Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations

Edithe Selwa, Eddy Elisée, Agustin Zavala, Bogdan Iorga

Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations

Edithe Selwa,[1,‡] Eddy Elisee,[1,‡] Agustin Zavala,[1,‡] Bogdan I. Iorga[1,]*

[1] Institut de Chimie des Substances Naturelles, CNRS UPR 2301, LabEx LERMIT, 91198 Gif-sur-Yvette, France

Corresponding Author

[*] Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

Author Contributions

[‡] These authors contributed equally.

KEYWORDS: docking ; scoring function ; Gold ; Vina ; Autodock ; Farnesoid X receptor ; FXR ; D3R ; Drug Design Data Resource ; Grand Challenge 2

ABSTRACT

Our participation to the D3R Grand Challenge 2 involved a protocol in two steps, with an initial analysis of the available structural data from the PDB allowing the selection of the most appropriate combination of docking software and scoring function. Subsequent docking calculations showed that the pose prediction can be carried out with a certain precision, but this is dependent on the specific nature of the ligands. The correct ranking of docking poses is still a problem and cannot be successful in the absence of good pose predictions. Our free energy calculations on two different subsets provided contrasted results, which might have the origin in non-optimal force field parameters associated with the sulfonamide chemical moiety.

INTRODUCTION

Drug Design Data Resource (D3R, https://drugdesigndata.org/) organizes, on a regular basis, blinded prediction challenges with the aim to evaluate the performance of tools and protocols that are used in real-life computer-aided drug discovery projects. To achieve this, datasets presenting different levels of difficulty are presented to the community, which is asked to predict, in "blind" conditions, the binding modes and the relative affinities of compounds.

The D3R Grand Challenge 2, which was held in 2016, was focused on a single protein, farnesoid X receptor (FXR, Figure 1), a target with multiple potential applications that has received much attention during the recent years [1-30].

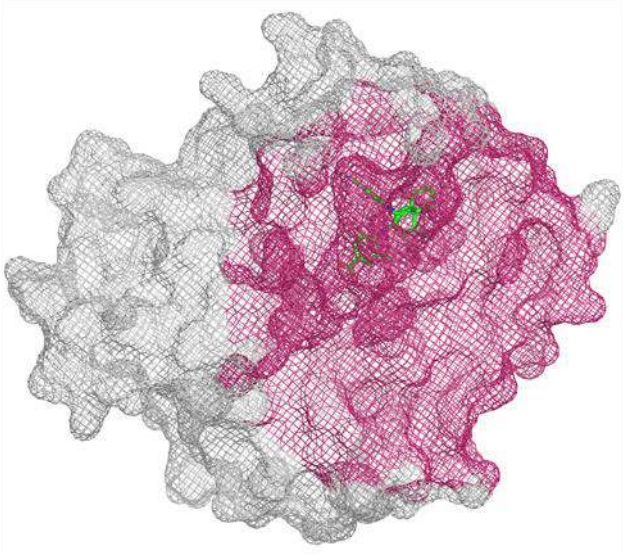
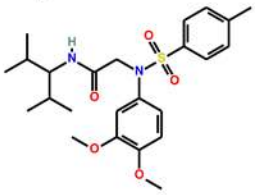


Figure 1. Mesh surface representation of a representative crystal structure (PDB code 3OLF) of the FXR target. The binding site, as defined for our docking studies, is colored in red, and the ligand is colored in green.

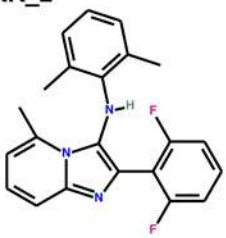
In Phase 1 the participants were asked to provide affinity predictions for 102 FXR ligands and pose predictions for 36 of them. In Phase 2 the participants were required to provide the same affinity predictions as in Phase 1, taking into account the additional structural data (36 new protein-ligand complexes) released at the end of Phase 1.

Figure 2 shows the chemical structures of compounds from FXR dataset for which the pose predictions were required. Most of the compounds included in this dataset can be organized in four homogeneous classes based on their chemical structures (benzimidazoles, isoxazoles, sulfonamides, spiro compounds), and the remaining ones presented inhomogeneous structures and were included in a group called miscellaneous. Biological activities data were available for some compounds from this dataset [31-33]. The exact composition of each group can be found in the Electronic Supplementary Material, as well as the structures of the entire FXR dataset, containing 102 ligands used for ranking prediction (Figure S1).

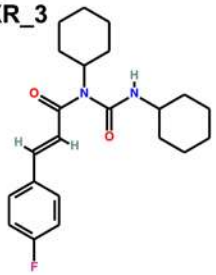
FXR_1



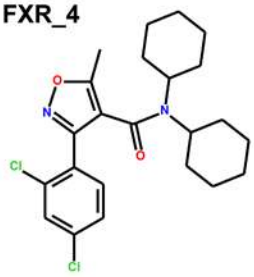
FXR_2



FXR_3

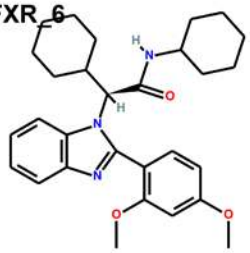


FXR_4



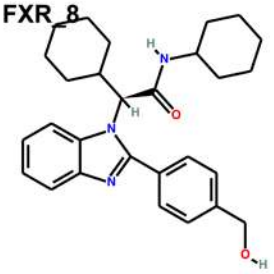
FXR_5

FXR_6

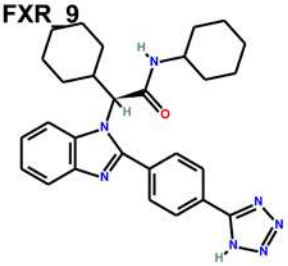


FXR_7

FXR_8



FXR_9

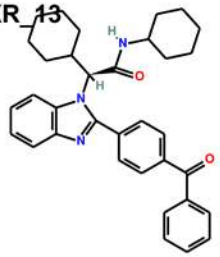


FXR_10

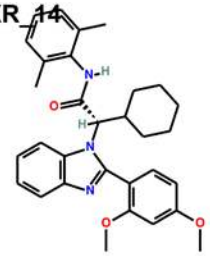
FXR_11

FXR_12

FXR_13

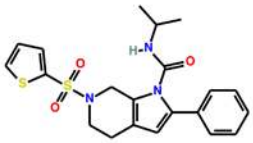


FXR_14



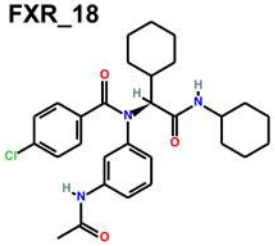
FXR_15

FXR_16

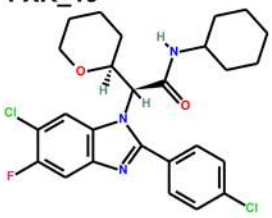


FXR_17

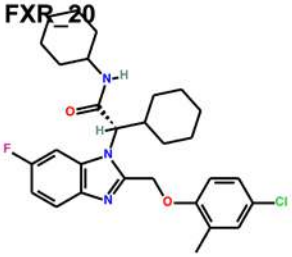
FXR_18



FXR_19



FXR_20



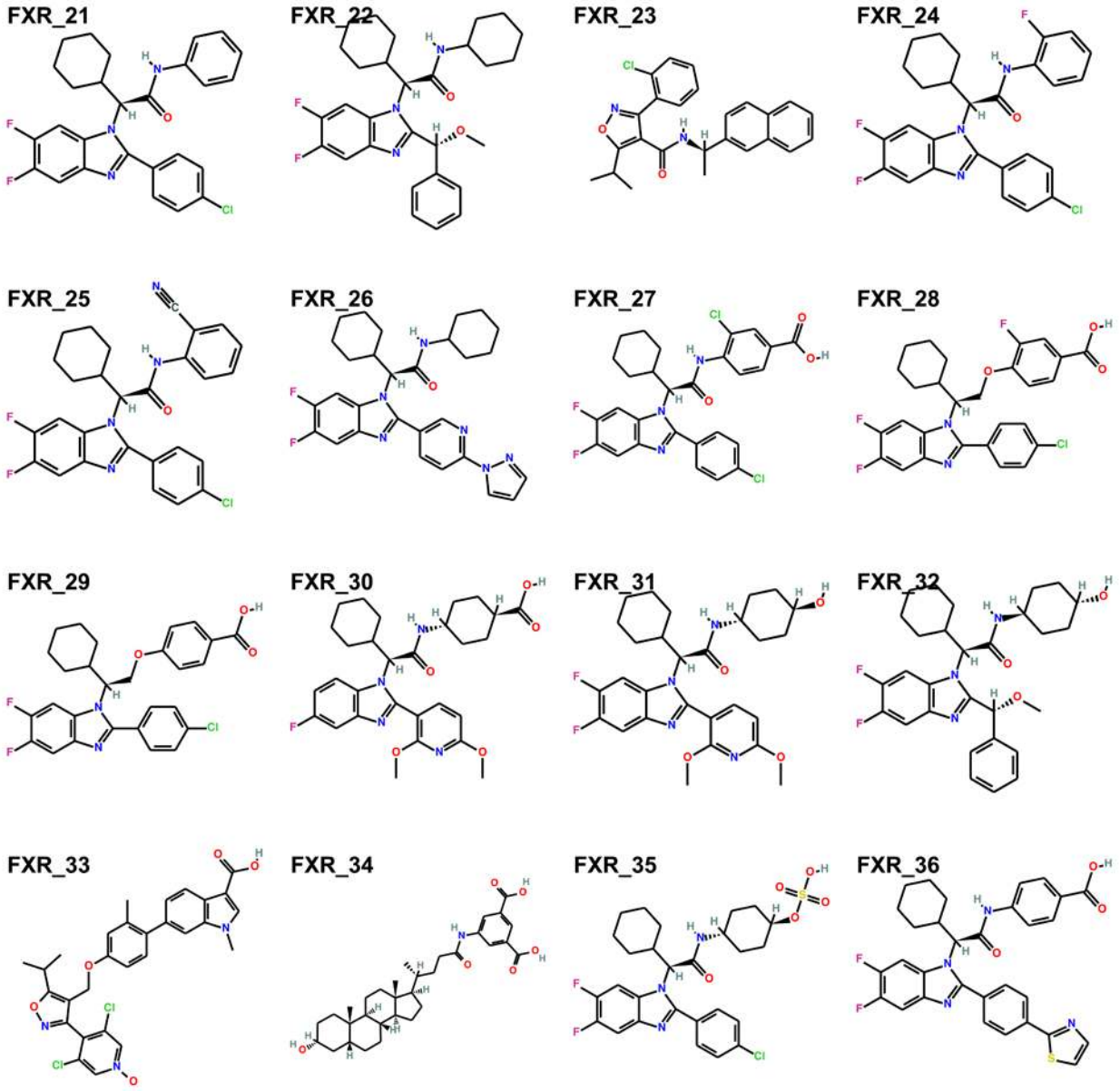


Figure 2. Chemical structures of the 36 FXR ligands included in Phase 1 for pose prediction (compound **FXR_33** was ultimately retired from the pose prediction analysis).

Additionally, the participants were asked to predict the relative affinities for two homogeneous subsets of compounds that are suited for free energy calculations. The structures of compounds (count of 15 and 18, respectively) included in the two free energy subsets are presented in Figures 3 and 4.

