak	Yes
Arg350	Yes	No	No
Met364	Yes	No	Yes
His449	Yes	Weak	Yes
Tyr473	Yes	Weak	Weak

## 5.5 Discussion

The MD data in our study meshes with data observed by others who have conducted simulations with PPAR $\gamma$  and other nuclear receptors. SMD simulations with other NHRs and each suggested multiple entry/exit pathways for ligand binding and unbinding (229, 230). Genest et al. performed pulling simulations with a partial agonist-bound PPAR $\gamma$  structure and identified at least three possible entry and exit points (231). All of these studies suggested ligands might enter the binding cavity on the H12 side of the protein. In the active conformation, the helix blocks access to the cavity (229-231). But this is not the case in the inactive form where the position of H12 away from the AF-2 domain leaves the binding cavity accessible to ligand entry (229). Perakyla suggested that the H12 surface on vitamin D receptors is the entry point for the inactive form, which closes off after activation and leaves the opening near the H2'-H3 loop as a solvent-accessible exit point (229). Our data showed motions of H12 away from the main body of the protein in the simulations of the inactive forms, which seems to agree with the idea of a H12-gated entry point in the inactive conformation.

In a 2008 study, Teotico et al. showed that the helices that make up the AF-2 domain of PXR, ER $\alpha$ , and PPAR $\gamma$  move in a correlated manner in the presence of ligand, and in an anti-correlated fashion when a ligand is not present (208). In the simulations of the active orientations, our data showed some correlated motion between areas of H3 and H12, but independent motion between H3 and the H3' to H4 region. More anti-correlated motion was seen in the simulations of the inactive forms for these regions. Although we didn't see strong indicators of correlated motion between the regions highlighted by Teotico et al., the observation of anti-correlated motion in the AI simulations with a transition to more independent motion in the BI simulations suggested congruence with their work. The shifts in conformational sampling suggested a concerted effort of helix rearrangement between the states. Some correlated motion may occur between the helices of the AF-2 region, but longer simulation time or a reduction in the contributions of the loop regions to the major components of motion might be required to see this activity. It is not clear which, if either, condition might contribute to the differences between our study and theirs, but we have established longer simulation times than the published work. Further, Teotico et al. suggested that H3 of PXR acts as a bridge between the AF-2 region and the  $\beta$ -sheet that sits near the cavity opening (208). Our data suggested connections in motion between the loop regions of PPAR $\gamma$ , and that H3 may serve as the mediator of long-distance loop motions across the LBD, which agrees with their conclusions.

The data presented here show that minor conformational differences translate to major differences in motion across the LBD. We have also shown that differences in conformational sampling can be achieved and observed over a relatively short simulation time. Longer simulation times would be beneficial, particularly for the BI system to see if any common conformations arise between systems. Simulations performed with larger systems would also be of interest. The active form of PPAR $\gamma$  is a part of a larger complex with RXR $\alpha$  and co-activator proteins. It is possible that some of the loop dynamics would change in the presence of these other proteins. Chandra et al. published a PDB structure in 2008 of the active RXR $\alpha$ -PPAR $\gamma$ -SRC-1 complex associated with DNA (56). The

structure indicates that the H2-s1 loop sits close to DNA and the H9-H10 loop sits above the DNA-binding domain of RXR $\alpha$ . The proximity of each of these could significantly affect the motion of these regions, and with this additional simulation data, the interplay between the loops might be better understood.

The SMD portion of this study has proven helpful for identifying specific residues necessary for ligand binding. Sets of residues have been proposed as advantageous to binding of each tested ligand. Additional analysis of the energy terms and more SMD simulations with different PPAR $\gamma$ -ligand complexes of varying activity types would facilitate building a more extensive list of key binding residues. The partial agonists in particular would be of great interest given the differences in binding that arise within this activity type. Studies have shown PPAR $\gamma$  partial agonists can mediate responses similar to those seen with TZDs while excluding downstream effects that can result in undesired side effects (210) and references therein). Recent work has suggested partial agonist fatty acids in particular can improve insulin sensitization without turning on the adipogenesis processes typically activated by TZDs that lead to excessive weight gain and perpetuation of the chronic inflammation state (210, 232). Additional pharmacophore analysis and MD work with an emphasis on naturally derived partial agonists would contribute to understanding the intricacies of partial agonism and subsequent dynamic changes.

Some insight to the agonism process is provided here, but very little is known about antagonism. Antagonists bind in such a way that H12 sits in an active-like position, but co-repressors are recruited instead of co-activators (56, 76). Therefore, the antagonists are capable of influencing H12 dynamics while maintaining an inactive PPAR $\gamma$  conformation. Few antagonists have been identified, which limits the information available on the binding patterns for this activity class. These compounds sit close to H12 similar to full agonists, but do not appear to interact with the remainder of the binding cavity, which the agonists do (56, 76). The dynamics sampled in this study would benefit from simulations of an antagonist-bound system to elucidate the dynamics that contribute to stabilization of the inactive form of PPAR $\gamma$ . As a whole, the study presented here serves as a first major step toward better understanding of the dynamics that direct differences in binding patterns for different activity types.

## **5.6 Acknowledgements**

The authors would like to thank Drs. Justin A. Lemkul and William J. Allen for their assistance with experimental design.

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>

The authors would like to acknowledge the National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health (NIH) for funding this work (Grant number 1F31DK091186-01A1).



## 6 Conclusions

Despite the wealth of knowledge available through the numerous publications on PPAR $\gamma$ , a gap still exists regarding the intricacies of the ligand binding process. The major issue with appropriately assessing binding is reconciling how binding drives the many roles that PPARs can play throughout the body. There is a level of specificity that is necessary for turning certain genes on and off. Tissue specificity and differences in expression patterns contribute to differential regulation, and in many cases, factors and proteins that are co-expressed with PPAR $\gamma$  also influence downstream effects (211).

The range of activity classes seen for PPAR $\gamma$  ligands further complicates the challenge of binder identification. The many processes that require regulation also require differences in ligand-induced conformational changes. Subtype-specific full agonists, partial agonists, dual agonists, pan agonists, and antagonists are individual classes of activity that result in differences in dynamic responses upon binding. The partial agonists as a class contain additional subclasses of activation that result in reduced co-activator recruitment relative to full agonists, and divergent regulation patterns. The structure-based pharmacophore modeling indicated that the partial-agonist containing crystal structures possessed more diverse binding cavities than the full agonist counterparts did. A variety of responses is seen experimentally with partial agonists, which seem to be directly influenced by the ligand. For example, the partial agonist KR-62776 results in differential expression of 42 genes compared to rosiglitazone (233). With these differences in expression, an increase in insulin sensitization without activation of adipogenesis processes can be seen due to the up-regulation of extracellular signal-regulated kinase (ERK)1/2 (233). Compare this to GW0072, which also increases insulin sensitization without increased adipogenesis, but does so by preventing adipocyte differentiation through selective recruitment of co-activators (87). As more studies have been done to determine the mechanism of action of TZDs, it has become clear that these compounds, which are believed to be full agonists, might be better termed partial agonists. The TZDs exhibit partial agonism in the selectivity of processes up-regulated and a level of partial antagonism in precluding binding of other agonists (211, 212). Endogenous ligands may be the only true PPAR $\gamma$  full agonists (211, 212). As the field of PPAR $\gamma$  drug discovery and development moves forward, there appears to be a need to move away from the TZD model of agonism and focus more on the fatty acid model. TZDs have served as a means to assess therapeutic potential. A new model that combines the mechanisms behind fatty acid binding with the highly efficacious gene regulatory and dynamic differences seen with synthetics may prove a better approach to therapeutic development. Alternatively, greater knowledge of the specifics of endogenous ligand binding with a focus on therapeutic development to mimic fatty acid binding would be the best means of maintaining the combination of regulatory processes that result in metabolic homeostasis. The VS process described here is poised to approach therapeutic development with a fatty acid-focused model that includes TZD-like efficacy assessment.

In a 2006 review, Klebe outlined the steps and considerations necessary for virtual screening with large databases (95). The nine major considerations explained in the study were target selection, druggability of the selected protein, target geometry, protein flexibility and adaptability, protonation state, assessment of protein “hot spots” for

binding, the role of water in ligand binding, ligand selection, and identification of binders through scoring, ranking, and validation methods (95). The preliminary steps of target selection (3, 6, 127, 163), druggability assessment (234), and “hot spots” identification (148, 199) have been considered and established for PPAR $\gamma$  by others. The outstanding issues were addressed in this project.

We have maintained and regularly updated an internal database of PPAR $\gamma$  crystal structures for docking and dynamics studies. PPAR $\gamma$  has been well established as a target for therapeutic development, and research has shown benefits of targeting this protein in various diseases (2, 130, 136, 235-237). The LBD is the primary location for small-molecule association, and the LBD contains a large binding cavity with broad affinity for various compounds, but with a propensity for fatty acid binding given the role of this protein in fatty acid sensing and regulation.

Addressing the issue of target geometry and protein flexibility can be handled by incorporating multiple structure models into the docking procedure. Rueda et al. performed a benchmark study using ensembles of crystal structures for binder identification (238). They concluded that docking with multiple structure models resulted in less uncertainty in the identification of binders compared to single model tests (238). They also suggested that apo structures should be excluded from the ensemble, and structures that contain a large ligand provide a better binding cavity for ligand position sampling (238). Their study did not include PPAR $\gamma$ , and other benchmarking studies have used a single PPAR $\gamma$  structure (169, 170). This study included aspects of the methods in these previous studies with PPAR $\gamma$  as a single target of interest.

The issue of establishing the protonation state for the binding cavity has been addressed by focusing on the protonation of residues within the binding cavity assuming the biological pH of 7.0. Crystal structure assessment indicated that water is not involved in interactions that result in full agonism and therefore water was excluded as a factor. The use of supercomputers to perform the screening reduced the immediate need for stringent ligand selection. Computer innovations have made screening of compound databases numbering in the tens of thousands possible in a feasible amount of time. Further, it is possible to use freely available software and scripting languages to perform VS. Supercomputing power can expedite the process, which can be obtained through federally funded resources made available to the research community. There are more tools available with commercial software packages that can improve efficiency and productivity. MOE (198), as an example, has proven helpful for pharmacophore searching, sorting, and modeling. Lastly, the use of pharmacophore filtering serves two purposes: as a means to address the scoring and ranking issue by pairing compounds with structures that can sterically fit them to improve the quality of predicted poses, and as a post-processing means of proposing activity based on what is known about existing PPAR $\gamma$  ligands.

When conducting virtual screening with a difficult target, one has to understand the limits of the data available. Screening for particular types of binders can be done given what is known about key interactions. Part of the screening process should pertain to matching the docking data to what is known about key interactions. Proposing binders can be done better in a setting where the different activity classes can be compared so that data pertaining to an undesired activity class can be excluded (239). Over the course of the

studies conducted here, it has become apparent that comparisons work better for assessing hits, as it is easier to identify binders if one also knows what should not bind or does not satisfy the desired criteria.

Free energy of binding alone is not sufficient for assessing docking success when screening diverse ligands. Free energy calculations for AutoDock4, and presumably other docking programs, are highly influenced by the number of atoms in the ligand. Arbitrary scores, such as those used in the DOCK 6 program, are only useful if the molecules being tested are molecularly similar. Molecule size may not heavily influence interaction information. Instead, interactions depend on the presence of key molecular groups with the characteristics necessary for binding. Molecule size may play a larger role in driving the process of binding and filling the cavity to promote interactions. Furthermore, it is possible to enrich the division of binders and non-binders given the characteristics of the screened ligands and the interactions that are presented based on those characteristics.

The need for screening to find alternative PPAR $\gamma$  agonists is pressing despite recent developments in the push to lift or reevaluate restrictions placed on Avandia®. The original 2007 meta-analysis performed by Nissen and Wolski to assess any risk associated with rosiglitazone use has come into question given a 2009 clinical trial designated RECORD funded by GlaxoSmithKline (240, 241). The meta-analysis suggested an increase in cardiovascular events, specifically myocardial infarction, in patients taking the rosiglitazone-containing medication (240). The RECORD trial indicated that no significant link between rosiglitazone and myocardial infarction could be determined, but there was some slight increase in risk for heart failure and potentially other cardiovascular events (241). Nissen and Wolski updated their meta-analysis in 2010 by combining the RECORD study data with data from other studies (242). The authors noted that in all cases, an increase in low-density lipoprotein cholesterol was found that could contribute to the cardiovascular issues and the overall risk for any cardiovascular problems was high (242). The authors concluded that the existence of other insulin sensitizing drugs with lower risk-benefits ratios for cardiovascular mortality warrants the continued restriction on rosiglitazone (242).

The larger concern that should be considered for PPAR $\gamma$  therapeutic discovery that was suggested by Nissen and Wolski's conclusions is the risk associated with currently available insulin sensitizing agents. The existence of risk that can lead to death or development of severe side effects validates the need for additional screening for PPAR $\gamma$  binders. Assessing novelty, toxicity, considering downstream effects, and including derivation of compounds from endogenous and fatty acid agonists may be the ideal approach for finding novel agonists with disease benefits and reduced risk. We have proposed focusing on natural compounds that serve as PPAR $\gamma$  agonists to introduce endogenous-like compounds to restore homeostasis. As observed in the study in Chapter 3, the efficacy of natural compounds can be lower than synthetic compounds because of differences in affinity or differences in therapeutic drug levels. Existing drugs can be repurposed, but patents can limit which compounds can be considered for therapeutic efficacy in alternate diseases. Additionally, there is a wealth of knowledge about toxic groups that can be screened against to aid in identifying compounds that might be less toxic and result in fewer severe side effects. A third avenue worth exploring is use of VS to identify multiple targets for established compounds, especially cases where side effects

are known. Often side effects are the result of a compound binding to other proteins and subsequently mediating responses other than those of interest. In the case of TZDs, affinity values, therapeutic blood levels, alternate binding partners, and activation patterns for these compounds contribute to the severe side effects seen in patients that take these medications (211, 243). Rosiglitazone may also play a role in the phosphorylation mechanisms that control PPAR $\gamma$  (243). Selected hits can be excluded or retained accordingly depending on the desired biological outcomes.

The original goal of the project that led to this dissertation work was to identify compounds that bind to PPAR $\gamma$  and pipe these potential binders into *in vitro* and pre-clinical experiments for efficacy in a T2D model. PPAR $\gamma$  as a screening target proved to be more complex and involved than originally imagined. As this project has unfolded, the challenges that arose have provoked interesting questions. Although the goal of finding a novel T2D therapeutic was not achieved, we have verified the effectiveness of our virtual screening process by identifying compounds that show benefits in a variety of chronic-inflammation associated diseases. The methods developed herein are a novel approach toward efficient identification of PPAR $\gamma$ -dependent therapeutics for treating chronic inflammation-related diseases.

## 7 Bibliography

1. Martin H. Role of PPAR-gamma in inflammation. Prospects for therapeutic intervention by food components. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2009;669:1-7.
2. Tontonoz P, Spiegelman BM. Fat and Beyond: The Diverse Biology of PPAR $\gamma$ . *Annual Review of Biochemistry*. 2008;77:289-312.
3. Guri AJ, Hontecillas R, Bassaganya-Riera J. Peroxisome proliferator-activated receptors: Bridging metabolic syndrome with molecular nutrition. *Clinical Nutrition*. 2006;25:871-85.
4. Bassaganya-Riera J, Ferrer G, Casagran O, Sanchez S, de Horna A, Duran E, et al. F4/80hiCCR2hi macrophage infiltration into the intra-abdominal fat worsens the severity of experimental IBD in obese mice and DSS colitis. e-SPEN, *European e-Journal of Clinical Nutrition and Metabolism*. 2009;4:e90-e7.
5. Hontecillas R, Diguardo M, Duran E, Orpi M, Bassaganya-Riera J. Catalpic acid decreases abdominal fat deposition, improves glucose homeostasis and upregulates PPAR  $\alpha$  expression in adipose tissue. *Clinical Nutrition*. 2008;27:764-72.
6. Bassaganya-Riera J, Hontecillas R, Beitz DC. Colonic anti-inflammatory mechanisms of conjugated linoleic acid. *Clinical Nutrition*. 2002;21(6):451-9.
7. Boussetta T, Raad H, Letteron P, Gougerot-Pocidallo M-A, Marie J-C, Driss F, et al. Punicic Acid a Conjugated Linolenic Acid Inhibits TNF $\alpha$ -Induced Neutrophil Hyperactivation and Protects from Experimental Colon Inflammation in Rats. *PLoS One*. 2009;4(7). doi: 10.1371/journal.pone.0006458.
8. Mohapatra SK, Guri AJ, Climent M, Vives C, Carbo A, Horne WT, et al. Immunoregulatory Actions of Epithelial Cell PPAR $\gamma$  at the Colonic Mucosa of Mice with Experimental Inflammatory Bowel Disease. *PLoS One*. 2010;5(4):e10215.
9. Ramakers JD, Verstege MI, Thuijls G, Te Velde AA, Mensink RP, Plat J. The PPAR $\gamma$  Agonist Rosiglitazone impairs colonic inflammation in mice with experimental colitis. *Journal of Clinical Immunology*. 2007;27(3):275-83.
10. Festa A, D'Agostino Jr. R, Tracy RP, Haffner SM. Elevated Levels of Acute-Phase Proteins and Plasminogen Activator Inhibitor-1 Predict the Development of Type 2 Diabetes; The Insulin Resistance Atherosclerosis Study. *Diabetes*. 2002;51(4):1131-7.
11. Bassaganya-Riera J, Guri A, King J, Hontecillas R. Peroxisome Proliferator-Activated Receptors: The Nutritionally Controlled Molecular Networks that Integrate Inflammation, Immunity and Metabolism. *Current Nutrition and Food Science*. 2005;1:179-87.
12. Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States. Atlanta, GA: U.S. Department of Health and Human Services; 2007.
13. Centers for Disease Control and Prevention. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States. Atlanta, GA: U.S. Department of Health and Human Services; 2011.

14. Lakatos P. Recent trends in the epidemiology of inflammatory bowel diseases: up or down? *World Journal of Gastroenterology*. 2006;12(38):6102-8.
15. Loftus EJ. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology*. 2004;126(6):1504-17.
16. Centers for Disease Control and Prevention. Inflammatory Bowel Disease 2007 [updated July 15, 2011; cited 2013]. Available from: <http://www.cdc.gov/ibd/>.
17. Butcher L. Digestive diseases: Epidemiology, Economics and the Pipeline. *Biotechnology and Healthcare*. 2008;5(4):12-3.
18. Guilherme A, Virbasius JV, Puri V, Czech MP. Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. *Nature Reviews Molecular Cell Biology*. 2008;9:367-77.
19. Kelly DP. Irisin, Light My Fire. *Science*. 2012;336(6077):42-3.
20. Harris PK, Kletzien RF. Localization of a pioglitazone response element in the adipocyte fatty acid-binding protein gene. *Mol Pharmacol*. 1994;45(3):439-45. Epub 1994/03/01. PubMed PMID: 8145730.
21. Benoit G, Cooney A, Giguere V, Ingraham H, Lazar M, Muscat, et al. International Union of Pharmacology. LXVI. Orphan nuclear receptors. *Pharmacological Reviews*. 2006;58(4):798-836.
22. Huang TH-W, Teoh AW, Lin B-L, Lin DS-H, Roufogalis B. The role of herbal PPAR modulators in the treatment of cardiometabolic syndrome. *Pharmacological Research*. 2009;60:195-206.
23. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*. 1984;81(8):3684-91.
24. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*. 1983;79(2):926-36.
25. Krey G, Keller H, Mahfoudi A, Medin J, Ozato K, Dreyer C, et al. Xenopus peroxisome proliferator activated receptors: genomic organization, response element recognition, heterodimer formation with retinoid X receptor and activation by fatty acids. *Journal of Steroid Biochemistry and Molecular Biology*. 1993;47(1-6):65-73.
26. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*. 2008;4(3):435-47.
27. Dreyer C, Krey G, Keller H, Givel F, Helftenbein G, Wahli W. Control of the peroxisomal beta-oxidation pathway by a novel family of nuclear hormone receptors. *Cell*. 1992;68(5):879-87.
28. Peng YH, Coumar M, Leou J, Wu J, Shiao H, Lin C, et al. Crystal structure of human PPAR-gamma ligand binding domain complex with a potency improved agonist. To be published. 2011.
29. Escher P, Wahli W. Peroxisome proliferator-activated receptors: insight into multiple cellular functions. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2000;448(2):121-38.
30. Lewis JD, Lichtenstein GR, Deren JJ, Sands BE, Hanauer SB, Katz JA, et al. Rosiglitazone for Active Ulcerative Colitis: A Randomized Placebo-Controlled Trial. *Gastroenterology*. 2008;134:688-59.

