Molecular Dynamics Free Energy Simulation Study to Rationalize the Relative Activities of PPARδ Agonists

Woojin Lee, Hwangseo Park,†,* and Sangyoub Lee*

*Department of Chemistry, Seoul National University, Seoul 151-747, Korea. *E-mail: sangyoub@snu.ac.kr*
*†Department of Bioscience and Biotechnology, Sejong University, Seoul 143-747, Korea. *E-mail: hspark@sejong.ac.kr*
Received November 6, 2007

As a computational method for the discovery of the effective agonists for PPARδ, we address the usefulness of molecular dynamics free energy (MDFE) simulation with explicit solvent in terms of the accuracy and the computing cost. For this purpose, we establish an efficient computational protocol of thermodynamic integration (TI) that is superior to free energy perturbation (FEP) method in parallel computing environment. Using this protocol, the relative binding affinities of GW501516 and its derivatives for PPARδ are calculated. The accuracy of our protocol was evaluated in two steps. First, we devise a thermodynamic cycle to calculate the absolute and relative hydration free energies of test molecules. This allows a self-consistent check for the accuracy of the calculation protocol. Second, the calculated relative binding affinities of the selected ligands are compared with experimental IC_{50} values. The average deviation of the calculated binding free energies from the experimental results amounts at the most to 1 kcal/mol. The computational efficiency of current protocol is also assessed by comparing its execution times with those of the sequential version of the TI protocol. The results show that the calculation can be accelerated by 4 times when compared to the sequential run. Based on the calculations with the parallel computational protocol, a new potential agonist of GW501516 derivative is proposed.

Key Words : Thermodynamic integration, Binding free energy, Molecular dynamics, PPARδ, Lead optimization

Introduction

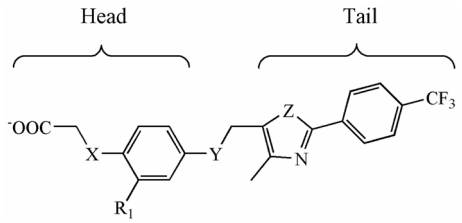
Peroxisome proliferators-activated receptors (PPARs) are ligand-activated transcription factors that regulate cellular and physiological fatty acid and carbohydrate metabolism. PPARγ is the target of the marketed glitazones, which lower blood glucose and are used to treat Type II diabetes. Another subtype, PPARα, is the target of marketed fibrate drugs, which are used to lower triglycerides and raise high-density lipoprotein cholesterol (HDL-C). In contrast, biological role of PPAR was unclear[1] due, in part, to its broad tissue expression[2] and the lack of good chemical tools. Recently, a potent and selective PPAR agonists, GW501516 (phase II clinical trial), was developed[3] which was shown to increase plasma high-density lipoprotein cholesterol (HDL-C) levels and decrease low-density lipoprotein cholesterol (LDL-C) and triglycerides in obese and dyslipidemic rhesus monkeys.[4] As individuals, who are obese or who have type 2 diabetes and the metabolic syndrome, are at the high risk for cardiovascular disease, the regulation of PPARs by synthetic ligands has medical importance.[5] Among the medicinal strategies, such as manipulation of food intake centrally or peripherally, blocking fat absorption in intestines, modulation of fat burning, and control of adaptive thermogenesis, stimulation of thermogenesis becomes a popular choice to treat obesity.[6] PPARδ is also a key regulator of fat burning and thus a novel therapeutic target for anti-obesity drugs.

We have been collaborating with an experimental group developing modified PPARδ agonists,[7] and have found that molecular docking with rigid receptor technique is not sufficient for predict the drug efficacy. One of the limitations is polar hydrogen orientation. The structure of PPARδ shows that this protein has a large entrance, and the channel entrance to binding pocket is surrounded by polar amino acids.[8] Hence it is expected that reorientation of sidechains and interaction with solvent water molecules should be taken into account.

Although MDFE simulation with explicit solvent molecules is expensive in terms of computing resource, the method has already been used in many drug discovery processes.[9-11] In this study, we calculate the relative binding affinities of selected ligands that are used for optimization process of GW501516, and the accuracy of the results are compared with experimental values of IC_{50} for the ligands. We aim to establish and optimize the calculation protocol utilizing TI implemented in Amber8 package in conjunction with molecular docking. For the efficiency of the calculation, we try to utilize the parallel nature of TI calculation. To improve the accuracy, we analyze the free energy profiles for alchemical changes in TI calculations with test molecules. The alchemical changes between test molecules are similar to those between PPARδ ligands, except for environments. As will be described in the *Computational Methods* section, the accuracy of free energy calculations for the test cases will be checked in a self-consistent manner.

The TI protocol validated with the test cases is then applied to obtain relative binding affinities of selected ligands[3] (Table 1) in conjunction with molecular docking

Table 1. Ligand structures investigated in this work



Lig	X	R_1	Y	Z	EC_{50} (nM)	IC_{50} (nM)
1	O	CH_3	S	S	1	1
2	O	CH_3	S	O	10	3
3	O	CH_3	O	S	NA	NA
4	O	CH_3	O	O	30	30
5	CH_2	CH_3	O	S	30	60
6	O	H	S	S	NA	NA
7	O	H	S	O	420	90
8	O	H	O	O	470	340

Experimental data are taken from ref. 3.

calculations based on the X-ray crystal structure of PPARδ. The accuracy of the calculations is estimated and analyzed by comparing with the experimental results. Finally, we propose some minor structural modifications to the current lead compound (GW501516) which leads to an improved affinity for PPARδ.

Computational Methods

System preparation. As crystal structures are unavailable for the ligands under investigation, we choose a crystal structure that contains a ligand with similar moiety as the current test set. The protein structure used in the calculation is based on the co-crystal structure of hPPARδ-LBD and GW2433 (PDB entry 1GWX)[12] obtained from RCSB Protein Data Bank.[13] The co-crystalized ligand, GW2433, and crystallographic water molecules are removed. Missing heavy atoms and hydrogen atoms are added in Leap module in Amber8 package.[14] The standard ionization state at neutral pH is considered for all ionizable residues and the terminal amino acids is set to be charged.

