Computational Identification of Potential Chemoprophylactic Agents according to Dynamic Behavior of Peroxisome Proliferator-Activated Receptor gamma

Zhiwei Yang [1, 2, 3, *], Yizhen Zhao [1], Dongxiao Hao [1], He Wang [1], Shengqing Li [4], Lintao Jia [5], Xiaohui Yuan [6], Lei Zhang [1], Lingjie Meng [2, 7], Shengli Zhang [1, *]

[1] *MOE Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, School of Physics, Xi'an Jiaotong University, Xi'an 710049, China*

[2] *MOE Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter, School of Chemistry, Xi'an Jiaotong University, Xi'an 710049, China*

[3] *School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China*

[4] *Department of Pulmonary and Critical Care Medicine, Huashan Hospital, Fudan University, Shanghai 200041, China*

[5] *State Key Laboratory of Cancer Biology, Department of Biochemistry and Molecular Biology, Air Force Medical University, Xi'an 710032, China*

[6] *Institute of Biomedicine, Jinan University, Guangzhou 510632, China*

[7] *Instrumental analysis center, Xi'an Jiao Tong University, Xi'an 710049, China*

Materials and Methods

1. Representative Structure Setup

To generate a representative set of PPAR$_\gamma$ ligand-binding domain (PPAR$_\gamma$-LBD) structures, four crystal structures were retrieved from RCSB Protein Data Bank: 1PRG [1] (apo-active form, abbreviated to **Apo-active**), 1FM6 [2] (active form with full agonist rosiglitazone (RSG) and coactivator peptide, abbreviated to **Full**), 3B3K [3] (active form with partial agonist S-enantiomer of (2S)-2-(biphenyl-4-yloxy)-3-phenylpropanoic acid (LRG^S), abbreviated to **Partial-1**), and 3D6D [3] (active form with partial agonist R-enantiomer of (2S)-2-(biphenyl-4-yloxy)-3-phenylpropanoic acid (LRG^R), abbreviated to **Partial-2**). Besides, apo-inhibition form of PPAR$_\gamma$-LBD with corepressor peptide was conducted by the MODELER module [4, 5], with the templates of PPAR$_\alpha$-LBD (1KKQ, in inhibition state) [6] and 1PRG [1] (abbreviated to **Apo-inhibited**, see Figure S1). The reliability of homology model was evaluated by Profile-

---

*Corresponding authors. Tel & Fax: +86-29-82660915. E-mails: yzws-123@xjtu.edu.cn (ZWY) & zhangsl@xjtu.edu.cn (SLZ)

3D module [5] and Procheck program [7]. All the hetero-atoms were removed, and missing hydrogen atoms were added using Discovery Studio [5], based on the expected charge distributions of amino acids at neutral pH. The energy minimizations of the five structures were performed with Charmm27 force field [8], until converged to 0.01 kcal·mol^{-1}·Å^{-1}.

2. Molecular Dynamics (MD) Simulation

The five energy-minimized structures were equilibrated by MD simulations, using GROMACS5.1.4 program [9] and Charmm27 force field [8]. Details of the MD simulation setup are in agreement with our previous works [10-12]. In brief, each system was solvated in a cubic box of SPC/E (simple-point-charge) water molecules extending at least 9.0 Å from any solute atom. Na$^+$ counter-anions were placed to neutralize the system. To mimic physiological conditions, the NPT ensemble was applied at constant pressure (1 atm) and 300 K [13]. Particle-mesh Ewald (PME) method [14] and LINCS algorithm [15] were applied to handle long-range electrostatics and constrain all covalent bonds. The cutoff radii for coulomb and van der Waals interactions were set to 8.0 Å. Free dynamics were performed using a 2.0 fs time step, and coordinates were collected every 10.0 ps.

3. Extract Representative Conformations

Root mean-square difference (RMSD) conformational clustering was performed using g_cluster tool that implemented in GROMACS [16]. The resulting trajectory conformations for each simulation were superimposed to remove overall rotation and translation, and then clustered into batches of similar configurations using all backbone atoms and a cutoff in the range 1.1-1.6 Å. For each cluster, the conformation with the smallest RMSD value was chosen as the representative. Except **Full** with four clusters, one cluster was obtained for each structure. For each simulation, most dominant conformation represented over 60 % of the ensemble.

4. Receptor-based Screening with Multiple Conformations

The 'Drugs-Now' and 'TCM Database @ Taiwan' subsets of ZINC database (downloaded on 28 August 2018) [17] were used in virtual screening, and firstly filtered by rules formulated by Lipinski (Rule of Five) [18] and Veber [19]. There are over 7.4 million compounds with molecular weight between 150 and 500, LogP and polar surface area less than 5 and 140 Å^2, as well as numbers of rotatable bonds, hydrogen bond donors / acceptors (sum) being no more than 10, 5/10 (12). Geometry and partial atomic charges of selected compounds were conducted throughout 'Prepare Ligands' and 'Minimize Ligands' tools [5] using the Charmm force field [20], especially correct

ionization and low-energy conformers (converged to 0.001 kcal·mol^{-1}·Å^{-1}).

Virtual screening process was performed via two programs, LibDock [21] and cDocker [22]. The former is a high-throughput algorithm using protein site features (polar and apolar), and the latter is a grid-based method that the residues are held rigid and ligands are allowed to flex during the filtering process. In our virtual screening protocol, the compounds were firstly evaluated (LibDock) across MD-generated representative conformation of **Full**, assigned with a sphere of 15.0 Å binding site. The compounds, that have LibDock Scores larger than those of full agonist RSG and partial agonist LRG^S (≥ 118), were selected for the second filter. In terms of cDocker algorithm, MD-generated representative conformations of **Full** and **Partial-1** were both adopted, and the binding site sphere was assigned with a sphere of 10.0 Å. The optimal orientations of compounds within receptors were probed on the basis of interactions with binding residues and geometrical matching qualities [12, 23], and then energy-minimized with a convergence criterion of 0.01 kcal·mol^{-1}·Å^{-1}. The binding poses of best ten compounds were selected and further refined by 100.0-ns explicit solvent MD simulations, using above described methods in Section "Molecular Dynamics (MD) Simulation".