31. Gale EA. Troglitazone: the lesson that nobody learned? *Diabetologia*. 2006;49(1):1-6. Epub 2005/12/20. doi: 10.1007/s00125-005-0074-6. PubMed PMID: 16362281.
32. Lewis JD, Ferrara A, Peng T, Hedderson M, Bilker WB, Quesenberry CPJ, et al. Risk of Bladder Cancer Among Diabetic Patients treated with Pioglitazone. *Diabetes Care*. 2011;34(4):916-22.
33. Azoulay L, Yin H, Filion KB, Assayag J, Majdan A, Pollak MN, et al. The use of pioglitazone and the risk of bladder cancer in people with type 2 diabetes: nested case-control study. *British Medical Journal*. 2012;344:e3645.
34. Henney JE. Withdrawal of Troglitazone and Cisapride. *JAMA: Journal of the American Medical Association*. 2000;283(17):2228.
35. Riley K. FDA significantly restricts access to the diabetes drug Avandia. *FDA News Release*. 2010.
36. Harris G. Research Ties Diabetes Drug to Heart Woes. *The New York Times*. 2010.
37. Fisher L. Adverse Diabetes Drug News Sends Warner-Lambert Down. *The New York Times*. 1997.
38. Cohen J. Risks of troglitazone apparent before approval in USA. *Diabetologia*. 2006;49(6):1454-5.
39. Waku T, Shiraki T, Oyama T, Fujimoto Y, Maebara K, Kamiya N, et al. Structural Insight into PPAR $\gamma$  Activation Through Covalent Modification with Endogenous Fatty Acids. *Journal of Molecular Biology*. 2009;385:188-99.
40. Yasui Y, Hosokawa M, Sahara T, Suzuki R, Ohgiya S, Kohno H, et al. Bitter gourd seed fatty acid rich in 9c, 11t, 13t-conjugated linolenic acid induces apoptosis and up-regulates the GADD45, p53, and PPAR $\gamma$  in human colon cancer Caco-2 cells. *Prostaglandins, Leukotrienes, and Essential Fatty Acids*. 2005;73:113-9.
41. Itoh T, Fairall L, Amin K, Inaba Y, Szanto A, Balint BL, et al. Structural basis for the activation of PPAR $\gamma$  by oxidized fatty acids. *Nature Structural & Molecular Biology*. 2008;15(9):924-31.
42. Bassaganya-Riera J, Hontecillas R. CLA and n-3 PUFA differentially modulate clinical activity and colonic PPAR-responsive gene expression in a pig model of experimental IBD. *Clinical Nutrition*. 2006;25:454-65.
43. Uppenberg J, Svensson C, Jaki M, Bertilsson G, Jendeberg L, Berkenstam A. Crystal Structure of the Ligand Binding Domain of the Human Nuclear Receptor PPAR $\gamma$ . *Journal of Biological Chemistry*. 1998;273(47):31108-12.
44. Zhang BB, Moller DE. New approaches in the treatment of type 2 diabetes. *Current Opinion in Chemical Biology*. 2000;4:461-7.
45. Gampe RT, Montana VG, Lambert MH, Miller AB, Bledsoe RK, Milburn MV, et al. Asymmetry in the PPAR $\gamma$ /RXR $\alpha$  Crystal Structure Reveals the Molecular Basis of Heterodimerization among Nuclear Receptors. *Molecular Cell*. 2000;5:545-55.
46. Braissant O, Fougère F, Scotto C, Dauca M, Wahli W. Differential Expression of Peroxisome Proliferator-Activated Receptors (PPARs): Tissue Distribution of PPAR- $\alpha$ , - $\beta$ , and - $\gamma$  in the Adult Rat. *Endocrinology*. 1996;137(1):354-66.
47. Desvergne B, Wahli W. Peroxisome Proliferator-Activated Receptors: Nuclear Control of Metabolism. *Endocrine Reviews*. 1999;20(5):649-88.
48. Mokdad AH, Bowman BA, Ford ES, Marks JS, Koplan JP. The Continuing Epidemics of Obesity and Diabetes in the United States. *Journal of American Medical Association*. 2001;286(10):1195-200.

49. Nesto RW, Bell D, Bonow RO, Fonseca V, Grundy SM, Horton ES, et al. Thiazolidinedione Use, Fluid Retention, and Congestive Heart Failure. *Circulation*. 2003;108:2941-8.
50. Weatherman RV, Fletterick RJ, Scanlan TS. Nuclear-receptor ligands and ligand-binding domains. *Annual Review of Biochemistry*. 1999;68:599-81.
51. Kallenberger BC, Love JD, Krishna V, Chatterjee K, Schwabe JWR. A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. *Nature*. 2003;10(2):136-40.
52. Nolte RT, Wisely GB, Westin S, Cobb JE, Lambert MH, Kurokawa R, et al. Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor- $\gamma$ . *Nature*. 1998;395:137-43.
53. Rosen ED, Spiegelman BM. PPAR $\gamma$ : a Nuclear Regulator of Metabolism, Differentiation, and Cell Growth. *The Journal of Biological Chemistry*. 2001;276(41):37731-4.
54. Pochetti G, Godio C, Mitro N, Caruso D, Galmozzi A, Scurati S, et al. Insights into the Mechanism of Partial Agonism: Crystal Structures of the Peroxisome Proliferator-Activated Receptor  $\gamma$  Ligand-Binding Domain in the Complex with Two Enantiomeric Ligands. *Journal of Biological Chemistry*. 2007;282(23):17314-24.
55. Bruning JB, Chalmers MJ, Prasad S, Busby SA, Kamenechka TM, He Y, et al. Partial Agonists Activate PPAR $\gamma$  Using a Helix 12 Independent Mechanism. *Structure*. 2007;15:1258-71.
56. Chandra V, Huang P, Hamuro Y, Raghuram S, Wang Y, Burris TP, et al. Structure of the intact PPAR- $\gamma$ -RXR- $\alpha$  nuclear receptor complex on DNA. *Nature*. 2008;456:350-7.
57. Xu HE, Stanley TB, Montana VG, Lambert MH, Shearer BG, Cobb JE, et al. Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPAR $\alpha$ . *Nature*. 2002;415:813-7.
58. Hamuro Y, Coales SJ, Morrow JA, Molnar KS, Tuske SJ, Southern MR, et al. Hydrogen/deuterium-exchange (H/D-Ex) of PPAR $\gamma$ LBD in the presence of various modulators. *Protein Science*. 2006;15:1883-92.
59. Berman HM, Battistuz T, Bhat TN, Bluhm W, Bourne PE, Burkhardt K, et al. The Protein Data Bank. *Acta Crystallographica Section D: Biological Crystallography*. 2002;D58:899-907.
60. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Research*. 2000;28:235-42.
61. Casimiro-Garcia A, Bigge CF, Davis JA, Padalino T, Pulaski J, Ohren JF, et al. Effects of modifications of the linker in a series of phenylpropanoic acid derivatives: Synthesis, evaluation as PPAR $\alpha/\gamma$  dual agonists, and X-ray crystallographic studies. *Bioorganic & Medicinal Chemistry*. 2008;16(9):4883-907.
62. Cronet P, Petersen JF, Folmer R, Blomberg N, Sjoblom K, Karlsson U, et al. Structure of the PPAR $\alpha$  and - $\gamma$  ligand binding domain in complex with AZ 242; ligand selectivity and agonist activation in the PPAR family. *Structure*. 2001;9(8):699-706. Epub 2001/10/06. PubMed PMID: 11587644.
63. Moras D, Gronemeyer H, Gaal V, Scheen A, Cock T-A, Houten S, et al. High Resolution Structure of Peroxisome Proliferator-Activated Receptor gamma and



- Characterisation of its Interaction with the Co-activator Transcriptional Intermediary Factor 2. To be published.
64. Xu HE, Lambert MH, Montana VG, Plunket KD, Moore LB, Collins JL, et al. Structural determinants of ligand binding selectivity between the peroxisome proliferator-activated receptors. *Proceedings of the National Academy of Science*. 2001;98(24):13919-24.
  65. Sauerberg P, Pettersson I, Jeppesen L, Bury PS, Mogensen JP, Wassermann K, et al. Novel tricyclic- $\alpha$ -alkyloxyphenylpropionic acids: dual PPAR $\alpha/\gamma$  agonists with hypolipidemic and antidiabetic activity. *J Med Chem*. 2002;45(4):789-804. Epub 2002/02/08. PubMed PMID: 11831892.
  66. Ebdrup S, Pettersson I, Rasmussen HB, Deussen HJ, Frost Jensen A, Mortensen SB, et al. Synthesis and biological and structural characterization of the dual-acting peroxisome proliferator-activated receptor  $\alpha/\gamma$  agonist ragaglitazar. *J Med Chem*. 2003;46(8):1306-17. Epub 2003/04/04. doi: 10.1021/jm021027r. PubMed PMID: 12672231.
  67. Haffner C, Lenhard J, Miller A, McDougald D, Dwornik K, Ittoop O, et al. Structure-based design of potent retinoid X receptor alpha agonists. *Journal of medicinal chemistry*. 2004;47:2010-29.
  68. Ostberg T, Svensson S, Selen G, Uppenberg J, Thor M, Sundbom M, et al. A new class of peroxisome proliferator-activated receptor agonists with a novel binding epitope shows antidiabetic effects. *Journal of Biological Chemistry*. 2004;279:41124-30.
  69. Shi GQ, Dropinski JF, McKeever BM, Xu S, Becker JW, Berger JP, et al. Design and synthesis of  $\alpha$ -aryloxyphenylacetic acid derivatives: a novel class of PPAR $\alpha/\gamma$  dual agonists with potent antihyperglycemic and lipid modulating activity. *Journal of medicinal chemistry*. 2005;48(13):4457-68. Epub 2005/06/25. doi: 10.1021/jm0502135. PubMed PMID: 15974597.
  70. Li Y, Choi M, Suino K, Kovach A, Daugherty J, Kliewer SA, et al. Structural and biochemical basis for selective repression of the orphan nuclear receptor liver receptor homolog 1 by small heterodimer partner. *Proceedings of the National Academy of Science*. 2005;102(27):9505-10.
  71. Mahindroo N, Huang CF, Peng YH, Wang CC, Liao CC, Lien TW, et al. Novel indole-based peroxisome proliferator-activated receptor agonists: design, SAR, structural biology, and biological activities. *J Med Chem*. 2005;48(26):8194-208. Epub 2005/12/22. doi: 10.1021/jm0506930. PubMed PMID: 16366601.
  72. Mahindroo N, Wang C-C, Liao C-C, Huang C-F, Lu I-L, Lien T-W, et al. Indol-1-yl Acetic Acids as Peroxisome Proliferator-Activated Receptor Agonists: Design, Synthesis, Structural Biology, and Molecular Docking Studies. *Journal of medicinal chemistry*. 2006;49:1212-6.
  73. Burgermeister E, Schnoebelen A, Flament A, Benz J, Stihle M, Gsell B, et al. A Novel Partial Agonist of Peroxisome Proliferator-Activated Receptor- $\gamma$  (PPAR $\gamma$ ) Recruits PPAR $\gamma$ -Coactivator-1 $\alpha$ , Prevents Triglyceride Accumulation, and Potentiates Insulin Signaling in Vitro. *Molecular Endocrinology*. 2006;20(4):809-30.
  74. Lu I-L, Huang C-F, Peng Y-H, Lin Y-T, Hsieh H-P, Chen C-T, et al. Structure-Based Drug Design of a Novel Family of PPAR $\gamma$  Partial Agonists: Virtual Screening,

- X-ray Crystallography, and in Vitro/in Vivo Biological Activities. *Journal of medicinal chemistry*. 2006;49(9):2703-12.
75. Kuhn B, Hilpert H, Benz J, Binggell A, Grether U, Humm R, et al. Structure-based design of indole propionic acids as novel PPAR $\alpha/\gamma$  co-agonists. *Bioorganic & Medicinal Chemistry Letters*. 2006;16(15):4016-20.
  76. Hopkins CR, O'Neil S V, Laufersweiler MC, Wang Y, Pokross M, Mekel M, et al. Design and synthesis of novel N-sulfonyl-2-indole carboxamides as potent PPAR- $\gamma$  binding agents with potential application to the treatment of osteoporosis. *Bioorg Med Chem Lett*. 2006;16(21):5659-63. Epub 2006/08/22. doi: 10.1016/j.bmcl.2006.08.003. PubMed PMID: 16919947.
  77. Montanari R, Saccoccia F, Scotti E, Crestani M, Godio C, Gilardi F, et al. Crystal structure of the peroxisome proliferator-activated receptor  $\gamma$  (PPAR $\gamma$ ) ligand binding domain complexed with a novel partial agonist: a new region of the hydrophobic pocket could be exploited for drug design. *Journal of medicinal chemistry*. 2008;51(24):7768-76.
  78. Mahindroo N, Peng Y-H, Lin C-H, Tan UK, Prakash E, Lien TW, et al. Structural basis for the structure-activity relationships of peroxisome proliferator-activated receptor agonists. *Journal of medicinal chemistry*. 2006;49(21):6421-4.
  79. Zhang H, Ryono DE, Devasthale P, Wang W, O'Malley K, Farrelly D, et al. Design, synthesis and structure-activity relationships of azole acids as novel, potent dual PPAR  $\alpha/\gamma$  agonists. *Bioorg Med Chem Lett*. 2009;19(5):1451-6. Epub 2009/02/10. doi: 10.1016/j.bmcl.2009.01.030. PubMed PMID: 19201606.
  80. Pochetti G, Montanari R, Mazza F, Crestani M, Godio C, Loiodice F, et al. Crystal structures of PPAR $\gamma$  ligand binding domain complexed with new partial and full agonists: a new region of the hydrophobic pocket could be exploited for drug design. To be published.
  81. Li Y, Kovach A, Suino-Powell K, Martynowski D, Xu HE. Structural and biochemical basis for the binding selectivity of peroxisome proliferator-activated receptor  $\gamma$  to PGC-1 $\alpha$ . *Journal of Biological Chemistry*. 2008;283(27):19132-9.
  82. Greschik H, Althage M, Flaig R, Sato Y, Chavant V, Peluso-Iltis C, et al. Communication between the ERR $\alpha$  homodimer interface and the PGC-1 $\alpha$  binding surface via the helix 8-9 loop. *J Biol Chem*. 2008;283(29):20220-30. Epub 2008/04/29. doi: 10.1074/jbc.M801920200. PubMed PMID: 18441008.
  83. Ambrosio AL, Dias SM, Polikarpov I, Zurier RB, Burstein SH, Garratt RC. Ajulemic acid, a synthetic nonpsychoactive cannabinoid acid, bound to the ligand binding domain of the human peroxisome proliferator-activated receptor  $\gamma$ . *J Biol Chem*. 2007;282(25):18625-33. Epub 2007/04/28. doi: 10.1074/jbc.M702538200. PubMed PMID: 17462987.
  84. Einstein M, Akiyama TE, Castriota GA, Wang CF, McKeever B, Mosley RT, et al. The differential interactions of peroxisome proliferator-activated receptor  $\gamma$  ligands with Tyr473 is a physical basis for their unique biological activities. *Mol Pharmacol*. 2008;73(1):62-74. Epub 2007/10/18. doi: 10.1124/mol.107.041202. PubMed PMID: 17940191.
  85. Trump R, Cobb J, Shearer B, Lambert M, Nolte R, Wilson TM, et al. Co-crystal structure guided array synthesis of PPAR $\gamma$  inverse agonists. *Science Direct* 2007;17. doi: 10.1016/j.bmcl.2007.04.111.

86. Artis DR, Lin JJ, Wang W, Mehra U, Perreault M, Erbe D, et al. Scaffold-based discovery of indeglitazar, a PPAR pan-active anti-diabetic agent. *PNAS*. 2009;106(1):262-7.
87. Oberfield J, Collins J, Holmes C, Goreham D, Cooper J, Cobb J, et al. A peroxisome proliferator-activated receptor  $\gamma$  ligand inhibits adipocyte differentiation. *Proceedings of the National Academy of Sciences, USA*. 1999;96:6102-6.
88. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nature Reviews Drug Discovery*. 2004;3:935-49.
89. Kellenberger E, Rodrigo J, Muller P, Rognan D. Comparative Evaluation of Eight Docking Tools for Docking and Virtual Screening Accuracy. *Proteins*. 2004;57(255-242).
90. Cavasotto CN, Abagyan A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *Journal of Molecular Biology*. 2004;337:209-25.
91. Lyne PD. Structure-based virtual screening: an overview. *Drug Discovery Today*. 2002;7(20):1047-55.
92. Stahura FL, Bajorath J. New Methodologies for Ligand-Based Virtual Screening. *Current Pharmaceutical Design*. 2005;11:1189-202.
93. Schneider G, Bohm H-J. Virtual Screening and fast automated docking methods. *Combinatorial Chemistry*. 2002;7(1):64-70.
94. Walters WP, Stahl MT, Murcko MA. Virtual Screening - an overview. *Drug Discovery Today*. 1998;3(4):160-78.
95. Klebe G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today*. 2006;11(13-14):580-94.
96. Brenk R, Irwin JJ, Schoichet BK. Here Be Dragons: Docking and Screening in an Uncharted Region of Chemical Space. *Journal of Biomolecular Screening*. 2005;10(7):667-74.
97. Elowe NH, Blanchard JE, Cechetto JD, Brown ED. Experimental screening of dihydrofolate reductase yields a "test set" of 50,000 small molecules for a computational data-mining and docking competition. *Journal of Biomolecular Screening*. 2005;10:653-7.
98. Zolli-Juran M, Cechetto JD, Hartlena R, Daigle DM, Brown ED. High throughput screening identifies novel inhibitors of *Escherichia coli* dihydrofolate reductase that are competitive with dihydrofolate. *Bioorganic and Medicinal Chemistry Letters*. 2003;13:2493-6.
99. Kaya T, Mohr SC, Waxman DJ, Vajda S. Computational Screening of Phthalate Monoesters for Binding to PPAR $\gamma$ . *Chemical Research in Toxicology*. 2006;19:999-1009.
100. Scarsi M, Podvinec M, Roth A, Hug H, Kersten S, Albrecht H, et al. Sulfonylureas and Glinides Exhibit Peroxisome Proliferator-Activated Receptor Activity: A combined Virtual Screening and Biological Assay Approach. *Molecular Pharmacology*. 2007;71(2):398-406.
101. Xu XY, Cheng F, Shen JH, Luo XM, Chen LL, Yue LD, et al. Agonists-PPAR $\gamma$  Interactions: Molecular Modelling Study with Docking Approach. *International Journal of Quantum Chemistry*. 2003;93:405-10.

102. Marshall TG, Lee RE, Marshall FE. Common angiotensin receptor blockers may directly modulate the immune system via VDR, PPAR, and CCR2b. *Theoretical Biology and Medical Modelling*. 2006;3:1.
103. Salam NK, Huang TH-W, Kota BP, Kim MS, Li Y, Hibbs DE. Novel PPAR-gamma Agonists Identified from a Natural Product Library: A Virtual Screening, Induced-Fit Docking and Biological Assay Study. *Chemical Biology & Drug Design*. 2008;71:57-70.
104. Schapira M, Abagyan R, Totrov M. Nuclear Hormone Receptor Targeted Virtual Screening. *Journal of medicinal chemistry*. 2003;46(14):3045-59.
105. Yu C, Chen L, Luo H, Chen J, Cheng F, Gui C, et al. Binding analysis between Human PPAR $\gamma$ -LBD and ligands. *European Journal of Biochemistry*. 2004;271:386-97.
106. Shay NF, Banz WJ. Regulation of Gene Transcription by Botanicals: Novel Regulatory Mechanisms. *Annual Review of Nutrition*. 2005;25:297-315.
107. Hussein Z, Wentworth JM, Nankervis AJ, Proietto J, Colman PG. Effectiveness and side effects of thiazolidinediones for type 2 diabetes: real-life experience from a tertiary hospital. *Medical Journal of Australia*. 2004;181(10):536-9.
108. Jachak SM, Saklani A. Challenges and opportunities in drug discovery from plants. *Current Science*. 2007;92(9):1251-7.
109. Liu Y, Wang M-W. Botanical drugs: Challenges and opportunities; Contribution to Linnaeus Memorial Symposium 2007. *Life Sciences*. 2008;82:445-9.
110. Rollinger JM, Haupt S, Stuppner H, Langer T. Combining Ethnopharmacology and Virtual Screening for Lead Structure Discovery: COX-inhibitors as Application Example. *Journal of Chemical Information and Modeling*. 2004;44(2):480-8.
111. Leach AR. *Molecular Modelling: Principles and Applications*. 2nd. ed. Dorset: Pearson Education Limited; 2001.
112. Wahli W. A gut feeling of the PXR, PPAR, and NF-kB connection. *Journal of Internal Medicine*. 2008;263:613-9.
113. Dubuquoy L, Rousseaux C, Thuru X, Peyrin-Biroulet L, Romano O, Chavatte P, et al. PPAR $\gamma$  as a new therapeutic target in inflammatory bowel diseases. *Gut*. 2006;55(9):1341-9. PubMed Central PMCID: PMCPMC1860011.
114. Gani OABSM, Sylte I. Molecular recognition of long chain fatty acids by peroxisome proliferator-activated receptor  $\alpha$ . *Medicinal Chemistry Research*. 2009;18:8-19. doi: 10.1007/s00044-008-9102-7.
115. Guri AJ, Hontecillas R, Ferrer G, Casagran O, Wankhade U, Noble AM, et al. Loss of PPAR $\gamma$  in immune cells impairs the ability of abscisic acid to improve insulin sensitivity by suppressing monocyte chemoattractant protein-1 expression and macrophage infiltration into white adipose tissue. *Journal of Nutritional Biochemistry*. 2008;19:216-28.
116. Reiss AB, Vagell ME. PPAR $\gamma$  Activity in the Vessel Wall: Anti-Atherogenic Properties. *Current Medicinal Chemistry*. 2006;13(26):3227-38.
117. Wu GD. Is There a Role for PPAR $\gamma$  in IBD? Yes, No, Maybe. *Gastroenterology*. 2003;124(5):1538-42.
118. Bassaganya-Riera J, Reynolds K, Martino-Catt S, Cui Y, Hennighausen L, Gonzalez F, et al. Activation of PPAR $\gamma$  and  $\delta$  by Conjugated Linoleic Acid Mediates

- Protection From Experimental Inflammatory Bowel Disease. *Gastroenterology*. 2004;127:777-91.
119. Ricote M, Li A, Willson TM, Kelly C, Glass C. The peroxisome proliferator-activated receptor- $\gamma$  is a negative regulator of macrophage activation. *Nature*. 1998;391:79-82.
  120. Ljung T, Karlen P, Schmidt D, Hellstrom PM, Lapidus A, Janczewska I, et al. Infliximab in inflammatory bowel disease: clinical outcome in a population based cohort from Stockholm County. *Gut*. 2004;53:849-53.
  121. Hanauer SB, Feagan BG, Lichtenstein GR, Mayer LF, Schreiber S, Colombel JFC, et al. Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. *The Lancet*. 2002;359(9317):1541-9.
  122. Goldsmith P, McGarity M, Walls AF, Church MK, Millward-Sadler GH, Robertson DAF. Corticosteroid treatment reduces mast cell numbers in inflammatory bowel disease. *Digestive Diseases and Sciences*. 1990;35(11):1409-13.
  123. Lichtenstein GR, Abreu MT, Cohen R, Tremaine W. American Gastroenterological Association Institute Technical Review on Corticosteroids, Immunomodulators, and Infliximab in Inflammatory Bowel Disease. *Gastroenterology*. 2006;130:940-87.
  124. Compston JE. Review article: osteoporosis, corticosteroids, and inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics*. 1995;9(3):237-50.
  125. Guri AJ, Hontecillas R, Bassaganya-Riera J. Abscisic acid ameliorates experimental IBD by downregulating cellular adhesion molecule expression and suppressing immune cell infiltration. *Clinical Nutrition*. 2010;In Press.
  126. Hontecillas R, Wannemeulher M, Zimmerman D, Hutto D, Wilson J, et al. Nutritional regulation of porcine bacterial-induced colitis by conjugated linoleic acid. *Journal of Nutrition*. 2002;132:2019-27.
  127. Hontecillas R, O'Shea M, Einerhand A, Diguardo M, Bassaganya-Riera J. Activation of PPAR $\gamma$  and  $\alpha$  by Punicic Acid Ameliorates Glucose Tolerance and Suppresses Obesity-Related Inflammation. *Journal of the American College of Nutrition*. 2009;28(2):184-95.
  128. Badami R, Patil K. Structure and occurrence of unusual fatty acids in minor seed oils. *Progress in Lipid Research*. 1980;19(3-4):119-53.
  129. Tsuzuki T, Kawakami Y. Tumor angiogenesis suppression by  $\alpha$ -eleostearic acid, a linoleic acid isomer with a conjugated triene system, via peroxisome proliferator-activated receptor  $\gamma$ . *Carcinogenesis*. 2008;29(4):797-806.
  130. Moon H-S, Guo D-D, Lee H-G, Choi Y-J, Kang J-S, Jo K, et al. Alpha-eleostearic acid suppresses proliferation of MCF-7 breast cancer cells via activation of PPAR $\gamma$  and inhibition of ERK1/2. *Cancer Science*. 2009;101(2):396-402.
  131. Tsuzuki T, Tokuyama Y, Igarashi M, Miyazawa T. Tumor growth suppression by  $\alpha$ -eleostearic acid, a linoleic acid isomer with a conjugated triene system, via lipid peroxidation. *Carcinogenesis*. 2004;25(8):1417-25.
  132. Eom J-M, Seo M-J, Baek J-Y, Chu H, Han SH, Min TS, et al. Alpha-eleostearic acid induces autophagy-dependent cell death through targeting AKT/mTOR and ERK1/2 signal together with the generation of reactive oxygen species. *Biochemical and Biophysical Research Communications*. 2009;391(1):903-8.