The receptor input file of docking calculation is then prepared. For consistency with subsequent MD simulations, the force field parameters for receptor are taken from the Amber99 force field (ff99).[15] This requires the assignment of the corresponding charges to the ligands as well. Solvation parameters are added by Addsol module of Autodock. The affinity grids for docking are centered on the geometric center of the GW2433 which is originally complexed with the receptor. The grid has a dimension of $70 \times 70 \times 70$ grid points with grid spacing of 0.375 Å. By visual inspection, we confirm that this box practically covers the known binding pocket of the receptor. The affinity grid maps are then generated for all atom types of a ligand. In calculating the Coulomb interactions, a sigmoidal distance-dependent dielectric function[16] is used.

The ligand structures are generated by MMFFs94 force field[17] as ionized forms. For consistency of partial charges with receptor, single-point calculations with GAMESS[18] (Nov. 2004) are performed to obtain the electrostatic potential using HF/6-31G* level of theory. Fitting charges to the electrostatic potential is then performed according to two-step restrained electrostatic potential (RESP) protocol.[19] Equivalent atoms are given equal partial charges. Ligands input files for docking calculations are prepared by Autotors module in Autodock3.0. Based on preliminary docking calculations, more reasonable and converged conformations have been obained by fixing a rotatable bond between thia(oxa)zole and phenyl ring in hydrophobic tail of each ligand. The other torsions are set free to rotate. Atomic solvation parameters are determined based on atom types.[20]

Docking. AutoDock version 3.0.5 is used to obtain starting conformation for MD simulation of PPAR-ligand complexes.[23] The Lamarckian genetic algorithm (LGA) is used for conformation searching in the binding site. For each compound, the following running parameters are used: random initial position, orientation and dihedral, population size of 50, elitism of 1, mutation rate of 0.02, crossover rate of 0.8, local search rate of 0.06, 4 million energy evaluations and 200 trials of dockings. Docking results by Genetic algorithms (GA) depend on pseudorandom numbers, so multiple runs usually need to be done (200 trials in this case), and there is no guarantee that the global best solution can be found. However, good results are usually found much more quickly than a purely random search or a systematic search. Docked conformations from the 200 trials for each ligand are clustered by use of a tolerance of 1.0 Å root-mean-square deviation (RMSD) and sorted in terms of docking energy. We performed molecular docking for all ligands to compare the accuracy of the calculated free energies, although we only need to perform docking small subset of the ligands generally.

The best solutions of each ligand by docking calculations are used as starting configurations of the subsequent MD/TI calculations in protein and in solution. Amber99 force field[15] for the protein and gaff[21] for the ligands are used in all MD calculations. The atomic partial charges borne by the ligands are those from RESP protocol described above. The protein-ligand complexes are surrounded by a octahedron box of pre-equilibrated TIP3P[22] water molecules at least 10 Å away from the protein atoms. One Cl^- ion is added to neutralize the system. For unbound state simulation, each ligand is solvated by a 10 Å layer of TIP3P water molecules and one Na^+ ion is added to neutralize system.

Equilibration of MD simulations. All MD simulations presented in this work are performed using Sander module in Amber8 package.[3] For all MD simulations, 1 fs time step is used, and SHAKE[24] restraints with a tolerance of 10^{-6} are applied to all bond lengths involving hydrogen atom. Particle Mesh Ewald[25] is employed to calculate the long-range electrostatic interactions. A 10 Å cutoff is used for the nonbonded interactions with updates of the pair list every 25 steps. All dynamics simulations are carried out in the isothermal-isobaric ensemble. The temperature and the pres-

sure of the system are maintained using Berendsen weak coupling algorithm.[26] Each complex is subject to following sequential equilibration protocol. First, the system is minimized holding water molecules added by Leap module, then minimize the system holding the protein and the ligand. Next, the whole system is minimized 500 steps of steepest descent followed by 500 steps of conjugate gradients. After 10 ps of equilibration of water molecules holding solute molecules, the system is gradually heated to 300 K with a 10 ps interval per 100 K. Finally the system is equilibrated for 100 ps at 300 K and the pressure is maintained to 1 atm. For each unbound state, similar equilibration protocol is applied. During 300 K equilibration run, trajectory is collected every 1 ps for subsequent analysis of the complex.

Free energy calculations. According to the thermodynamic cycle[27,28] shown in Figure 1, binding free energy difference can be calculated as follows:

$$\Delta\Delta G_{12} = \Delta G_2^{\text{bind}} - \Delta G_1^{\text{bind}} = \Delta G_{\text{b}} - \Delta G_{\text{ub}}. \quad (1)$$

As calculation of ΔG_1 and ΔG_2 is generally very difficult in computer simulations, we calculate the vertical legs of the thermodynamic cycle by alchemical mutation to estimate the free energy change by thermodynamic integration (TI) method[29]:

$$\Delta G = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda \cong \sum_{i=1}^{N} \left(\left\langle \frac{\partial V(X,\lambda)}{\partial \lambda} \right\rangle_{\lambda(i)} + \left\langle \frac{\partial V(X,\lambda)}{\partial \lambda} \right\rangle_{\lambda(i+1)} \right) \times \frac{\lambda(i+1)-\lambda(i)}{2}. \quad (2)$$