5. Free energy calculation

The binding free energies (ΔG_{bind}) were evaluated by the molecular mechanics generalized born surface area method (g_mmpbsa) [24]. Details of parameters are similar to those performed the previous works [10, 11].

ΔG_{bind} was estimated by using

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{GB} + \Delta G_{SA} - T\Delta S \tag{1}$$

where ΔE_{MM} represents the molecular mechanical contribution consisting of internal energy ($\Delta E_{internal}$), electrostatic (ΔE_{ele}) and van der Waals (ΔE_{vdw}). ΔG_{GB} and ΔG_{SA} are the polar and nonpolar contributions to solvation free energies. $T\Delta S$ represents the entropic contribution, which can be neglected due to less difference in the same protein system and high computational costs [10, 11]. All values were calculated in averages over 200 snapshots evenly extracted from the 60~100 ns MD trajectories.

6. Geometric analysis

Apart from the standard methods, the secondary structures, volumes and binding pocket size of various PPAR$_\gamma$-LBD conformations were separately determined by the defined secondary structure of proteins (DSSP) method (do_dssp) [25], Discovery studio client [5] and Fpocket program [26]. Structural plotting and visualization were accomplished by Discovery studio client [5].

Principal component analysis (PCA) was calculated using Bio3D package [27], which can transform a series of potentially coordinated observations into a set of orthogonal vectors called principal components (PCs). In this, the positional covariance matrix between the C_α atoms of any two residues generated by the fitted trajectory is defined as follows [28]:

$$\sigma_{ij} = <(x_i - <x_i>)(x_j - <x_j>)> (i,j = 1,2,3,...,3N) \qquad (2)$$

where $x_i(x_j)$ represents the Cartesian coordinate of the ith(jth) C_α atom, $<x_i>$ or $<x_j>$ is the time average over all sampled conformations, and N is the number of the C_α atoms. The symmetrical covariance matrix σ is diagonalized to produce eigenvectors γ_n (namely principal component PC_n) and the corresponding eigenvalues λ_n.

The pairwise cross-correlation coefficient of residues were explored by the dynamic cross-correlation map (DCCM) of C_α atoms through GROMACS implemented tools [9]. The cross-correlation coefficient C_{ij} between the C_α atoms of the ith and jth residues is given by

$$C_{ij} = \frac{<\Delta r_i \cdot \Delta r_j>}{\sqrt{<\Delta r_i \cdot \Delta r_i><\Delta r_j \cdot \Delta r_j>}} \qquad (3)$$

where Δr_i and Δr_j are the displacement vectors of the ith and jth residues. The angle bracket $<...>$ represents the time average over the trajectory. Positive value of C_{ij} shows that the motion of two correlated residues is in the same direction, while negative value represents the opposite way.

Table S1. Docking results of top ten compounds (based on cDocker interaction energies) [a]

#	ZINC No.	Vendors	Catalog number	Full	Partial-1	Average
1	ZINC03775146	AK Scientific	K529	−79.54	−90.00	−84.77
2	ZINC03831462	3B Scientific Corporation	3B2-0795	−72.56	−85.44	−79.00
3	ZINC15120682	--	--	−66.69	−80.88	−73.79
4	ZINC12381030	Vitas-M	STK170412	−64.54	−81.85	−73.19
5	ZINC14087743	BioSynth	M-7200	−73.78	−66.51	−70.14
6	ZINC03874917	3B Scientific Corporation	3B3-013268	−74.87	−65.16	−70.02
7	ZINC17719775	Acros Organics	44948	−69.78	−70.11	−69.94
8	ZINC03874915	3B Scientific Corporation	3B3-013268	−66.95	−72.31	−69.63
9	ZINC58581064	AK Scientific	X7595	−56.73	−81.55	−69.14
10	ZINC85569445	--	--	−66.57	−71.10	−68.84
	RSG [b]	Vitas-M	STL350047	−61.29	−51.57	−56.43
	LRG^S [c]	--	--	−50.01	−66.62	−58.31

a Energy units in kcal mol^{-1}, obtained by the cDocker module;

b Rosiglitazone, control for full agonist;

c S-(2S)-2-(biphenyl-4-yloxy)-3-phenylpropanoic acid, control for partial agonist.

Table S2. Binding free energies and their components of compound-**Full** complexes [a]

#	ZINC No.	ΔE_{ele}	ΔE_{vdw}	ΔG_{sur}	ΔG_{GB}	ΔG_{bind}
1	ZINC03775146	−240.02±3.48	−42.00±1.07	−6.52±0.05	164.21±3.82	−124.17±1.49
2	ZINC03831462	−0.28±1.06	−58.90±0.73	−6.83±0.05	61.83±1.69	−4.16±1.03
3	ZINC15120682	−25.98±5.69	−35.03±4.80	−3.76±0.52	64.16±8.68	−0.92±2.08
4	ZINC12381030	−17.94±4.78	−34.74±4.95	−3.86±0.51	71.37±10.38	14.60±1.07
5	ZINC14087743	−12.40±0.94	−59.64±0.71	−6.78±0.05	54.18±1.38	−24.58±0.97
6	ZINC03874917	−17.58±4.78	−27.98±5.12	−3.28±0.62	66.60±11.40	18.02±1.33
7	ZINC17719775	−10.31±0.98	−58.26±2.81	−6.51±0.32	45.47±2.60	−29.52±1.62
8	ZINC03874915	−17.06±1.27	−47.84±1.05	−5.84±0.05	105.81±0.96	35.25±0.99
9	ZINC58581064	−27.57±1.07	−47.57±0.84	−5.03±0.05	86.88±0.97	6.69±1.05
10	ZINC85569445	−24.49±2.01	−48.93±0.63	−5.83±0.07	55.29±2.47	24.85±1.64

[a] All values are given in kcal mol^{-1}, and behind "±" are their standard deviations (S.D.).