133. Yuan G-F, Wahlqvist ML, Yuan J-Q, Wang Q-M, Li D. Effect of Punicic Acid Naturally Occuring in Food on Lipid Peroxidation in Healthy Young Humans. *Journal of the Science of Food and Agriculture*. 2009;89(13):2331-5.
134. Vroegrijk I, van Diepen J, van den Berg S, Westbroek I, Keizer H, Gambelli L, et al. Pomegranate seed oil, puniceic acid prevents diet-induced obesity and insulin resistance in mice. *Food and Chemical Toxicology*. 2011;In press.
135. Meerts IA, Verspeek-Rip CM, Buskens CA, Keizer HG, Bassaganya-Riera J, Jouni ZE, et al. Toxicology evaluation of pomegranate seed oil. *Food and Chemical Toxicology*. 2009;47:1085-92.
136. Bassaganya-Riera J, DiGuardo M, Climent M, Vives C, Carbo A, Jouni Z, et al. Activation of PPAR  $\gamma$  and  $\delta$  by dietary puniceic acid ameliorates intestinal inflammation in mice. *British Journal of Nutrition*. 2011;106(6):878-86.
137. Saha SS, Ghosh M. Comparative study of antioxidant activity of  $\alpha$ -eleostearic acid and puniceic acid against oxidative stress generated by sodium arsenite. *Food and Chemical Toxicology*. 2009;47:2551-6.
138. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*. 1998;19(14):1639-62.
139. Trott O, Olson AJ. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of Computational Chemistry*. 2010;31(2):455-61.
140. van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E, Meulenhoff PJ, et al. *Gromacs User Manual version 3.3*. 2005.
141. Pettersen E, Goddard T, Huang C, Couch G, Greenblatt D, Meng E, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. 2004;25(13):1605-12.
142. Tanrikulu Y, Rau O, Schwarz O, Proschak E, Siems K, Muller-Kuhrt L, et al. Structure-Based Pharmacophore Screening for Natural-Product-Derived PPAR $\gamma$  Agonists. *ChemBioChem*. 2009;10:75-8.
143. Kuroda M, Mimaki Y, Honda S, Tanaka H, Yokota S, Mae T. Phenolics from *Glycyrrhiza glabra* roots and their PPAR- $\gamma$  ligand-binding activity. *Bioorganic & Medicinal Chemistry*. 2010;18:962-70. doi: 10.1016/j.bmc.2009.11.027.
144. Markt P, Schuster D, Kirchmair J, Laggner C, Langer T. Pharmacophore modeling and parallel screening for PPAR ligands. *Journal of Computer-Aided Molecular Design*. 2007;21:575-90. doi: 10.1007/s10822-007-9140-0.
145. Schuettelkopf A, van Aalten D. PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica D60*. 2004:1355-63.
146. Moreland JL, Gramada A, Buzko OV, Zhang Q, Bourne PE. The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics*. 2005;6(21):1-7. doi: 10.1186/1471-2105-6-21.
147. Schwartz RL, Phoenix T, Foy BD. *Learning Perl*: O'Reilly Media, Inc.; 2008.
148. Bassaganya-Riera J, Guri AJ, Lu P, Climent M, Carbo A, Sobral B, et al. Abscisic Acid Regulates Inflammation via Ligand-Binding Domain-Independent Activation of PPAR  $\gamma$ . *Journal of Biological Chemistry*. 2011;286(4):2504-16. doi: 10.1074/jbc.M110.160077.

149. Guri AJ, Hontecillas R, Si H, Liu D, Bassaganya-Riera J. Dietary abscisic acid ameliorates glucose tolerance and obesity-related inflammation in db/db mice fed high-fat diets. *Clinical Nutrition*. 2007;26(1):107-16.
150. Chang MW, Ayeni C, Breuer S, Torbett BE. Virtual Screening for HIV Protease Inhibitors: A Comparison of AutoDock 4 and Vina. *PLoS One*. 2010;5(8):e11955. doi: 10.1371/journal.pone.0011955.
151. Bassaganya-Riera J, Guri AJ, Hontecillas R. Treatment of Obesity-Related Complications with Novel Classes of Naturally Occurring PPAR Agonists. *Journal of Obesity*. 2011;2011:1-7. doi: 10.1155/2011/897894.
152. Shah Y, Morimura K, Gonzalez F. Expression of peroxisome proliferator-activated receptor- $\gamma$  in macrophage suppresses experimentally induced colitis. *American Journal of Physiology, Gastrointestinal and Liver Physiology*. 2007;292(2):G657-66.
153. Zhang T, Gao Y, Mao Y, Zhang Q, Lin C, Lin P, et al. Growth inhibition and apoptotic effect of alpha-eleostearic acid on human breast cancer cells. *Journal of Natural Medicines*. 2011;1-8. doi: 10.1007/s11418-011-0556-4.
154. Popovich DG, Lee Y, Li L, Zhang W. *Momordica charantia* Seed Extract Reduced Pre-Adipocyte Viability, Affects Lactate Dehydrogenase Release, and Lipid Accumulation in 3T3-L1 Cells. *Journal of Medicinal Food*. 2011;14(3):201-8.
155. Hennessy AA, Ross RP, Devery R, Stanton C. The Health Promoting Properties of the Conjugated Isomers of  $\alpha$ -Linoleic Acid. *Lipids*. 2011;46:105-19.
156. Tran HNA, Bae S-Y, Song B-H, Lee B-H, Bae Y-S, Kim Y-H, et al. Pomegranate (*Punica granatum*) seed linolenic acid isomers: Concentration-dependent modulation of estrogen receptor activity. *Endocrin Research*. 2010;35(1):1-16.
157. Hayhurst GP, Lee Y-H, Lambert G, Ward JM, Gonzalez FJ. Hepatocyte Nuclear Factor 4a (Nuclear Receptor 2A1) is Essential for Maintenance of Hepatic Gene Expression and Lipid Homeostasis. *Molecular and Cellular Biology*. 2001;21(4):1393-403. doi: 10.1128/MCB.21.4.1393-1403.2001.
158. Nagao K, Yanagita T. Bioactive lipids in metabolic syndrome. *Progress in Lipid Research*. 2008;47(2):127-46. doi: 10.1016/j.plipres.2007.12.002.
159. Rosen ED, Hsu C-H, Wang X, Sakai S, Freeman MW, Gonzalez FJ, et al. C/EBP $\alpha$  induces adipogenesis through PPAR $\gamma$ : a unified pathway. *Genes and Development*. 2002;16(1):22-6.
160. Shao D, Rangwala S, Bailey S, Krakow S, Reginato M, Lazar M. Interdomain communication regulating ligand binding by PPAR $\gamma$ . *Nature*. 1998;396(6709):377-80.
161. Straus DS, Glass CK. Anti-inflammatory actions of PPAR ligands: new insights on cellular and molecular mechanisms. *TRENDS in Immunology*. 2007;28(12):551-8.
162. Bassaganya-Riera J, Mikiyak S, Guri AJ, Hontecillas R. PPAR  $\gamma$  is highly expressed in F4/80hi adipose tissue macrophages and dampens adipose-tissue inflammation. *Cellular Immunology*. 2009;258:138-46.
163. Lewis SN, Bassaganya-Riera J, Bevan DR. Virtual Screening as a Technique for PPAR Modulator Discovery. *PPAR Research*. 2010;2010.
164. Farce A, Renault N, Chavatte P. Structural Insight into PPAR $\gamma$  Ligands Binding. *Current Medicinal Chemistry*. 2009;16(14):1768-89.

165. DeGrazia MJ, Thompson J, Vanden Heuvel JP, Peterson BR. Synthesis of a High-Affinity Fluorescent PPAR $\gamma$  Ligand for High-Throughput Fluorescence Polarization Assays. *Bioorganic & Medicinal Chemistry*. 2003;11:4325-32.
166. Wu B, Gao J, Wang M-w. Development of a complex scintillation proximity assay for high-throughput screening of PPAR $\gamma$  modulators. *Acta Pharmacologica Sinica*. 2005;26(3):339-44.
167. Brozell SR, Mukherjee S, Balius TE, Roe DR, Case DA, Rizzo RC. Evaluation of DOCK 6 as a post generation and database enrichment tool. *Journal of Computer-Aided Molecular Design*. 2012;26(6):749-73.
168. Irwin J, Shoichet B, Mysinger M, Huang N, Colizzi F, Wassam P, et al. Automated docking screens: a feasibility study. *Journal of medicinal chemistry*. 2009;52(18):5712-20.
169. Huang N, Shoichet BK, Irwin JJ. Benchmarking Sets for Molecular Docking. *Journal of medicinal chemistry*. 2006;49(23):6789-801. doi: 10.1021/jm0608356.
170. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of medicinal chemistry*. 2013;55:6582-94.
171. Zhang X, Wong SE, Lightstone FC. Message Passing Interface and Multithreading Hybrid for Parallel Molecular Docking of Large Databases on Petascale High Performance Computing Machines. *Journal of Computational Chemistry*. 2013;34:915-27.
172. Sundriyal S, Bharatam PV. Important pharmacophoric features of pan PPAR agonists: Common chemical feature analysis and virtual screening. *European Journal of Medical Chemistry*. 2009;44:3488-95.
173. Al-Najjar BO, Wahab HA, Muhammad TST, Shu-Chien AC, Noruddin NAA, Taha MO. Discovery of new nanomolar peroxisome proliferator-activated receptor  $\gamma$  activators via elaborate ligand-based modeling. *European Journal of Medical Chemistry*. 2011;46:2513-29.
174. Paliwal S, Yadav D, Yadav R, Paliwal S. In silico structure-based drug design approach to develop novel pharmacophore model of human peroxisome proliferator-activated receptor  $\gamma$  agonists. *Medicinal Chemistry Research*. 2011;20:656-9.
175. Hartenfeller M, Proschak E, Schuller A, Schneider G. Concept of Combinatorial De Novo Design of Drug-like Molecules by Particle Swarm Optimization. *Chemical Biology and Drug Design*. 2008;72:16-26.
176. Shon Y-s, Lee Y, Park C, Hwang S, Kim S, Baek A, et al. Pharmacophore Identification for Peroxisome Proliferator-Activated Receptor Gamma Agonists. *Bulletin of the Korean Chemical Society*. 2011;32(1):201-7.
177. Sonawane LV, Bari SB. Ligand-based in silico 3D-QSAR study of PPAR- $\gamma$  agonists. *Medicinal Chemistry Research*. 2011;20:1005-14.
178. Maltarollo V, Honorio K. Ligand- and structure-based drug design strategies and PPAR $\delta/\alpha$  selectivity. *Chemical Biology and Drug Design*. 2012;80(4):533-44. doi: 10.1111/j.1747-0285.2012.01424.x.
179. Schneidman-Duhovny D, Dror O, Inbar Y, Nussinov R, Wolfson H. PharmaGist: a webserver for ligand-based pharmacophore detection. *Nucleic Acids Research*. 2008;36:W223-W8.



180. Yang S-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discovery Today*. 2010;15(11-12):444-50.
181. Harris R, Olson AJ, Goodsell DS. Automated prediction of ligand-binding sites in proteins. *Proteins*. 2007. doi: 10.1002/prot.21645.
182. Zoete V, Grosdidier A, Michielin O. Peroxisome proliferator-activated receptor structures: Ligand specificity, molecular switch and interactions with regulators. *Biochimica et Biophysica Acta*. 2007;1771:915-25.
183. Oyama T, Toyota K, Waku T, Hirakawa Y, Nagasawa N, Kasuga J, et al. Adaptability and selectivity of human peroxisome proliferator-activated receptor (PPAR) pan agonists revealed from crystal structures. *Acta Crystallographica Section D: Biological Crystallography*. 2009;65:786-95.
184. Waku T, Shiraki T, Oyama T, Maebara K, Nakamori R, Morikawa K. The nuclear receptor PPAR $\gamma$  individually responds to serotonin- and fatty acid-metabolites. *EMBO Journal*. 2010;29:3395-407.
185. Ohashi M, Oyama T, Nakagome I, Satoh M, Nishio Y, Nobusada H, et al. Design, synthesis, and structural analysis of phenylpropanoic acid-type PPAR $\gamma$ -selective agonists: discovery of reversed stereochemistry-activity relationship. *Journal of medicinal chemistry*. 2011;54:331-41.
186. Tomioka D, Hashimoto H, Sato M, Shimizu T. Crystal Structure of human PPAR gamma in complex with MCC555. To be published. 2011.
187. Wakabayashi K, Hayashi S, Matsui Y, Matsumoto T, Furukawa A, Kuroha M, et al. Pharmacology and in vitro profiling of novel peroxisome proliferator-activated receptor  $\gamma$  ligand, Cerco-A. *Biological & Pharmaceutical Bulletin*. 2011;34(7):1094-104.
188. Motani A, Wang Z, Weiszmann J, McGee L, Lee G, Liu Q, et al. INT131: a selective modulator of PPAR gamma. *Journal of Molecular Biology*. 2009;386(1301-1311).
189. Lin C, Peng Y, Coumar M, Chittimalla SK, Liao CC, Lyn P, et al. Design and structural analysis of novel pharmacophores for potent and selective peroxisome proliferator-activated receptor  $\gamma$  agonists. *Journal of medicinal chemistry*. 2009;52:2618-22.
190. Connors RV, Wang Z, Harrison M, Zhang A, Wanska M, Hiscock S, et al. Identification of a PPAR $\delta$  agonist with partial agonistic activity on PPAR $\gamma$ . *Bioorganic & Medicinal Chemistry Letters*. 2009;19(13):3550-4. doi: <http://dx.doi.org/10.1016/j.bmcl.2009.04.151>.
191. Fracchiolla G, Laghezza A, Piemontese L, Tortorella P, Mazza F, Montanari R, et al. New 2-Aryloxy-3-phenyl-propanoic Acids As Peroxisome Proliferator-Activated Receptors  $\alpha/\gamma$  Dual Agonists with Improved Potency and Reduced Adverse Effects on Skeletal Muscle Function. *Journal of medicinal chemistry*. 2009;52(20):6382-93. doi: 10.1021/jm900941b.
192. Li Y, Wang Z, Furukawa N, Escaron P, Weiszmann J, Lee G, et al. T2384, a novel antidiabetic agent with unique peroxisome proliferator-activated receptor  $\gamma$  binding properties. *Journal of Biological Chemistry*. 2008;283(14):9168-76.
193. Furukawa A, Arita T, Satoh S, Wakabayashi K, Hayashi S, Matsui Y, et al. Discovery of a novel selective PPAR $\gamma$  modulator from (-)-Cercosporamide

- derivatives. *Bioorganic & Medicinal Chemistry Letters*. 2010;20(7):2095-8. doi: <http://dx.doi.org/10.1016/j.bmcl.2010.02.073>.
194. Riu A, Grimaldi M, le Maire A, Bey G, Phillips K, Boulahtouf A, et al. Peroxisome proliferator-activated receptor  $\gamma$  is a target for halogenated analogs of bisphenol A. *Environmental Health Perspectives*. 2011;119(9):1227-32.
  195. Riu A, le Maire A, Grimaldi M, Audebert M, Hillenweck A, Bourguet W, et al. Characterization of Novel Ligands of ER $\alpha$ , Er $\beta$ , and PPAR $\gamma$ : The Case of Halogenated Bisphenol A and Their Conjugated Metabolites. *Toxicological Sciences*. 2011;122(2):372-82. doi: 10.1093/toxsci/kfr132.
  196. Mueller J, Schupp M, Unger T, Kintscher U, Heinemann U. Binding diversity of pioglitazone by peroxisome proliferator-activated receptor-gamma. To be published. 2011.
  197. Inbar Y, Schneidman-Duhovny D, Dror O, Nussinov R, Wolfson H. Deterministic Pharmacophore Detection via Multiple Flexible Alignment of Drug-Like Molecules. In Proc of RECOMB. 2007;3692 of Lecture Notes in Computer Science:423-34.
  198. Molecular Operating Environment, 2012.10. Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2012.
  199. Lewis SN, Brannan L, Guri AJ, Lu P, Hontecillas R, Bassaganya-Riera J, et al. Dietary  $\alpha$ -Eleostearic Acid Ameliorates Experimental Inflammatory Bowel Disease in Mice by Activating Peroxisome Proliferator-Activated Receptor- $\gamma$ . *PLoS One*. 2011;6(8):e24031. doi: 10.1371/journal.pone.0024031. PubMed Central PMCID: PMC3164124.
  200. Irwin JJ, Shoichet BK. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *Journal of Chemical Information and Modeling*. 2005;45:177-82.
  201. Sushko I, Salmina E, Potemkin V, Poda G, Tetko I. ToxAlerts: A web server of structural alerts for toxic chemicals and compounds with potential adverse reactions. *Journal of Chemical Information and Modeling*. 2012;52(8):2310-6.
  202. Kazius J, McGuire R, Bursi R. Derivation and validation of toxicophores for mutagenicity prediction. *J Med Chem*. 2005;48(1):312-20. Epub 2005/01/07. doi: 10.1021/jm040835a. PubMed PMID: 15634026.
  203. Oprea T. Property distribution of drug-related chemical databases. *Journal of Computer-Aided Molecular Design*. 2000;14(3):251-64. doi: 10.1023/a:1008130001697.
  204. Triballeau N, Acher F, Brabet I, Pin J-P, Bertrand H-O. Virtual Screening Workflow Development Guided by the "Receiver Operating Characteristic" Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *Journal of medicinal chemistry*. 2005;48(7):2534-47.
  205. Lehmann J, Lenhard J, Oliver B, Ringold G, Kliewer S. Peroxisome proliferator-activated receptors  $\alpha$  and  $\gamma$  are activated by indomethacin and other non-steroidal anti-inflammatory drugs. *Journal of Biological Chemistry*. 1997;272(6):3406-10.
  206. Wiggins J, Rajapakse R. Balsalazide: a novel 5-aminosalicylate prodrug for the treatment of active ulcerative colitis. *Expert Opinion on Drug Metabolism and Toxicology*. 2009;5(10):1279-84.
  207. Patil SA, Moss AC. Balsalazide disodium for the treatment of ulcerative colitis. *Expert Review of Gastroenterology and Hepatology*. 2008;2(2):177-84.

208. Teotico DG, Frazier ML, Ding F, Dokholyan NV, Temple BR, Redinbo M. Active nuclear receptors exhibit highly correlated AF-2 domain motions. *PLoS Computational Biology*. 2008;4(7):e1000111.
209. Chen S, Johnson BA, Li Y, Aster S, McKeever B, Mosley R, et al. Both coactivator LXXLL motif-dependent and -independent interactions are required for peroxisome proliferator-activated receptor  $\gamma$  (PPAR $\gamma$ ) function. *J Biol Chem*. 2000;275(6):3733-6. Epub 2000/02/08. PubMed PMID: 10660518.
210. Liberato MV, Nascimento AS, Ayers SD, Lin JZ, Cvoro A, Silveira RL, et al. Medium Chain Fatty Acids Are Selective Peroxisome Proliferator Activated Receptor (PPAR)  $\gamma$  Activators and Pan-PPAR Partial Agonists. *PLoS One*. 2012;7(5):e36297. doi: 10.1371/journal.pone.0036297.
211. Spiegelman BM. PPAR-g: Adipogenic regulator and thiazolidinedione receptor. *Diabetes*. 1998;47:507-14.
212. Olefsky JM. Treatment of insulin resistance with peroxisome proliferator-activated receptor  $\gamma$  agonists. *The Journal of Clinical Investigation*. 2000;106(4):467-72. doi: 10.1172/jci10843.
213. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Research*. 2012;40:D109-D14.
214. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 2000;28:27-30.
215. Michalik L, Zoete V, Krey G, Grosdidier A, Gelman L, Chodanowski P, et al. Combined Simulation and Mutagenesis Analyses Reveal the Involvement of Key Residues for Peroxisome Proliferator-activated Receptor $\alpha$  Helix 12 Dynamic Behavior. *The Journal of Biological Chemistry*. 2007;282(13):9666-77.
216. Aranda A, Pascual A. Nuclear Hormone Receptors and Gene Expression. *Physiological Reviews*. 2001;81(3):1269-304.
217. Pronk S, Pail S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013;29(7):845-54.
218. Sorin EJ, Pande VS. Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophysical Journal*. 2005;88(4):2472-93.
219. DePaul AJ, Thompson EJ, Patel SS, Haldeman K, Sorin EJ. Equilibrium conformational dynamics in an RNA tetraloop from massively parallel molecular dynamics. *Nucleic Acids Research*. 2010;38(14):4856-67.
220. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type preception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*. 2006;25:247-60.
221. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Antechamber, an accessory software package for molecular mechanical calculations. *Journal of Chemical Information and Computer Science*. 2004;25:1157-74.
222. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*. 2007;126(1):014101.
223. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*. 1981;52(12):7182-91.