A coupling parameter λ smoothly interpolates the potential energy function for the system between state 1 and state 2. We calculate the work for changing the system from state 1 to state 2 along a unphysical path defined by λ:

$$V(\lambda) = (1-\lambda)V_1 + \lambda V_2. \quad (3)$$

$\langle \partial V/\partial \lambda \rangle_\lambda$ is numerically integrated to give the free energy change by using simple trapezoidal rule or Gaussian quadrature according to sampling protocol. Single topology method is used to change one ligand to another in protein (b) and in solution (ub). It is known that when the atom creation/annihilation is involved, the $\partial V/\partial \lambda$ at end point cannot be determined.[30] To circumvent this problem and the instability caused by the finite charges borne by the atoms with very small Lennard-Jones parameters, we utilize the electrostatic decoupling scheme[31] with nonlinear mixing rule[32] for potential functions as implemented in current version of Amber8:

$$V(\lambda) = (1-\lambda)^k V_1 + [1-(1-\lambda)^k]V_2. \quad (4)$$

During a simulation, the $\partial V/\partial \lambda$ is calculated by following equation:

$$\partial V/\partial \lambda = k(1-\lambda)^{k-1}(V_2-V_1). \quad (5)$$

With $k = 1$, the above equation corresponds to the linear mixing of unperturbed and perturbed potential functions. As Eq. (4) does not prevent the end point singularity for atom creation case, the mutations are performed in the direction of atom annihilation if any atom vanishes or appears.

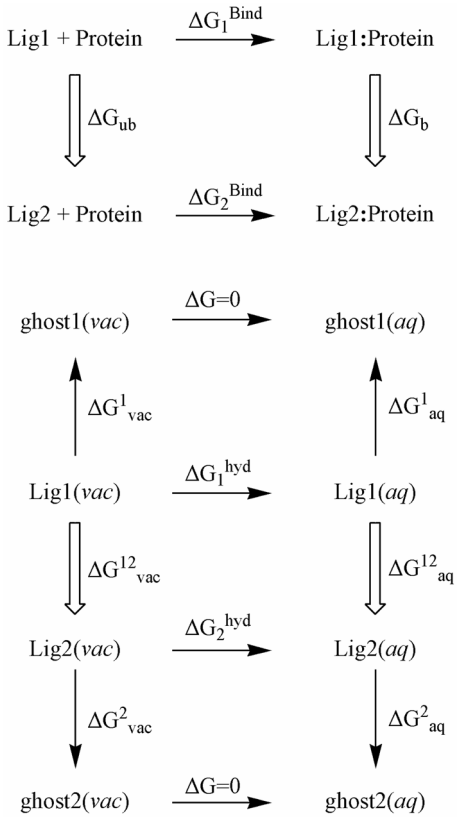


Figure 1. Thermodynamic cycles utilized in the free energy calculations.

We use the 3-point Gaussian quadrature for scaling the electrostatic (*chg*) contributions of the system and 16 nonuniform λ points (0.0, 0.025, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95) for scaling van der Waals (*vdw*) contributions with $k = 6$ for potential mixing. In addition, c-spline fitting is applied to each van der Waals profile to reduce integration error. 50 ps equilibration followed by 150 ps sampling is performed at each λ point in NPT ensemble. To maximize the parallel computation efficiency and to reduce the total run time, each calculation is performed in the following way as summarized in Figure 2. For bound state run, we sequentially perform the three equilibration runs of *chg* scaling using parallel version of Sander. From the configuration obtained by the third equilibration run, 16 *vdw* scaling runs at each λ are performed concurrently with same initial momentum distribution. At the same time, the remaining sampling runs for *chg* scaling are performed from each equilibrated configuration. For unbound state run, *chg* scaling is performed sequentially, followed by *vdw* scaling at each λ point. Unlike the free energy perturbation (FEP), there is no finite difference term in the ensemble average in TI. Therefore, the difference between the potential surfaces of neighboring λ states is not an explicit issue.[33] However, we examined this by comparing the results from different protocol.

Before tackling the main problem, we apply the protocol to simple test cases including that introduced in Amber

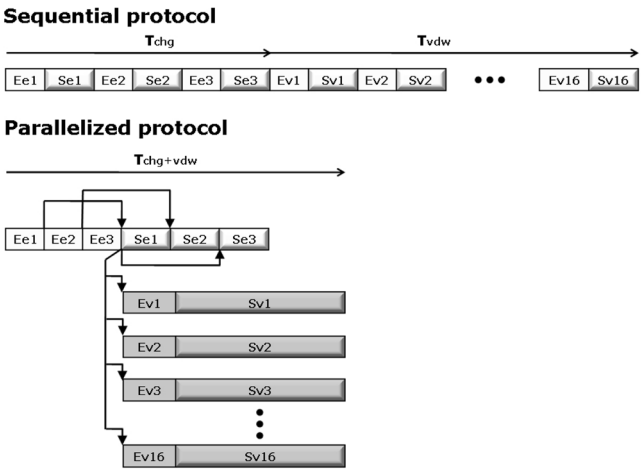


Figure 2. Calculation protocol used in this work. E{e,v}1: Equilibration run at λ_1, S{e,v}1: sampling run at λ_1. In the parallelized protocol, each bent arrow shows the starting conformation used in the subsequent sampling runs. Gray boxes mean calculations with single CPU version of Sander module, otherwise all calculations are performed with parallel version of Sander module with 16 CPUs. T_{chg} or T_{vdw} is CPU time.

tutorial for TI calculation.[34] The structures of the test molecules are available in supporting information. By utilizing the lower thermodynamic cycles in Figure 1, we can obtain absolute free energy of hydration of ligand **1** or ligand **2** along the vertical line arrows:

$$\Delta G_1^{\text{Hyd}} = \Delta G^1_{\text{vac}} - \Delta G^1_{\text{aq}} \text{ and } \Delta G_2^{\text{Hyd}} = \Delta G^2_{\text{vac}} - \Delta G^2_{\text{aq}}. \quad (6)$$

The relative free energy of hydration between ligand **1** and ligand **2** can be obtained from the difference between the above values. Another pathway to obtain the relative free energy of hydration is possible along the boxed arrows of the lower cycle in Figure 1:

$$\Delta\Delta G^{12}_{\text{hyd}} = \Delta G^2_{\text{hyd}} - \Delta G^1_{\text{hyd}} = \Delta G^{12}_{\text{aq}} - \Delta G^{12}_{\text{vac}}. \quad (7)$$

The calculations of relative free energy change (boxed arrow in the lower cycles in Figure 1) of these cases contain a bond-breaking, while each vertical line arrow corresponds to a case in which only non-bond interactions are changed with other internal parameters being maintained. Therefore, the results provide an internal consistency check for parameter change during the mutation.