Table S3. Binding free energies and their components of compound-**Partial-1** complexes [a]

#	ZINC No.	ΔE_{ele}	ΔE_{vdw}	ΔG_{sur}	ΔG_{GB}	ΔG_{bind}
1	ZINC03775146	−250.59±3.24	−43.75±0.75	−6.44±0.05	206.68±3.78	−93.90±1.70
2	ZINC03831462	−36.79±0.68	−60.68±0.61	−6.65±0.06	92.02±0.67	−12.09±1.12
3	ZINC15120682	1.23±2.71	−29.87±5.85	−3.03±0.62	47.08±7.95	15.31±2.15
4	ZINC12381030	−20.51±5.37	−37.28±3.86	−4.19±0.48	83.11±7.35	20.74±2.41
5	ZINC14087743	−30.84±1.50	−55.99±0.69	−6.92±0.05	72.60±1.73	−21.29±1.04
6	ZINC03874917	−22.87±4.60	−30.58±3.88	−4.10±0.48	86.44±10.06	29.13±1.76
7	ZINC17719775	−9.13±0.73	−62.16±0.60	−6.55±0.04	37.51±0.96	−40.29±0.65
8	ZINC03874915	−53.24±2.45	−42.65±0.97	−5.56±0.04	137.43±2.52	36.13±1.10
9	ZINC58581064	−41.10±0.91	−45.75±0.55	−4.76±0.04	100.55±0.97	9.12±0.69
10	ZINC85569445	−25.46±1.24	−58.32±0.73	−6.15±0.05	119.80±1.42	29.74±1.30

[a] All values are given in kcal mol^{-1}, and behind "±" are their standard deviations (S.D.).
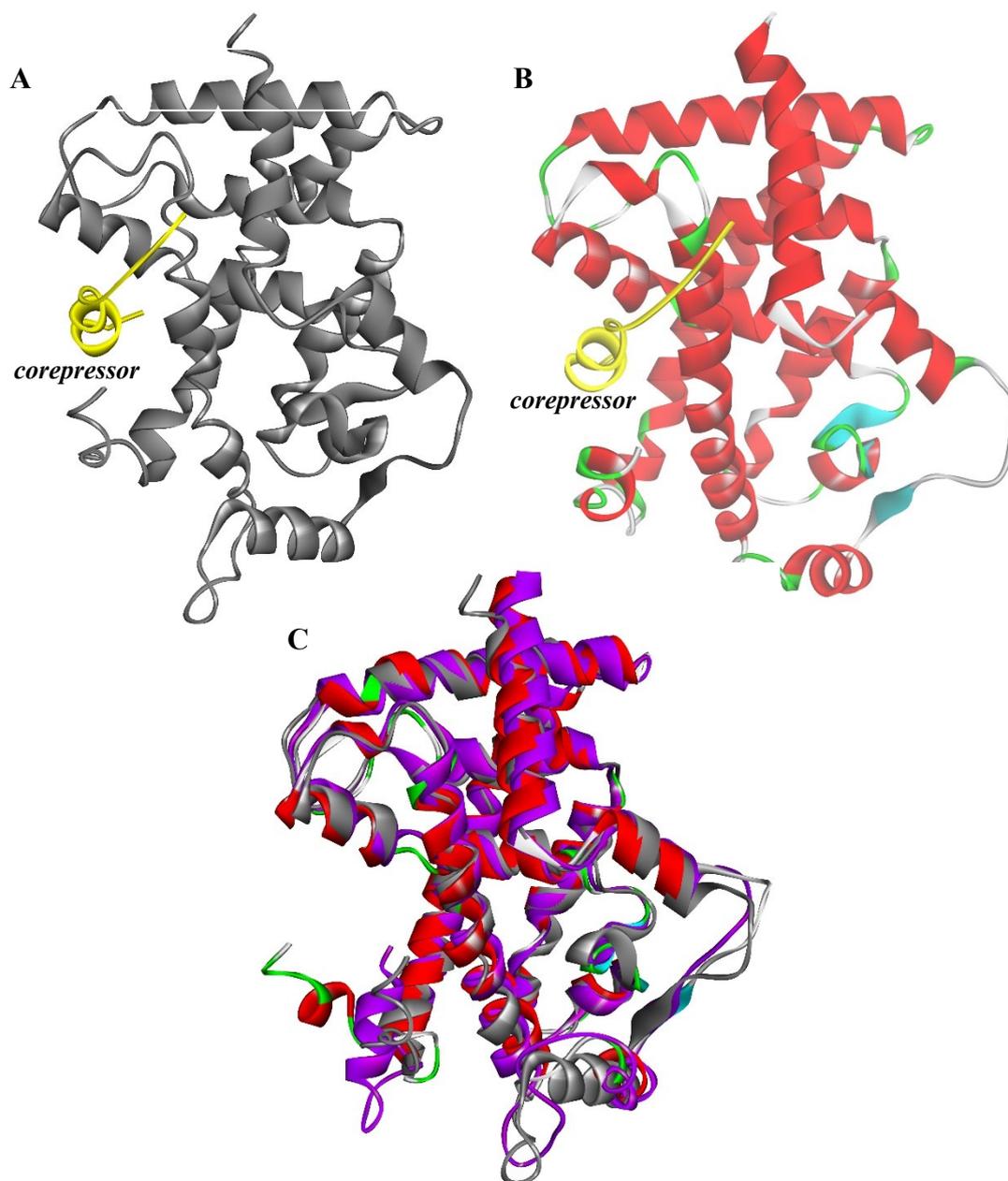
Figure S1. (A) PPAR$_\alpha$-LBD with corepressor (1KKQ chainA [6]), (B) apo-inhibition form of PPAR$_\gamma$-LBD (**Apo-inhibited**) and (C) contrast of **Apo-inhibited** with the two templates (1KKQ [6] and 1PRG [1]). **Apo-inhibited** is in ribbon and conducted by the MODELER module [4, 5], with the templates of 1PRG chainB (purple) [1] and PPAR$_\alpha$-LBD with corepressor (1KKQ chainA, gray). The colors of the ribbons distinguish between helices (red), β-sheets (cyan), hydrogen-bonded turns (green), and random coils (white). The homology modeling structure is almost entirely accord with the two templates, associated with 92 % residues exhibiting reasonable folding (Profile-3D program [5]) and 91% residues being in allowed region of Ramachandran plot (Procheck program [7]). Structural plotting and visualization are accomplished by Discovery studio