224. Nose S, Klein ML. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*. 1983;50(5):1055-76.
225. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*. 1995;103(19):8577-93.
226. Grace, Beaverton, OR, 2002.
227. Jorgensen WL. Drug Discovery: Pulled from a protein's embrace. *Nature*. 2010;466:42-3.
228. Colizzi F, Perozzo R, Scapozza L, Recanatini M, Cavalli A. Single-Molecule Pulling Simulations Can Discern Active from Inactive Enzyme Inhibitors. *Journal of the American Chemical Society*. 2010;132(21):7361-71. doi: 10.1021/ja100259r.
229. Perakyla M. Ligand unbinding pathways from the vitamin D receptor studied by molecular dynamics simulations. *European Biophysics Journal*. 2009;38:185-98.
230. Martinez L, Polikarpov I, Skaf MS. Only subtle protein conformational adaptations are required for ligand binding to thyroid hormone receptors: Simulations using a novel multipotent steered molecular dynamics approach. *Journal of Physical Chemistry B*. 2008;112:10741-51.
231. Genest D, Garnier N, Arrault A, Marot C, Morin-Allory L, Genest M. Ligand-escape pathways from the ligand-binding domain of PPAR $\gamma$  receptor as probed by molecular dynamics simulations. *European Biophysics Journal*. 2008;37:369-79.
232. Malapaka RRV, Khoo S, Zhang J, Choi JH, Zhou XE, Xu Y, et al. Identification and Mechanism of 10-Carbon Fatty Acid as Modulating Ligand of Peroxisome Proliferator-activated Receptors. *Journal of Biological Chemistry*. 2012;287(1):183-95. doi: 10.1074/jbc.M111.294785.
233. Kim J, Han D, Kim J, Lee S, Kim S, Woo J, et al. PPAR  $\gamma$  partial agonist, KR-62776, inhibits adipocyte differentiation via activation of ERK. *Cellular and Molecular Life Sciences*. 2009;66(10):1766-81.
234. Schmidtke P, Barril X. Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *Journal of medicinal chemistry*. 2010;53:5858-67.
235. Murphy G, Holder J. PPAR- $\gamma$  agonists: Therapeutic role in diabetes, inflammation and cancer. *Trends in Pharmacological Sciences*. 2000;21(12):469-74.
236. Michalik L, Wahli W. PPARs Mediate Lipid Signaling in Inflammation and Cancer. *PPAR Research*. 2008;2008:1-15. doi: 10.1155/2008/134059.
237. Carter AB, Misyak SA, Hontecillas R, Bassaganya-Riera J. Dietary Modulation of Inflammation-Induced Colorectal Cancer through PPAR $\gamma$ . *PPAR Research*. 2009;2009:1-9. doi: 10.1155/2009/498352.
238. Rueda M, Bottegoni G, Abagyan R. Recipes for the Selection of Experimental Protein Conformations for Virtual Screening. *Journal of Chemical Information and Modeling*. 2010;50:186-93.
239. Guasch L, Sala E, Castell-Auví A, Cedó L, Liedl KR, Wolber G, et al. Identification of PPAR $\gamma$  Partial Agonists of Natural Origin (I): Development of a Virtual Screening Procedure and In Vitro Validation. *PLoS One*. 2012;7(11):e50816. doi: 10.1371/journal.pone.0050816.

240. Nissen SE, Wolski K. Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes. *New England Journal of Medicine*. 2007;356(24):2457-71.
241. Home PD, Pocok SJ, Beck-Nielsen H, Curtis PS, Gomis R, Hanefeld M, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. *The Lancet*. 2009;373(9681):2125-35.
242. Nissen SE, Wolski K. Rosiglitazone Revisited: An updated meta-analysis of risk for myocardial infarction and cardiovascular mortality. *Archives of Internal Medicine*. 2010;170(14):1191-201.
243. Trivedi M, Lokhandwala M. Rosiglitazone restores renal D1A receptor-Gs protein coupling by reducing receptor hyperphosphorylation in obese rats. *American Journal of Physiology, Renal Physiology*. 2005;289(2):F298-304.

## 8 Appendices

### 8.1 Appendix A: Executive summary

#### Executive summary of additional published works

The articles summarized below were written and published in peer reviewed journals during the course of the predoctoral period. As the articles do not pertain directly to completion of the dissertation work, they are summarized here below the full citation for the papers.

Bassaganya-Riera J, Guri AJ, Lu P, Climent M, Carbo A, Sobral BW, Evans C, Horne WT, **Lewis SN**, Bevan DR, and Hontecillas R. (2011) Abscisic acid Regulates Inflammation via Ligand-Binding Domain-Independent Activation of PPAR $\gamma$ . *J. Biol. Chem.* 286(4):2504-16. DOI: 10.1074/jbc.M110.160077.

Abscisic acid (ABA) was shown to ameliorate T2D and inflammation in a seeming PPAR $\gamma$ -dependent manner that did not involve direct binding to PPAR $\gamma$ . Docking was used to complement experimental results showing ABA does not bind to PPAR $\gamma$ . Experiments suggested ABA up-regulates PPAR $\gamma$  expression and this may be mediated by lanthionine synthetase C-like 2 (LANCL2). Docking of ABA into LANCL2 suggested favorable binding sites. It was suggested that ABA binds to LANCL2 as part of a bifurcated pathway that activates PPAR $\gamma$  for anti-inflammatory responses. My role for this study was docking of ABA into PPAR $\gamma$ . I performed the docking and analysis, and wrote the sections pertaining to this component for the paper. I also was involved in reviewing and editing the full manuscript before and during the review process. This paper is the most relevant to my primary research of all the papers included in the appendices, but will not be included as a chapter given I was not the primary author.

**Lewis SN**, Nsoesie E, Weeks C, Qiao D, Zhang, L. (2011) Prediction of Disease and Phenotype Associations from Genome-Wide Association Studies. *PLoS ONE*. 6(11): e27175. doi: 10.1371/journal.pone.0027175.

**Lewis SN**, Nsoesie E, Weeks C, Qiao D, Zhang, L. (2011) Meta-analysis of Genome-Wide Association Studies to Understand Disease Relatedness. *Type I Diabetes / Book 3: InTech*. p. 199-212. ISBN 979-953-307-127-4.

Both of these publications derived from use of bioinformatics, genetics, and computer science techniques to mine genetic variation data for similarities between 61 diseases and phenotypes. The goal of the parent study was to collect, catalog, and compare single nucleotide polymorphism (SNP) data to see if similarities could be quantified for four levels: SNP, gene, protein, and pathway. This approach was adopted as a means to assess relatedness potentially ignored or overlooked by genome-wide association studies alone. As data was grouped for each successive level, the clarity of connections between diseases was more pronounced. Multiple sclerosis, type 1 diabetes, and rheumatoid arthritis were identified as three diseases with predominant similarity for all levels. Our

data meshed with peripheral medical observations published mostly as editorial letters for various medical journals that speculate relationships exist between pairs of these diseases.

The book chapter included detailed statistical analysis of the data set with added genomic data from online bioinformatics databases. Emphasis was placed on type 1 diabetes relatedness. Author Contributions: Dr. Zhang conceived and designed the original study with suggestions on improvements from myself and the other three authors. All authors contributed equally to performing the screening, scripting, analysis of the data, and writing of the manuscript. Dr. Zhang, Dr. Nsoesie, and I composed and revised the manuscript and book chapter during the review process.

Lu P, Bevan DR, **Lewis SN**, Hontecillas R, Bassaganya-Riera J. (2010) Molecular modeling of lanthionine synthetase component C-like protein 2: a potential target for the discovery of novel type 2 diabetes prophylactics and therapeutics. *J Mol Model*. 17(3):543-553. DOI: 10.1007/s00894-010-0748-y.

Lu P, Hontecillas R, Horne WT, Carbo A, Viladomiu M, Pedragosa M, Bevan DR, **Lewis SN**. (2012) Computational Modeling-Based Discovery of Novel Classes of Anti-Inflammatory Drugs That Target Lanthionine Synthetase C-Like Protein 2. *PLoS ONE* 7(4): e34643. doi:10.1371/journal.pone.0034643.

The above papers were computational modeling studies of LANCL2 using methods similar to those mentioned within the dissertation for PPAR $\gamma$ . The first study proposed ABA and diabetic thiazolidinedione drugs bind to LANCL2 and possess a role in quenching T2D phenotypes. This paper was a follow up to the Bassaganya-Riera et al. 2011 article in which it was proposed that ABA binds to LANCL2 instead of PPAR $\gamma$ . The second study involved the use of virtual screening to mine compounds from the National Cancer Institute database for LANCL2 binders potentially effective at treatment of inflammatory diseases. One compound, NSC61610, was identified and tested in vitro as a LANCL2 binder that ameliorated chemically-induced colitis in mice. My primary role was consultant for the docking and dynamics procedures and analysis for both studies. I was involved in the review process as well.

## 8.2 Appendix B: PPAR $\gamma$ PDB table

The following is a list of PPAR $\gamma$  crystal structures available through the Protein Data Bank (PDB) website. The list was last updated on June 3, 2013. The structures most recently added (deposition date between 2011 and 2013) do not have a ligand type designations.

PDB ID	Chain ID	Ligand ID	Ligand Type	Ki (nM)	Kd (nM)	EC50 (nM)	IC50 (nM)	Dep. Date	Resolution	Citation Author	Title
<b>1FM6</b>	D	<b>BRL</b>	full	* 8-440 (BDB)		* 4-2880 (BDB)	* 30-50000 (BDB)	2000-08-16	2.10	Gampe Jr., R.T., Montana, V.G., Lambert, M.H., Miller, A.B., Bledsoe, R.K., Milburn, M.V., Kliewer, S.A., Willson, T.M., Xu, H.E.	Asymmetry in the PPAR $\gamma$ /RXR $\alpha$ crystal structure reveals the molecular basis of heterodimerization among nuclear receptors.
<b>1FM9</b>	D	<b>570</b>	full	* 1.1 (BDB) * 1 (BMO AD_64 56)		* 0.34-0.6 (BDB)	* 217 (BDB)	2000-08-16	2.10	Gampe Jr., R.T., Montana, V.G., Lambert, M.H., Miller, A.B., Bledsoe, R.K., Milburn, M.V., Kliewer, S.A., Willson, T.M., Xu, H.E.	Asymmetry in the PPAR $\gamma$ /RXR $\alpha$ crystal structure reveals the molecular basis of heterodimerization among nuclear receptors.
<b>1I7I</b>	A	<b>AZ2</b>	dual	* 18-200 (BDB)		* 13-3528 (BDB)	* 350 (BDB)	2001-03-09	2.35	Cronet, P., Petersen, J.F., Folmer, R., Blomberg, N., Sjoblom, K., Karlsson, U., Lindstedt, E.L., Bamberg, K.	Structure of the PPAR $\alpha$ and - $\gamma$ ligand binding domain in complex with AZ 242; ligand selectivity and agonist activation in the PPAR family.



<b>1K74</b>	D	<b>544</b>	full	* 1 (BDB)		* 0.2- 2.7 (BDB)		2001- 10-18	2.30	Xu, H.E., Lambert, M.H., Montana, V.G., Plunket, K.D., Moore, L.B., Collins, J.L., Oplinger, J.A., Kliewer, S.A., Gampe Jr., R.T., McKee, D.D., Moore, J.T., Willson, T.M.	Structural determinants of ligand binding selectivity between the peroxisome proliferator-activated receptors.
<b>1KNU</b>	A	<b>YPA</b>	dual		* 170 (BDB)	* 170 (BDB)		2001- 12-19	2.50	Sauerberg, P., Pettersson, I., Jeppesen, L., Bury, P.S., Mogensen, J.P., Wassermann, K., Brand, C.L., Sturis, J., Woldike, H.F., Fleckner, J., Andersen, A.-S.T., Mortensen, S.B., Svensson, L.A., Rasmussen, H.B., Lehmann, S.V., Polivka, Z., Sindelar, K., Panajotova, V., Ynddal, L., Wulff, E.M.	Novel tricyclic-alpha- alkyloxyphenylpropionic acids: dual PPARalpha/gamma agonists with hypolipidemic and antidiabetic activity
<b>1NYX</b>	A	<b>DRF</b>	dual	* 90 (BDB)		* 570- 600 (BDB)	* 92 (BDB )	2003- 02-14	2.65	Ebdrup, S., Pettersson, I., Rasmussen, H.B., Deussen, H.-J., Frost Jensen, A., Mortensen, S.B., Fleckner, J., Pridal, L., Nygaard, L., Sauerberg, P.	Synthesis and biological and structural characterization of the dual-acting peroxisome proliferator-activated receptor alpha/gamma agonist ragaglitazar

<b>1PRG</b>	A		apo					1998-07-02	2.20	Nolte, R.T., Wisely, G.B., Westin, S., Cobb, J.E., Lambert, M.H., Kurokawa, R., Rosenfeld, M.G., Willson, T.M., Glass, C.K., Milburn, M.V.	Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma.
<b>1RDT</b>	D	<b>570</b>	full	* 1.1 (BDB)	* 0.34-0.6 (BDB)	* 217 (BDB)		2003-11-06	2.40	Haffner, C.D., Lenhard, J.M., Miller, A.B., McDougald, D.L., Dwornik, K., Ittoop, O.R., Gampe Jr., R.T., Xu, H.E., Blanchard, S., Montana, V.G., Consler, T.G., Bledsoe, R.K., Ayscue, A., Croom, D.	Structure-based design of potent retinoid X receptor alpha agonists.
<b>1WM0</b>	X	<b>PLB</b>	partial					2004-07-01	2.90	Ostberg, T., Svensson, S., Selen, G., Uppenberg, J., Thor, M., Sundbom, M., Sydow-Backman, M., Gustavsson, A.L., Jendeberg, L.	A new class of peroxisome proliferator-activated receptor agonists with a novel binding epitope shows antidiabetic effects
<b>1ZEO</b>	A	<b>C01</b>	dual			* 210 (BDB) * 210 (BM OAD _3325 4)	* 390 (BDB)	2005-04-19	2.50	Shi, G.Q., Dropinski, J.F., McKeever, B.M., Xu, S, Becker, J.W., Berger, J.P., MacNaul, K.L., Elbrecht, A., Zhou, G., Doebber, T.W., Wang, P., Chao, Y.-S., Forrest, M., Heck, J.V., Moller, D.E., Jones, B.A.	Design and Synthesis of alpha-Aryloxyphenylacetic Acid Derivatives: A Novel Class of PPAR alpha/gamma Dual Agonists with Potent Antihyperglycemic and Lipid Modulating Activity

<b>1ZGY</b>	A	<b>BRL</b>	full	* 8-440 (BDB)	* 4-2880 (BDB)	* 30-50000 (BDB) * 500 (BM OAD _7439)	2005-04-22	1.80	Li, Y., Choi, M., Suino, K., Kovach, A., Daugherty, J., Kliewer, S.A., Xu, H.E.	Structural and biochemical basis for selective repression of the orphan nuclear receptor liver receptor homolog 1 by small heterodimer partner.
<b>2ATH</b>	A	<b>3EA</b>	full		* 230 (BDB)	* 152 (BDB) * 152 (BM OAD _3325)	2005-08-25	2.28	Mahindroo, N., Huang, C.-F., Peng, Y.-H., Wang, C.-C., Liao, C.-C., Lien, T.-W., Chittimalla, S.K., Huang, W.-J., Chai, C.-H., Prakash, E., Chen, C.-P., Hsu, T.-A., Peng, C.-H., Lu, I.-L., Lee, L.-H., Chang, Y.-W., Chen, W.-C., Chou, Y.-C., Chen, C.-T., Goparaju, C.M.V., Chen, Y.-S., Lan, S.-J., Yu, M.-C., Chen, X., Chao, Y.-S., Wu, S.-Y., Hsieh, H.-P.	Novel indole-based peroxisome proliferator-activated receptor agonists: design, SAR, structural biology, and biological activities
<b>2F4B</b>	A	<b>EHA</b>	pan		* 70 (BDB)	* 50 (BDB) * 50 (BM OAD _3325)	2005-11-23	2.07	Mahindroo, N., Wang, C.C., Liao, C.C., Huang, C.F., Lu, I.L., Lien, T.W., Peng, Y.H., Huang, W.J., Lin, Y.T., Hsu, M.C., Lin, C.H., Tsai, C.H., Hsu, J.T., Chen, X., Lyu, P.C., Chao, Y.S., Wu, S.Y., Hsieh, H.P.	Indol-1-yl Acetic Acids as Peroxisome Proliferator-Activated Receptor Agonists: Design, Synthesis, Structural Biology, and Molecular Docking Studies

<b>2FVJ</b>	A	<b>RO0</b>	partial					2006-01-31	1.99	Burgermeister, E., Schnoebelen, A., Flament, A., Benz, J., Stihle, M., Gsell, B., Rufer, A., Ruf, A., Kuhn, B., Maerki, H.P., Mizrahi, J., Sebkova, E., Niesor, E., Meyer, M.	A novel partial agonist of peroxisome proliferator-activated receptor-gamma (PPARgamma) recruits PPARgamma-coactivator-1alpha, prevents triglyceride accumulation, and potentiates insulin signaling in vitro
<b>2G0G</b>	A	<b>SP0</b>	partial				* 512 (BDB )	2006-02-13	2.54	Lu, I.L., Huang, C.F., Peng, Y.H., Lin, Y.T., Hsieh, H.P., Chen, C.T., Lien, T.W., Lee, H.J., Mahindroo, N., Prakash, E., Yueh, A., Chen, H.Y., Goparaju, C.M., Chen, X., Liao, C.C., Chao, Y.S., Hsu, J.T., Wu, S.Y.	Structure-Based Drug Design of a Novel Family of PPARgamma Partial Agonists: Virtual Screening, X-ray Crystallography, and in Vitro/in Vivo Biological Activities
<b>2G0H</b>	A	<b>SP3</b>	partial				* 22.7 (BDB ) * 22.7 (BM OAD _3325 2)	2006-02-13	2.30	Lu, I.L., Huang, C.F., Peng, Y.H., Lin, Y.T., Hsieh, H.P., Chen, C.T., Lien, T.W., Lee, H.J., Mahindroo, N., Prakash, E., Yueh, A., Chen, H.Y., Goparaju, C.M., Chen, X., Liao, C.C., Chao, Y.S., Hsu, J.T., Wu, S.Y.	Structure-Based Drug Design of a Novel Family of PPARgamma Partial Agonists: Virtual Screening, X-ray Crystallography, and in Vitro/in Vivo Biological Activities
<b>2GTK</b>	A	<b>208</b>	dual				* 250 (BM OAD _3325 3)	2006-04-28	2.10	Kuhn, B., Hilpert, H., Benz, J., Binggeli, A., Grether, U., Humm, R., Meyer, M., Mohr, P.	Structure-based design of indole propionic acids as novel PPARalpha/gamma co-agonists

<b>2HFP</b>	A	<b>NSI</b>	antagonist			* 3 (BDB) * 3 (BM OAD_33801)	2006-06-25	2.00	Hopkins, C.R., O'neil, S.V., Laifersweiler, M.C., Wang, Y., Pokross, M., Mekel, M., Evdokimov, A., Walter, R., Kontoyianni, M., Petrey, M.E., Sabatakos, G., Roesgen, J.T., Richardson, E., Demuth Jr., T.P.	Design and synthesis of novel N-sulfonyl-2-indole carboxamides as potent PPAR-gamma binding agents with potential application to the treatment of osteoporosis.
<b>2HWQ</b>	A	<b>DRY</b>	dual			* 2210 (BDB)	2006-08-01	1.97	Mahindroo, N., Peng, Y.H., Lin, C.H., Tan, U.K., Prakash, E., Lien, T.W., Lu, I.L., Lee, H.J., Hsu, J.T.A., Chen, X., Liao, C.C., Lyu, P.C., Chao, Y.S., Wu, S.Y., Hsieh, H.P.	Structural basis for the structure-activity relationships of peroxisome proliferator-activated receptor agonists
<b>2HWR</b>	A	<b>DRD</b>	dual			* 210 (BDB)	2006-08-01	2.34	Mahindroo, N., Peng, Y.H., Lin, C.H., Tan, U.K., Prakash, E., Lien, T.W., Lu, I.L., Lee, H.J., Hsu, J.T.A., Chen, X., Liao, C.C., Lyu, P.C., Chao, Y.S., Wu, S.Y., Hsieh, H.P.	Structural basis for the structure-activity relationships of peroxisome proliferator-activated receptor agonists

<b>2I4J</b>	A	<b>DRJ</b>	full	* 88 (BDB)	* 684.8 (BDB)	* 73.3 (BDB)		2006-08-22	2.10	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.	Insights into the mechanism of partial agonism: crystal structures of the peroxisome proliferator-activated receptor gamma ligand-binding domain in the complex with two enantiomeric ligands
<b>2I4P</b>	A	<b>DRH</b>	partial	* 971 (BDB)	* 1978 (BDB)	* 593 (BDB)		2006-08-22	2.10	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.	Insights into the mechanism of partial agonism: crystal structures of the peroxisome proliferator-activated receptor gamma ligand-binding domain in the complex with two enantiomeric ligands.
<b>2I4Z</b>	A	<b>DRH</b>	partial	* 971 (BDB)	* 1978 (BDB)	* 593 (BDB)		2006-08-23	2.25	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.	Insights into the mechanism of partial agonism: crystal structures of the peroxisome proliferator-activated receptor gamma ligand-binding domain in the complex with two enantiomeric ligands
<b>2OM9</b>	A	<b>AJA</b>	partial					2007-01-21	2.80	Ambrosio, A.L.B., Dias, S.M.G., Polikarpov, I., Zurier, R.B., Burstein, S.H., Garratt, R.C.	Ajulemic Acid, a Synthetic Nonpsychoactive Cannabinoid Acid, Bound to the Ligand Binding Domain of the Human Peroxisome Proliferator-activated Receptor gamma