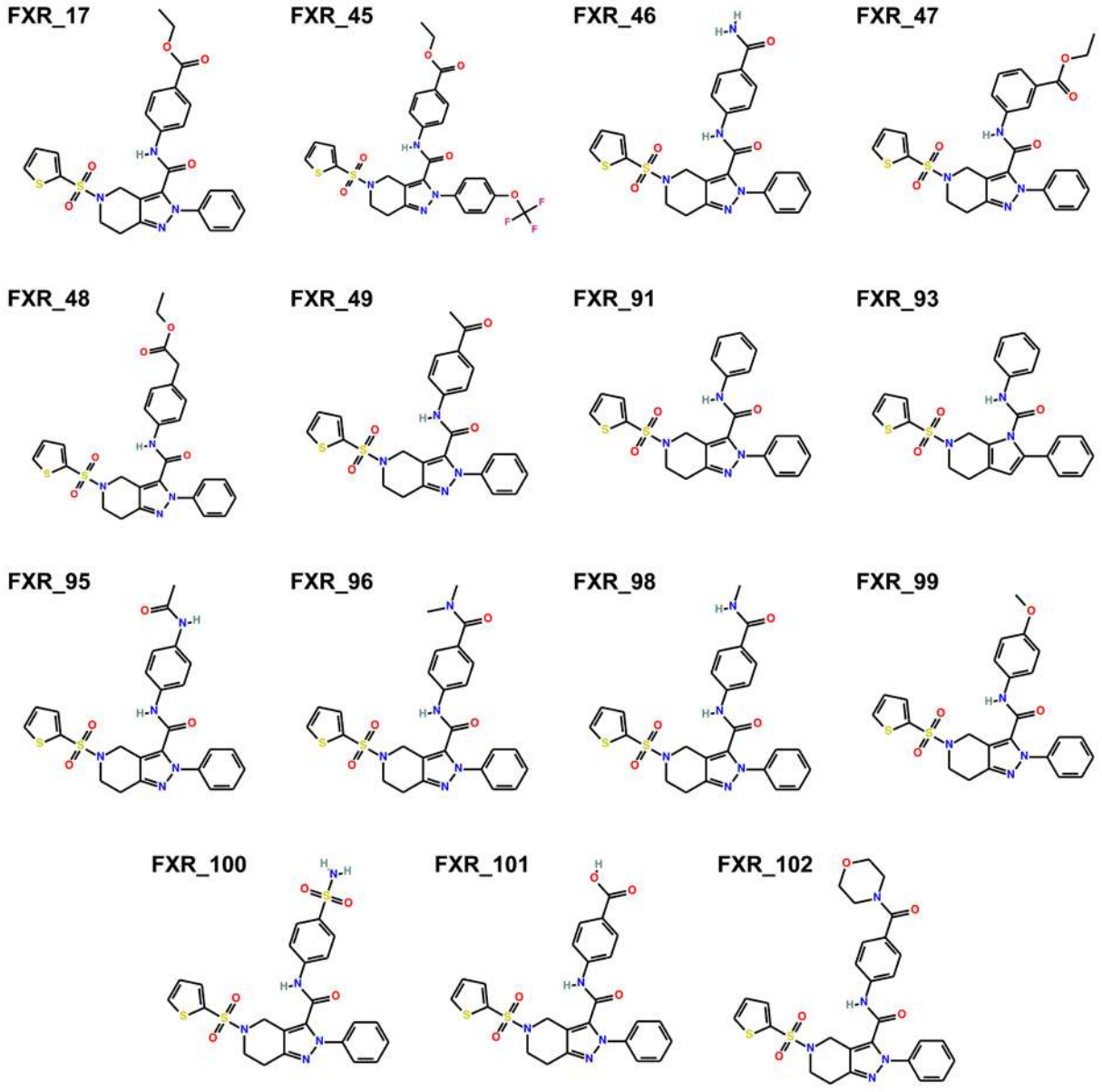


Figure 3. Chemical structures of the 15 FXR ligands included in free energy set1 (sulfonamides).

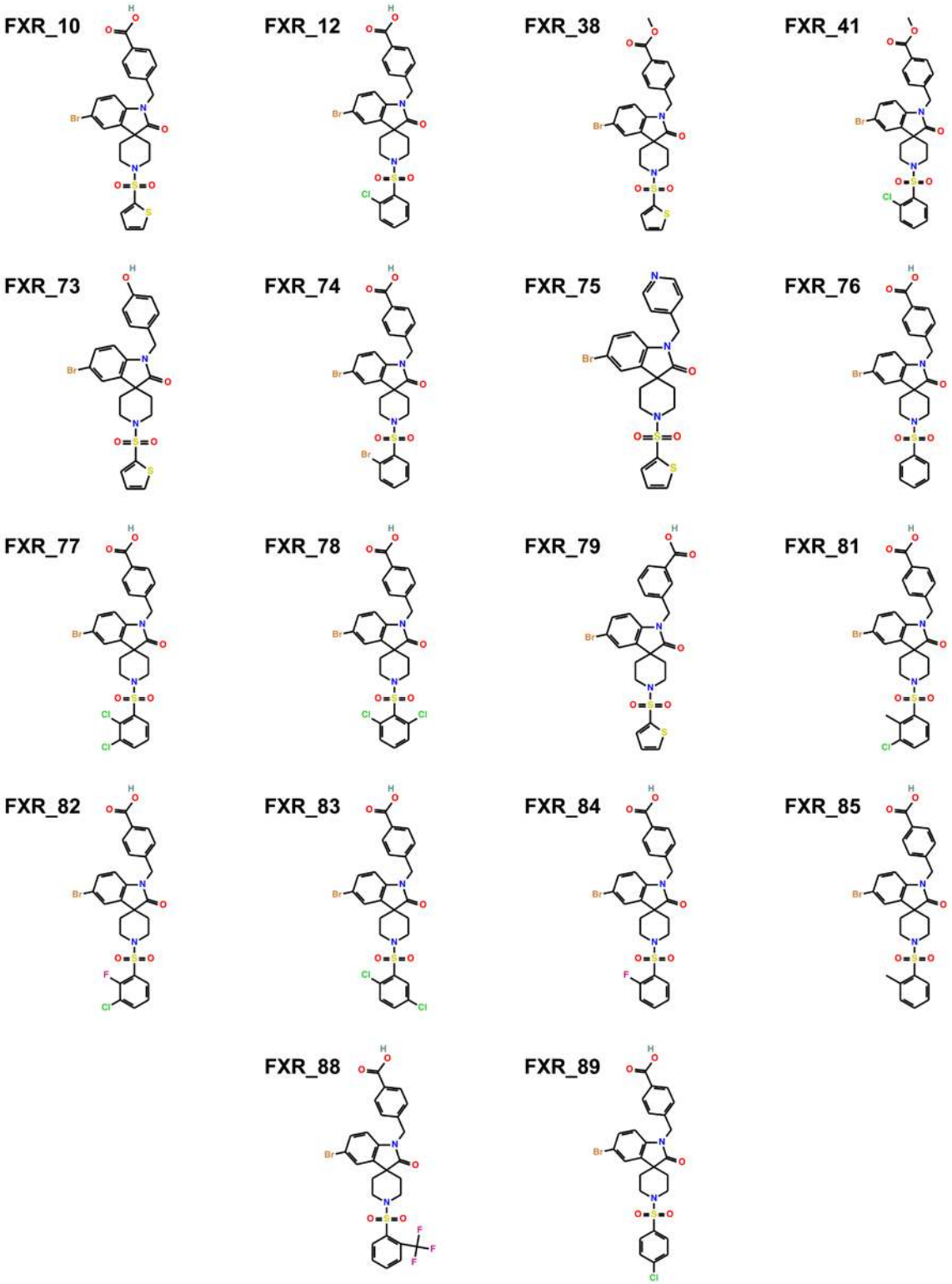


Figure 4. Chemical structures of the 18 FXR ligands included in free energy set2 (spiro compounds).

METHODS

Protein structures. We found 27 crystal structures available in the Protein Data Bank (PDB) [34] for FXR (see the Electronic Supplementary Material for the complete list). These structures constituted our evaluation dataset. All ligands, ions and solvent molecules that were present were manually removed, then the structures were superimposed on the reference structure apo FXR provided by the D3R Grand Challenge organizers, in order to conserve the same coordinate system through the whole process. Missing residues in the structures were added using Modeller 9v12 [35]. Hydrogen atoms were added using Hermes, the graphical interface of Gold v5.2.2 [36] software, or with AutoDock Tools [37] prior to docking.

Ligands. Ligand structures from the evaluation dataset were retrieved from PDB in the SMILES format and they were converted into three-dimensional MOL2 files using CORINA v3.60 (http://www.molecular-networks.com/). This protocol was used instead of retrieving directly the three-dimensional coordinates from the PDB in order to avoid any bias in the docking process that might be related to the initial coordinates of the ligands. Ligand structures used in Phase 1 were obtained from the SMILES strings provided by organizers upon conversion into three-dimensional MOL2 files using CORINA. Three-dimensional coordinates of the ligands used in Phase 2 were built using UCSF Chimera [38], by superimposing their common backbone on the released FXR_17, FXR_10 and FXR_12 crystal structures and by manual addition of the appropriate substituents. In all cases, the protonation state for all compounds was adjusted at physiological pH using LigPrep (Schrödinger, http://www.schrodinger.com/).

Docking. In the preliminary analysis step, several docking software and scoring functions have been tested (re-docking and cross-docking) for their ability to reproduce the protein-ligand complexes from the evaluation dataset: Gold [36] with the GoldScore, ChemScore, ChemPLP and ASP scoring functions, Vina [39] and AutoDock [37]. Default parameters were used in all cases for docking, except with Gold, where a search efficiency of 200% was used in order to better explore the conformational space. The binding sites were considered with Gold as spheres with a 20 Å radius around the Cα atom of Ala288 (numbering from the 1OSV structure). With Vina and AutoDock, the binding sites were defined as a 40 x 40 x 40 $Å^3$ cube centered on the same atom. The protein was considered to be either rigid, or with a few key residues (Leu291/Asn297/Met332/Arg335/Ser336/His451/Trp458 or Arg335 only) from the binding site as flexible. As a result of preliminary analysis, Gold with the ASP scoring function and the rigid protein was used in the Phase 1 predictions. In Phase 2, the rescoring of the FXR complexes was carried out using Gold with the ASP scoring function. For submission, the protein structures were converted into PDB format using UCSF Chimera [38], and the docking poses were converted into MOL format using CORINA (the MOL format corresponds to the SDF output format in CORINA). Unfortunately, the ligand conversion with CORINA was carried out initially without the option "-d no3d", which led to the generation of new coordinates and therefore invalid conformations in our "rjyhz" submission. The results reported here (named "rjyhz_revised") represent the correct poses, obtained with the option "-d no3d".

Free energy calculations. The protein used for free energy calculation was taken from PDB database with PDB entry corresponding to 3FLI. Three-dimensional coordinates of ligands were built using UCSF Chimera [38], by superimposing common backbone on released FXR_17, FXR_10 and FXR_12 structures. In the set2, the structure solved by X-ray crystallography for

FXR_12 had two alternative positions for the aromatic ring substituent (AA and AB). Both of them were considered in our calculations, and the one leading to the most favorable energy chosen for the submission. Alchemical free energies were calculated using Gromacs [17] and OPLS-AA force field [40,41], some scripts from the PMX software [42-44] and some in house developed scripts. The main steps of this protocol are presented in Figure 5. Hybrid structures and topologies were built using a modified version of the MOL2FF package developed in our team. Hybrid topologies represent simultaneously both ligands, the contribution of each structure being controlled by a parameter λ. For example, a λ value of 0 represents exclusively the ligand A, a λ value of 1 represents exclusively the ligand B, and a λ value of 0.3 represents a contribution of 70% of the ligand A and a contribution of 30% of the ligand B. FXR_91 was used as reference structure for set1, and FXR_10 or FXR_12 for set2. Equilibrium 10 ns MD simulations were performed for the two states (corresponding to lambda 0 and 1), using Gromacs [17] and OPLS-AA force field [40,41]. Snapshots from the equilibrium runs were extracted to spawn 100 simulations of 50 ps each to alchemically morph between the two states of the system. The work values over every non-equilibrium transition were extracted and further used to estimate the free energy differences relying on the Crooks Fluctuation Theorem and utilizing Crooks-Gaussian Intersection as estimator. When the charge was the same in the two ligands considered for the alchemical transformation, separate calculations were carried out for the transformation A into B of the protein-ligand complex and of the ligand alone, the relative free energy of binding being the difference between the corresponding work for these two transformations (see Figures S2 and S3 in the Electronic Supplementary Material file). When the charge was different between the two ligands, a single box containing the protein-ligand complex and the ligand alone, separated by 30 Å, was considered. The ligand from the complex was

converted from state A into B, whereas the ligand alone was converted simultaneously from state B into A. In these conditions, the overall charge of the system was conserved during the whole simulation, and the relative free energy of binding between ligands corresponds to the global work for this system (see Figure S4 in the Electronic Supplementary Material file).