client [5].
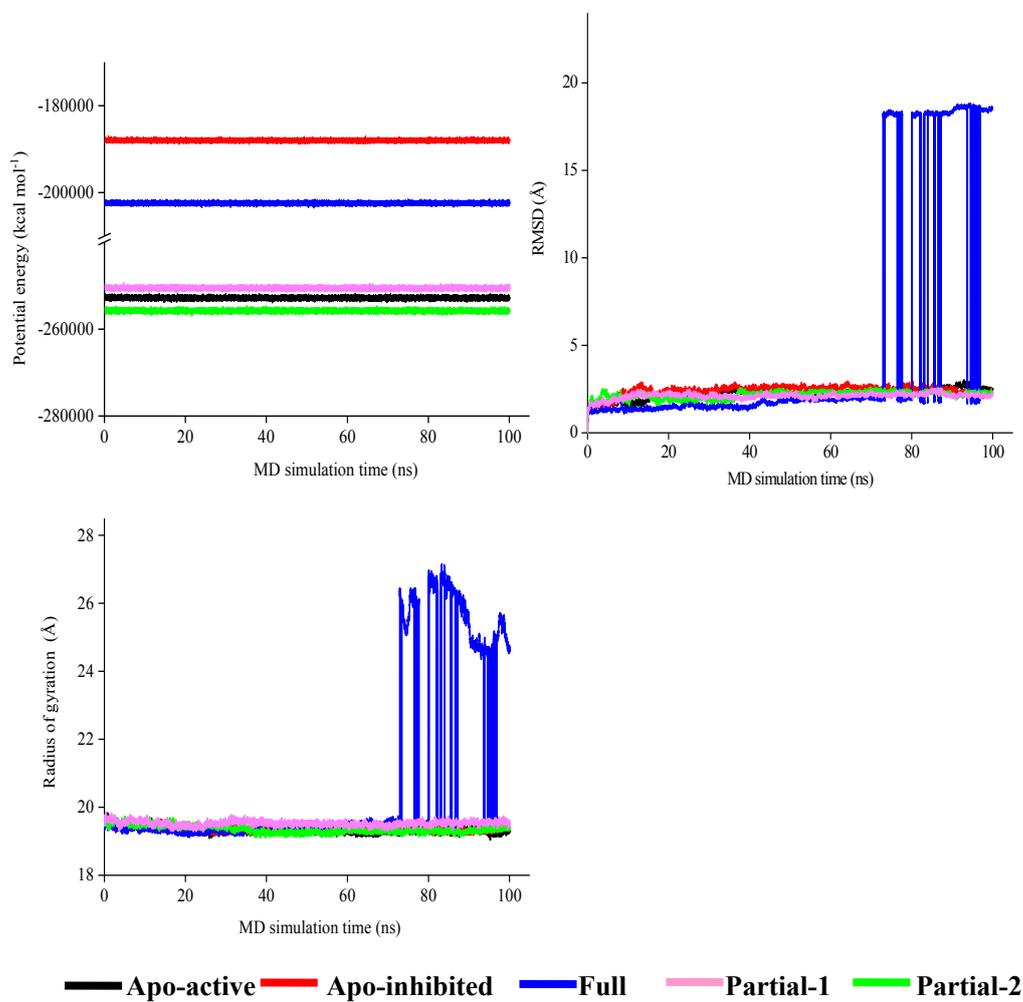
Figure S2. Variation of the potential energy, backbone-atom root-mean-square deviations (RMSD) and backbone radius of gyration (Rg) for various PPAR$_\gamma$-LBD structures during 100-ns MD simulations.
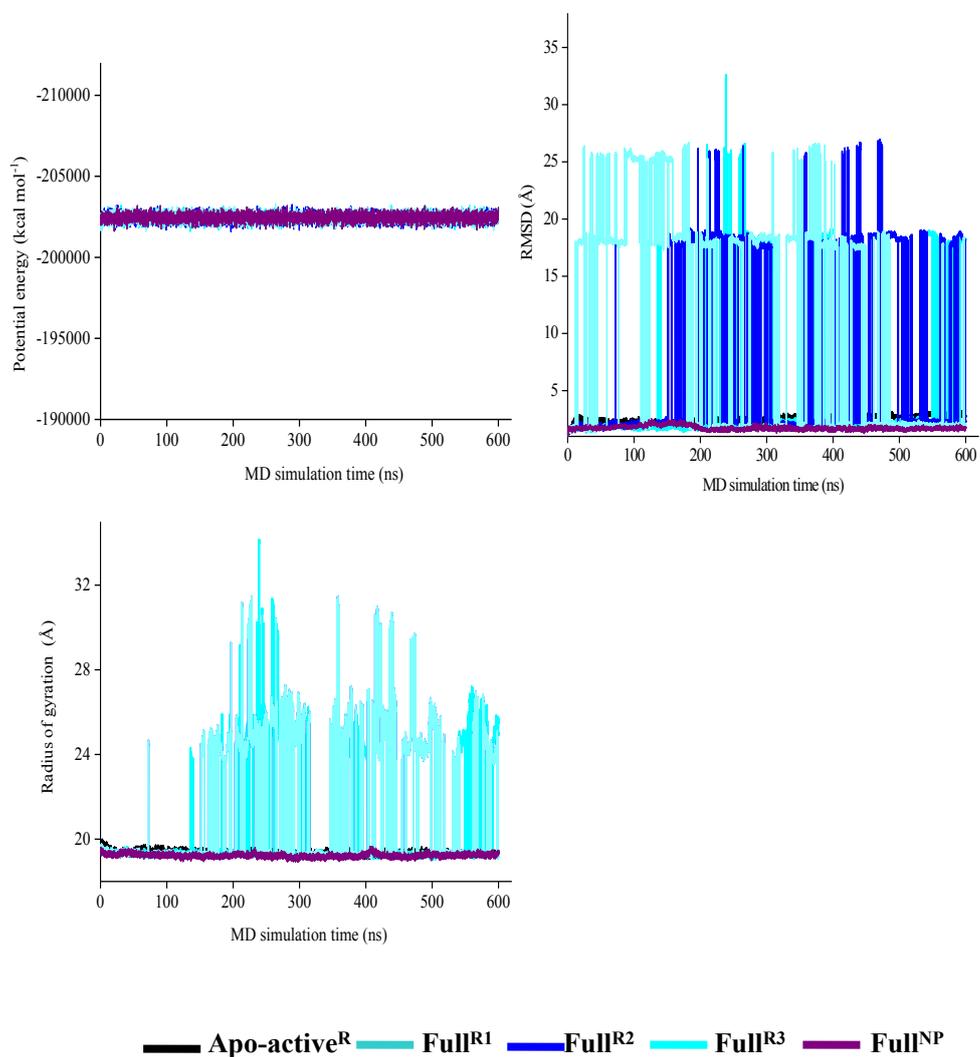
Figure S3. Variation of the potential energy, backbone-atom root-mean-square deviations (RMSD) and backbone radius of gyration (Rg) for various PPAR$_\gamma$-LBD structures during 600-ns MD simulations.
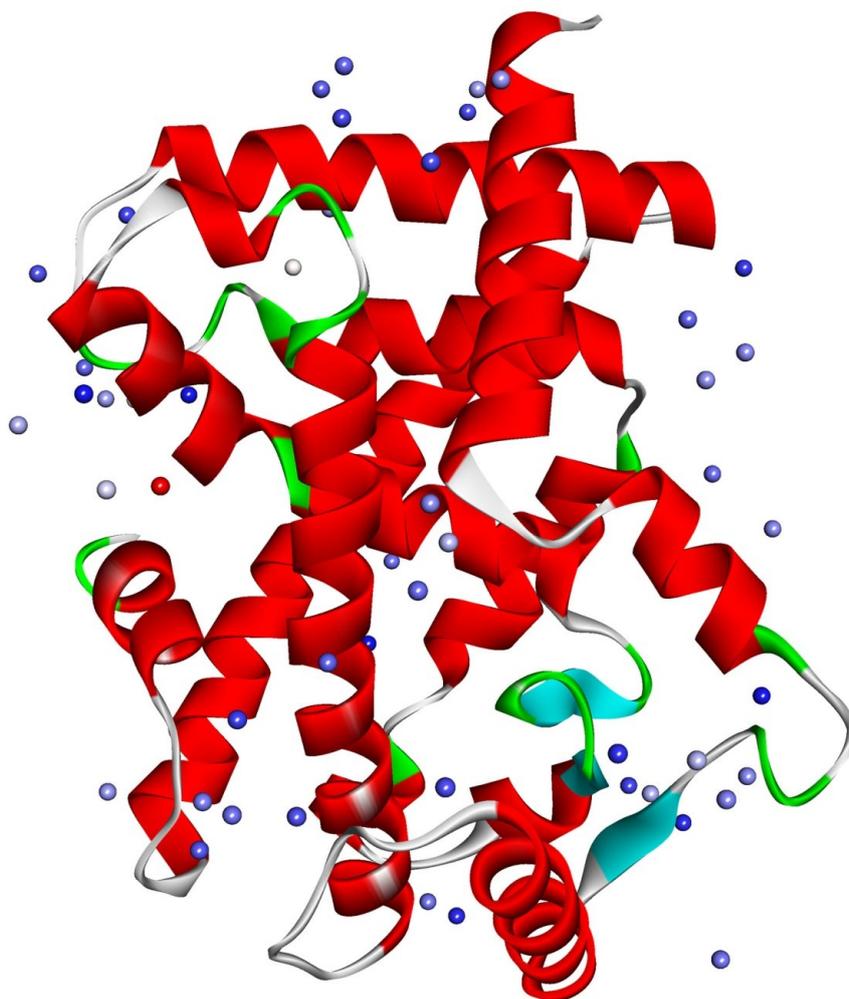