<b>2P4Y</b>	A	<b>C03</b>	partial	* 1 (BMO AD_47 728)				2007- 03-13	2.25	Einstein, M., Akiyama, T.E., Castriota, G.A., Wang, C.F., McKeever, B., Mosley, R.T., Becker, J.W., Moller, D.E., Meinke, P.T., Wood, H.B., Berger, J.P.	The differential interactions of peroxisome proliferator-activated receptor gamma ligands with Tyr473 is a physical basis for their unique biological activities.
<b>2POB</b>	A	<b>GW4</b>	full					2007- 04-26	2.30	Trump, R.P., Cobb, J.E., Shearer, B.G., Lambert, M.H., Nolte, R.T., Willson, T.M., Buckholtz, R.G., Zhao, S.M., Leesnitzer, L.M., Iannone, M.A., Pearce, K.H., Billin, A.N., Hoekstra, W.J.	Cocrystal structure guided array synthesis of PPARgamma inverse agonists
<b>2PRG</b>	A	<b>BRL</b>	full	* 8- 440 (BDB)		* 4- 2880 (BDB)	* 30- 50000 (BDB )	1998- 08-14	2.30	Nolte, R.T., Wisely, G.B., Westin, S., Cobb, J.E., Lambert, M.H., Kurokawa, R., Rosenfeld, M.G., Willson, T.M., Glass, C.K., Milburn, M.V.	Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma.
<b>2Q59</b>	A	<b>240</b>	full				* 2 (BDB )	2007- 05-31	2.20	Bruning, J.B., Chalmers, M.J., Prasad, S., Busby, S.A., Kamenecka, T.M., He, Y., Nettles, K.W., Griffin, P.R.	Partial Agonists Activate PPARgamma Using a Helix 12 Independent Mechanism
<b>2Q5P</b>	A	<b>241</b>	partial			* 1-2 (BDB)	* 1-2 (BDB )	2007- 06-01	2.30	Bruning, J.B., Chalmers, M.J., Prasad, S., Busby, S.A., Kamenecka, T.M., He, Y., Nettles, K.W.,	Partial Agonists Activate PPARgamma Using a Helix 12 Independent Mechanism

										Griffin, P.R.	
<b>2Q5S</b>	A	<b>NZA</b>	partial			* 55 (BDB)	* 26 (BDB )	2007- 06-01	2.05	Bruning, J.B., Chalmers, M.J., Prasad, S., Busby, S.A., Kamenecka, T.M., He, Y., Nettles, K.W., Griffin, P.R.	Partial Agonists Activate PPARgamma Using a Helix 12 Independent Mechanism
<b>2Q61</b>	A	<b>SF1</b>	partial					2007- 06-04	2.20	Bruning, J.B., Chalmers, M.J., Prasad, S., Busby, S.A., Kamenecka, T.M., He, Y., Nettles, K.W., Griffin, P.R.	Partial Agonists Activate PPARgamma Using a Helix 12 Independent Mechanism
<b>2Q6R</b>	A	<b>SF2</b>	partial					2007- 06-01	2.41	Bruning, J.B., Chalmers, M.J., Prasad, S., Busby, S.A., Kamenecka, T.M., He, Y., Nettles, K.W., Griffin, P.R.	Partial Agonists Activate PPARgamma Using a Helix 12 Independent Mechanism
<b>2Q6S</b>	B	<b>PLB</b>	partial					2007- 06-01	2.40	Bruning, J.B., Chalmers, M.J., Prasad, S., Busby, S.A., Kamenecka, T.M., He, Y., Nettles, K.W., Griffin, P.R.	Partial Agonists Activate PPARgamma Using a Helix 12 Independent Mechanism



<b>2Q8S</b>	A	<b>L92</b>	dual				* 140 (BDB)	* 185 (BDB ) * 185 (BM OAD _4773 0)	2007- 06-11	2.30	Casimiro-Garcia, A., Bigge, C.F., Davis, J.A., Padalino, T., Pulaski, J., Ohren, J.F., McConnell, P., Kane, C.D., Royer, L.J., Stevens, K.A., Auerbach, B.J., Collard, W.T., McGregor, C., Fakhoury, S.A., Schaum, R.P., Zhou, H.	Effects of modifications of the linker in a series of phenylpropanoic acid derivatives: Synthesis, evaluation as PPARalpha/gamma dual agonists, and X-ray crystallographic studies.
<b>2QMV</b>	A		NMR						2007- 07-17		Moras, D., Gronemeyer, H., Gaal, V.L., Scheen, A.J., Cock, T.-A., Houten, S.M., Auwerx, Stumvoll, M., Haering, H., Uppenberg, J., Svensson, C., Jaki, M., Bertilsson, G., Jendeberg, L., Berkenstam, A.	High Resolution Structure of Peroxisome Proliferation-Activated Receptor gamma and Characterisation of its Interaction with the Co- activator Transcriptional Intermediary Factor 2
<b>2VSR</b>	A	<b>9HO</b>	full/fatt y acid						2008- 04-29	2.05	Itoh, T., Fairall, L., Amin, K., Inaba, Y., Szanto, A., Balint, B.L., Nagy, L., Yamamoto, K., Schwabe, J.W.R.	Structural Basis for the Activation of Pparg by Oxidised Fatty Acids
<b>2VST</b>	A	<b>243</b>	full/fatt y acid						2008- 04-29	2.35	Itoh, T., Fairall, L., Amin, K., Inaba, Y., Szanto, A., Balint, B.L., Nagy, L., Yamamoto, K., Schwabe, J.W.R.	Structural Basis for the Activation of Pparg by Oxidised Fatty Acids

<b>2VV0</b>	A	<b>HXA</b>	full/fatty acid					2008-06-02	2.55	Itoh, T., Fairall, L., Amin, K., Inaba, Y., Szanto, A., Balint, B.L., Nagy, L., Yamamoto, K., Schwabe, J.W.R.	Structural Basis for the Activation of Pparg by Oxidised Fatty Acids
<b>2VV1</b>	A	<b>4HD</b>	full/fatty acid					2008-06-02	2.20	Itoh, T., Fairall, L., Amin, K., Inaba, Y., Szanto, A., Balint, B.L., Nagy, L., Yamamoto, K., Schwabe, J.W.R.	Structural Basis for the Activation of Pparg by Oxidised Fatty Acids
<b>2VV2</b>	A	<b>5HE</b>	full/fatty acid					2008-06-02	2.75	Itoh, T., Fairall, L., Amin, K., Inaba, Y., Szanto, A., Balint, B.L., Nagy, L., Yamamoto, K., Schwabe, J.W.R.	Structural Basis for the Activation of Pparg by Oxidised Fatty Acids
<b>2VV3</b>	A	<b>4R8</b>	covalent full					2008-06-02	2.85	Itoh, T., Fairall, L., Amin, K., Inaba, Y., Szanto, A., Balint, B.L., Nagy, L., Yamamoto, K., Schwabe, J.W.R.	Structural Basis for the Activation of Pparg by Oxidised Fatty Acids
<b>2VV4</b>	B	<b>6OB</b>	covalent full					2008-06-02	2.35	Itoh, T., Fairall, L., Amin, K., Inaba, Y., Szanto, A., Balint, B.L., Nagy, L., Yamamoto, K., Schwabe, J.W.R.	Structural Basis for the Activation of Pparg by Oxidised Fatty Acids
<b>2VV4</b>	A	<b>6OC</b>	covalent full					2008-06-02	2.35	Itoh, T., Fairall, L., Amin, K., Inaba, Y., Szanto, A., Balint, B.L., Nagy, L., Yamamoto, K., Schwabe, J.W.R.	Structural Basis for the Activation of Pparg by Oxidised Fatty Acids
<b>2XKW</b>	A	<b>P1B</b>	full					2010-07-14	2.02	Mueller, J.J., Schupp, M., Unger, T., Kintscher, U., Heinemann, U.	Binding Diversity of Pioglitazone by Peroxisome Proliferator-Activated Receptor-

											Gamma
<b>2YFE</b>	A	<b>YFE</b>	full					2011-04-05	2.00	Weidner, C., De Groot, J.C., Prasad, A., Freiwald, A., Quedenau, C., Kliem, M., Witzke, A., Kodelja, V., Han, C.-T., Giegold, S., Baumann, M., Klebl, B., Siems, K., Mueller-Kuhrt, L., Schuermann, A., Schueller, R., Pfeiffer, A.F.H., Schroeder, F.C., Buessow, K., Sauer, S.	Amorfrutins are potent antidiabetic dietary natural products.
<b>2ZK0</b>	A		apo					2008-03-12	2.36	Waku, T., Shiraki, T., Oyama, T., Fujimoto, Y., Maebara, K., Kamiya, N., Jingami, H., Morikawa, K.	Structural insight into PPARgamma activation through covalent modification with endogenous fatty acids
<b>2ZK1</b>	A	<b>PTG</b>	full					2008-03-12	2.61	Waku, T., Shiraki, T., Oyama, T., Fujimoto, Y., Maebara, K., Kamiya, N., Jingami, H., Morikawa, K.	Structural insight into PPARgamma activation through covalent modification with endogenous fatty acids
<b>2ZK2</b>	A	<b>PTG</b>	full					2008-03-12	2.26	Waku, T., Shiraki, T., Oyama, T., Fujimoto, Y., Maebara, K., Kamiya, N., Jingami, H., Morikawa, K.	Structural insight into PPARgamma activation through covalent modification with endogenous fatty acids
<b>2ZK3</b>	A	<b>OCX</b>	covalent full					2008-03-12	2.58	Waku, T., Shiraki, T., Oyama, T., Fujimoto, Y., Maebara, K., Kamiya, N., Jingami,	Structural insight into PPARgamma activation through covalent modification with

										H., Morikawa, K.	endogenous fatty acids
<b>2ZK4</b>	A	<b>OCR</b>	covalent full					2008- 03-12	2.57	Waku, T., Shiraki, T., Oyama, T., Fujimoto, Y., Maebara, K., Kamiya, N., Jingami, H., Morikawa, K.	Structural insight into PPARgamma activation through covalent modification with endogenous fatty acids
<b>2ZK5</b>	A	<b>NRO</b>	covalent full					2008- 03-12	2.45	Waku, T., Shiraki, T., Oyama, T., Fujimoto, Y., Maebara, K., Kamiya, N., Jingami, H., Morikawa, K.	Structural insight into PPARgamma activation through covalent modification with endogenous fatty acids
<b>2ZK6</b>	A	<b>C08</b>	partial					2008- 03-12	2.41	Waku, T., Shiraki, T., Oyama, T., Maebara, K., Nakamori, R., Morikawa, K.	The nuclear receptor PPARgamma individually responds to serotonin- and fatty acid-metabolites
<b>2ZNO</b>	A	<b>S44</b>	full					2008- 04-30	2.40	Oyama, T., Toyota, K., Waku, T., Hirakawa, Y., Nagasawa, N., Kasuga, J., Hashimoto, Y., Miyachi, H., Morikawa, K.	Adaptability and selectivity of human peroxisome proliferator-activated receptor (PPAR) pan agonists revealed from crystal structures
<b>2ZVT</b>	A	<b>PTG</b>	full					2008- 11-19	1.90	Waku, T., Shiraki, T., Oyama, T., Morikawa, K.	Atomic structure of mutant PPARgamma LBD complexed with 15d-PGJ2: novel modulation mechanism of PPARgamma/RXRalpha function by covalently bound ligands

<b>3ADS</b>	A	<b>IMN</b>	full (x2)			*	50000 (BDB)	2010-01-29	2.25	Waku, T., Shiraki, T., Oyama, T., Maebara, K., Nakamori, R., Morikawa, K.	The nuclear receptor PPARgamma individually responds to serotonin- and fatty acid-metabolites
<b>3ADT</b>	A	<b>HID</b>	full (x2)					2010-01-29	2.70	Waku, T., Shiraki, T., Oyama, T., Maebara, K., Nakamori, R., Morikawa, K.	The nuclear receptor PPARgamma individually responds to serotonin- and fatty acid-metabolites
<b>3ADU</b>	A	<b>MYI</b>	full (x2)					2010-01-29	2.77	Waku, T., Shiraki, T., Oyama, T., Maebara, K., Nakamori, R., Morikawa, K.	The nuclear receptor PPARgamma individually responds to serotonin- and fatty acid-metabolites
<b>3ADV</b>	A	<b>SRO</b>	full (x2)					2010-01-29	2.27	Waku, T., Shiraki, T., Oyama, T., Maebara, K., Nakamori, R., Morikawa, K.	The nuclear receptor PPARgamma individually responds to serotonin- and fatty acid-metabolites
<b>3ADW</b>	A	<b>MYI</b>	full					2010-01-29	2.07	Waku, T., Shiraki, T., Oyama, T., Maebara, K., Nakamori, R., Morikawa, K.	The nuclear receptor PPARgamma individually responds to serotonin- and fatty acid-metabolites
<b>3ADW</b>	A	<b>OCR</b>	partner					2010-01-29	2.07	Waku, T., Shiraki, T., Oyama, T., Maebara, K., Nakamori, R., Morikawa, K.	The nuclear receptor PPARgamma individually responds to serotonin- and fatty acid-metabolites
<b>3ADX</b>	A	<b>IMN</b>	full			*	50000 (BDB)	2010-01-29	1.95	Waku, T., Shiraki, T., Oyama, T., Maebara, K., Nakamori, R., Morikawa, K.	The nuclear receptor PPARgamma individually responds to serotonin- and fatty acid-metabolites
<b>3ADX</b>	A	<b>NRO</b>	partner					2010-01-29	1.95	Waku, T., Shiraki, T., Oyama, T., Maebara, K., Nakamori, R., Morikawa, K.	The nuclear receptor PPARgamma individually responds to serotonin- and fatty acid-metabolites

<b>3AN3</b>	A	<b>M7S</b>	select full					2010-08-30	2.30	Ohashi, M., Oyama, T., Nakagome, I., Satoh, M., Nishio, Y., Nobusada, H., Hirono, S., Morikawa, K., Hashimoto, Y., Miyachi, H.	Design, Synthesis, and Structural Analysis of Phenylpropanoic Acid-Type PPAR gamma-Selective Agonists: Discovery of Reversed Stereochemistry-Activity Relationship
<b>3AN4</b>	A	<b>M7R</b>	full					2010-08-30	2.30	Ohashi, M., Oyama, T., Nakagome, I., Satoh, M., Nishio, Y., Nobusada, H., Hirono, S., Morikawa, K., Hashimoto, Y., Miyachi, H.	Design, Synthesis, and Structural Analysis of Phenylpropanoic Acid-Type PPAR gamma-Selective Agonists: Discovery of Reversed Stereochemistry-Activity Relationship
<b>3B0Q</b>	A	<b>MC5</b>	full					2011-06-13	2.10	Tomioka, D., Hashimoto, H., Sato, M., Shimizu, T.	Crystal structure of human PPAR gamma in complex with MCC555
<b>3B0R</b>	A	<b>GW9</b>	antagonist					2011-06-13	2.15	Shimizu, T.	Human PPAR gamma ligand binding domain complexed with GW9662 in a covalent bonded form
<b>3B1M</b>	A	<b>KRC</b>	partial					2011-07-05	1.60	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.	Pharmacology and in Vitro Profiling of a Novel Peroxisome Proliferator-Activated Receptor gamma Ligand, Cerco-A

<b>3B3K</b>	A	<b>LRG</b>	full			* 480-5930 (BDB)		2007-10-22	2.60	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.	Crystal Structure of the Peroxisome Proliferator-Activated Receptor gamma (PPARgamma) Ligand Binding Domain Complexed with a Novel Partial Agonist: A New Region of the Hydrophobic Pocket Could Be Exploited for Drug Design
<b>3BC5</b>	A	<b>ZAA</b>	dual (a/g)			* 4 (BDB)	* 5 (BDB)	2007-11-12	2.27	Zhang, H., Ryono, D.E., Devasthale, P., Wang, W., O'Malley, K., Farrelly, D., Gu, L., Harrity, T., Cap, M., Chu, C., Locke, K., Zhang, L., Lippy, J., Kunselman, L., Morgan, N., Flynn, N., Moore, L., Hosagrahara, V., Zhang, L., Kadiyala, P., Xu, C., Doweiko, A.M., Bell, A., Chang, C., Muckelbauer, J., Zahler, R., Hariharan, N., Cheng, P.T.	Design, synthesis and structure-activity relationships ofazole acids as novel, potent dual PPAR alpha/gamma agonists.
<b>3CDP</b>	A	<b>YRG</b>	partial			* 2700-7940 (BDB)		2008-02-27	2.80	Pochetti, G., Montanari, R., Mazza, F., Crestani, M., Godio, C., Loiodice, F., Fracchiolla, G., Lavecchia, A., Novellino, E.	Crystal structures of PPARgamma ligand binding domain complexed with new partial and full agonists: a new region of the hydrophobic pocket could be exploited for drug design

<b>3CDS</b>	A	<b>GRR</b>	full			* 3600 (BDB)		2008-02-27	2.65	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.	Crystal Structure of the Peroxisome Proliferator-Activated Receptor gamma (PPARgamma) Ligand Binding Domain Complexed with a Novel Partial Agonist: A New Region of the Hydrophobic Pocket Could Be Exploited for Drug Design
<b>3CS8</b>	A	<b>BRL</b>	full	* 8-440 (BDB)		* 4-2880 (BDB)	* 30-50000 (BDB)	2008-04-09	2.30	Li, Y., Kovach, A., Suino-Powell, K., Martynowski, D., Xu, H.E.	Structural and biochemical basis for the binding selectivity of peroxisome proliferator-activated receptor gamma to PGC-1alpha.
<b>3CWD</b>	A	<b>LNA</b>	partial/full				* 1610 (BDB)	2008-04-21	2.40	Li, Y., Zhang, J., Schopfer, F.J., Martynowski, D., Garcia-Barrio, M.T., Kovach, A., Suino-Powell, K., Baker, P.R., Freeman, B.A., Chen, Y.E., Xu, H.E.	Molecular recognition of nitrated fatty acids by PPAR gamma.
<b>3CWD</b>	A	<b>LNB</b>	partial/full			* 45-70 (BDB)	* 410 (BDB)	2008-04-21	2.40	Li, Y., Zhang, J., Schopfer, F.J., Martynowski, D., Garcia-Barrio, M.T., Kovach, A., Suino-Powell, K., Baker, P.R., Freeman, B.A., Chen, Y.E., Xu, H.E.	Molecular recognition of nitrated fatty acids by PPAR gamma.



<b>3D6D</b>	A	<b>LRG</b>	full			* 480-5930 (BDB)		2008-05-19	2.40	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.	Crystal Structure of the Peroxisome Proliferator-Activated Receptor gamma (PPARgamma) Ligand Binding Domain Complexed with a Novel Partial Agonist: A New Region of the Hydrophobic Pocket Could Be Exploited for Drug Design
<b>3DZU</b>	D	<b>PLB</b>	partial					2008-07-30	3.20	Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T.P., Rastinejad, F.	Structure of the intact PPAR-gamma-RXR-alpha nuclear receptor complex on DNA.
<b>3DZY</b>	D	<b>BRL</b>	full	* 8-440 (BDB)		* 4-2880 (BDB)	* 30-50000 (BDB)	2008-07-30	3.10	Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T.P., Rastinejad, F.	Structure of the intact PPAR-gamma-RXR-alpha nuclear receptor complex on DNA.
<b>3E00</b>	D	<b>GW9</b>	antagonist					2008-07-30	3.10	Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T.P., Rastinejad, F.	Structure of the intact PPAR-gamma-RXR-alpha nuclear receptor complex on DNA.