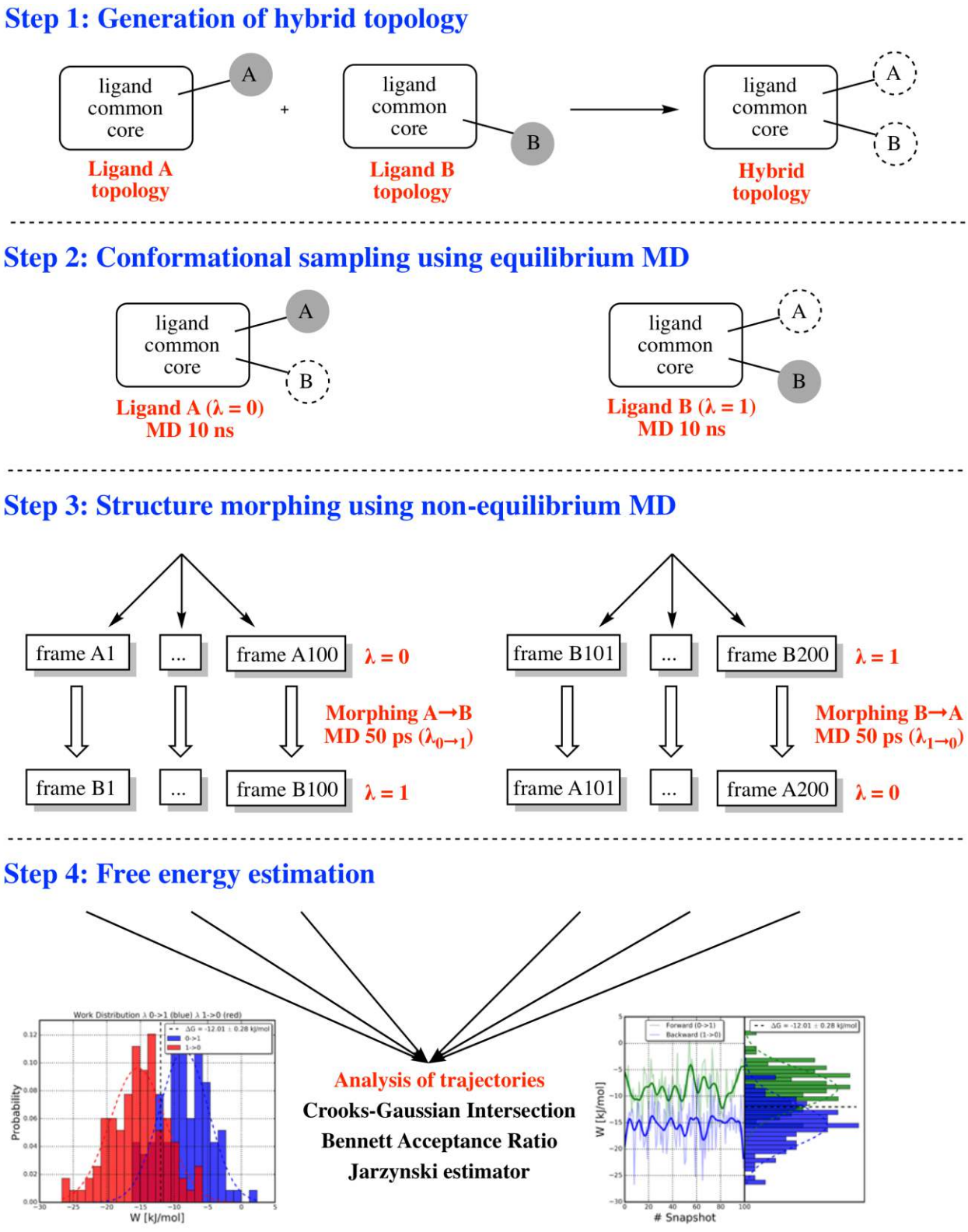


Figure 5. The main steps of the protocol used for the calculation of alchemical free energies.

Graphics. Chemical structures were depicted using CACTVS Chemoinformatics Toolkit v3.409 (Xemistry, http://www.xemistry.com/), images for protein structures were generated using PyMol 1.8.1 (Schrödinger, http://www.pymol.org/) and histograms were obtained using the R package (http://www.r-project.org).

Statistics. Statistics were computed using the online tools available at http://www.sthda.com/english/rsthda/correlation.php.

Chemoinformatics. Tanimoto similarities were computed using CACTVS Chemoinformatics Toolkit v3.409 (Xemistry, http://www.xemistry.com/).

RESULTS AND DISCUSSION

In our previous participations to the SAMPL3 (2011) [45], SAMPL4 (2013) [46], CSAR (2014) [47] and D3R Grand Challenge (2015) [48] docking and virtual screening challenges we followed an approach involving two steps. The first step consists in a preliminary analysis of information available in literature (structural, and in some cases enzymatic data), which allows the identification of the best combination of docking software and scoring function that are suited for studying the system of interest. In the second step, the combination of docking software and scoring function is used to predict the binding modes (pose prediction) and the relative affinities of ligands (scoring). As our previous studies [45-48] highlighted the importance of using enhanced genetic algorithm parameters for docking (a search efficiency of 200%), in this work we used the same parameters in order to ensure an adequate conformational sampling of docking conformations.

Preliminary analysis

We found 27 crystal structures of FXR that were available in the Protein Data Bank (PDB) [34]. These structures were organized in five distinct groups, according to their conformation and the ligand present in the binding site (see the Electronic Supplementary Material for a complete list of these structures and the exact composition of each group). A representative structure was selected from each group, based on the crystal structure resolution and the lack of missing residues. The three-dimensional structure of protein in these structures is well conserved, with the exception of two fragments (residues 258-285 and 335-358) that are very flexible (Figure 6). These five structures, together with the apo FXR structure provided by the organizers, constitute our evaluation dataset, which was used in re-docking and cross-docking calculations using the FXR ligands from all the 27 structures available and several combinations of docking software and scoring functions: Gold with the GoldScore, ChemScore, ChemPLP and ASP scoring functions, Vina and Autodock.

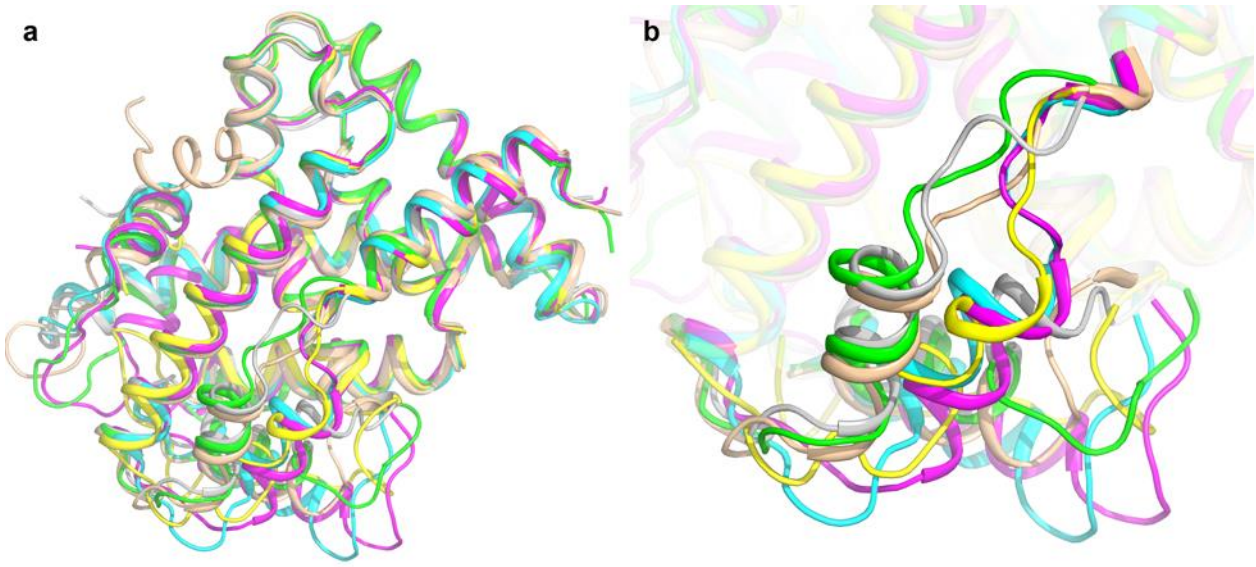


Figure 6. Representative 6 FXR PDB structures superimposed: a) general view, showing a very good global conservation of structural features; b) zoom on residues 258-285 and 335-358, highlighting the conformational flexibility of these fragments. The structures are represented as

follows: 1OSV (green), 3FLI (cyan), 3OLF (magenta), 4WVD (yellow), FXR apo (wheat), 3HC5 (gray).

RMSD values compared with the native ligands from the crystallographic structures were calculated for all docking poses. In order to evaluate the accuracy of docking and scoring, we have considered the lowest RMSD value and the RMSD value of the best ranking pose for each combination protein-ligand-(docking software)-(scoring function). AutoDock provided very poor results, with most of the docking conformations positioned outside the binding site, whereas Gold/ASP, followed by Gold/GoldScore and Vina, could reproduce rather well the native protein-ligand complexes, especially in the cross-docking calculations. As expected, the re-docking results outperformed the cross-docking results. It was also observed that the combination of a protein and a ligand belonging to the same group was more favorable than a combination of a protein and a ligand from different groups.

Pose prediction and scoring

The 102 FXR ligands from the D3R Grand Challenge 2 dataset containing 180 ligands were docked on the 6 representative FXR structures shown in Figure 6 using Gold with the ASP scoring function. For each ligand, the best-ranked docking conformation was selected and the overall ranking was submitted, as well as the coordinates for the ligands FXR_1 to FXR_36. For the ligands belonging to a group for which crystal structures were available (e.g. benzimidazoles, isoxazoles), the RMSD was calculated using the largest common fragment, and the conformations with the best RMSD were selected for a second submission. The RMSD calculation was realized using an in house developed script based on CACTVS Chemoinformatics Toolkit. The poses from the spiro and sulfonamides groups were visually

inspected using UCSF Chimera. Only the 3FLI and the APO structures provided docking poses with a carboxylate group (that is present in most spiro structures and in FXR_101 from the sulfonamides group) interacting with Arg335. This was considered as the correct orientation, since most of the crystalized ligands show the same kind of interaction. Overall, poses obtained with the structure 3OLF were selected for benzimidazoles, with 1OSV for steroids, with 3HC5 for isoxazoles and with 3FLI for all others.

The performance of submissions for pose prediction (best RMSD and RMSD of pose 1) is presented in Figure 7, showing a relatively good result that we obtained in this category compared with the other participants.

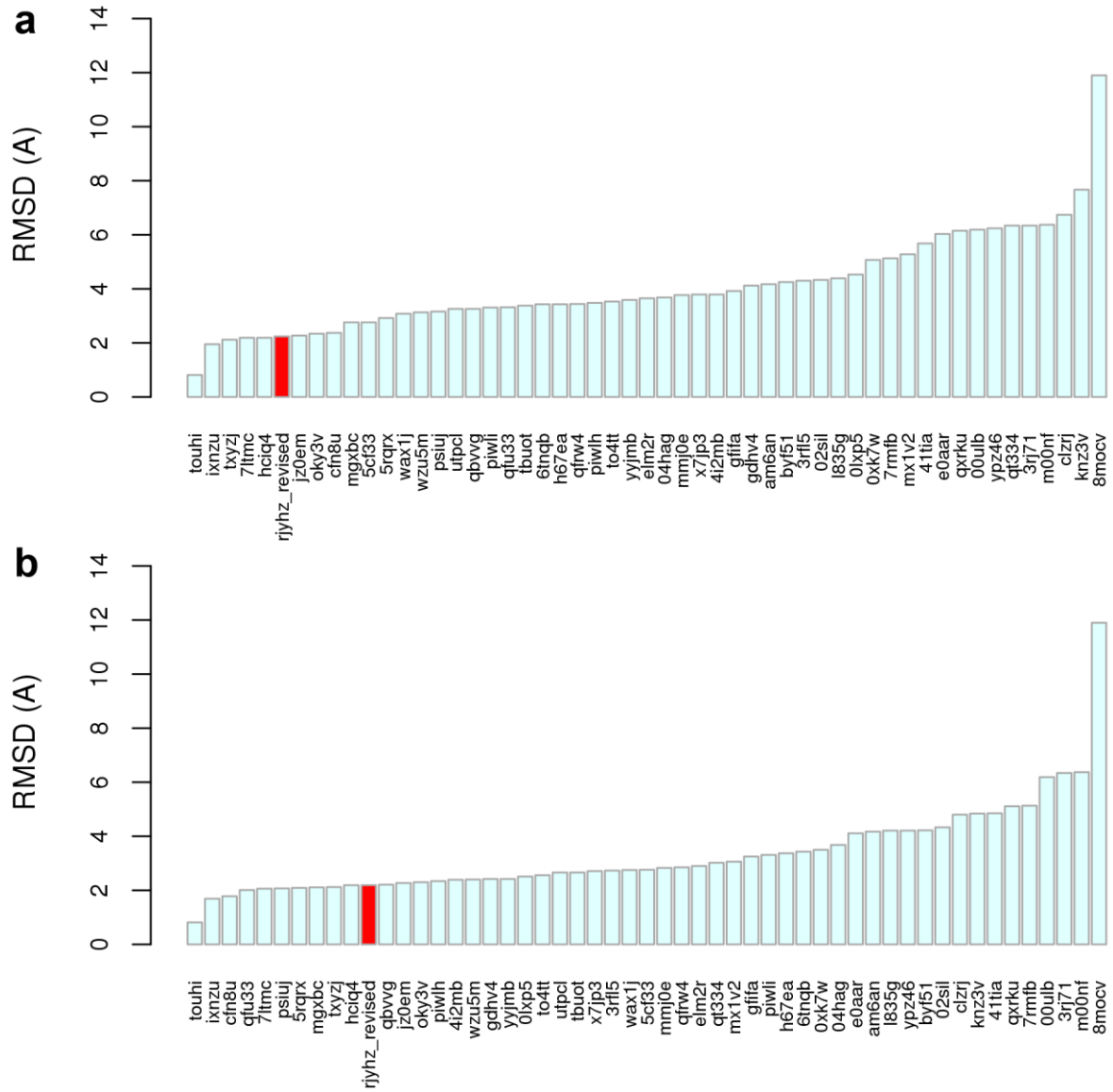


Figure 7. Performance of Phase 1 pose prediction submissions (Kendall Tau) for the FXR D3R Grand Challenge 2 dataset: best RMSD (**a**) and RMSD of pose 1 (**b**). Our submissions are colored in red (see text for details).