Figure S4. The speculative locations of coactivator peptide within PPAR$_\gamma$-LBD (**Full**) during 600-ns MD simulations. The location of coactivator peptide is represented by ball model. The origin location of coactivator peptide within the crystal structure is in red, and the shifting locations during the MD simulations are in blue, with the odds being represented by the depths of blue colour (gradually decrease from blue to white). Structural plotting and visualization are accomplished by Discovery studio client [5].
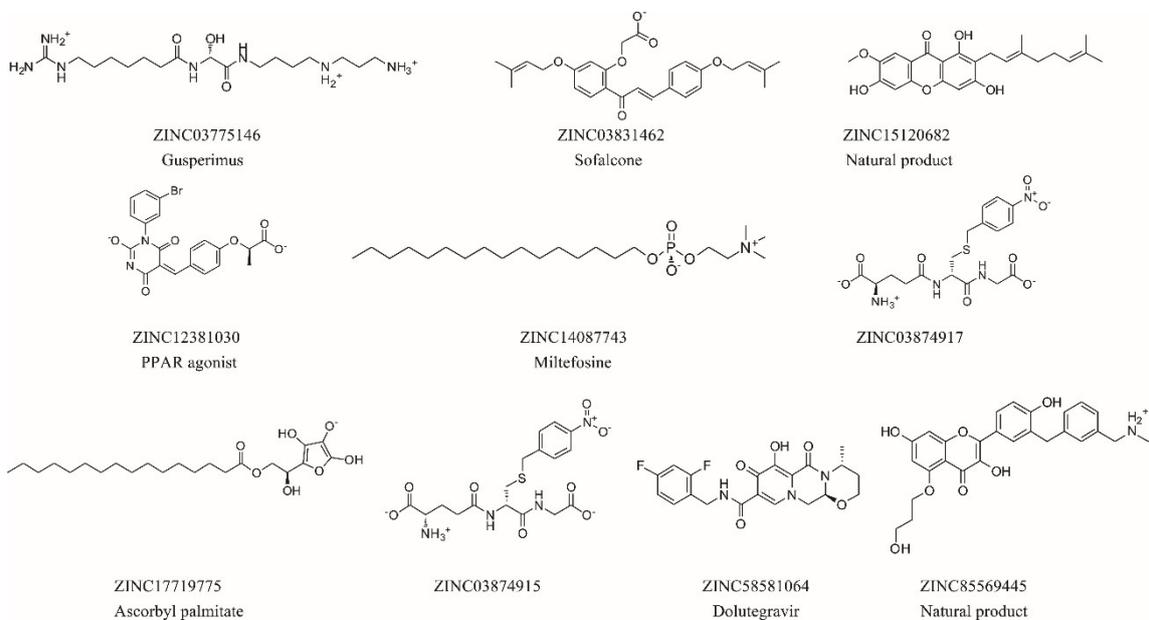
Scheme S1. Structure of the compounds selected after visual inspection of the screening results.