<b>3ET0</b>	A	<b>ET0</b>	full							2008-10-06	2.40	Artis, D.R., Lin, J.J., Zhang, C., Wang, W., Mehra, U., Perreault, M., Erbe, D., Krupka, H.I., England, B.P., Arnold, J., Plotnikov, A.N., Marimuthu, A., Nguyen, H., Will, S., Signaevsky, M., Kral, J., Cantwell, J., Settachatgull, C., Yan, D.S., Fong, D., Oh, A., Shi, S., Womack, P., Powell, B., Habets, G., West, B.L., Zhang, K.Y., Milburn, M.V., Vlasuk, G.P., Hirth, K.P., Nolop, K., Bollag, G., Ibrahim, P.N., Tobin, J.F.	Scaffold-based discovery of indeglitazar, a PPAR pan-active anti-diabetic agent
-------------	---	------------	------	--	--	--	--	--	--	------------	------	---	---

<b>3ET3</b>	A	<b>ET1</b>	pan					2008-10-06	1.95	Artis, D.R., Lin, J.J., Zhang, C., Wang, W., Mehra, U., Perreault, M., Erbe, D., Krupka, H.I., England, B.P., Arnold, J., Plotnikov, A.N., Marimuthu, A., Nguyen, H., Will, S., Signaevsky, M., Kral, J., Cantwell, J., Settachatgull, C., Yan, D.S., Fong, D., Oh, A., Shi, S., Womack, P., Powell, B., Habets, G., West, B.L., Zhang, K.Y., Milburn, M.V., Vlasuk, G.P., Hirth, K.P., Nolop, K., Bollag, G., Ibrahim, P.N., Tobin, J.F.	Scaffold-based discovery of indeglitazar, a PPAR pan-active anti-diabetic agent
<b>3FEJ</b>	A	<b>CTM</b>	full	* 740 (BMO AD_54 157)				2008-11-30	2.01	Grether, U., Benardeau, A., Benz, J., Binggeli, A., Blum, D., Hilpert, H., Kuhn, B., Maerki, H.P., Meyer, M., Mohr, P., Puntener, K., Raab, S., Ruf, A., Schlatter, D.	Design and biological evaluation of novel, balanced dual PPARalpha/gamma agonists

<b>3FUR</b>	A	<b>Z12</b>	partial	* 10 (BMO AD_54 272)				2009- 01-14	2.30	Motani, A., Wang, Z., Weiszmann, J., McGee, L.R., Lee, G., Liu, Q., Staunton, J., Fang, Z., Fuentes, H., Lindstrom, M., Liu, J., Biermann, D.H.T., Jaen, J., Walker, N.P., Learned, R.M., Chen, J.-L., Li, Y.	INT131: a selective modulator of PPAR gamma
<b>3G9E</b>	A	<b>RO7</b>	dual (a/g)				* 19 (BDB ) * 19 (BM OAD _5415 6)	2009- 02-13	2.30	Benardeau, A., Benz, J., Binggeli, A., Blum, D., Boehringer, M., Grether, U., Hilpert, H., Kuhn, B., Marki, H.P., Meyer, M., Puntener, K., Raab, S., Ruf, A., Schlatter, D., Mohr, P.	Aleglitazar, a new, potent, and balanced dual PPARalpha/gamma agonist for the treatment of type II diabetes.
<b>3GBK</b>	A	<b>2PQ</b>	full				* 50 (BDB)	2009- 02-19	2.30	Lin, C.-H., Peng, Y.-H., Coumar, M.S., Chittimalla, S.K., Liao, C.-C., Lyn, P.-C., Huang, C.-C., Lien, T.- W., Lin, W.-H., Hsu, J.T.-A., Cheng, J.-H., Chen, X., Wu, J.-S., Chao, Y.-S., Lee, H.-J., Juo, C.-G., Wu, S.-Y., Hsieh, H.-P.	Design and structural analysis of novel pharmacophores for potent and selective peroxisome proliferator-activated receptor gamma agonists

<b>3H0A</b>	D	<b>D30</b>	partial	* 33 (BDB) * 33 (BMO AD_54 154)	* 160 (BDB)	2009- 04-08	2.10	Connors, R.V., Wang, Z., Harrison, M., Zhang, A., Wanska, M., Hiscock, S., Fox, B., Dore, M., Labelle, M., Sudom, A., Johnstone, S., Liu, J., Walker, N.P., Chai, A., Siegler, K., Li, Y., Coward, P.	Identification of a PPARdelta agonist with partial agonistic activity on PPARgamma.
<b>3H00</b>	A	<b>DKD</b>	full		* 400- 860 (BDB)	2009- 06-01	2.60	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.	New 2-Aryloxy-3-phenylpropanoic Acids As Peroxisome Proliferator-Activated Receptors alpha/gamma Dual Agonists with Improved Potency and Reduced Adverse Effects on Skeletal Muscle Function
<b>3H0D</b>	A	<b>ZZH</b>	full		* 572- 580 (BDB)	2009- 06-02	2.10	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.	New 2-Aryloxy-3-phenylpropanoic Acids As Peroxisome Proliferator-Activated Receptors alpha/gamma Dual Agonists with Improved Potency and Reduced Adverse Effects on Skeletal Muscle Function

<b>3IA6</b>	A	<b>UNT</b>	dual (a/g)			* 3 (BDB ) * 3 (BM OAD _5426 7)	2009- 07-13	2.31	Casimiro-Garcia, A., Bigge, C.F., Davis, J.A., Padalino, T., Pulaski, J., Ohren, J.F., McConnell, P., Kane, C.D., Royer, L.J., Stevens, K.A., Auerbach, B., Collard, W., McGregor, C., Song, K.	Synthesis and evaluation of novel alpha-heteroaryl- phenylpropanoic acid derivatives as PPARalpha/gamma dual agonists.
<b>3K8S</b>	A	<b>Z27</b>	partial				2009- 10-14	2.55	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.M., Chen, J.-L.	T2384, a novel antidiabetic agent with unique peroxisome proliferator- activated receptor gamma binding properties
<b>3KMG</b>	A	<b>538</b>	partial/f ull				2009- 11-10	2.10	Lamotte, Yann, Martres, Paul, Faucher, Nicolas, Laroze, Alain, Grillot, Didier, Ancellin, Nicolas, Saintillan, Yannick, Beneton, Veronique, Gampe, Robert	Synthesis and biological activities of novel indole derivatives as potent and selective PPAR-gamma modulators
<b>3LMP</b>	A	<b>CEK</b>	partial				2010- 01-31	1.90	Furukawa, A., Arita, T., Satoh, S., Wakabayashi, K., Hayashi, S., Matsui, Y., Araki, K., Kuroha, M., Ohsumi, J.	Discovery of a novel selective PPARgamma modulator from (-)- Cercosporamide derivatives

<b>3NOA</b>	A	<b>5BC</b>	full					2010-06-25	1.98	Peng, Y.H., Coumar, M.S., Leou, J.S., Wu, J.S., Shiao, H.Y., Lin, C.H., Lyu, P.C., Hsieh, H.P., Wu, S.Y.	Crystal structure of human PPAR-gamma ligand binding domain complex with a potency improved agonist
<b>3OSI</b>	A	<b>XDH</b>	partial					2010-09-09	2.70	Riu, A., Grimaldi, M., le Maire, A., Bey, G., Phillips, K., Boulahtouf, A., Perdu, E., Zalko, D., Bourguet, W., Balaguer, P.	Peroxisome Proliferator-Activated Receptor Gamma is a Target for Halogenated Analogues of Bisphenol-A.
<b>3OSW</b>	A	<b>XDI</b>	partial					2010-09-10	2.55	Riu, A., Grimaldi, M., le Maire, A., Bey, G., Phillips, K., Boulahtouf, A., Perdu, E., Zalko, D., Bourguet, W., Balaguer, P.	Peroxisome Proliferator-Activated Receptor Gamma is a Target for Halogenated Analogues of Bisphenol-A.
<b>3PBA</b>	A	<b>ZXG</b>	partial					2010-10-20	2.30	Riu, A., le Maire, A., Grimaldi, M., Audebert, M., Hillenweck, A., Bourguet, W., Balaguer, P., Zalko, D.	Characterization of Novel Ligands of ER{alpha}, Er{beta}, and PPAR{gamma}: The Case of Halogenated Bisphenol A and Their Conjugated Metabolites.
<b>3PRG</b>	A		apo					1998-08-24	2.90	Uppenberg, J., Svensson, C., Jaki, M., Bertilsson, G., Jendeberg, L., Berkenstam, A.	Crystal structure of the ligand binding domain of the human nuclear receptor PPARgamma.
<b>3QT0</b>	A	<b>486</b>						2011-02-22	2.50	Lin, S., Han, Y., Rong, H., Zheng, S., Lin, S.-C., Li, Y.	Revealing a steroid receptor ligand as a unique PPARgamma agonist

<b>3R5N</b>	A	<b>MLO</b>						2011-03-18	2.00	Zhang, H., Xu, X., Chen, L., Chen, J., Hu, L., Jiang, H., Shen, X.	Crystal structure of PPARgammaLBD complexed with the agonist magnolol
<b>3R8A</b>	A	<b>HIG</b>				* 591 (BDB)		2011-03-23	2.41	Casimiro-Garcia, A., Filzen, G.F., Flynn, D., Bigge, C.F., Chen, J., Davis, J.A., Dudley, D.A., Edmunds, J.J., Esmail, N., Geyer, A., Heemstra, R.J., Jalaie, M., Ohren, J.F., Ostroski, R., Ellis, T., Schaum, R.P., Stoner, C.	X-ray crystal structure of the nuclear hormone receptor PPAR-gamma in a complex with a compound with dual PPAR gamma agonism and Angiotensin II Type I receptor antagonism activity
<b>3R8I</b>	A	<b>XCX</b>						2011-03-24	2.30	Porcelli, L., Gilardi, F., Laghezza, A., Piemontese, L., Mitro, N., Azzariti, A., Altieri, F., Cervoni, L., Fracchiolla, G., Giudici, M., Guerrini, U., Lavecchia, A., Montanari, R., Di Giovanni, C., Paradiso, A., Pochetti, G., Simone, G.M., Tortorella, P., Crestani, M., Loiodice, F.	Crystal Structure of PPARgamma with an achiral ureidofibrate derivative (RT86)
<b>3S9S</b>	A	<b>M0T</b>						2011-06-02	2.55	Sime, M., Allan, A.C., Chapman, P., Fieldhouse, C., Giblin, G.M., Healy, M.P., Lambert, M.H., Leesnitzer, L.M., Lewis, A., Merrihew,	Ligand binding domain of PPARgamma complexed with a benzimidazole partial agonist



										R.V., Rutter, R.A., Sasse, R., Shearer, B.G., Wilson, T.M., Xu, R.X., Virley, D.J.	
<b>3SZ1</b>	A	<b>KNA</b>						2011-07-18	2.30	Puhl, A.C., Bernardes, A., Silveira, R.L., Yuan, J., Campos, J.L., Saidemberg, D.M., Palma, M.S., Cvaro, A., Ayers, S.D., Webb, P., Reinach, P.S., Skaf, M.S., Polikarpov, I.	Human PPAR gamma ligand binding domain in complex with luteolin and myristic acid
<b>3SZ1</b>	B	<b>LU2</b>						2011-07-18	2.30	Puhl, A.C., Bernardes, A., Silveira, R.L., Yuan, J., Campos, J.L., Saidemberg, D.M., Palma, M.S., Cvaro, A., Ayers, S.D., Webb, P., Reinach, P.S., Skaf, M.S., Polikarpov, I.	Human PPAR gamma ligand binding domain in complex with luteolin and myristic acid
<b>3SZ1</b>	B	<b>MYR</b>						2011-07-18	2.30	Puhl, A.C., Bernardes, A., Silveira, R.L., Yuan, J., Campos, J.L., Saidemberg, D.M., Palma, M.S., Cvaro, A., Ayers, S.D., Webb, P., Reinach, P.S., Skaf, M.S., Polikarpov, I.	Human PPAR gamma ligand binding domain in complex with luteolin and myristic acid
<b>3T03</b>	A	<b>3T0</b>						2011-07-19	2.10	Amato, A.A., Rajagopalan, S., Lin, J.Z., Carvalho, B.M., Figueira, A.C., Lu, J., Ayers, S.D., Mottin, M., Silveira, R.L., Souza, P.C., Mourao, R.H., Saad, M.J., Togashi, M., Simeoni,	Crystal structure of PPAR gamma ligand binding domain in complex with a novel partial agonist GQ-16

										L.A., Abdalla, D.S., Skaff, M.S., Polikarpov, I., Lima, M.C., Galdino, S.L., Brennan, R.G., Baxter, J.D., Pitta, I.R., Webb, P., Phillips, K.J., Neves, F.A.		
<b>3TY0</b>	A	<b>082</b>					* 83.1 (BDB ) * 83.1 (BM OAD _4832 ) * 83.1 (PDB bind)	* 10- 360 (BDB)	2011- 09-23	2.00	Liu, W., Lau, F., Liu, K., Wood, H.B., Zhou, G., Chen, Y., Li, Y., Akiyama, T.E., Castriota, G., Einstein, M., Wang, C., McCann, M.E., Doebber, T.W., Wu, M., Chang, C.H., McNamara, L., McKeever, B., Mosley, R.T., Berger, J., Meinke, P.T.	Structure of PPARgamma ligand binding domain in complex with (R)-5-(3-((3-(6-methoxybenzo[d]isoxazol-3-yl)-2-oxo-2,3-dihydro-1H-benzo[d]imidazol-1-yl)methyl)phenyl)-5-methyloxazolidine-2,4-dione
<b>3U9Q</b>	A	<b>DKA</b>				* 41700 (BMO AD_48 52)			2011- 10-19	1.52	Malapaka, R.R., Khoo, S., Zhang, J., Choi, J.H., Zhou, X.E., Xu, Y., Gong, Y., Li, J., Yong, E.L., Chalmers, M.J., Chang, L., Resau, J.H., Griffin, P.R., Chen, Y.E., Xu, H.E.	Ligand binding domain of PPARgamma complexed with Decanoic Acid and PGC-1a peptide
<b>3V9T</b>	A	<b>17L</b>						* 240 (BDB)	2011- 12-28	1.65	Furukawa, A., Arita, T., Fukuzaki, T., Satoh, S., Mori, M., Honda, T., Matsui, Y., Wakabayashi, K., Hayashi, S., Araki, K., Ohsumi, J.	Crystal structure of the PPARgamma-LBD complexed with a cercosporamide derivative modulator

<b>3V9V</b>	A	<b>21L</b>				* 130 (BDB)		2011-12-28	1.60	Furukawa, A., Arita, T., Fukuzaki, T., Satoh, S., Mori, M., Honda, T., Matsui, Y., Wakabayashi, K., Hayashi, S., Araki, K., Ohsumi, J.	Crystal structure of the PPARgamma-LBD complexed with a cercosporamide derivative modulator
<b>3V9Y</b>	A	<b>24L</b>				* 67 (BDB)		2011-12-28	2.10	Furukawa, A., Arita, T., Fukuzaki, T., Satoh, S., Mori, M., Honda, T., Matsui, Y., Wakabayashi, K., Hayashi, S., Araki, K., Ohsumi, J.	Crystal structure of the PPARgamma-LBD complexed with a cercosporamide derivative modulator
<b>3VJH</b>	A	<b>J35</b>						2011-10-20	2.22	Kuwabara, N., Oyama, T., Tomioka, D., Ohashi, M., Yanagisawa, J., Shimizu, T., Miyachi, H.	Human PPAR GAMMA ligand binding domain in complex with JKPL35
<b>3VJI</b>	A	<b>J53</b>						2011-10-20	2.61	Kuwabara, N., Oyama, T., Tomioka, D., Ohashi, M., Yanagisawa, J., Shimizu, T., Miyachi, H.	Human PPAR gamma ligand binding domain in complex with JKPL53
<b>3VN2</b>	A	<b>TLS</b>				* 1520-4060 (BDB)	* 12200 (BDB)	2011-12-21	2.18	Amano, Y., Yamaguchi, T., Ohno, K., Niimi, T., Orita, M., Sakashita, H., Takeuchi, M.	Crystal Structure of PPARgamma complexed with Telmisartan
<b>3VSO</b>	A	<b>EK1</b>						2012-04-30	2.00	Ohashi, M., Oyama, T., Putranto, E.W., Waku, T., Nobusada, H., Kataoka, K., Matsuno, K., Yashiro, M., Morikawa, K., Huh,	Human PPAR gamma ligand binding domain in complex with a gamma selective agonist mekt21

										N.H., Miyachi, H.	
<b>3VSP</b>	A	<b>EK8</b>						2012-04-30	2.40	Oyama, T., Waku, T., Ohashi, M., Morikawa, K., Miyachi, H.	Human PPAR gamma ligand binding domain in complex with a gamma selective agonist mekt28
<b>4A4V</b>	A	<b>YFD</b>						2011-10-20	2.00	De Groot, J.C., Weidner, C., Krausze, J., Kawamoto, K., Schroeder, F.C., Sauer, S., Buessow, K.	Ligand binding domain of human PPAR gamma in complex with amorfrutin 2
<b>4A4W</b>	A	<b>YFB</b>						2011-10-20	2.00	De Groot, J.C., Weidner, C., Krausze, J., Kawamoto, K., Schroeder, F.C., Sauer, S., Buessow, K.	Ligand binding domain of human PPAR gamma in complex with amorfrutin B
<b>4E4K</b>	A	<b>RRG</b>						2012-03-13	2.50	Laghezza, A., Pochetti, G., Lavecchia, A., Fracchiolla, G., Faliti, S., Piemontese, L., Di Giovanni, C., Iacobazzi, V., Infantino, V., Montanari, R., Capelli, D., Tortorella, P., Loiodice, F.	Crystal Structure of PPARgamma with the ligand JO21
<b>4E4Q</b>	A	<b>RRH</b>						2012-03-13	2.50	Laghezza, A., Pochetti, G., Lavecchia, A., Fracchiolla, G., Faliti, S., Piemontese, L., Di Giovanni, C., Iacobazzi, V., Infantino, V., Montanari, R., Capelli, D., Tortorella, P., Loiodice, F.	Crystal structure of PPARgamma with the ligand FS214
<b>4EM9</b>	A	<b>KNA</b>						2012-04-11	2.10	Liberato, M.V., Nascimento, A.S.,	Human PPAR gamma in complex with nonanoic

										Ayers, S.D., Lin, J.Z., Cvoro, A., Silveira, R.L., Martinez, L., Souza, P.C., Saidemberg, D., Deng, T., Amato, A.A., Togashi, M., Hsueh, W.A., Phillips, K., Palma, M.S., Neves, F.A., Skaf, M.S., Webb, P., Polikarpov, I.	acids
<b>4EMA</b>	A	<b>BRL</b>						2012-04-11	2.55	Liberato, M.V., Nascimento, A.S., Ayers, S.D., Lin, J.Z., Cvoro, A., Silveira, R.L., Martinez, L., Souza, P.C., Saidemberg, D., Deng, T., Amato, A.A., Togashi, M., Hsueh, W.A., Phillips, K., Palma, M.S., Neves, F.A., Skaf, M.S., Webb, P., Polikarpov, I.	Human peroxisome proliferator-activated receptor gamma in complex with rosiglitazone
<b>4F9M</b>	A	<b>FCM</b>				* 0.94-12 (BDB)		2012-05-19	1.90	Furukawa, A., Arita, T., Fukuzaki, T., Mori, M., Honda, T., Satoh, S., Matsui, Y., Wakabayashi, K., Hayashi, S., Nakamura, K., Araki, K., Kuroha, M., Tanaka, J., Wakimoto, S., Suzuki, O., Ohsumi, J.	Crystal structure of the PPARgamma-LBD complexed with a cercosporamide derivative modulator
<b>4FGY</b>	A	<b>0W3</b>						2012-06-05	2.84	Zheng, W., Feng, X., Qiu, L., Pan, Z., Wang, R., Lin, S., Hou, D.,	Identification of a unique PPAR ligand with an unexpected binding mode

										Jin, L., Li, Y.	and antibiotic activity
<b>4PRG</b>	A	<b>072</b>		* 70 (PDBb ind)				1999- 05-07	2.90	Oberfield, J.L., Collins, J.L., Holmes, C.P., Goreham, D.M., Cooper, J.P., Cobb, J.E., Lenhard, J.M., Hull-Ryde, E.A., Mohr, C.P., Blanchard, S.G., Parks, D.J., Moore, L.B., Lehmann, J.M., Plunket, K., Miller, A.B., Milburn, M.V., Kliewer, S.A., Willson, T.M.	0072 PARTIAL AGONIST PPAR GAMMA COCRYSTAL

## 8.3 Appendix C: Virtual screening process

### 8.3.1 Steps for process

The basic steps for performing VS using the methods divulged in this dissertation are as follows:

1. Download PDB structure complexed with a ligand.
2. Perform re-docking to test predictability of the structure.
3. Based on a catalogue of known active compounds (either from the PDB or from DUD), perform additional docking (cross-docking) to assess predictability with non-native ligands.
4. Repeat steps 1 through 3 for additional structure models under consideration for unknown screening. If using multiple structure models, superimpose the structures before performing docking.
5. Perform ligand-based pharmacophore screening using PharmaGist or Molecular Operating Environment (MOE).
  - a. Given a multi-mol2 of ligands, PharmaGist (<http://bioinfo3d.cs.tau.ac.il/PharmaGist/>) will provide sets of features that can then be used to generate 3D models in MOE by selecting the pivot molecule as the representative for the feature set. The pharmacophore tools in MOE will automatically generate features based on the ligand provided and that set can be adjusted to match the common features proposed by PharmaGist.
  - b. If a multi-mol2 file of ligands is imported as a database in MOE (see tutorials in MOE for making a database), the pharmacophore elucidation tool can be used to find common features. Groups of ligands with similar activity type should be used to do this (similar to the PharmaGist step).
6. Compare the created 3D pharmacophore models to the database of unknown compounds in MOE. MOE will generate a database of the compounds that match the pharmacophore model, which you can open and save as a text file for further manipulation. A separate database would need to be generated for each 3D model compared.
7. Using the scripts and commands provided below (step 1 and step 2), prepare the unknown ligands for screening by creating PDBQT files, upload these and the PDBQT files for the receptor models to the supercomputer. A copy of the AutoDock Vina (Vina) executable should be uploaded as well.
8. The grid box tools in AutoDock Tools (ADT) should be used to determine the grid box center and size with a 1.0 Å grid space. These values should be recorded into a configuration file (config.txt) to use for docking with Vina. The configuration file should be uploaded to the super computer as well.
9. Use the shell script in step 4 as a guide for running the docking on a supercomputer. Multiple scripts would need to be created for dockings with the dataset of ligands divided into multiple directories.
10. The scripts in step 5 are for post processing of the docking output files. The table of energies and list of scored poses relative to key residues should be downloaded to determine binders. This data combined should be combined with the ligand

- pharmacophore lists to determine binders. The data can be imported into a spreadsheet program (like Microsoft Excel) to sort the data.
- a. Sort by presence of key interactions.
  - b. Sort by free energy of binding (most negative to least negative)
11. Use MOE and/or the ToxAlert server (<https://ochem.eu/home/show.do>) to determine toxicity for the top binders on the list.
  12. An online search for patents and publications that reference the top binders and the protein of interest will determine novelty of the selected binders.

### 8.3.2 Scripts for performing virtual screening

Below are the Shell and Perl scripts used to perform VS with a large ligand database. The scripts were designed to work with a ligand database numbering in the thousands and more than one protein receptor model into which all the ligands were docked. The scripts can be adjusted for use with smaller databases or a single receptor model. Shells scripts for execution on a supercomputer have the #PBS special character in the header of the script. Headers also include a short description of the script and any additional scripts or files necessary for execution. Shell scripts that should be executed on an individual machine (not the supercomputer) are executed with the prefix “sh”. Perl scripts require the “perl” prefix to run.