Our scoring results for the two submissions in Phase 1 were very modest, with Kendall Tau values of 0.13 and 0.072. Table S1 from the Supplementary Information file shows the rank of the best RMSD pose for compounds with existing reference structural data (53 compounds out of 102 compounds from the dataset). A mean value of 4.68 (out of 10 poses in each case) was

obtained for this rank, which is quite low. If we also consider that for the remaining 49 compounds with no reference structure available we have no information about the docking reliability, these data altogether might explain the incorrect scoring prediction.

The crystallographic structures of the 36 FXR complexes proposed for pose prediction were released at the end of Phase 1. A comparison of several representative docking poses and the corresponding crystallographic conformations is provided in Figure 8. We predicted well the conformation of most benzimidazoles, but the other three groups (isoxazoles, spiro compounds and sulfonamides) were more challenging, and we could predict correctly only the overall orientation of the ligand, but not the details of the interaction with the binding site. The compounds from the miscellaneous group were even more difficult, and in some cases our prediction was completely opposite compared to the crystal structure.

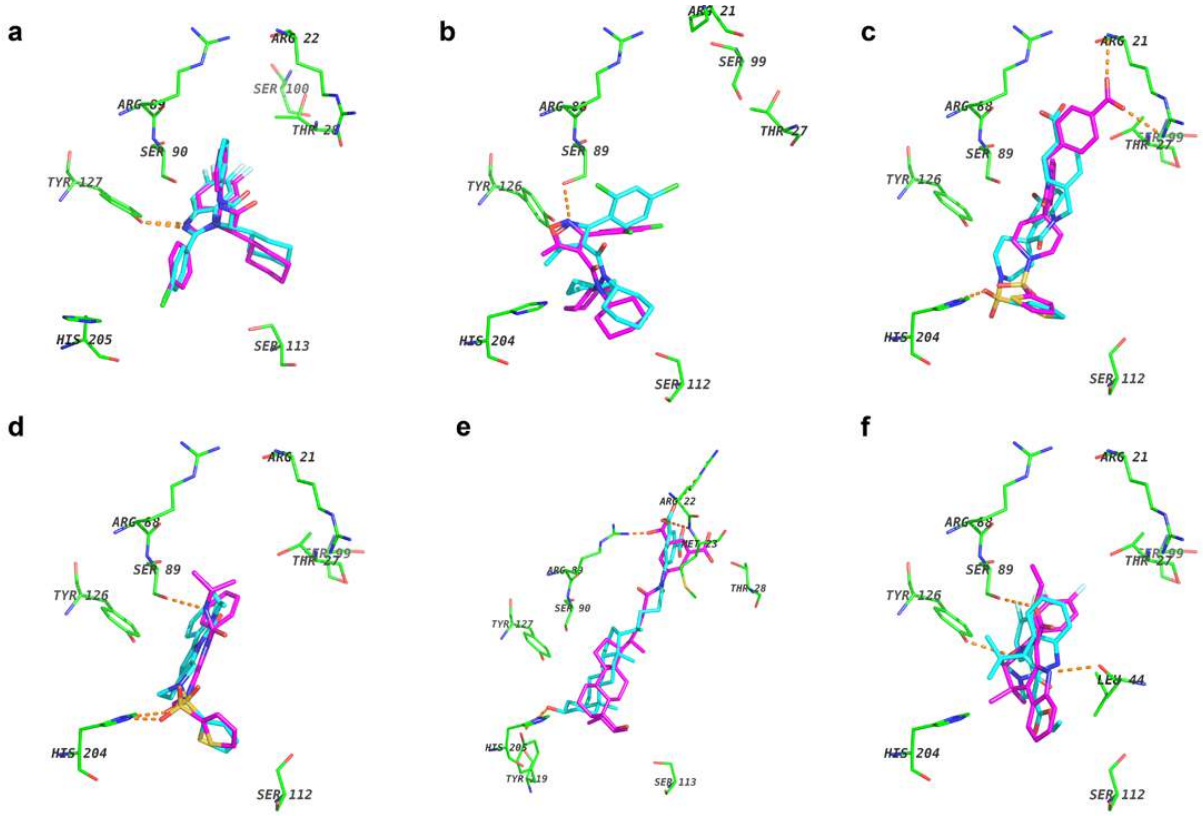


Figure 8. Comparison of our docking poses (cyan) with crystal structure conformations (magenta) for representative FXR ligands from different families: a) benzimidazoles (FXR_21/3OLF, RMSD 0.97 Å); b) isoxazoles (FXR_4/3HC5, RMSD 3.87 Å); c) spiro compounds (FXR_10/3FLI, RMSD 2.85 Å); d) sulfonamides (FXR_16/3FLI, RMSD 2.03 Å); e) miscellaneous (FXR_34/1OSV, RMSD 3.76 Å); f) miscellaneous (FXR_5/3FLI, RMSD 4.70 Å).

In Phase 2, the three-dimensional coordinates of the ligands FXR_37 to FXR_102 were built using UCSF Chimera [38], by superimposing their common backbone on the released FXR_17, FXR_10 and FXR_12 crystal structures. The protein-ligand complexes of these ligands, together with the 36 ligands from the crystal structures, were rescored using Gold with the ASP scoring function and the results were slightly improved compared with Phase 1, with a Kendall Tau value of 0.17.

Free energy calculations

The free energy calculations were carried out using a protocol adapted from the methodology implemented within the PMX software [42-44]. An important advantage of our procedure is the possibility to simulate transformations involving charge modification, which is relatively difficult or even impossible using other protocols (see the Methods section and Figures S2, S3 and S4 in the Electronic Supplementary Material for more details).

We obtained very good results for the free energy prediction of the set1 (sulfonamides), our submission **nszkx** being ranked #2. However, the corresponding submission **2ytv8** for set2 (spiro compounds) was not at all competitive, being ranked #20 (Figure 9). After the end of the D3R Grand Challenge 2 we have recomputed all data after fixing a bug in the hybrid topologies, and

also using docking poses instead of crystal structures (equivalent of Phase 1 calculations carried out retrospectively) and using AMBER/GAFF force field instead of OPLS-AA (Figure 9).

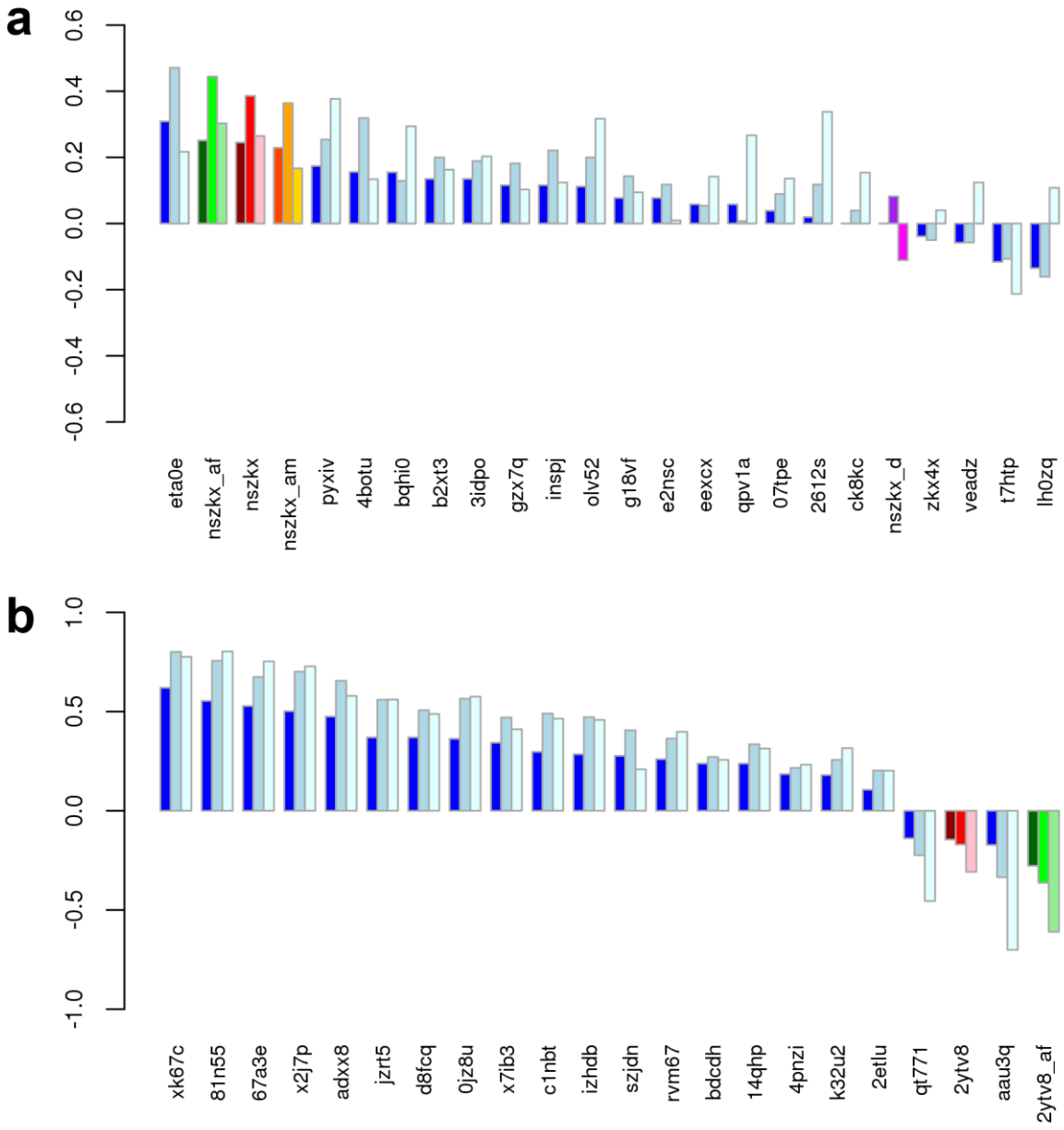


Figure 9. Performance of Phase 2 free energy submissions for set1 (**a**) and set2 (**b**). The correlation coefficients are represented as follows: Kendall tau in blue, Spearman rho in light blue and Pearson r in cyan. Our submissions are colored in dark red, red and pink, respectively (nszkx and 2ytv8). The results obtained on recomputed simulations after fixing a bug in the

hybrid topologies are represented in different shades of green (nszkx_af and 2ytv8_af). The results from simulations using docking poses instead of crystal structures (equivalent of Phase 1 calculations carried out retrospectively) are represented in magenta (nszkx_d). The results from simulations using AMBER/GAFF force field instead of OPLS-AA are represented in orange (nszkx_am).

Tables 1 and 2 contain the detailed computed values for set1 and set2, respectively, together with the corresponding statistics (Kendall's rank correlation tau, Spearman's rank correlation rho and Pearson's product-moment correlation r).

For set1, similar results were obtained with OPLS-AA before and after correction, as well as with AMBER/GAFF force field. However, when docking poses were used as initial coordinates (which represents Phase 1 calculations carried out retrospectively), no correlation was obtained. This corroborates with the pose prediction results, showing that no good quality predictions can be obtained from inaccurate docking poses.

Table 1. Free energies computed for set1. FXR_91 was used as reference compound.

		OPLS-AA before correction (nszkx)		OPLS-AA after correction (nszkx_af)	
	Experimental IC50 (μM)	ΔΔG (kJ/mol)	ΔΔG error (kJ/mol)	ΔΔG (kJ/mol)	ΔΔG error (kJ/mol)
FXR_17	0.79	-9.32	1.16	-18.35	1.12
FXR_45	28.85	-3.14	2.33	-21.21	2.15
FXR_46	62.37	-3.81	21.82	-12.98	1.26
FXR_47	20.96	NA	NA	-18.61	2.41
FXR_48	100.00	NA	NA	5.36	1.75
FXR_49	100.00	-3.73	1.48	-9.47	1.02
FXR_91	29.63	0.00	0.00	0.00	0.00
FXR_93	46.66	-8.98	1.26	-4.04	0.39
FXR_95	32.17	-8.39	1.45	-6.63	1.21
FXR_96	58.86	-19.12	1.62	-21.36	2.15
FXR_98	13.14	-21.07	1.42	-12.95	0.97
FXR_99	100.00	-6.34	1.59	-11.07	0.56
FXR_100	19.14	-15.86	2.49	-25.81	3.05
FXR_101	27.64	-36.24	3.38	-27.94	3.69
FXR_102	29.23	-0.15	2.85	13.45	2.17
		Correlation coefficient	p-value	Correlation coefficient	p-value

Kendall's rank correlation tau	0.2452	0.2455	0.2512	0.1962
Spearman's rank correlation rho	0.3851	0.1937	0.4444	0.0970
Pearson's product-moment correlation r	0.2648	0.3819	0.3029	0.2726

For set2, the results were quite disappointing, with negative correlations with OPLS-AA before and after the correction. Apparent better correlations were obtained with OPLS-AA using the docking poses and with AMBER/GAFF, but they are not representative since they were computed only for 12 and 6 values, all of them belonging to the FXR_12 subset (see Table 2 for the compounds belonging to the FXR_10 and FXR_12 subsets), so we decided not to represent them in Figure 9.

Table 2. Free energies computed for set2. The FXR_10 subset contains the five compounds marked with a star, and the FXR_12 subset contains the remaining compounds. FXR_10 and FXR_12 were used as reference compounds for each subset, then all free energies from subset FXR_10 were translated relative to FXR_12.