Figure S5. Top ten compounds (compound 1-10) superposed in two PPAR$_\gamma$-LBD conformations (A) **Full** and (B) **Partial-1** and views of their interactions with the key residues. Key residues are represented by stick models. Compounds are represented by ball and stick models. The O, N, C, S atoms are colored in red, blue, green and dark yellow. The important H-bonding interactions are labeled in the green dotted lines.
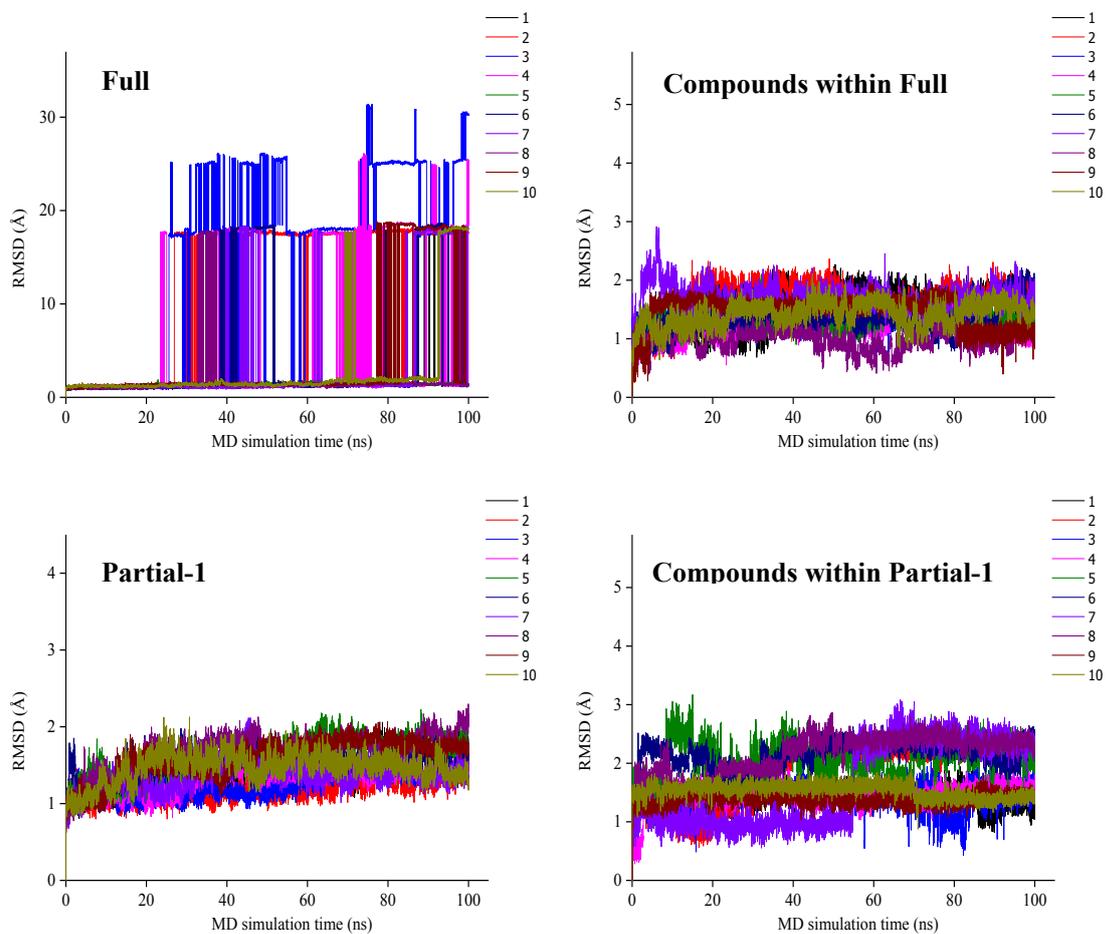
Figure S6. Backbone-atom RMSD of the docked complexes and heavy-atom RMSD of the compounds in 100 ns MD simulation.
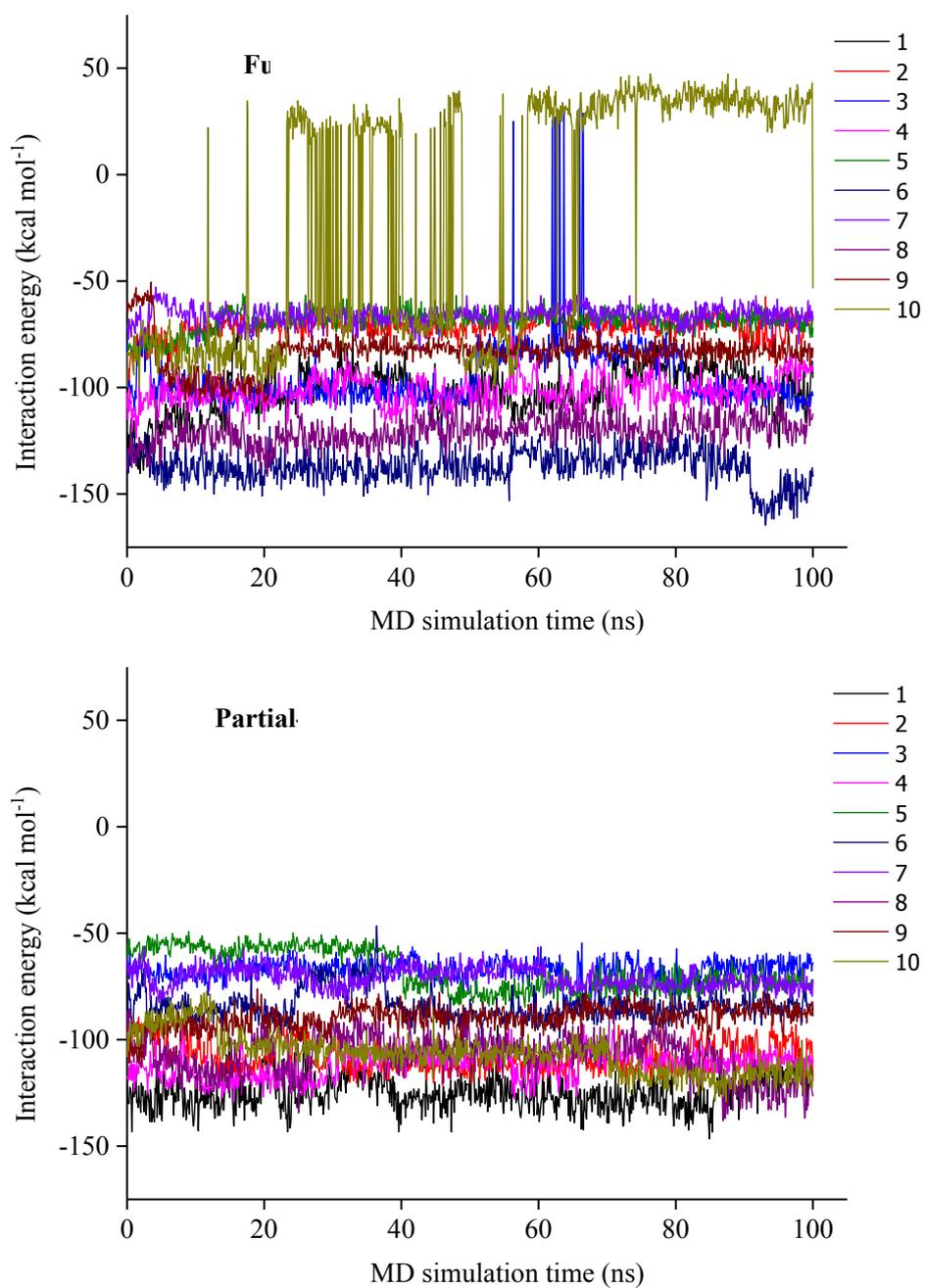
Figure S7. Profiles of interaction energies (short-range energy components) between PPARγ-LBD conformations (**Full** and **Partial-1**) and top ten compounds during the 100 ns MD simulation.