Results and Discussion

Docking simulations. As ligands investigated in this study share much similar scaffold, we expect that the best solution for each ligand shows similar binding pattern. This is also required for subsequent free energy simulation. The use of the protein structure of 1GWX is likely to lead to successful docking simulations, although a rigid receptor

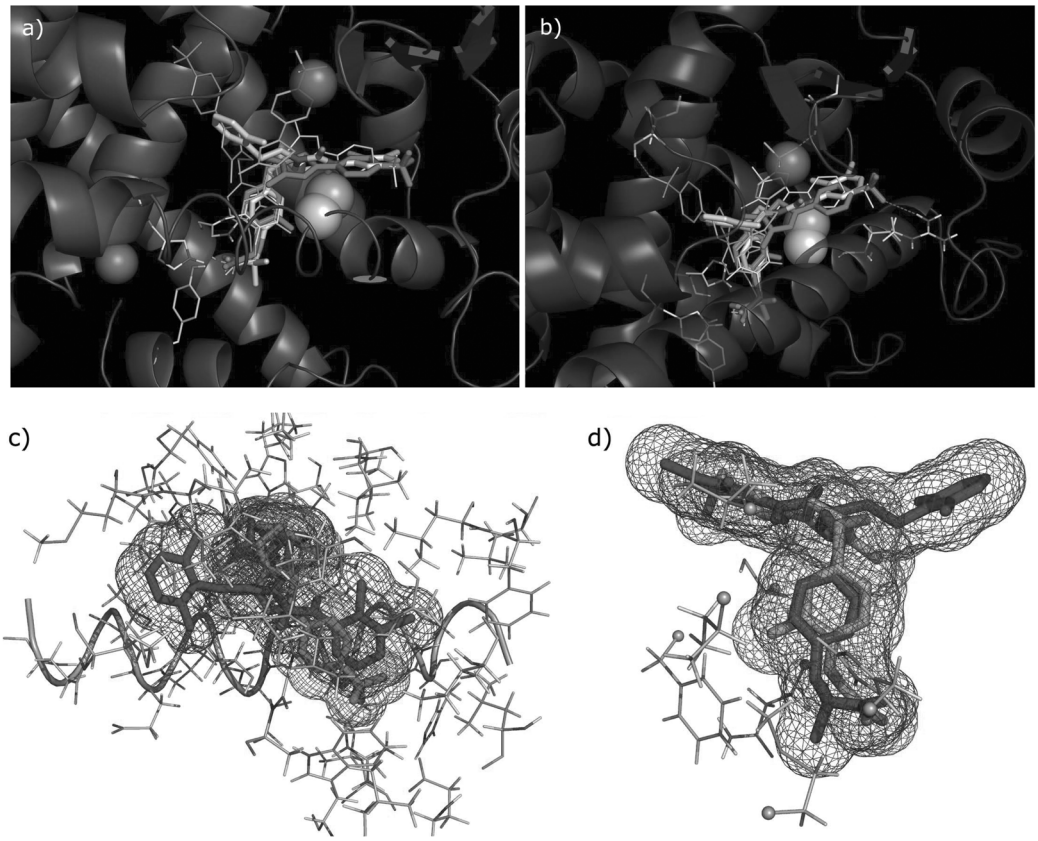


Figure 3. Representative binding poses of ligand **1** by docking calculation: (a) Front view with protein, (b) Side view, (c) GW2433 (blue) and docked GW501516 (red) with mesh representation, (d) Rear view. The red spheres are crystallographic waters in top loop region and the yellow sphere is the sulfur atom of Cys285.

model is used.[35] Because of the little structural difference between the ligands under investigation and GW2433 for head moiety, we obtain similar binding poses for the head moiety.

Docking results for ligand **1** are summarized in supporting information, and the representative binding modes are shown in Figure 3.

The reference structure is the GW2433 co-crystallized with the receptor. The head moiety of the lowest energy conformer of ligand **1** (red in Figure 3) occupies a similar position with respect to GW2433, but shows little shift compared to GW2433. The RMSD between the geometric centers of the head phenyl rings of GW2433 and ligand **1** is 0.49 Å. From the head moiety, hydrophobic tail is placed in the right pocket of the binding site and leans against Helix3 of the receptor. This binding pose results in some void in the backside of the right pocket where the two chlorine atoms in 2,3-dichloro phenyl group of GW2433 reside (Figure 3c). The two-atom spacer consisting of sulfur and methyl connecting head and tail of the ligand occupies similar volume as the three-atom spacer connecting head phenyl and ureido group of GW2433 (Figure 3d). Considering the reduced torsional degrees of freedom, the change from three-atom spacer to two-atom spacer might contribute to enhance the binding free energy while maintaining steric contacts.

The binding position of the ortho methyl group at the head phenyl ring reveals a small cavity surrounded by Phe282, Cys285, Gln286 and Ile157. As we use the rigid receptor strategy in docking calculations using the chain A of the GW2433/hPPARδ co-crystal structure, we may first attribute the small shift of the head moiety to the existence of the methyl group in head phenyl group. Then the binding mode of ligand **6** should result in more similar position of phenyl ring to that of GW2433. However, we obtain very similar binding poses for ligand **1** and ligand **6**. The RMSD of the head phenyl group between ligand **1** and ligand **6** is 0.0066 Å (0.1 Å for all scaffold atoms). Hence the methyl substitution is believed to have little effect on the shift of the head moiety. This is not surprising because there is already an enough room to accommodate methyl group without moving nearby protein residues. From these docking results, we observe increased binding affinity by the methyl substitution, which was already verified in experiment[3] (Table 2: ligand {**1**,**2**,**4**} *vs*. ligand {**6**,**7**,**8**}). The shift in the binding mode seems to be a compromise between interactions of the head and the tail moieties with nearby protein residues.