		OPLS-AA before correction (2ytv8)		OPLS-AA after correction (2ytv8_af)	
	Experimental IC50 (μM)	ΔΔG (kJ/mol)	ΔΔG error (kJ/mol)	ΔΔG (kJ/mol)	ΔΔG error (kJ/mol)
FXR_10*	5.64	-10.69	0.81	-4.85	0.48
FXR_12	0.06	0.00	0.00	0.00	0.00
FXR_38*	100.00	-27.49	–	-18.86	–
FXR_41	100.00	-15.19	4.15	-3.24	4.27
FXR_73*	11.22	10.77	–	-14.83	–
FXR_74	0.66	-3.16	–	-2.57	0.16
FXR_75*	100.00	-71.30	–	-32.41	–
FXR_76	41.83	-0.01	0.62	-1.78	66.29
FXR_77	0.25	-2.35	0.35	-2.47	0.23
FXR_78	0.03	-15.94	1.08	-5.89	19.75
FXR_79*	4.15	2.39	–	14.84	–
FXR_81	2.69	-11.09	0.73	-10.97	18.40
FXR_82	0.18	7.14	0.85	2.60	0.58
FXR_83	0.33	-1.75	0.31	5.02	0.77
FXR_84	4.54	-5.87	0.66	4.09	17.88
FXR_85	0.30	-9.67	0.40	-5.10	90.68
FXR_88	0.54	-3.73	0.40	0.81	0.33
FXR_89	0.74	-10.31	75.22	-4.15	75.34
		Correlation coefficient	p-value	Correlation coefficient	p-value
Kendall's rank correlation tau		-0.2772	0.1107	-0.2772	0.1107
Spearman's rank correlation rho		-0.3320	0.1784	-0.3630	0.1387
Pearson's product-moment correlation r		-0.6900	0.0015	-0.6105	0.0071

We tried to find a rational explanation for the discrepancy of the results obtained for set1 and set2, using the same protocol. Among the possible hypotheses, we can mention: i) the intrinsic greater structural diversity in set2 compared with set1; ii) incorrect force field parameters and iii) insufficient conformation sampling of ligands.

To validate the first hypothesis, we computed the Tanimoto similarity matrix for set1 and set2 (see Tables S2 and S3 in Electronic Supplementary Material). The global mean values of Tanimoto similarity for the two datasets are very close, 0.88 and 0.86, respectively, suggesting a similar degree of diversity. However, a visual inspection of the two datasets shows that set1 is quite homogeneous, with variations on the substitution pattern of a single phenyl ring. On the other hand, compounds from set2 contain variations on two fragments: one can be a diversely substituted phenyl ring, and the other can be either a thienyl ring or a diversely substituted phenyl ring. According to the presence or not of the thienyl ring, set2 can be divided into two subsets, which have FXR_10 and FXR_12 as representative compounds. We computed the statistics separately on these two subsets and the results are presented in Table 3. Compared with the whole set2, only a small improvement in the correlation with experimental data is observed for the FXR_12 subset. However, for the FXR_10 subset we observe almost a perfect anticorrelation with the experimental data. Overall, this analysis shows that the differences between the structural diversity of set1 and set2 are too small to be discriminated by descriptors such as Tanimoto similarity, but the set2 is more diverse and can be divided in two subsets. One of these subsets, containing a thienyl substituent, has a major negative impact in the prediction of free energies for set2.

Table 3. Statistics computed for the subsets FXR_10 and FXR_12 of set2. See Table 2 and text for the list of compounds included in each subset.

	OPLS-AA before correction		OPLS-AA after correction	
FXR_10 subset	Correlation coefficient	p-value	Correlation coefficient	p-value
Kendall's rank correlation tau	-0.5270	0.2065	-0.9487	0.0230
Spearman's rank correlation rho	-0.6669	0.2189	-0.9747	0.0048
Pearson's product-moment correlation r	-0.8377	0.0766	-0.7737	0.1247
FXR_12 subset	Correlation coefficient	p-value	Correlation coefficient	p-value
Kendall's rank correlation tau	-0.2564	0.2519	-0.0513	0.8577
Spearman's rank correlation rho	-0.2692	0.3733	-0.0330	0.9206
Pearson's product-moment correlation r	-0.3261	0.2769	-0.0902	0.7696

To evaluate the pertinence of the second hypothesis, we analyzed the conformational distribution of the ligands FXR_17, FXR_10 and FXR_12, as representative structures for set1 and the two subsets of set2, in two force fields, OPLS-AA and AMBER/GAFF. In each case, we extracted and superimposed all the 501 conformations from the 10 ns molecular dynamics simulation of the ligand alone in water. The result is presented in Figure S5 (Electronic Supplementary Material).

For compound FXR_17, we observe 4 main differences between the distributions OPLS-AA (**a**) and AMBER/GAFF (**b**): i) the phenyl ring is mostly parallel with the bicyclic system in **a**, and perpendicular in **b**; ii) the amide group is mostly perpendicular with the bicyclic system in **a**, and parallel in **b**; iii) the distribution of the thienyl ring around the dihedral C-N-S-C is restricted to a very narrow window in **a**, whereas in **b** there are two larger windows in opposite positions, showing in the latter case an unrestricted exchange between these two positions; iv) in **a** the thienyl ring shows equivalent populations of both faces, whereas in **b** the rotation around the dihedral N-S-C-S is very much restricted. However, as the predictions of set1 using either OPLS-AA or AMBER/GAFF are very similar (see Figure 9a and Table 1), these differences should not have a major contribution or, more probably, should cancel mutually.

For compound FXR_10, we observe in **a** a restricted rotation around the C-N-S-C dihedral and, in opposition with FXR_17, an impossible rotation around the N-S-C-S dihedral, probably because of the close proximity of the amide oxygen. In the case of **b**, we observe a free rotation around the C-N-S-C dihedral and a restricted rotation around the N-S-C-S dihedral, similar with FXR_17.

Finally, we observe for FXR_12 a restricted rotation around the C-N-S-CA dihedral in **a** and a free rotation around the N-S-CA-CA dihedral (the chlorine substituent is positioned equally on both sides), whereas in **b** the rotation around the C-N-S-CA dihedral is relatively free in the conditions of simulations, but the rotation around the N-S-CA-CA dihedral is almost completely restricted.

These results suggest the possible existence of two non-optimal dihedrals associated with the sulfonamide group, similarly with a recent report regarding the incorrect conformational sampling of linezolid [49]. For set1, their influence might be compensated by two other dihedrals, which is not the case for set2. Additionally, in the FXR_12 subset there are two atropoisomers that can contribute to the overall binding energy, whereas in our calculations we have considered only one, the most favorable.

The third hypothesis, insufficient conformation sampling of ligands, is not very probable given the length of our molecular dynamics simulations and the relative rigidity of the ligands. If the conformation space is not sampled correctly, this should be more due to inadequate force field parameters than to insufficient length of simulations. Along the same lines, in a few specific cases, the standard deviation of our predictions is unusually high (see the "ΔΔG error" columns in Tables 1 and 2), especially for set2.

CONCLUSIONS

We used in this work a protocol in two steps, involving an initial analysis of the available structural data from the PDB, which allows the selection of the most appropriate combination of docking software and scoring function. Subsequent docking calculations showed that the pose prediction can be carried out with a certain precision, but this is dependent on the specific nature of the ligands. The correct ranking of docking poses is still a problem and cannot be successful in the absence of good pose predictions. Our free energy calculations on two different subsets provided contrasted results, which might have the origin in non-optimal force field parameters associated with the sulfonamide chemical moiety.

Electronic Supplementary Material. The Electronic Supplementary Material contains the chemical structures of the entire FXR dataset, the rank of best RMSD poses, conformational distribution of representative ligands, Tanimoto similarity matrices and a schematic description of the systems and thermodynamic cycles used for free energy calculations.

Acknowledgments. We thank Prof. Bert de Groot for helpful discussions. The comments and suggestions of the reviewers are also acknowledged, as they greatly contributed to improve the manuscript. This work was supported by the Laboratory of Excellence in Research on Medication and Innovative Therapeutics (LERMIT) [grant number ANR-10-LABX-33], by the JPIAMR transnational project DesInMBL [grant number ANR-14-JAMR-0002] and by the Région Ile-de-France (DIM Malinf).

REFERENCES

1. Bishop-Bailey D (2004) FXR as a novel therapeutic target for vascular disease. Drug news & perspectives 17 (8):499-504

2. Claudel T, Sturm E, Kuipers F, Staels B (2004) The farnesoid X receptor: a novel drug target? Expert opinion on investigational drugs 13 (9):1135-1148. doi:10.1517/13543784.13.9.1135

3. Pellicciari R, Costantino G, Fiorucci S (2005) Farnesoid X receptor: from structure to potential clinical applications. J Med Chem 48 (17):5383-5403. doi:10.1021/jm0582221

4. Westin S, Heyman RA, Martin R (2005) FXR, a therapeutic target for bile acid and lipid disorders. Mini reviews in medicinal chemistry 5 (8):719-727

5. Cai SY, Boyer JL (2006) FXR: a target for cholestatic syndromes? Expert opinion on therapeutic targets 10 (3):409-421. doi:10.1517/14728222.10.3.409

6. Lee FY, Lee H, Hubbert ML, Edwards PA, Zhang Y (2006) FXR, a multipurpose nuclear receptor. Trends in biochemical sciences 31 (10):572-580. doi:10.1016/j.tibs.2006.08.002

7. Cariou B, Staels B (2007) FXR: a promising target for the metabolic syndrome? Trends Pharmacol Sci 28 (5):236-243. doi:10.1016/j.tips.2007.03.002

8. Wang YD, Chen WD, Huang W (2008) FXR, a target for different diseases. Histology and histopathology 23 (5):621-627. doi:10.14670/hh-23.621

9. Zimber A, Gespach C (2008) Bile acids and derivatives, their nuclear receptors FXR, PXR and ligands: role in health and disease and their therapeutic potential. Anti-cancer agents in medicinal chemistry 8 (5):540-563

10. Crawley ML (2010) Farnesoid X receptor modulators: a patent review. Expert Opin Ther Pat 20 (8):1047-1057. doi:10.1517/13543776.2010.496777

11. Fiorucci S, Mencarelli A, Distrutti E, Palladino G, Cipriani S (2010) Targetting farnesoid-X-receptor: from medicinal chemistry to disease treatment. Current medicinal chemistry 17 (2):139-159

12. Mencarelli A, Fiorucci S (2010) FXR an emerging therapeutic target for the treatment of atherosclerosis. Journal of cellular and molecular medicine 14 (1-2):79-92. doi:10.1111/j.1582-4934.2009.00997.x

13. Teodoro JS, Rolo AP, Palmeira CM (2011) Hepatic FXR: key regulator of whole-body energy metabolism. Trends in endocrinology and metabolism: TEM 22 (11):458-466. doi:10.1016/j.tem.2011.07.002

14. Adorini L, Pruzanski M, Shapiro D (2012) Farnesoid X receptor targeting to treat nonalcoholic steatohepatitis. Drug Discov Today 17 (17-18):988-997. doi:10.1016/j.drudis.2012.05.012

15. Fiorucci S, Mencarelli A, Distrutti E, Zampella A (2012) Farnesoid X receptor: from medicinal chemistry to clinical applications. Future medicinal chemistry 4 (7):877-891. doi:10.4155/fmc.12.41

16. Fiorucci S, Zampella A, Distrutti E (2012) Development of FXR, PXR and CAR agonists and antagonists for treatment of liver disorders. Current topics in medicinal chemistry 12 (6):605-624

17. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and

highly parallel open source molecular simulation toolkit. Bioinformatics (Oxford, England) 29 (7):845-854. doi:10.1093/bioinformatics/btt055

18. Carotti A, Marinozzi M, Custodi C, Cerra B, Pellicciari R, Gioiello A, Macchiarulo A (2014) Beyond bile acids: targeting Farnesoid X Receptor (FXR) with natural and synthetic ligands. Current topics in medicinal chemistry 14 (19):2129-2142

19. Fiorucci S, Distrutti E, Ricci P, Giuliano V, Donini A, Baldelli F (2014) Targeting FXR in cholestasis: hype or hope. Expert opinion on therapeutic targets 18 (12):1449-1459. doi:10.1517/14728222.2014.956087

20. Gege C, Kinzel O, Steeneck C, Schulz A, Kremoser C (2014) Knocking on FXR's door: the "hammerhead"-structure series of FXR agonists - amphiphilic isoxazoles with potent in vitro and in vivo activities. Current topics in medicinal chemistry 14 (19):2143-2158

21. Huang H, Xu Y, Zhu J, Li J (2014) Recent advances in non-steroidal FXR antagonists development for therapeutic applications. Current topics in medicinal chemistry 14 (19):2175-2187

22. Lamers C, Schubert-Zsilavecz M, Merk D (2014) Medicinal chemistry and pharmacological effects of Farnesoid X Receptor (FXR) antagonists. Current topics in medicinal chemistry 14 (19):2188-2205