#### 8.3.3 Step 1: Prepare the ligand files for docking (prepare\_ligand.sh)

```
# Must have split_multi_mol2_file.py and prepare_ligand4.py in
# working directory.
# These scripts can be found within the AutoDock4 python scripts
# or in the online tutorial: http://autodock.scripps.edu/faqs-
help/tutorial/using-autodock4-for-virtual-screening

# Ask user for path in which files are to be created.
# Path should also contain the multi-mol2 file with all the
# ligands for docking
echo "Enter the path of the VS Directory:"
read VSRROOT
cd $VSRROOT

# Make directory for the ligand PDBQT files
mkdir Ligands
echo "Enter the name of the multi-mol2 file (EXCLUDING the .mol2 extension):"
read ligs
cd $VSRROOT/Ligands

# create links to the python script to split the multi-mol2 and
# the multi-mol2 into the Ligands directory
# This section can be commented out if the ligands are already
# split into individual mol2 files.
ln -s $VSRROOT/split_multi_mol2_file.py .
ln -s $VSRROOT/"$ligs".mol2 .
# Path for the pythonsh executable may be different depending on
# how MGLTools was installed. This should be checked before
```



```

# running, or is a good place to start for debugging if
# the script doesn't work.
/Library/MGLTools/1.5.6rc2/bin/pythonsh split_multi_mol2_file.py -i
"$ligs".mol2
\rm "$ligs".mol2
\rm split_multi_mol2_file.py

# Use the prepare_ligand4.py to prepare a PDBQT for each ligand
for f in `ls *.mol2`; do
    echo $f
    /Library/MGLTools/1.5.6rc2/bin/pythonsh $VSRROOT/prepare_ligand4.py -l $f
-d $VSRROOT/ligand_dict.py
done

```

### 8.3.4 Step 2: command line items for dividing ligand files into separate directories for VS on supercomputer.

```

# This implies parallel serial runs of the AutoDock Vina program
# To incorporate this into the shell script, create a loop
# with a counter. The max value should be the total number of
# ligand files divided by 300. This will be the number of
# directories created.
#mv `ls | head -300` directory_name

# If lists are created first...
# (Helpful if docking stops unexpectedly on supercomputer and you need to pick
up where you left off.
# For this, list the completed dockings into a text file and delete these from
the full list of ligands.)
#cat dockings1_done_list.txt | xargs -I % mv % directory_name

# Ligands and receptor(s) files should be uploaded to
# the supercomputer. Lists of each ligand directory should be
# made to use with the lsvs_mpi.sh script.
# Create config.txt that contains the size and center location
# for the grid box.
# The box information should be established using ADT.

# Example config.txt
#center_x = 18.656
#center_y = -21.929
#center_z = 7.715
#size_x = 30
#size_y = 40
#size_z = 40

```

### 8.3.5 Step 3: Large-scale screening script for running serial dockings on a supercomputer (lsvs\_mpi.sh).

```

# The shell script should be copied, one for each directory of

```

```

# ligands.
#!/bin/bash

## The time limit for the job. If it runs over this it will be killed.
## -lwalltime=0:05:0 would be five minutes.
# A ligand directory containing 300 ligands should take about 48 hours on 6
processors.
# This estimate was calculated using HokieOne
#PBS -l walltime=00:48:00:00

# Set the number of nodes, and the number of processors per node (generally
should be 6)
# HokieOne: 6 processors per node
# BlueRidge: 16 processors per node
#PBS -l nodes=1:ppn=6

#PBS -W group_list=hokieone
#PBS -q normal_q
#PBS -A hokieone ## set account to bevanlab for BlueRidge

# Uncomment and add your email address to get an email when your job starts,
completes, or aborts
##PBS -M snl@vt.edu
##PBS -m bea

## Get in to the directory qsub was run from
cd $PBS_O_WORKDIR

# Run the MPI program mpiProg. The -np flag tells MPI how many processes to
use. $PBS_NP
# is an environment variable that holds the number of processes you requested.
So if you
# selected nodes=4:ppn=6 above, $PBS_NP will hold 24.
VSRROOT="/home/lewisn/msusdrugs_lsvs"
for f in `cat ligand1.list`; do
    LIG=`basename $f .pdbqt`

    for g in `cat receptor.list`; do
        REC=`basename $g .pdbqt`
        $VSRROOT/vina --config $VSRROOT/config.txt --receptor $VSRROOT/Receptors/"$g"
        --ligand $VSRROOT/Ligands/"$f" --log
        $VSRROOT/Dockings11/"${LIG}"_"${REC}"_log.txt --out
        $VSRROOT/Dockings11/"${LIG}"_"${REC}"_out.pdbqt --cpu $PBS_NP
    done
done
exit;

```

### 8.3.6 Step 4: Post processing of the docked poses (post-process.sh).

```

# Requires two text files: directory.list and receptor.list
# first contains list of directory names, second contains list of receptor
filenames

```

```

# Also need list of interactions for which to search relative to docked poses
(key_intxns_list.txt)
# list should be in the format: HIS 449 NE2 (3-letter residue name, residue
number, atom name)
# Script requires five perl scripts in working directory (included below):
# 1. vina_pdbqt2pdb_multi.pl
# 2. vina_parse_E.pl
# 3. all_energy_pose_multi.pl
# 4. calc_all_dist_ato.pl
# 5. weighted_score_key_intxns.pl

#!/bin/bash

## The time limit for the job. If it runs over this it will be killed.
## -lwalltime=0:05:0 would be five minutes.
#PBS -l walltime=00:12:00:00

# Set the number of nodes, and the number of processors per node (generally
should be 6)
#PBS -l nodes=1:ppn=6

#PBS -W group_list=hokieone
#PBS -q normal_q
#PBS -A hokieone

# Uncomment and add your email address to get an email when your job starts,
completes, or aborts
#PBS -M snl@vt.edu
#PBS -m bea

## Get in to the directory qsub was run from
cd $PBS_O_WORKDIR

# Run the MPI program mpiProg. The -np flag tells MPI how many processes to
use. $PBS_NP
# is an environment variable that holds the number of processes you requested.
So if you
# selected nodes=4:ppn=6 above, $PBS_NP will hold 24.

# recursive running of perl script to calculate distances for docked poses
relative to protein.

# needed paths
VSPATH=/home/lewisn/msusdrugs_lsvs
RECPATH=$VSPATH/Receptors

cd $VSPATH

# isolate lowest energy pose and create PDB file
for j in `cat $VSPATH/directory.list`; do
    perl vina_pdbqt2pdb_multi.pl "$j"
done

# generate list of energies from lowest energy poses
for i in `cat $VSPATH/directory.list`; do

```

```

    cd "$i"
    perl $VSPATH/vina_parse_E.pl pdb_files
    cd $VSPATH
done

# create table of energies
perl all_energy_pose_multi.pl directory.list table_energies.txt

# need to reorganize files into folders with receptor name
for b in `cat receptor.list`; do
    DIR=`basename $b .pdbqt`
    mkdir $DIR
    cd $VSPATH/$DIR
    mv $VSPATH/Dockings*/pdb_files/*_$DIR*.pdb ./
    cd $VSPATH
done

# remove pdb_files directories as no longer needed
rm -r $VSPATH/Dockings*/pdb_files/

# calculate distances
for g in `cat $VSPATH/receptor.list`; do
    DIR=`basename $g .pdbqt`
    cd $VSPATH/$DIR
    perl $VSPATH/calc_all_dist_auto.pl 4.0 $RECPATH/"$g" $VSPATH/$DIR
    cd $VSPATH
done

# find poses with desired interactions
for f in `cat $VSPATH/receptor.list`; do
    DIR=`basename $f .pdbqt`
    cd $VSPATH/$DIR
    cp $VSPATH/key_intxn_list.txt ./
    perl $VSPATH/weighted_score_key_intxns.pl key_intxn_list.txt distances
"$DIR"_scored.txt "$DIR"_success.txt
    rm ./key_intxn_list.txt
    cd $VSPATH
    echo "Finished interaction scoring for $DIR!"
done

exit;

```

### 8.3.7 vina\_pdbqt2pdb\_multi.pl

=head1

Objective: Write PDB formatted text file using PDBQT formatted text file.

Input: multi-PDBQT Vina output file. Coded to process one file at a time.

Can be changed to process multiple files in this format. Sections are commented out

Output: PDB file with same file name

Format for input:

```
MODEL 1
```

```

    HETATM    1  C7  735 A 469      25.676  17.012  -1.075  1.00 17.50
0.260 C
    HETATM    2  O8  735 A 469      24.450  17.131  -1.095  1.00 19.39  -
0.267 OA
    etc.
    ENDMDL

```

Execution: perl vina\_pdbqt2pdb.pl

1Apr11

Written by Nikki Lewis

18Dec12 added conditional to read in .pdb files as well as .pdbqt for parsing energy information.

2Jan12 edit to read in arguments with script execution for scripting on HOKIEONE

new usage: perl vina\_pdbqt2pdb\_multi.pl directory  
=cut

```

#!C:/Perl/bin -w  # for PC
#!/usr/bin/perl -w # for Unix

```

```
use strict;
```

```

#-- ask user for directory name
#-- read in all text file names into array
my $dir = $ARGV[0];
opendir DIR, $dir or die "Directory $dir could not be opened.\n"; # open
directory
my @file_array = grep{$_ =~ /\.(pdbqt)/} readdir DIR; # get file names

```

```

#-- report number of files
my $count = @file_array;
print "$count PDBQT files were found.\n";

```

```

#-- make directory to hold PDB files
my $path = 'pdb_files';
mkdir "$dir/$path" or print "$dir/$path exists\n";

```

```

for(my $i=0; $i<@file_array; $i++)
{
    my($output,$file_contents_ref);
    ($output,$file_contents_ref) = make_pdb($dir,$file_array[$i]);

    open(OUT, ">$dir/$path/$output");
    my @temp = @$file_contents_ref;
    for(my $j=0; $j<@temp; $j++)
    {
        print OUT $temp[$j];
    }
    print OUT "END\n";
    close OUT;
}

```

```
print "Finished!\n";
```

```

# subroutine input: input file name and path
# output: output file name and contents

sub make_pdb {
  my ($dir,$input) = @_ ;
  my @file_contents;

  # create output file for PDB using PDBQT file name
  my $output = $input;
  $output =~ s/pdbqt/pdb/;
  # IMPORTANT: each iteration of code will overwrite output previously in
  # named output file

  # open PDBQT file
  unless (open(TEXTFILE, "$dir/$input")) # open text file
  {
    print "Could not open file $input!\n";
  }

  my $line; # comment out this for multiple files
  # read in each line
  while($line = <TEXTFILE>)
  {
    if($line =~ /(MODEL 1)/)
    {
      push(@file_contents,"REMARK MODEL 1\n");
    }
    elsif($line =~ /(REMARK VINA RESULT)/)
    {
      push(@file_contents,$line);
    }
    # look for HETATM lines
    elsif($line =~ /(HETATM)/ || $line =~ /(ATOM)/)
    {
      my $charge = substr($line,70,6);
      $line =~ s/$charge/      /;
      push(@file_contents,$line);
    }
    elsif($line =~ /(ENDMDL)/)
    {
      last;
    }
  }

  # close text file after recording energy value
  close TEXTFILE;

  return ($output,\@file_contents);
}

exit;

```

### 8.3.8 vina\_parse\_E.pl

```
=head1
Print the model number and Vina energy value line from each PDBQT output file
in a directory.
25Oct12 NLewis
Usage: perl vina_parse_E.pl dir_name
=cut1

#!C:/Perl/bin -w    # for PC
#!/usr/bin/perl -w  # for Unix

use strict;

# read in directory name with script execution line
my $dir = $ARGV[0];

# open $dir and read in file contents with .pdbqt extention
opendir DIR, $dir or die "Directory $dir could not be opened.\n";
my @file_array = grep{$_ =~ /\.(pdbqt)|(.pdb)/} readdir DIR;

# report number of pdbqt files read
my $count = scalar(@file_array);
print "$count files found.\n";

# generate and open output file
open(OUT, ">parsed_energies.txt");

# open each file, record model and energies in array, print file name and
energy values to output file
my %file_energies;
for(my $i=0; $i<@file_array; $i++)
{
    unless(open(TEXTFILE, "$dir/$file_array[$i]"))
    {
        print "Could not open file $file_array[$i]!\n";
    }

    my @model;
    my @energy;
    while(my $line = <TEXTFILE>)
    {
        if($line =~ /^(REMARK VINA RESULT)/)
        {
            chomp $line;
            my @blah = split(':', $line);
            $blah[1] =~ s/^\s+//; # remove leading spaces
            my @temp = split(/\s+/, $blah[1]); # index [0] should have energy
value
            push(@energy, "$temp[0]\t"); # push energy value into array
        }
        else {}
    }
}
```

```

# print all to output file
#print OUT "$file_array[$i]\t@energy\n";

# print 1st/lowest energy to output
print OUT "$file_array[$i]\t$energy[0]\n";

    close TEXTFILE;
}

close OUT;
exit;

```

### 8.3.9 all\_energy\_pose\_multi.pl

=head

Script to read parsed energy lists for multiple docked poses and report most negative energy value and pose.

Specific to file structure with directories named by PDB ID.

Input: list of directory names containing "parsed\_energies.txt"

EXAMPLE: ligand\_protein\_out.pdb \t -#.# \n

Output: tab delimited table of all energy values

usage: perl all\_energy\_pose.pl dir\_name\_list.txt output\_name.txt

19Dec12 NLewis

2Jan13 original script adjusted for use on HOKIEONE where file structure is not dependent on PDB ID as directory name

new usage: perl all\_energy\_pose\_multi.pl directory.list output\_name.txt

=cut

```
#!/C:/Perl/bin -w
```

```
#!/usr/bin/perl -w
```

```
use strict;
```

```
# directory name list
```

```
my $input = $ARGV[0];
```

```
unless ( open(INFILE, "$input") )
```

```
{
```

```
    print "Could not open directory list $input\n";
```

```
}
```

```
my @directories = <INFILE>;
```

```
close INFILE;
```

```
chomp (@directories);
```

```
my $output = $ARGV[1];
```

```
my @array_addy;
```

```
my %pose_name;
```

```
my %rec_names;
```

```
# set output file name
```

```
open(OUT, ">$output");
```



```

for(my $i=0; $i<@directories; $i++)
{
    opendir DIR, $directories[$i] or die "Directory $directories[$i] could not
be opened.\n";

    # open energy list file
    unless (open (ENERGY, "$directories[$i]/parsed_energies.txt"))
    {
        print "Could not open energy file
$directories[$i]/parsed_energies.txt\n";
    }

    my $line;
    while ($line = <ENERGY>)
    {
        chomp $line;
        my ($pose, $energy) = split(/\t/, $line);
        my ($lig, $no, $pro, $out);
        my $new_lig;

        if($pose =~ /(\_d_)/)
        {
            ($lig, $no, $pro, $out) = split('_', $pose);
            $new_lig = $lig."_".$no;
        }
        else
        {
            ($lig, $pro, $out) = split('_', $pose);
            $no = '';
            $new_lig = $lig;
        }
        # now have ligand name

        if(exists $rec_names{$pro}) {} # record receptor names
        else
        { $rec_names{$pro} = 1; }

        if(exists $pose_name{$new_lig}) # look for ligand name
        {
            if(exists $pose_name{$new_lig}->{$pro}) # look for protein name
            { $pose_name{$new_lig}->{$pro} .= ",$energy"; }
            else
            { $pose_name{$new_lig}->{$pro} = $energy; }
        }
        else # if ligand not in hash, set hash for holding proteins and energies
        {
            my %hash;
            $hash{$pro} = $energy;
            $pose_name{$new_lig} = \%hash;
        }
    }
}

```

```

foreach my $rec (sort keys %rec_names)
{
    print OUT "\t$rec";
}
print OUT "\n";

foreach my $key (sort keys %pose_name) # sort by ligand name
{
    print OUT "$key";
    my $ref = $pose_name{$key};
    my %hash = %$ref;
    foreach my $k (sort keys %hash) # sort by protein name
    {
        print OUT "\t$hash{$k}";
    }
    print OUT "\n";
}

close OUT;

exit;

```

### 8.3.10 calc\_all\_dist\_auto.pl

=head1

Calculate the distance between the docked conformation of a ligand and all residues of the macromolecule within a certain threshold.  
 Edited version of calc\_atom\_dist\_auto\_2Nov10.pl

Input File Format Examples:

Ligand file

```

HETATM    1  OAB DRG    1      9.181  -6.482  37.145  1.00 20.00
puni O
or
ATOM ...etc...

```

Protein file

```

ATOM      1  N  MET A 206      14.807   5.817   3.398  1.00 65.59
N

```

Edit: 21Dec10

No longer need interaction file as script will calculate all distances for all atoms.

This script will probably take considerably longer to run.

NOTE: Make sure protein PDB file DOES NOT contain hydrogen atoms. The script does not contain a conditional for exclusion of these atoms.

5Jul12 Added conditional that will ignore hydrogen atoms in protein.

18Dec12 Edited to run with execution through shell script  
 usage perl calc\_all\_dist\_auto.pl 4.0 protein.pdb results\_dir  
 =cut

```

#!C:/Perl/bin -w
#!/usr/bin/perl -w

use strict;

#22Feb12 include mkdir command for distances folder
mkdir 'distances' or print "Directory distances exists.\n";

# get threshold value from user
#print "threshold value for distance measurements: ";
#my $threshold = <STDIN>;
#chomp $threshold;
my $threshold = $ARGV[0];

# read in coordinates for protein atoms
# read in lines with 'ATOM'
# input: protein PDB

#print "protein file name (no hydrogen atoms): ";
#my $protein = <STDIN>;
#chomp $protein;
my $protein = $ARGV[1];

    unless (open (PRO, "$protein"))
    {
        print "Could not open file $protein!\n";
        exit;
    }

# VARIABLES
my $pro_line; # variable for while loop reading in file
my ($bx, $by, $bz); # atom coordinates
my ($res_name, $res_num, $res_atom_name); # residue atom attributes
my %pro_atom_coordinates; # hash to hold relevant residues
# these match atoms read from interaction file
my $res_id; # combined $res_name and $res_num
my $atom_id; # combined $res_atom_name and 3 coordinates

# pull out reisdue name, residue number, atom name, and each of 3 coordinates
while($pro_line = <PRO>)
{
    if($pro_line =~ /^(ATOM)/)
    {
        $res_atom_name = substr($pro_line,12,4); # position 12 for 4 characters
        # atom names should be only 3 characters long
        $res_atom_name =~ s/\s//g; # remove whitespace

        # conditional for hydrogen atoms
        # do nothing if atom type is "H*"
        if ($res_atom_name =~ /^H/)
        {}
        else
        {
            $res_name = substr($pro_line,17,3); # position 17 for 3 characters
            # residues names should be only 3 characters long

```

```

$res_num = substr($pro_line,23,3);
# residue numbers should be a max of 3 digits
$res_num =~ s/\s//g; # remove whitespace

# set up residue id (residue name and number)
$res_id = "$res_name\t$res_num\t$res_atom_name"; # use $res_id as key
for hash

# save $res_id --> coordinates to %pro_atom_coordinates if match found
# coordinates
$bx = substr($pro_line,30,8);
$bx =~ s/\s//g; # remove whitespace
$by = substr($pro_line,38,8);
$by =~ s/\s//g; # remove whitespace
$bz = substr($pro_line,46,8);
$bz =~ s/\s//g; # remove whitespace

# associate coordinates with $res_atom_name
$atom_id = "$bx,$by,$bz";

# put all atom-coordinate pairs into hash by residue ID
if(exists $pro_atom_coordinates{$res_id})
{
}
else
{
    $pro_atom_coordinates{$res_id} = $atom_id;
}
}
}
}
# Now have hash for protein information
# $res_id = $res_name,$res_num
# $atom_id = $res_atom_name,$bx,$by,$bz;
# $res_id --> $atom_id
# Example HIS\t449\tNE2 --> x,y,z

close PRO;
my $pro_coord_ref = \%pro_atom_coordinates; # reference to send to subroutines

#-----

# get input files names
# a list should be created of all the pose PDB files to compare to the protein
file
# should be list of complete file names (with .pdb)

# get list of all poses in directory
#print "result directory name: ";
#my $dir = <STDIN>;
#chomp $dir;
my $dir = $ARGV[2];
opendir DIR, $dir or die "Directory $dir could not be opened.\n"; # comment
out this and subsequent 2 lines for specific list option above

```

```

my @res_list = grep{$_ =~ /\.pdb/} readdir DIR;
my $result_list = \@res_list;

# send results list and reference for protein coordinates to subroutine
get_ligand ($dir,$result_list,$pro_coord_ref);

print "****Finished distance calculation for $dir!****\n";
# END OF MAIN

#-----
#subroutine
# Get coordinate information for posed ligand.
# Send ligand information to another subroutine (calc_dist_out) for distance
# calculation
# and printing.
# Input: $results (list of pose PDBs), $pro_coord_ref (reference for hash
# with protein coordinates)
# Return: N/A
#-----
sub get_ligand {
    my ($dir,$result_list,$pro_ref) = @_;

    # split results list into array for loop
    my @ligand = @$result_list;
    # my @ligand = split(/\n/, @$result_list);

    # loop for opening ligand and parsing coordinates
    for(my $i=0; $i<@ligand; $i++)
    {
        unless (open (LIG, "$dir/$ligand[$i]")) # open file
        {
            print "Could not open file $ligand[$i]!\n";
            exit;
        }

        # hash for ligand atom coordinates
        my %atom_coordinates;

        # variables for getting coordinates
        my $lig_line;
        my ($ax, $ay, $az);
        my ($lig_name, $atom_name);

        # read in line with 'HETATM'
        # pull out ligand name and atom name, each of 3 coordinates
        while($lig_line = <LIG>)
        {
            if($lig_line =~ /^(HETATM)/ || $lig_line =~ /^(ATOM)/)
            {
                $atom_name = substr($lig_line,12,4); # position 13 for 3 characters
                $atom_name =~ s/\s//g;

                $lig_name = substr($lig_line,17,3); # position 17 for 3 characters
                # ligand names should be only 3 characters long
            }
        }
    }
}