Table 2. Docking results for ligands **1-8**

Ligand	ΔG (exp)	c-rank	Nconf	E (low)
1	−12.28	1	7	−13.36
2	−11.63	2	8	−12.84
3	NA	2	5	−12.34
4	−10.26	2	7	−11.96
5	−9.85	4	2	−12.71
6	NA	3	11	−12.71
7	−9.61	3	6	−12.16
8	−8.82	3	6	−11.38

Δ*G*(exp)'s are calculated from experimental data[3] as pLog(IC_{50}), in kcal/mol. c-rank is the rank of conformation cluster which contains a probable binding mode. Nconf is the number of conformers in the cluster.

Figure 3a and 3b also show other binding modes found by docking calculation of ligand **1** except false-positives. Here, we exclude the configurations without a hydrogen bond between the carboxyl group of a ligand and the anchor triad (His323, His449 and Tyr473) which is known to be important in activating the receptor.[1] The lowest energy conformer of the second cluster (orange in Figure 3a, 3b) shows a different orientation of the hydrophobic tail of the ligand. The tail moiety of the second conformational cluster passes through the 2-chloro-6-fluoro-phenyl ring of GW2433 located in the left pocket, and reaches the small cavity surrounded by the loop connecting helix1 and helix2 and helix5. As many polar residues are located in this cavity, a possible H-bond is observed between the end group of ligand and Ser332, in addition to the shape complementarity. A little shift of the head moiety is observed for these conformers. We guess that this shift results from the overestimated favorable interaction of the ligand tail, especially from the H-bond of the trifluoromethyl group. The difference in docking energies between red and orange conformers is small, whereas those between average docking energies of each cluster are rather large.

The yellow conformer shows an opposite location of the methyl group at the head phenyl ring, with an overturned tail conformation compared to the red conformer. The yellow conformer reveals a cavity near the methyl group at thiazole ring. Nearby protein residues are Leu330, Leu339, Leu353, Ile364, Lys367 and Phe368 that are mostly liphophilic. The methyl group at thiazole ring of the yellow conformer is well fitted to this small cavity and this is a main reason for such a binding pose. Although the difference in docking energy from the red conformer is not large, this conformation results in the distortion of the head group. Hence this conformer is rejected as a proper binding mode for ligand **1**.

The trifluoromethyl group of the green conformer is located in the small cavity surrounded by a β turn (Ala342 to Ser345 in 4:4 β hairpin) and Met228, and results in an overall upright binding pose. A hydrogen bond can be formed between the trifluoromethyl group and Gly344 in addition to shape complementarity. But a water molecule is found in this region in the X-ray structure (red sphere in Figure 3a, 4b), and this water molecule is considered to stabilize the nearby protein residues through H-bond networks. We have surveyed available published X-ray crystal structures of hPPARδ (1GWX, 2GWX, 3GWX, 1Y0S) for the existence of a water molecule and corresponding B-factor in this region, and the results are summarized in supporting information. The low B-factor of the water molecule and nearby protein residues implies an uneasy desolvation cost for this region.

In addition to the uneasy desolvation, there is another reason for excluding those conformers. Although there is no experimental information on the dynamics of the protein, we

can refer to the B-factor for the relative flexibility of the protein. Among the residues near the entrance of the binding pocket, the loop between helix2' and helix3 shows very high B-factors.[1] This is also true for apo form of same protein (PDB entry, 2GWX). Hence if we assume that a ligand enter through this entrance,[8] the red conformation would be the probable one with an aid of the flexible loop. On the other hand, the green conformation is not probable unless the relatively stable regions are moved. Therefore, the green conformer and the others in this conformation cluster would not be involved in the real binding process.

The LGA docking results of all ligands to hPPARδ are summarized in Table 2. The experimental ΔG values are calculated from the experimental IC_{50} values.[3] The reported docking energy of each ligand corresponds to the docking energy of a conformer which holds a similar pose with the red conformer of ligand **1** and that is selected by the same criteria used for ligand **1**. For ligand {**2**,**3**,**4**}, false positives rank first, while the binding mode like red conformer of ligand **1** being found in the second conformation cluster. The head carboxyl groups of those false positives are located in the cavity where the trifluoromethyl group of orange conformer of ligand **1** occupies. Such binding modes reveal the possibility of hydrogen bond already discussed for orange conformer of ligand **1**. Similar binding modes are also found for the other ligands.

The docking energies display a tendency that is partly consistent with experimental data. As previously mentioned, we have obtained the increased binding affinities with the methyl substitution at head phenyl group. On the other hand, decreases in binding affinities are expected with the changes from thiazole to oxazole and from thiophenoxy to phenoxy group (X atom in Table 1). This result shows the liphophilicity of the nearby protein residues.

Free energy calculations. As we described in *Computational Methods* section, we first apply the TI protocol to get relative hydration free energies of test molecules. The agreements in $\Delta\Delta G$hyd's are quite satisfactory. Since the actual values of ΔG_{hyd}'s are not main concern of this study, we report the calculation results in the supporting information.