23. Ali AH, Carey EJ, Lindor KD (2015) Recent advances in the development of farnesoid X receptor agonists. Annals of translational medicine 3 (1):5. doi:10.3978/j.issn.2305-5839.2014.12.06

24. Carr RM, Reid AE (2015) FXR agonists as therapeutic agents for non-alcoholic fatty liver disease. Current atherosclerosis reports 17 (4):500. doi:10.1007/s11883-015-0500-2

25. Koutsounas I, Theocharis S, Delladetsima I, Patsouris E, Giaginis C (2015) Farnesoid x receptor in human metabolism and disease: the interplay between gene polymorphisms, clinical phenotypes and disease susceptibility. Expert opinion on drug metabolism & toxicology 11 (4):523-532. doi:10.1517/17425255.2014.999664

26. Sanyal AJ (2015) Use of farnesoid X receptor agonists to treat nonalcoholic fatty liver disease. Digestive diseases (Basel, Switzerland) 33 (3):426-432. doi:10.1159/000371698

27. Sepe V, Distrutti E, Fiorucci S, Zampella A (2015) Farnesoid X receptor modulators (2011 - 2014): a patent review. Expert Opin Ther Pat 25 (8):885-896. doi:10.1517/13543776.2015.1045413

28. Sepe V, Distrutti E, Limongelli V, Fiorucci S, Zampella A (2015) Steroidal scaffolds as FXR and GPBAR1 ligands: from chemistry to therapeutical application. Future medicinal chemistry 7 (9):1109-1135. doi:10.4155/fmc.15.54

29. Alawad AS, Levy C (2016) FXR Agonists: From Bench to Bedside, a Guide for Clinicians. Digestive diseases and sciences 61 (12):3395-3404. doi:10.1007/s10620-016-4334-8

30. De Magalhaes Filho CD, Downes M, Evans RM (2017) Farnesoid X Receptor an Emerging Target to Combat Obesity. Digestive diseases (Basel, Switzerland) 35 (3):185-190. doi:10.1159/000450909

31. Feng S, Yang M, Zhang Z, Wang Z, Hong D, Richter H, Benson GM, Bleicher K, Grether U, Martin RE, Plancher JM, Kuhn B, Rudolph MG, Chen L (2009) Identification of an N-oxide

pyridine GW4064 analog as a potent FXR agonist. Bioorganic & medicinal chemistry letters 19 (9):2595-2598. doi:10.1016/j.bmcl.2009.03.008

32. Richter HG, Benson GM, Bleicher KH, Blum D, Chaput E, Clemann N, Feng S, Gardes C, Grether U, Hartman P, Kuhn B, Martin RE, Plancher JM, Rudolph MG, Schuler F, Taylor S (2011) Optimization of a novel class of benzimidazole-based farnesoid X receptor (FXR) agonists to improve physicochemical and ADME properties. Bioorganic & medicinal chemistry letters 21 (4):1134-1140. doi:10.1016/j.bmcl.2010.12.123

33. Richter HG, Benson GM, Blum D, Chaput E, Feng S, Gardes C, Grether U, Hartman P, Kuhn B, Martin RE, Plancher JM, Rudolph MG, Schuler F, Taylor S, Bleicher KH (2011) Discovery of novel and orally active FXR agonists for the potential treatment of dyslipidemia & diabetes. Bioorganic & medicinal chemistry letters 21 (1):191-194. doi:10.1016/j.bmcl.2010.11.039

34. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28 (1):235-242

35. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234 (3):779-815. doi:10.1006/jmbi.1993.1626

36. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein–ligand docking using GOLD. Proteins Struct Funct Bioinf 52 (4):609-623. doi:10.1002/prot.10465

37. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30 (16):2785-2791. doi:10.1002/jcc.21256

38. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera -- a visualization system for exploratory research and analysis. J Comput Chem 25 (13):1605-1612. doi:10.1002/jcc.20084

39. Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31 (2):455-461. doi:10.1002/jcc.21334

40. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B 105 (28):6474-6487. doi:10.1021/jp003919d

41. Robertson MJ, Tirado-Rives J, Jorgensen WL (2015) Improved Peptide and Protein Torsional Energetics with the OPLSAA Force Field. J Chem Theory Comput 11 (7):3499-3509. doi:10.1021/acs.jctc.5b00356

42. Gapsys V, Michielssens S, Seeliger D, de Groot BL (2015) pmx: Automated protein structure and topology generation for alchemical perturbations. J Comput Chem 36 (5):348-354. doi:10.1002/jcc.23804

43. Gapsys V, Michielssens S, Peters JH, de Groot BL, Leonov H (2015) Calculation of binding free energies. Methods in molecular biology (Clifton, NJ) 1215:173-209. doi:10.1007/978-1-4939-1465-4_9

44. Gapsys V, Michielssens S, Seeliger D, de Groot BL (2016) Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. Angewandte Chemie (International ed in English) 55 (26):7364-7368. doi:10.1002/anie.201510054

45. Surpateanu G, Iorga BI (2012) Evaluation of docking performance in a blinded virtual screening of fragment-like trypsin inhibitors. J Comput Aided Mol Des 26 (5):595-601. doi:10.1007/s10822-011-9526-x

46. Colas C, Iorga BI (2014) Virtual screening of the SAMPL4 blinded HIV integrase inhibitors dataset. J Comput Aided Mol Des 28 (4):455-462. doi:10.1007/s10822-014-9707-5

47. Martiny VY, Martz F, Selwa E, Iorga BI (2016) Blind pose prediction, scoring, and affinity ranking of the CSAR 2014 dataset. Journal of chemical information and modeling 56 (6):996-1003. doi:10.1021/acs.jcim.5b00337

48. Selwa E, Martiny VY, Iorga BI (2016) Molecular docking performance evaluated on the D3R Grand Challenge 2015 drug-like ligand datasets. J Comput Aided Mol Des 30 (9):829-839. doi:10.1007/s10822-016-9983-3

49. Grunenberg J, Licari G (2016) Effective in silico prediction of new oxazolidinone antibiotics: Force field simulations of the antibiotic-ribosome complex supervised by experiment and electronic structure methods. Beilstein J Org Chem 12:415-428. doi:10.3762/bjoc.12.45

Electronic Supplementary Material

Blinded evaluation of farnesoid X receptor (FXR) ligands binding using molecular docking and free energy calculations

Edithe Selwa,[1,‡] Eddy Elisee,[1,‡] Agustin Zavala,[1,‡] Bogdan I. Iorga[1,]*

[1] Institut de Chimie des Substances Naturelles, CNRS UPR 2301, LabEx LERMIT, 91198 Gif-sur-Yvette, France

Corresponding Author

[*] Phone: +33 1 6982 3094; Fax: +33 1 6907 7247; Email: bogdan.iorga@cnrs.fr (B.I.I.).

Author Contributions

[‡] These authors contributed equally.

Table of contents

Protein Data Bank (PDB) structures available

27 structures were available in the PDB for FXR at the moment when the D3R Grand Challenge 2 took place. They were organized in 5 distinct groups, according to the type of ligand and the binding site conformation. The representative structure for each group (based on the crystal structure resolution and lack of missing residues) is colored in red.

Group 1 (isoxazoles): 3dct, 3dcu, 3gd2, **3hc5**, 3hc6, 3p88, 3p89, 3rut, 3ruu, 3rvf

Group 2 (benzimidazoles): 3okh, 3oki, **3olf**, 3omk, 3omm, 3oof, 3ook

Group 3 (FXR_5-like): 3l1b, **3fli**

Group 4 (steroid, FXR_34-like): 3bej, **1osv**, 1ot7, 4qe6

Group 5 (miscellaneous): 1osh, 4oiv, 4qe8, **4wvd**

FXR ligands regroupment according to their chemical structure

Isoxazoles: FXR_4, FXR_23, FXR_33, FXR_65

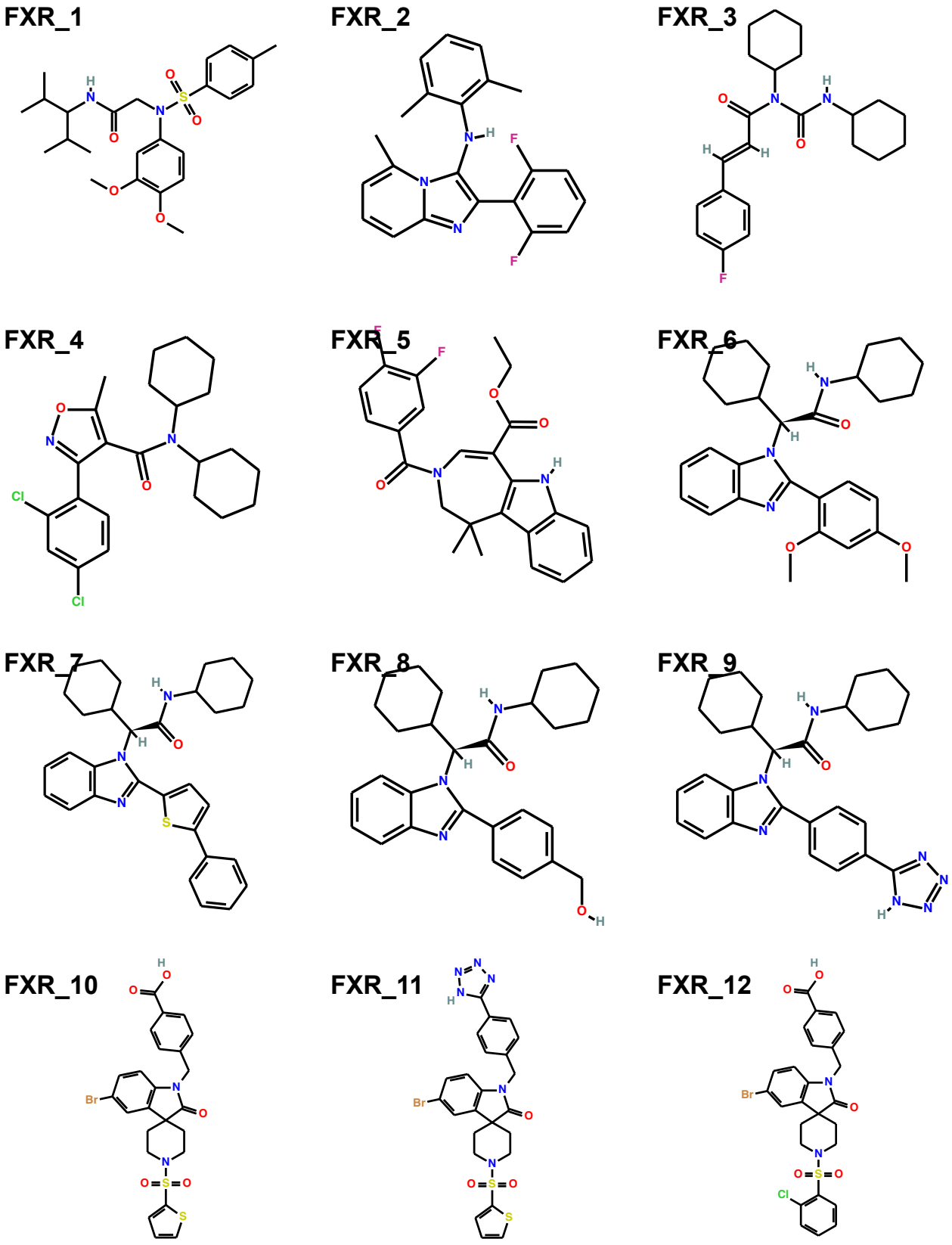
Benzimidazoles: FXR_6, FXR_7, FXR_8, FXR_9, FXR_13, FXR_14, FXR_19, FXR_20, FXR_21, FXR_22, FXR_24, FXR_25, FXR_26, FXR_27, FXR_28, FXR_29, FXR_30, FXR_31, FXR_32, FXR_35, FXR_36, FXR_37, FXR_39, FXR_40, FXR_42, FXR_50, FXR_51, FXR_52, FXR_53, FXR_54, FXR_55, FXR_56, FXR_57, FXR_58, FXR_59, FXR_60, FXR_61, FXR_62, FXR_63, FXR_64, FXR_66, FXR_67, FXR_68, FXR_69, FXR_70, FXR_71, FXR_72

Spiro compounds: FXR_10, FXR_11, FXR_12, FXR_38, FXR_41, FXR_73, FXR_74, FXR_75, FXR_76, FXR_77, FXR_78, FXR_79, FXR_80, FXR_81, FXR_82, FXR_83, FXR_84, FXR_85, FXR_86, FXR_87, FXR_88, FXR_89

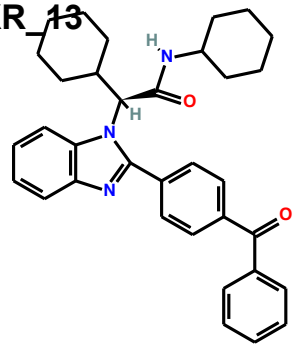
Sulfonamides: FXR_15, FXR_16, FXR_17, FXR_43, FXR_44, FXR_45, FXR_46, FXR_47, FXR_48, FXR_49, FXR_90, FXR_91, FXR_92, FXR_93, FXR_94, FXR_95, FXR_96, FXR_97, FXR_98, FXR_99, FXR_100, FXR_101, FXR_102

Miscellaneous: FXR_1, FXR_2, FXR_3, FXR_5, FXR_18, FXR_34

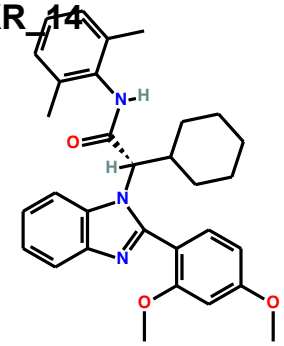
Figure S1. Chemical structures of the entire FXR dataset, containing 102 ligands used for ranking prediction.