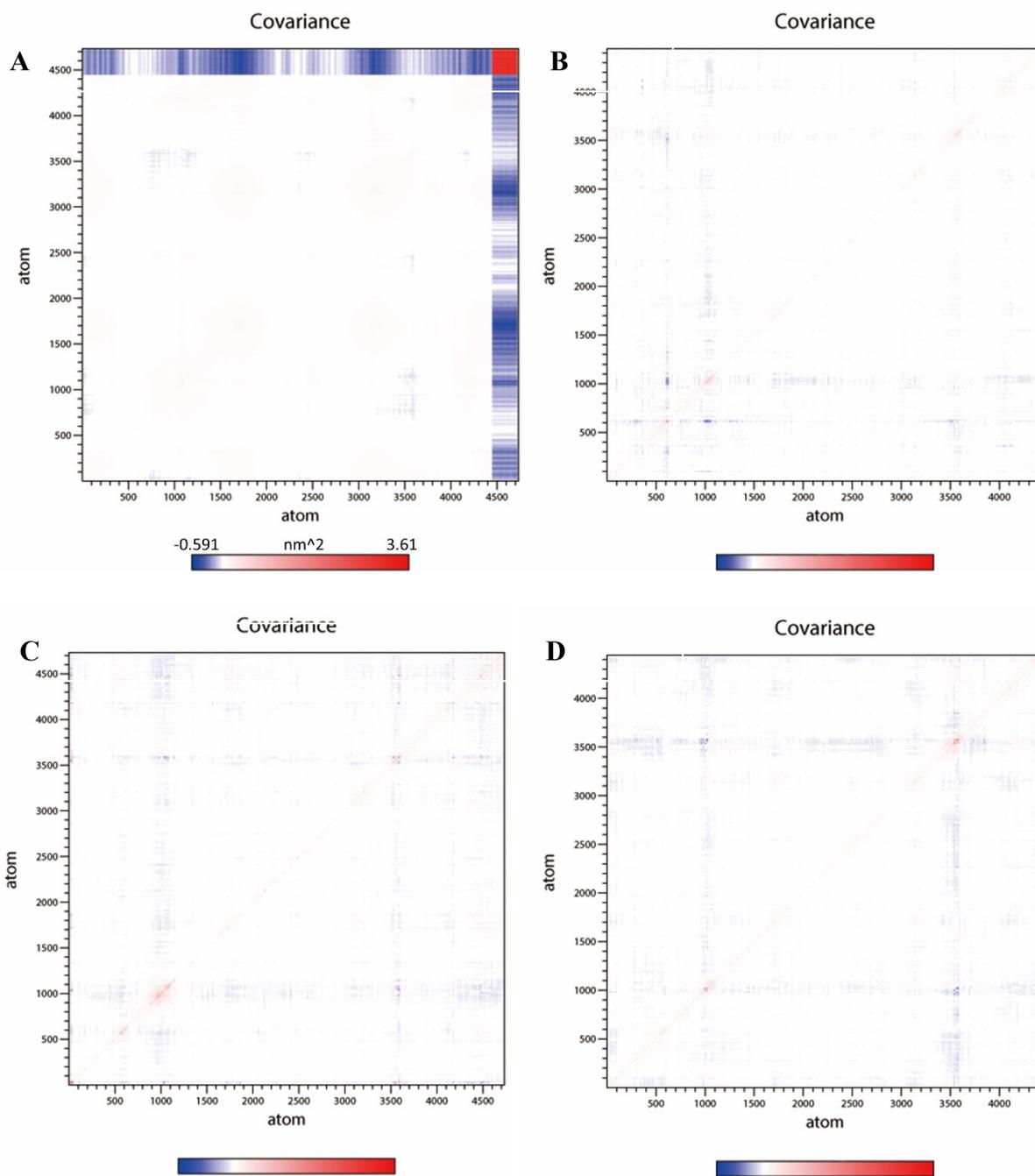
Figure S8. Dynamic cross-correlation maps (DCCM) reveal the extent of correlation for the PPAR$_\gamma$-LBD in the (A) ZINC03775146-**Full**, (B) ZINC03831462-**Partial-1**, (C) ZINC14087743--**Full** and (D) ZINC17719775-**Partial-1** complexes. The color-map covers correlation values between -0.591 and 3.61. Motion occurring along the same direction is represented by positive correlation (blue), while anti-correlated motion occurring along the opposite direction is represented by negative correlation (red). Dynamics cross-correlation matrices (DCCM) are calculated using GROMACS implemented tools [9].