Table 3. TI results for PPAR ligands (Energy in kcal/mol)

mutation	$\Delta\Delta G_{chg}$	$\Delta\Delta G_{vdw}$	$\Delta\Delta G_{total}$	$\Delta\Delta G_{TI}$	$\Delta\Delta G_{EXP}$	
1→1			0	0	0	
1→2	−0.35	1.81	1.47	1.67	0.65	1.02
2→1	−0.14	−1.73	−1.87			
1→3	0.52	1.32	1.84	2.19		
3→1	−0.83	−1.71	−2.54			
1→4	0.55	2.65	3.2	2.91	2.02	0.9
4→1	−0.79	−1.84	−2.63			
5→1	−2.98	−1.79	−4.76	4.62	2.43	2.19
5→1_2	−2.69	−1.78	−4.47			
1→6	−0.06	1.63	1.57	1.79		
1→6_2	−0.13	2.14	2.01			
1→7	0.09	1.71	1.8	2.89	2.67	0.22
1→7_2	0.05	3.94	3.99			
1→8	−0.11	3.82	3.71	3.94	3.45	0.49
1→8_2	0.38	3.79	4.17			

Each X→Y means practical direction of mutation from X to Y. X→Y_2 means second run from different initial condition. In the last column, the deviation from experiment is tabulated.

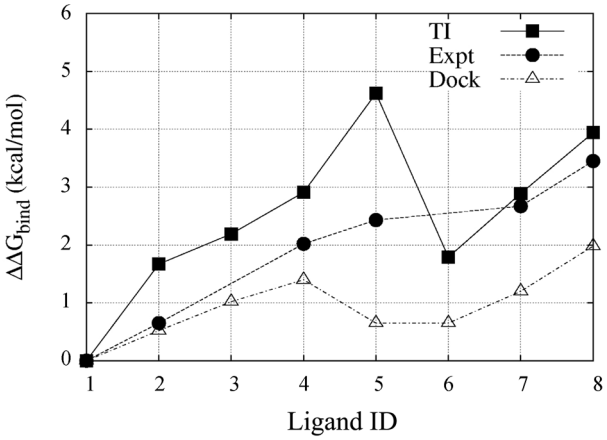


Figure 4. Comparison of the calculated binding affinities with experimental results. There are no experimental data for ligand 3 and ligand **6**.

TI results for relative binding affinities are summarized in Table 3 with respect to ligand **1**. For most mutations, ligand **1** is used as the initial state as we have obtained a highly converged and reproducible binding mode irrespective of running parameters of LGA docking runs (data not shown). We can obtain relative binding affinities for the mutation from ligand **1** to the other ligands except for ligand **5**. The parameter change from **1** to **5** is not supported in the current version of Amber8. Therefore, we obtain the relative free energy difference between **1** and **5** using the complex of protein and ligand **5** obtained by docking calculation.

As a crude error estimation in TI calculations, we carry out the reverse mutation (**X**→**1**) for possible cases by single topology method. While another TI calculation is performed after additional 50 ps equilibration from the starting configuration used for the first run if reverse mutation is not possible (**1**→**X_2** in Table 3). Considering that the reverse mutation starts from the respective binding pose by LGA docking, the discrepancies between the forward and reverse mutations are expected to be larger than the deviations between two runs of forward mutations. From the mutations between **1** and {**2**,**3**,**4**}, we can estimate a rough hysteresis in each environment. We obtain the largest deviation of 1.03 kcal/mol in protein environment for mutation between **1** and **4**, and 0.84 kcal/mol in water for mutation between **1** and **3** (data not shown). However, the deviations between the forward and reverse runs in terms of $\Delta\Delta G_{total}$ are smaller than 0.7 kcal/mol, as shown in Table 3. Relatively large deviations for mutation between **1** and {**3**, **4**} in protein or in water might result from the change in the linker position. The mutation between **1** and **2** involves only a change of an atom in a rigid ring, while the mutation **1**↔**3** and **1**↔**4** requires a change in linker atom. Changes in bond lengths and angle of central linker atom can affect the overall

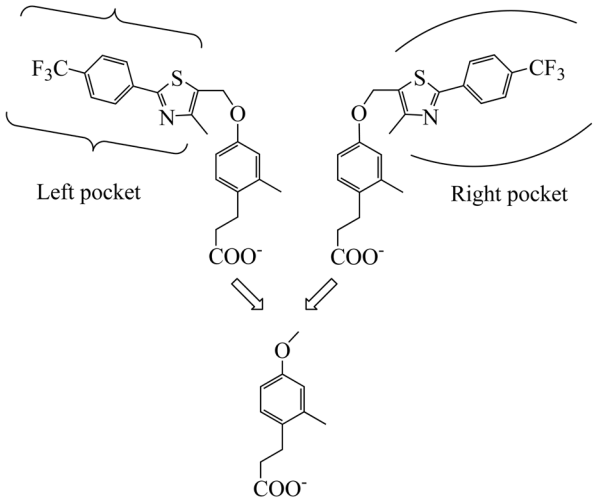


Figure 5. Schematic diagram to examine the different binding poses of Ligand **5**. From the two binding poses obtained by docking calculations, we mutate the hydrophobic tail into methyl group.

position of a ligand, especially in protein environment.

With the $\Delta\Delta G_{bind}$'s obtained by averaging two runs, we compare the calculated values with the data from experiments[3] in Table 3 and in Figure 4. The mean deviation from experiments is 0.96 kcal/mol for five pairs. The largest deviation from the experimental value occurs for **5**→**1**. We obtain 4.62 kcal/mol, but this underestimates severely the relative binding affinity of ligand **5** against ligand **1**. The starting conformation of ligand **5** has been selected from the fourth conformational cluster (Table 2). If a different binding mode with lower binding free energy is possible, the difference in $\Delta\Delta G_{bind}$ against ligand **1** can be reduced. The orange conformer shown in Figure 3 is also found in docking results of other ligands. To examine this aspect, we performed an additional TI calculation. The mutation strategy for this calculation is shown in Figure 5.