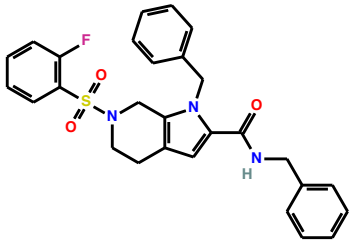
FXR_13



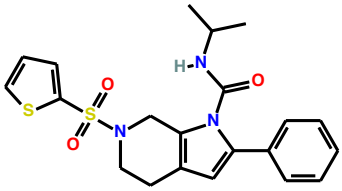
FXR_14



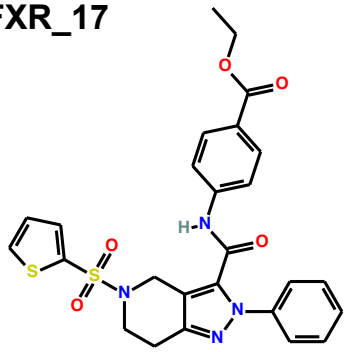
FXR_15



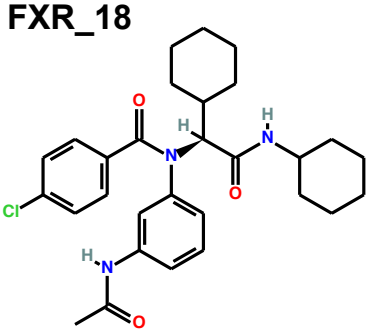
FXR_16



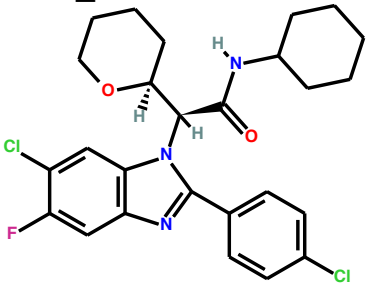
FXR_17



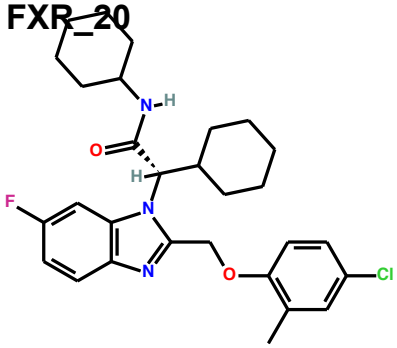
FXR_18



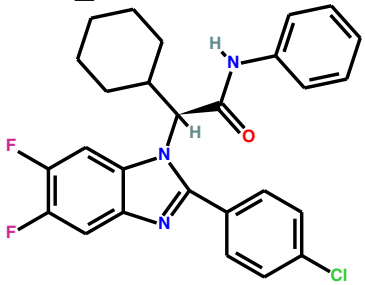
FXR_19



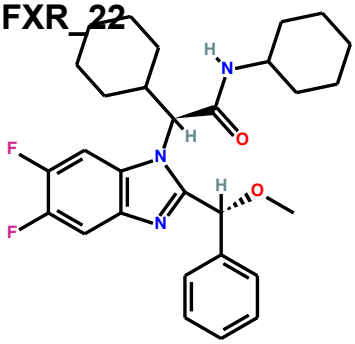
FXR_20



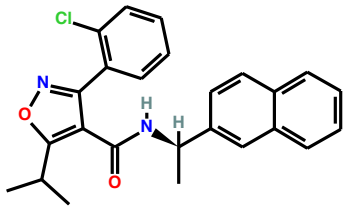
FXR_21



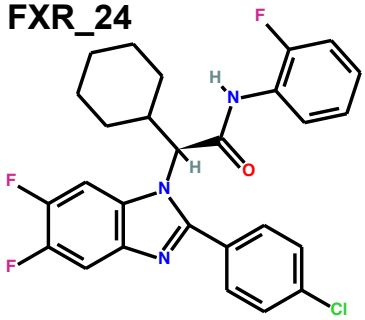
FXR_22



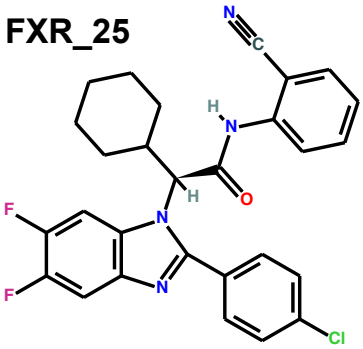
FXR_23



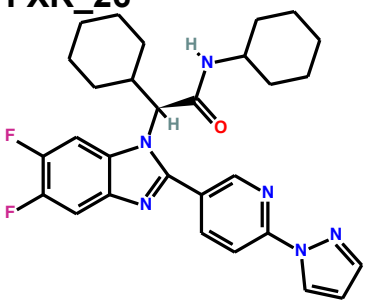
FXR_24



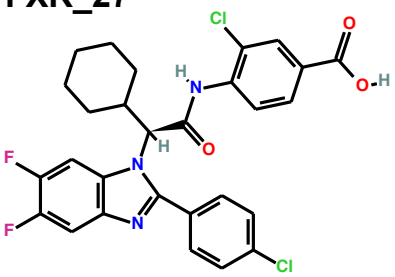
FXR_25



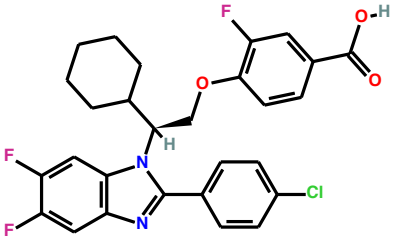
FXR_26



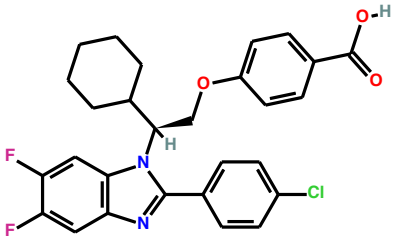
FXR_27



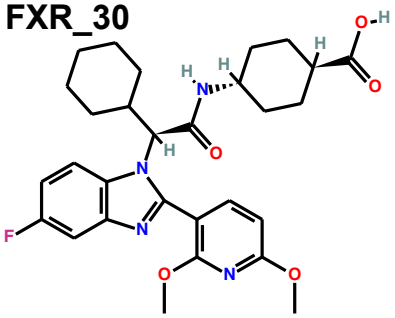
FXR_28



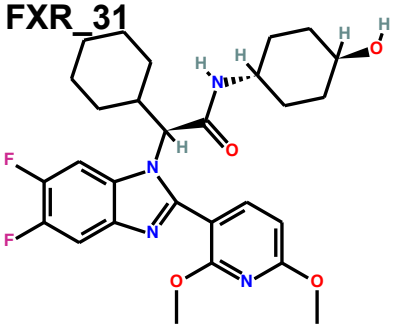
FXR_29



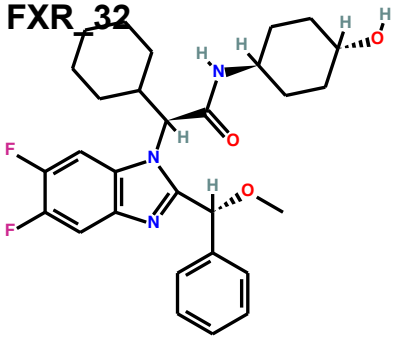
FXR_30



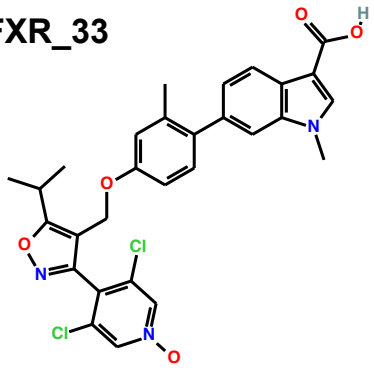
FXR_31



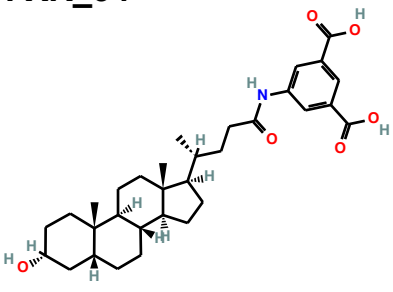
FXR_32



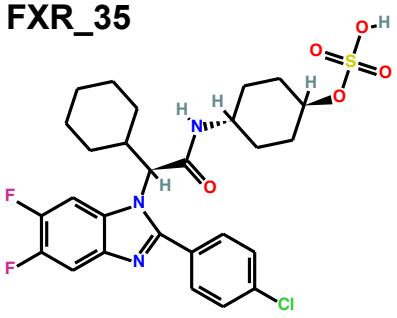
FXR_33



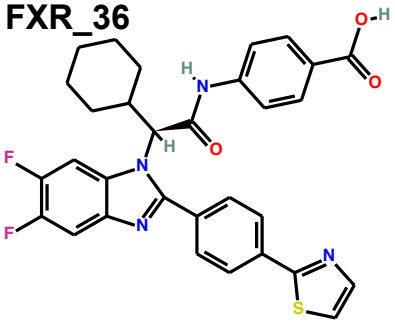
FXR_34



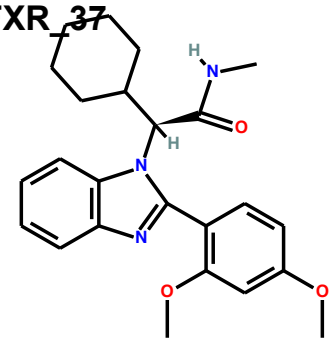
FXR_35



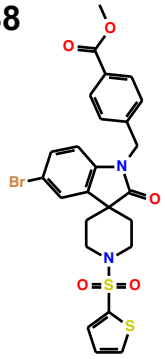
FXR_36



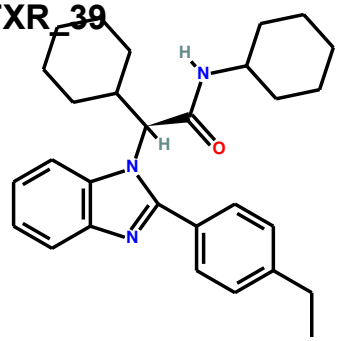
FXR_37



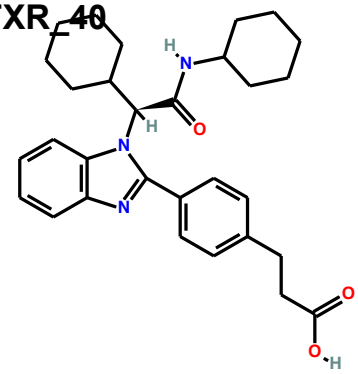
FXR_38



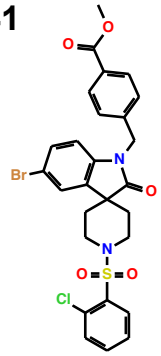
FXR_39



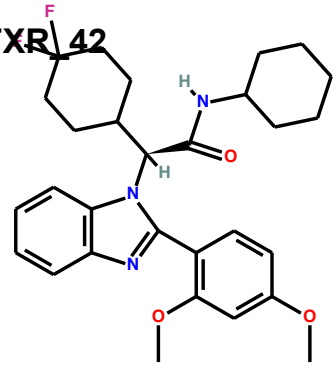
FXR_40



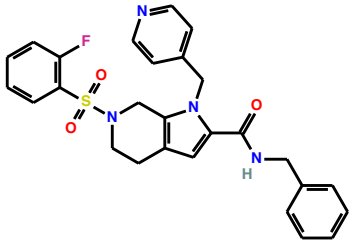
FXR_41



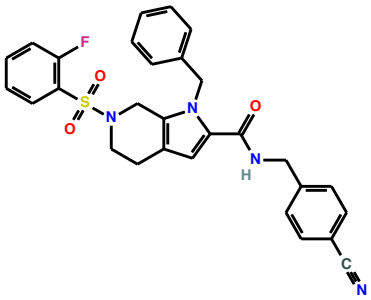
FXR_42



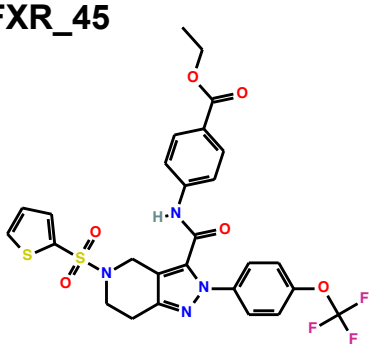
FXR_43



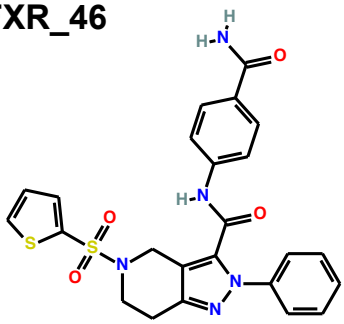
FXR_44



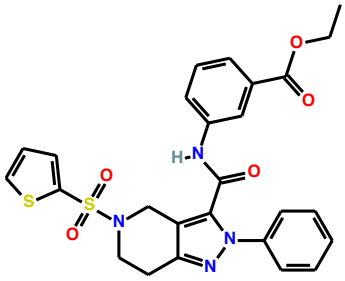