```

```

# coordinates
$ax = substr($lig_line,30,8);
$ax =~ s/\s//; # remove whitespace
$ay = substr($lig_line,38,8);
$ay =~ s/\s//; # remove whitespace
$az = substr($lig_line,46,8);
$az =~ s/\s//; # remove whitespace

if($atom_name !~ /^H/) # ignore hydrogens in ligand
{
    # put all atom coordinates into hash
    # all atom names should be unique
    my $lig_id = "$lig_name\t$atom_name";

    if(exists $atom_coordinates{$atom_name}) {}
    else { $atom_coordinates{$lig_id} = "$ax,$ay,$az"; }
}
}
}
# Now have hash for ligand information
# ligand_name --> $lig_name
# $atom_name --> $ax,$ay,$az
# Example: DRG\tC1 --> x,y,z
close LIG;

# Call subroutine to calculate distances and print to output file.
# Send reference for atom coordinate hash, ligand file name,
# and reference for protein coordinate hash.
# \%atom_coordinates, $ligand[$i], and $pro_ref
calc_dist_out(\%atom_coordinates, $ligand[$i], $pro_ref);
}
}

#-----
# subroutine
# Calculate distances and print data to output file.
# Input: \%atom_coordinates, $ligand[$i], and $pro_ref
# (atom coordinate hash reference, ligand pose file name, and protein
coordinate hash reference)
# Return: N/A
#-----
sub calc_dist_out {
    my ($lig_coord_ref, $lig_filename, $pro_coord_ref)=@_;

    # reference hashes
    #my %atom_coordinates = %$lig_coord_ref;
    my %pro_atom_coordinates = %$pro_coord_ref;

    # creat output file
    # remove .pdb extension
    $lig_filename =~ s/.pdb/.txt/;
    open (OUT, ">distances/intxns_$lig_filename");

    print OUT
"Ligand_Name\tAtom1\tRes_Name\tRes_Num\tAtom2\tccalc_dist\tinteraction\n";

```

```

# calculate distance between each protein and ligand atom
# only keep distances less than 4.000 Angstrom
my ($key,$value);
while(($key,$value) = each %pro_atom_coordinates)
{
    # set protein atom coordinates
    my @pro_xyz = split(',',$value); # split $value into separate
coordinates
    my $bx = $pro_xyz[0];
    my $by = $pro_xyz[1];
    my $bz = $pro_xyz[2];

    # measure protein atom against each ligand atom
    my ($k, $v);
    while (($k,$v) = each %$lig_coord_ref)
    {
        my @atm_xyz = split(',',$v);
        my $ax = $atm_xyz[0];
        my $ay = $atm_xyz[1];
        my $az = $atm_xyz[2];

        # calculate distance
        my $distance = sprintf("%.3f", (sqrt(($ax-$bx)**2 + ($ay-$by)**2 +
($az-$bz)**2)));

        # print distance if less than threshold in Angstroms
        my $phobic = 3.90;
        my $philic = 3.30;
        if($distance <= $phobic && $distance > $philic)
        {
            my @temp_pro = split(/\t/, $key); # [0]Res_Name [1]Res_No
[2]Atom2
            my @temp_lig = split(/\t/, $k); # [0]Lig_Name [1]Atom1

            if($temp_pro[2] =~ /^C/ && $temp_lig[1] =~ /^C/)
            {
                print OUT "$k\t$key\t$distance\thydrophobic\n";
            }
        }
        elsif($distance <= $philic)
        {
            # check that both atoms are not carbon atoms
            my @temp_pro = split(/\t/, $key); # [0]Res_Name [1]Res_No [2]Atom2
            my @temp_lig = split(/\t/, $k); # [0]Lig_Name [1]Atom1

            if($temp_pro[2] !~ /^C/ && $temp_lig[1] !~ /^C/)
            {
                print OUT "$k\t$key\t$distance\thydrogen bond\n";
            }
        }
        elsif($distance <= $threshold)
        {
            print OUT "$k\t$key\t$distance\tthreshold reached\n";
        }
    }
}

```

```

    }
}
close OUT;
}
exit;

```

### 8.3.11 weighted\_score\_key\_intxns.pl

=head  
Find specific residue interactions given directory of potential interaction lists.

INPUT::

Desired residue list format (e.g., key\_intxns.txt):

```
SER\t289\tOG
```

```
HIS 323 NE2
```

```
HIS 449 NE2
```

```
TYR 473 OH
```

Potential Interaction list format:

Ligand_Name	Atom1	Res_Name	Res_Num	Atom2	calc_dist	distance_type
4HD\t03\tHIS\t449\tNE2\t3.214						hydrogen bond
4HD	O2	HIS	449	NE2	2.954	hydrogen bond
4HD	C11	LEU	330	CD1	3.448	hydrophobic
4HD	C10	LEU	330	CD1	3.575	hydrophobic

OUTPUT::

List of each interaction with associated filenames

```
10Jan11
```

```
18Dec12: edited to work with shell script
```

```
usage: perl weighted_score_key_intxns_4Sep12.pl key_intxn_list.txt distances
scored.txt success.txt
```

```
=cut
```

```
#!/C:/Perl/bin -w
```

```
#!/usr/bin/perl -w
```

```
use strict;
```

```
#-----get interaction list-----
```

```
#print "key interaction file name: ";
```

```
#my $interactions = <STDIN>;
```

```
#chomp $interactions;
```

```
my $interactions = $ARGV[0];
```

```
unless (open (INT, "$interactions"))
```

```
{
```

```
    print "Could not open file $interactions!\n";
```

```
    exit;
```

```
}
```



```

# hash for interaction records
my %intxn_records;
my $intxn_count = 0;

# read in lines and set up hash with list of interactions
my $line;
while($line = <INT>)
{
    chomp $line;
    my @temp = split(/\s/, $line); # split line
    my $residue = "$temp[0]\t$temp[1]\t$temp[2]"; # recombine residue info

    if(!exists $intxn_records{$residue})
    {
        $intxn_records{$residue} = 1;
        $intxn_count++;
    }
}

# reference for interaction list
my $ref_intxns = \%intxn_records;

# Now have hash with interactions for comparison
# Example: HIS 449 NE2 -> 1

close INT;
#-----

#-----get list of interaction filenames-----
#print "directory name containing interaction lists: ";
#my $dir = <STDIN>;
#chomp $dir;
my $dir = $ARGV[1];
opendir DIR, $dir or die "Directory $dir could not be opened.\n"; # comment
out this and subsequent 2 lines for specific list option above
my @res_list = grep{$_ =~ /\.(txt)/} readdir DIR;
my $result_count = @res_list;
print "Found $result_count TXT files.\n";

# get score output filename
#print "score output file name: ";
#my $output = <STDIN>;
#chomp $output;
my $output = $ARGV[2];
open (OUT, ">$output");

# get successful pose output filename
#print "successful output file name: ";
#my $success = <STDIN>;
#chomp $success;
my $success = $ARGV[3];
open (SUC, ">$success");
#DEBUG
#print SUC "crystal residue count: $intxn_count\n";

```

```

my %file_intxn_info; # for each input file, address for hash with matching
interaction information
my %pose_counts; # counts for matching interactions for each input file

for(my $i=0; $i<@res_list; $i++)
{
    # send to subroutine
    # set %file_intxn_info key to file name and find interactions for each file
in directory
    ( $file_intxn_info{$res_list[$i]}, $pose_counts{$res_list[$i]} ) =
match_intxns($dir,$res_list[$i],$ref_intxns);
#DEBUG    print SUC "$res_list[$i] = $pose_counts{$res_list[$i]}\n";
}

# print output file header
print OUT "Filename";
foreach my $res (sort keys %intxn_records)
{
    $res =~ s/\t/\. /g;
    print OUT "\t$res";
}
print OUT "\n";

# print data collected from input files
my ($key,$value);
foreach $key (sort keys %file_intxn_info)
{
    $value = $file_intxn_info{$key}; # set value equal to list of filenames for
each interaction

    print OUT "$key"; # print filename to score output

    my $k;
    my %value_hash = %$value;
    foreach $k (sort keys %value_hash)
    {
        print OUT "\t$value_hash{$k}";
    }

    print OUT "\n"; # print newline

    if($intxn_count == $pose_counts{$key})
    {
        print SUC "$key\n"; # print filename to success output
    }
}

print "****Finished!****\n";
close OUT;
close SUC;
closedir DIR;
# END OF MAIN
#-----

```

```

#---subroutine-----
# objective: determine if the desired interactions are present in a file
# input: directory with lists, input file name, reference for hash of residues
to look for
# output: hash of interactions with corresponding weight (if found), counter
for number of interactions recorded
#-----
sub match_intxns {
  my ($dir,$filename,$ref_intxns) = @_;
  my %info_hash;

  # open file
  unless (open (INPUT, "$dir/$filename"))
  {
    print "Could not open file $filename!\n";
    exit;
  }

  my $line;
  my %file_contents;
  my $motif;
  while ($line = <INPUT>)
  {
    chomp $line; # remove newline
    my @temp = split(/\t/,$line);
    #BRL C5 GLY 284 C 3.992 threshold reached
    #[0] [1] [2] [3] [4] [5] [6]
    $motif = "$temp[2]\t$temp[3]\t$temp[4]";

    if(exists $file_contents{$motif}) # put file contents in hash for
comparison to reference file
    { $file_contents{$motif} .= ";$temp[6]"; }
    else
    { $file_contents{$motif} = "$temp[6]"; }
  }

  my $info_count = 0;
  my %deref_intxns = %$ref_intxns;
  foreach my $item (sort keys %deref_intxns) # check file contents against
reference list
  {

    my $marker = ''; # set marker for table
    if(exists $file_contents{$item}) # if key from reference hash exists in
file contents...
    {
      $info_count++;
      my $holder = $file_contents{$item}; # set value from contents hash
      if($holder =~ /;/)
      {
        my @temp = split(/;/,$holder); # split value if contains ;
        foreach my $a (@temp) # assign marker for each interaction from
file
        {
          if($a =~ /(hydrogen bond)/)

```

```

        { $marker .= 'x'; }
        elseif($a =~ /(hydrophobic)/)
        { $marker .= 'o'; }
        elseif($a =~ /(threshold reached)/)
        { $marker .= '/'; }
    }
}
else
{
    if($holder =~ /(hydrogen bond)/)
    { $marker = 'x'; }
    elseif($holder =~ /(hydrophobic)/)
    { $marker = 'o'; }
    elseif($holder =~ /(threshold reached)/)
    { $marker = '/'; }
}
}
else
# { } # do nothing if interaction does not match residues from reference
file
{ $marker = "-"; } # if key from reference not in file contents, insert
place holder

    if(exists $info_hash{$item})
    { $info_hash{$item} .= "$marker"; } # add marker if residue in hash
already
    else
        #(for instances with more than one
interaction with single residue)
    {
        $info_hash{$item} = $marker;

    } # else, add to hash and increment counter for success check
}

close INPUT;
return (\%info_hash,$info_count);
}
exit;

```

## 8.4 Appendix D: Molecular dynamics input files

This section contains example input files used for MD and SMD.

### 8.4.1 Energy minimization (EM)

```
; LINES STARTING WITH ';' ARE COMMENTS
title      = Minimization ; Title of run

; Parameters describing what to do, when to stop and what to save
integrator = steep      ; Algorithm (steep = steepest descent minimization)
emtol      = 1000.0    ; Stop minimization when the maximum force < 10.0
kJ/mol
emstep     = 0.01      ; Energy step size
nsteps     = 50000     ; Maximum number of (minimization) steps to perform
energygrps = Protein   ; Which energy group(s) to write to disk

; Parameters describing how to find the neighbors of each atom and how to
calculate the interactions
nstlist    = 1         ; Frequency to update the neighbor list and long range
forces
ns_type    = grid      ; Method to determine neighbor list (simple, grid)
rlist     = 1.0        ; Cut-off for making neighbor list (short range
forces)
coulombtype = PME      ; Treatment of long range electrostatic interactions
rcoulomb   = 1.0       ; long range electrostatic cut-off
rvdw      = 1.0        ; long range Van der Waals cut-off
pbc       = xyz        ; Periodic Boundary Conditions (yes/no)
```

### 8.4.2 Isochoric-isothermal ensemble (NVT)

```
title      = NVT equilibration
define     = -DPOSRES ; position restrain the protein and ligand
; Run parameters
integrator = md        ; leap-frog integrator
nsteps     = 50000     ; 2 * 50000 = 100 ps
dt         = 0.002     ; 2 fs
; Output control
nstxout    = 100       ; save coordinates every 0.2 ps
nstvout    = 100       ; save velocities every 0.2 ps
nstenergy  = 100       ; save energies every 0.2 ps
nstlog     = 100       ; update log file every 0.2 ps
energygrps = Protein
; Bond parameters
continuation = no      ; first dynamics run
constraint_algorithm = lincs ; holonomic constraints
```

```

constraints      = all-bonds      ; all bonds (even heavy atom-H bonds)
constrained
lincs_iter      = 1                ; accuracy of LINCS
lincs_order     = 4                ; also related to accuracy
; Neighborsearching
ns_type         = grid            ; search neighboring grid cells
nstlist        = 5                ; 10 fs
rlist          = 1.0              ; short-range neighborlist cutoff (in nm)
rcoulomb       = 1.0              ; short-range electrostatic cutoff (in nm)
rvdw          = 1.0              ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype    = PME              ; Particle Mesh Ewald for long-range
electrostatics
pme_order      = 4                ; cubic interpolation
fourierspacing = 0.16            ; grid spacing for FFT
; Temperature coupling
tcoupl        = V-rescale          ; modified Berendsen thermostat
tc-grps       = Protein non-Protein ; two coupling groups - more accurate
tau_t         = 0.1 0.1           ; time constant, in ps
ref_t         = 300 300           ; reference temperature, one for
each group, in K
; Pressure coupling
pcoupl        = no                ; no pressure coupling in NVT
; Periodic boundary conditions
pbc           = xyz              ; 3-D PBC
; Dispersion correction
DispCorr      = EnerPres         ; account for cut-off vdW scheme
; Velocity generation
gen_vel       = yes              ; assign velocities from Maxwell distribution
gen_temp      = 300              ; temperature for Maxwell distribution
gen_seed      = -1               ; generate a random seed

```

### 8.4.3 Isothermal-isobaric ensemble (NPT)

```

title          = NPT equilibration
define         = -DPOSRES        ; position restrain the protein and ligand
; Run parameters
integrator     = md              ; leap-frog integrator
nsteps        = 50000           ; 2 * 50000 = 100 ps
dt            = 0.002           ; 2 fs
; Output control
nstxout       = 100             ; save coordinates every 0.2 ps
nstvout       = 100             ; save velocities every 0.2 ps
nstenergy     = 100             ; save energies every 0.2 ps
nstlog        = 100             ; update log file every 0.2 ps
energygrps    = Protein
; Bond parameters
continuation   = yes            ; first dynamics run
constraint_algorithm = lincs     ; holonomic constraints
constraints    = all-bonds      ; all bonds (even heavy atom-H bonds)
constrained
lincs_iter    = 1                ; accuracy of LINCS
lincs_order   = 4                ; also related to accuracy

```

```

; Neighborsearching
ns_type      = grid      ; search neighboring grid cells
nstlist     = 5          ; 10 fs
rlist       = 1.0        ; short-range neighborlist cutoff (in nm)
rcoulomb    = 1.0        ; short-range electrostatic cutoff (in nm)
rvdw        = 1.0        ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype  = PME        ; Particle Mesh Ewald for long-range
electrostatics
pme_order   = 4          ; cubic interpolation
fourierspacing = 0.16    ; grid spacing for FFT
; Temperature coupling
tcoupl      = V-rescale   ; modified Berendsen thermostat
tc-grps     = Protein non-Protein ; two coupling groups - more accurate
tau_t       = 0.1 0.1    ; time constant, in ps
ref_t       = 300 300    ; reference temperature, one for
each group, in K
; Pressure coupling
pcoupl      = Parrinello-Rahman ; pressure coupling is on for NPT
pcoupltype  = isotropic    ; uniform scaling of box vectors
tau_p       = 2.0         ; time constant, in ps
ref_p       = 1.0         ; reference pressure, in bar
compressibility = 4.5e-5 ; isothermal compressibility of
water, bar^-1
refcoord_scaling = com
; Periodic boundary conditions
pbc         = xyz         ; 3-D PBC
; Dispersion correction
DispCorr    = EnerPres   ; account for cut-off vdW scheme
; Velocity generation
gen_vel     = no         ; velocity generation off after NVT

```

## 8.4.4 Production MD

```

title       = MD FOR PCA
; Run parameters
integrator  = md          ; leap-frog integrator
nsteps     = 50000000    ; 0.002 * 50000000 = 100000 ps (100 ns)
dt         = 0.002      ; 2 fs
; Output control
nstxout    = 0           ; suppress .trr output
nstvout    = 0           ; suppress .trr output
nstenergy  = 2500        ; save energies every 5 ps
nstlog     = 2500        ; update log file every 5 ps
nstxtcout  = 2500        ; write .xtc trajectory every 5 ps
energygrps = Protein
; Bond parameters
continuation = yes        ; first dynamics run
constraint_algorithm = lincs ; holonomic constraints
constraints  = all-bonds  ; all bonds (even heavy atom-H bonds)
constrained
lincs_iter  = 1          ; accuracy of LINCS
lincs_order = 4          ; also related to accuracy

```

```

; Neighborsearching
ns_type      = grid      ; search neighboring grid cells
nstlist      = 5         ; 10 fs
rlist        = 1.0       ; short-range neighborlist cutoff (in nm)
rcoulomb      = 1.0      ; short-range electrostatic cutoff (in nm)
rvdw         = 1.0      ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype   = PME      ; Particle Mesh Ewald for long-range
electrostatics
pme_order     = 4         ; cubic interpolation
fourierspacing = 0.16    ; grid spacing for FFT
; Temperature coupling
tcoupl        = V-rescale ; modified Berendsen thermostat
tc-grps       = Protein non-Protein ; two coupling groups - more accurate
tau_t         = 0.1 0.1   ; time constant, in ps
ref_t         = 300 300   ; reference temperature, one for
each group, in K
; Pressure coupling
pcoupl        = Parrinello-Rahman ; pressure coupling is on for NPT
pcoupltype    = isotropic ; uniform scaling of box vectors
tau_p         = 2.0       ; time constant, in ps
ref_p         = 1.0       ; reference pressure, in bar
compressibility = 4.5e-5 ; isothermal compressibility of
water, bar^-1
; Periodic boundary conditions
pbc           = xyz      ; 3-D PBC
; Dispersion correction
DispCorr      = EnerPres ; account for cut-off vdW scheme
; Velocity generation
gen_vel       = no       ; assign velocities from Maxwell distribution

```

### 8.4.5 SMD

```

title        = Umbrella pulling simulation
define       = -DPOSRES_B
; Run parameters
integrator    = md
dt           = 0.002
tinit        = 0
nsteps       = 375000    ; 750 ps
nstcomm      = 10
; Output parameters
nstxout      = 500       ; every 1 ps
nstvout      = 500
nstfout      = 500
nstxtcout    = 500       ; every 1 ps
nstenergy    = 500
; Bond parameters
constraint_algorithm = lincs
constraints   = all-bonds
continuation  = yes      ; continuing from NPT
; Single-range cutoff scheme
nstlist      = 5
ns_type      = grid

```



```

rlist      = 1.4
rcoulomb   = 1.4
rvdw       = 1.4
; PME electrostatics parameters
coulombtype = PME
fourierspacing = 0.12
fourier_nx   = 0
fourier_ny   = 0
fourier_nz   = 0
pme_order    = 4
ewald_rtol   = 1e-5
optimize_fft = yes
; Berendsen temperature coupling is on in two groups
Tcoupl       = Nose-Hoover
tc_grps      = Protein   Non-Protein
tau_t        = 0.5       0.5
ref_t        = 310       310
; Pressure coupling is on
Pcoupl       = Parrinello-Rahman
pcoupltype   = isotropic
tau_p        = 1.0
compressibility = 4.5e-5
ref_p        = 1.0
; Generate velocities is off
gen_vel      = no
; Periodic boundary conditions are on in all directions
pbc          = xyz
; Long-range dispersion correction
DispCorr     = EnerPres
; Pull code
pull         = umbrella
pull_geometry = direction
pull_vec1    = -1.018 0.036 0.993
pull_start   = yes          ; define initial COM distance > 0
pull_ngroups = 1
pull_group0  = r_365
pull_group1  = BRL
pull_rate1   = 0.008        ; 0.001 nm per ps = 10 nm per ns
pull_k1      = 1000         ; kJ mol-1 nm-2

```

#### 8.4.6 Steps for preparing SMD protein and ligand files

1. Add charges in UCSF Chimera
  - a. Add hydrogen atoms
  - b. Add charges: Tools --> Structure editing --> Add charges --> amberff\_03.r1, AM1-BCC
  - c. Set net charge (program predicts net charge)
2. Save mol2
3. Run: antechamber -i in.mol2 -fi mol2 -o .prep -fo prepi -c bcc -s 2
4. Run: parmchk -i .prep -f prepi -o .frcmod -p /usr/local/amber10/dat/leap/parm/gaff.dat

- a. The path has to be changed for the gaff.dat part to match settings on your computer.
  - b. Check the output for errors that can be fixed.
5. Edit/create leap\_ros.in text file, which should contain the following:
 

```
logfile log_2prg.log
source leaprc.ff03
source leaprc.gaff
frcmod = loadAmberParams ros.frcmod
loadAmberPrep ros.prep
saveamberparm BRL ros.top ros.crd
savepdb BRL ros.pdb
quit
```

NOTE: unit name (e.g. BRL) can't be numbers or get error that unit is double instead of characters. Have to change ligand ID in .prep file and set your ID to match as in leap\_ros.in above.
6. Run: xleap -f leap\_ros.in
7. Run: perl amb2gmx\_mask.pl --prmtop ros.top --crd ros.crd --outname ros\_top
  - a. Depending on the version of amber on the machine, the keywords section of the perl script may need to be masked by adding an asterisk (\*) before the words. The original amb2gmx.pl can be downloaded from the AMBER site.
8. Change ros\_top.top to .itp and edit the following:
  - a. Remove defaults section at beginning
  - b. Remove system and molecules section at end
  - c. Change "solute" to three-letter ligand ID