We obtain *ca*. 6 kcal/mol of relative stability for the conformer with the tail in the right pocket by this calculation. The detailed results are available in supporting information. Hence the largest underestimation of the binding affinity for ligand **5** might not result from an incorrect starting binding mode. One notable point for this mutation is the large partial charge changes in carboxylate. As benzyl instead of phenoxy substitution has an opposite inductive effect to the carboxylate group, we obtain *ca*. 40 kcal/mol for $\Delta G_{chg}^{5\to1}$ in protein and in solution. Due to the relatively large $\Delta G_{chg}^{5\to1}$, we have performed additional calculation with more sampling points, but no substantial change has been observed. Therefore, we guess that this results from the slight overestimation of the partial charges of the carboxylate of ligand **5**. We suppose that the carboxylate group always makes a hydrogen bond with nearby protein anchors or water molecules in current situation. Thus the inductive effect reflected on the partial charges of ligand **5** in gas phase may be slightly relieved with the hydrogen bond.

Although we have obtained a result that is inconsistent with experimental binding affinity[3] for ligand **5**, we can still analyze the effects of substitutions with the subset of the ligands. Increased affinities are observed for methyl substitution in the head group as found in docking. By comparing the sets {1,2,4} and {6,7,8}, we can observe the additive effects of methyl substitutions. This indicates that the receptor already has a enough room for accommodating this functional group.[3,4] Additive effects of the substitution at Y and Z positions are also observed. The change from ligand **4** to ligand 1 shows that a larger size atom in Y or Z position enhances the affinity for the receptor. We have included ligand **3**, which does not have experimental data,[3] to examine the additive effects at the positions Y and Z. We observe that ligand **2** has an enhanced affinity than ligand **4**. As in the two-atom spacer of GW2433, the sulfur at Y position may experience more steric contact with the nearby liphophilic residues (mainly Cys285) than oxygen. In addition, the change from oxazole to thiazole also results in increased affinities (from ligand **2** to ligand **1** and ligand **7** to ligand **6**). If we look at the binding pose of this scaffold, the linker Y can be regarded as a hinge (see Figure 4d) of a ligand. With the thizaole, this hinge atom is supposed to take a role of holding out the whole ligand to the binding pocket. The remaining hydrophobic tail is supposed to maintain the whole ligand more firmly to binding pocket. The ligands under investigation are agonists for the receptor. Hence it will be important that a ligand must have a key function to activate the receptor as well as high affinity. This key function is known to stabilize the AF-2 helix through hydrogen bond between ligand head group and anchor triad (His323, His449 and Tyr473) of the receptor.[1] In the absence of a ligand, water molecules are positioned near the anchor residues, as observed in the X-ray structure of the apo form of the receptor (pdb entry 2GWX). Hence the existence of the high affinity hinge that maintains the hydrophobic tail stable will contribute to enhance the ligand affinity (IC_{50}). But this aspect is not directly related to drug efficacy (EC_{50}). In Table 1, we can see the increase in EC_{50} and IC_{50} in both changes of ligand **4** to ligand **2** and ligand **8** to ligand **7**. These alterations modify the linker atom from oxygen to sulfur. The enhancements in IC_{50} are relatively larger than those in EC_{50} for these cases. Although direct prediction of drug efficacy (EC_{50}) from the binding affinities is not generally possible, we note that the change in hydration properties of the ligands can explain the trend in EC_{50}'s based on the current test set. With the same TI protocol used for test case molecules, we obtain the relative free energy differences of hydration for the ligands. While $\Delta\Delta G_{hyd}^{1\to2}$ is −0.04 kcal/mol, $\Delta\Delta G_{hyd}^{1\to3}$ and $\Delta\Delta G_{hyd}^{1\to4}$ are −5.60 kcal/mol and −5.30 kcal/mol, respectively. Therefore, ligand **2** should be less soluble than ligand **4**. We guess that this reduces the chance to encounter the receptor in cell based transactivation assay. The evaluation of hydration free energy requires an additional calculation involving the alchemical mutation in vacuum, but this requires a relatively small computing time. Therefore, in addition to the binding affinity calculation, evaluation of the ligand solvation properties should also play

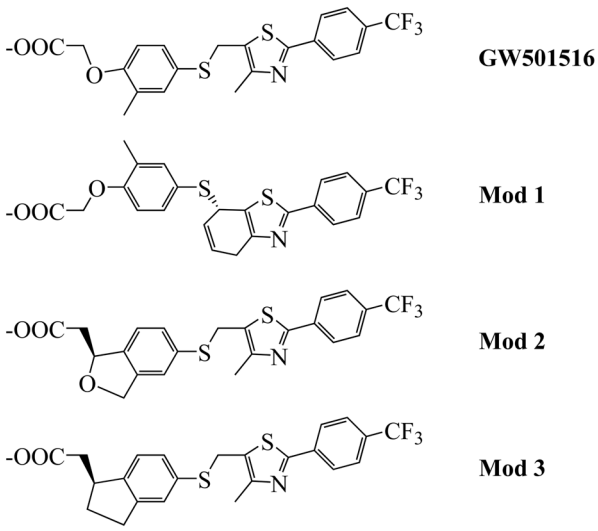


Figure 6. Structures of modified ligands with more rigidity.

an important role in designing the good agonists.

Design of modified ligands. As we have examined the experimentally verified agonists, we now suggest a new modification based on the current calculation method. In Figure 3d, we can see that the sulfur atom in thiophenoxy group occupies a similar volume as the two methyl group of GW2433. We thus focus our attention on reducing the torsional degrees of freedom to enhance the drug efficacy while maintaining the ligand size. Another point is that similar binding poses of the head group are observed across all ligands investigated. Hence we add constraints to a ligand to maintain the active pose if possible. As the PPARδ has a relative large portal and binding pocket, slightly rigid ligands also can enter the binding pocket.

The trial structures are shown in Figure 6. We first obtain the molecular docking results for the modified ligands shown in Figure 6. The **Mod 1** in Figure 6 shows unsatisfactory binding pose mainly due to the direction of the central linker atom that is positioned almost orthogonal to the thiazole ring. This results in the uplift of the hydrophobic tail of the ligands compared to ligand **1** (GW501516). Docking calculations of **Mod 2** and **3** show that they have very similar binding poses with each other and also with that of ligand **1**. But the estimated free energy of binding by Autodock scoring function shows that the binding affinity of **Mod 3** is lower by *ca*. 0.5 kcal/mol than **Mod 2**. This indicates the lipophilicity of the protein residues nearby the cyclized moiety of the modified ligand.