References

1. R. T. Nolte, G. B. Wisely, S. Westin, J. E. Cobb, M. H. Lambert, R. Kurokawa, M. G. Rosenfeld, T. M. Willson, C. K. Glass and M. V. Milburn, *Nature*, 1998, **395**, 137-143.

2. R. T. Gampe, Jr., V. G. Montana, M. H. Lambert, A. B. Miller, R. K. Bledsoe, M. V. Milburn, S. A. Kliewer, T. M. Willson and H. E. Xu, *Mol Cell*, 2000, **5**, 545-555.

3. R. Montanari, F. Saccoccia, E. Scotti, M. Crestani, C. Godio, F. Gilardi, F. Loiodice, G. Fracchiolla, A. Laghezza, P. Tortorella, A. Lavecchia, E. Novellino, F. Mazza, M. Aschi and G. Pochetti, *J Med Chem*, 2008, **51**, 7768-7776.

4. P. Larsson, B. Wallner, E. Lindahl and A. Elofsson, *Protein Sci*, 2008, **17**, 990-1002.

5. Accelrys, Discovery Studio 3.1, http://accelrys.com).

6. H. E. Xu, T. B. Stanley, V. G. Montana, M. H. Lambert, B. G. Shearer, J. E. Cobb, D. D. McKee, C. M. Galardi, K. D. Plunket, R. T. Nolte, D. J. Parks, J. T. Moore, S. A. Kliewer, T. M. Willson and J. B. Stimmel, *Nature*, 2002, **415**, 813-817.

7. R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein and J. M. Thornton, *Journal of biomolecular NMR*, 1996, **8**, 477-486.

8. B. R. Brooks, C. L. Brooks, 3rd, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J Comput Chem*, 2009, **30**, 1545-1614.

9. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *Softwarex*, 2015, **1-2**, 19-25.

10. J. Xia, L. Yang, L. Dong, M. Niu, S. Zhang, Z. Yang, G. Wumaier, Y. Li, X. Wei, Y. Gong, N. Zhu and S. Li, *Front Pharmacol*, 2018, **9**, 134.

11. Z. Yang, Y. Zhao, D. Hao, S. Ren, X. Yuan, L. Meng and S. Zhang, *J Biomol Struct Dyn*, 2020, **38**, 1918-1926.

12. Z. Li, S. Chen, C. Gao, Z. Yang, K. C. Shih, Z. Kochovski, G. Yang, L. Gou, M. P. Nieh, M. Jiang, L. Zhang and G. Chen, *J Am Chem Soc*, 2019, **141**, 19448-19457.

13. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *Journal of Chemical Physics*, 1984, **81**, 3684-3690.

14. T. Darden, D. York and L. Pedersen, *J Chem Phys*, 1993, **98**, 10089-10092.

15. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *Journal of Computational Chemistry*, 1997, **18**, 1463-1472.

16. X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *Angew Chem Int Edit*, 1999, **38**, 236-240.

17. J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, *J Chem Inf Model*, 2012, **52**, 1757-1768.

18. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv Drug Deliver Rev*, 1997, **23**, 3-25.

19. D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, *J Med Chem*, 2002, **45**, 2615-2623.

20. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *Journal of Computational Chemistry*, 1983, **4**, 187-217.

21. S. N. Rao, M. S. Head, A. Kulkarni and J. M. LaLonde, *J Chem Inf Model*, 2007, **47**, 2159-2171.

22. G. Wu, D. H. Robertson, C. L. Brooks, 3rd and M. Vieth, *J Comput Chem*, 2003, **24**, 1549-1562.

23. S. Lu, Q. Shen and J. Zhang, *Acc Chem Res*, 2019, **52**, 492-500.

24. R. Kumari, R. Kumar, C. Open Source Drug Discovery and A. Lynn, *J Chem Inf Model*, 2014, **54**, 1951-1962.

25. W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577-2637.

26. P. Schmidtke, V. Le Guilloux, J. Maupetit and P. Tuffery, *Nucleic Acids Res*, 2010, **38**, W582-589.

27. B. J. Grant, A. P. Rodrigues, K. M. ElSawy, J. A. McCammon and L. S. Caves, *Bioinformatics*, 2006, **22**, 2695-2696.

28. D. M. F. vanAalten, J. B. C. Findlay, A. Amadei and H. J. C. Berendsen, *Protein Engineering*, 1995, **8**, 1129-1135.