Therefore, we select **Mod 3** as a new candidate, and carry out TI calculations to mutate it into ligand **1**. For this mutation, we perform the sampling in two different ways. First run is done by the same protocol like others in this study (run-1), and the second run is performed in a sequential manner for all calculations (run-2). In the second run, we increase the sampling points for *chg* scaling from 3 to 5 points of Gaussian quadrature, while the sampling time at each λ point in *vdw* scaling is decreased from 150 ps to 100 ps. $\Delta\Delta G$(mod3→lig1) obtained from run-1 and run-2 are 1.51 kcal/mol and 1.60 kcal/mol, respectively. The detailed results are available in supporting information. The results from the two runs are in agreement with each other, but the run time of run-2 takes about four times longer than run-1. The relative free energy change from **Mod 3** to ligand **1** indicates increased binding affinity by the modification. It should also be emphasized that by using the concurrent sampling protocol used in this study the total run time can be reduced by a factor of 4.

Conclusions

We have proposed a computational pathway that can validate and predict the relative binding affinities of protein-ligand interactions in the receptor-based drug design. Comparison with experimental data shows that a more correct binding affinity prediction is possible with MD-based method than docking calculation. Although the MD-based method is time consuming, increased computing power, improvement of calculation software, and smart utilization of calculation protocol will make the method more viable. Unlike the initial stage of drug discovery, lead optimization process is generally performed for congeneric ligands that share similar scaffold. As in the PPARδ case, if the binding pocket has a relatively wide portal with many hydrophilic residues and thus the ligand experiences a small hindrance to enter the binding site, competition with water molecules for hydrogen bonding becomes an important factor. This situation can be treated properly by MD-based free energy calculations. We expect that the efficiency of the proposed calculation protocol as well as the numerical results on PPARδ will provide a helpful guidance for related researches. In subsequent works, we will report the results on the modified agonists that utilize the remaining hydrophobic pocket.

Acknowledgment. This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2005-070-C00065).

Supporting Information. Detailed calculation results are available at BKCS website (*www.kcsnet.or.kr/bkcs*).

References

1. Willson, T. M.; Wahli, W. *Curr. Opin. Chem. Biol.* **1997**, *1*, 235.
2. Braissant, O.; Foufelle, F.; Scotto, C.; Dauca, M.; Wahli, W. *Endocrinology* **1996**, *137*, 354.
3. Sznaidman, M. L.; Haffner, C. D.; Maloney, P. R.; Fivush, A.; Chao, E.; Goreham, D.; Sierra, M. L.; LeGrumelec, C.; Xu, H. E.; Montana, V. G.; Lambert, M. H.; Willson, T. M.; Oliver, W. R., Jr.; Sternbach, D. D. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1517.
4. Oliver, W. R.; Shenk, J. L.; Snaith, M. R.; Russell, C. S.; Plunket, K. D.; Bodkin, N. L.; Lewis, M. C.; Winegar, D. A.; Sznaidman, M. L.; Lambert, M. H.; Xu, H. E.; Sternbach, D. D.; Kliewer, S. A.; Hansen, B. C.; Willson, T. M. *PNAS* **2001**, *98*, 5306.
5. Wilson, T. M.; Brown, P. J.; Sternbach, D. D.; Henke, B. R. *J. Med. Chem.* **2000**, *43*, 527.
6. Wang, Y. X.; Lee, C. H.; Tiep, S.; Yu, R. T.; Ham, J.; Kang, H.; Evans, R. M. *Cell* **2003** *113*, 159.
7. Kang, H. *et al.* Gestational treatment with a novel PPARδ agonist

promotes running endurance (in preparation).
8. Hamilton, J. A. *Prog. Lipid Res.* **2004**, *43*, 177.
9. Miyamoto, S.; Kollman, P. *Proteins* **1993**, *16*, 226.
10. Oostenbrink, B. C.; Pitera, J. W.; van Lipzig, M. M. H.; Meerman, J. H. N.; van Gunsteren, W. F. *J. Med. Chem.* **2000**, *43*, 4594.
11. Steinbrecher, T.; Case, D. A.; Labahn, A. *J. Med. Chem.* **2006**, *49*, 1837.
12. Xu, H. E.; Lambert, M. H.; Montana, V. G.; Parks, D. J.; Blanchard, S. G.; Brown, P. J.; Sternbach, D. D.; Lehmann, J. M.; Wisely, G. B.; Willson, T. M.; Kliewer, S. A.; Milburn, M. V. *Mol. Cell.* **1999**, *3*, 397.
13. Berstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer Jr., E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
14. Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R. E.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, 2004.
15. Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
16. Mehler, E. L.; Solmajer, T. *Protein. Eng.* **1991**, *4*, 903.
17. Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 720.
18. Gamess 2004. Nov.
19. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
20. Stouten, P. F. W.; Frömmel, C.; Nakamura, H.; Sander, C. *Mol. Simul.* **1993**, *10*, 97.
21. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
22. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
23. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639.
24. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
25. Darden, D.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
26. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
27. Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281.
28. Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 3830.
29. Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.
30. Simonson, T. *Mol. Phys.* **1993**, *80*, 441.
31. Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564.
32. Nonlinear mixing rule used in amber8, http://amber.scripps.edu/doc8/amber8.pdf.
33. *Encyclopedia of Computational Chemistry*; Vol. 2 pp 1036-1061.
34. Amber tutorial for TI calculation http://amber.scripps.edu/tutorial/shirts/index.html.
35. Murray, C. W.; Baxter, C. A.; Frenkel, D. A. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 547.
36. Shirts, M. R.; Pietera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740.