FXR_45



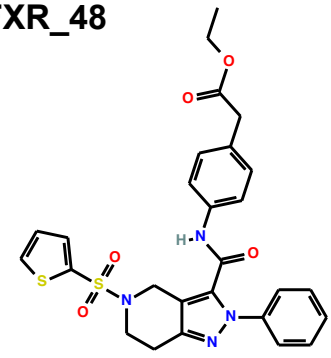
FXR_46



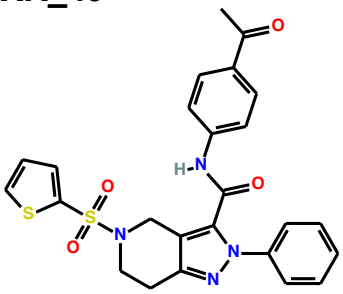
FXR_47



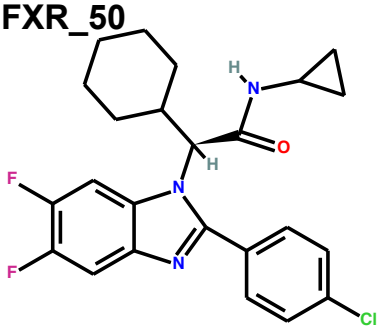
FXR_48



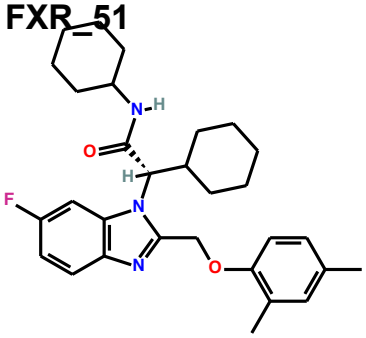
FXR_49



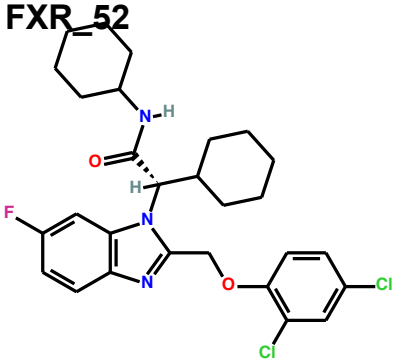
FXR_50



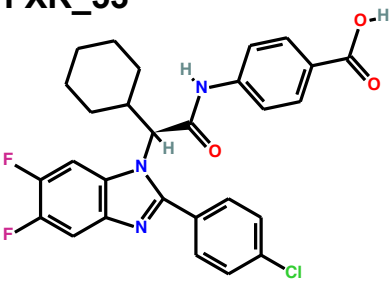
FXR_51



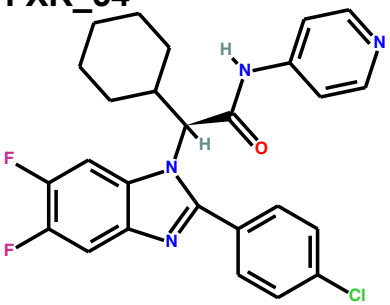
FXR_52



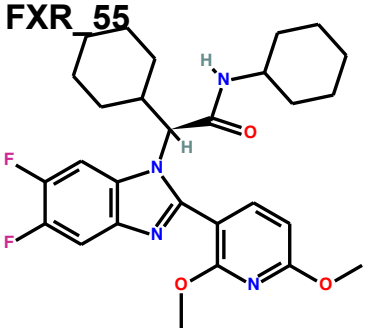
FXR_53



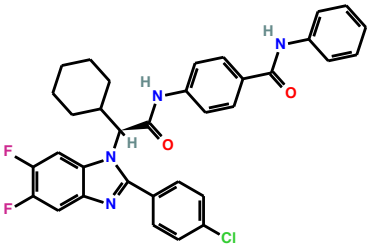
FXR_54



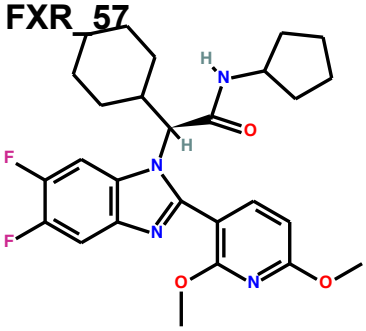
FXR_55



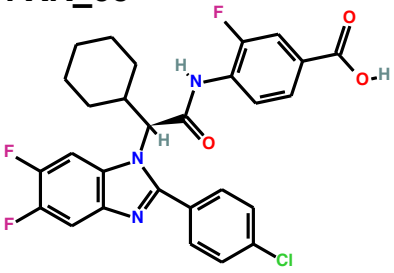
FXR_56



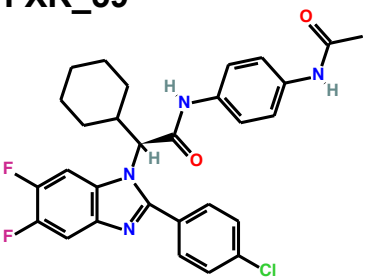
FXR_57



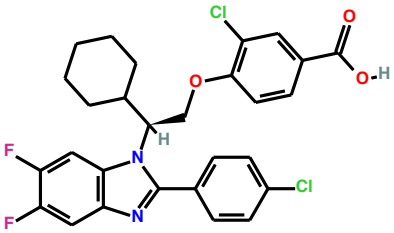
FXR_58



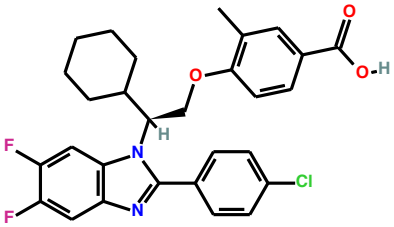
FXR_59



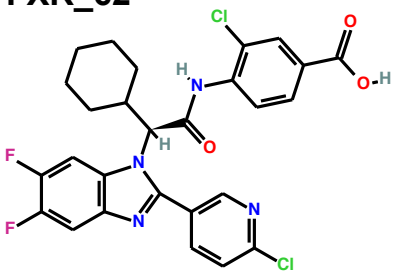
FXR_60



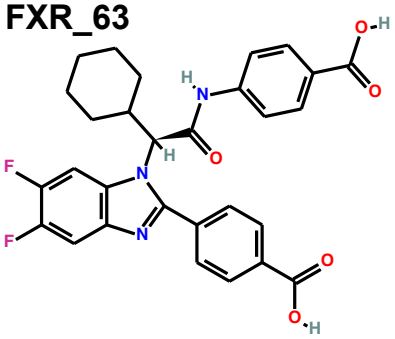
FXR_61



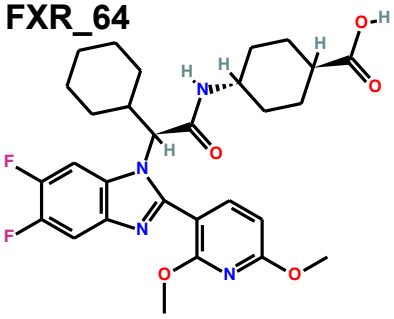
FXR_62



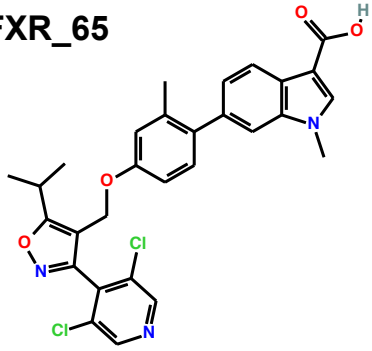
FXR_63



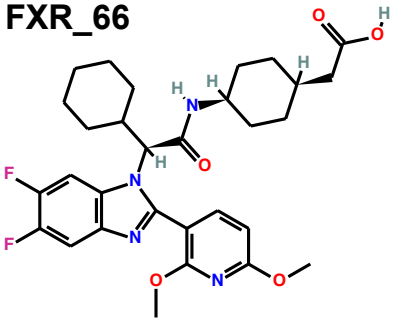
FXR_64



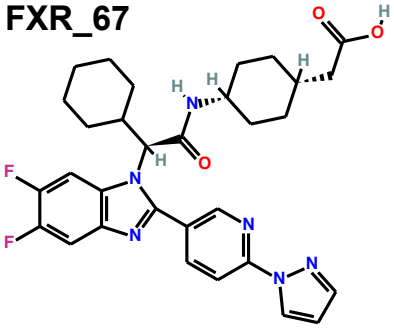
FXR_65



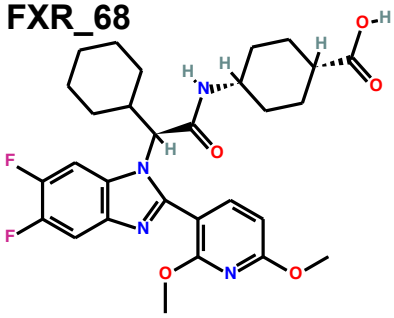
FXR_66



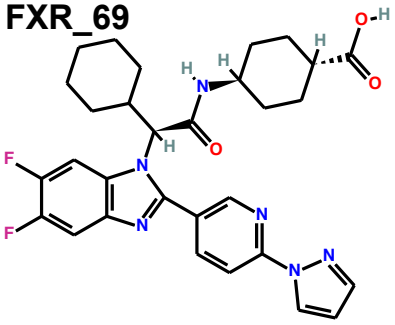
FXR_67



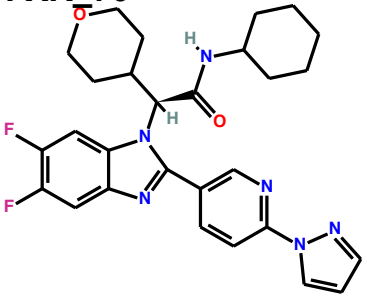
FXR_68



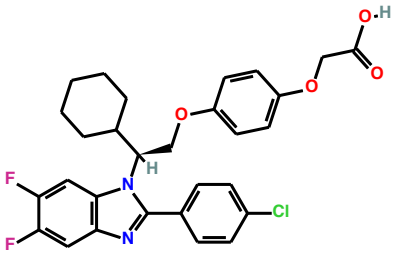
FXR_69



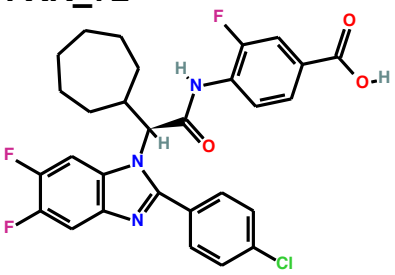
FXR_70



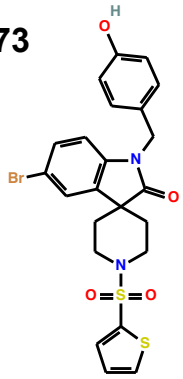
FXR_71



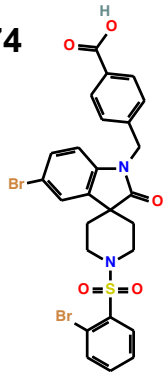
FXR_72



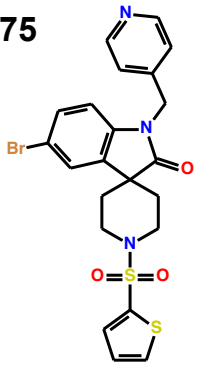
FXR_73



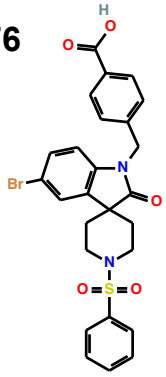
FXR_74



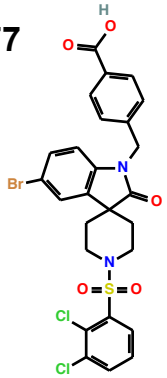
FXR_75



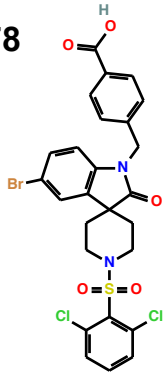
FXR_76



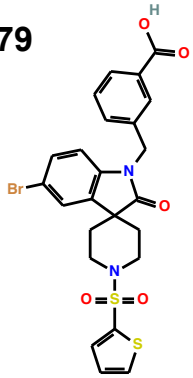
FXR_77



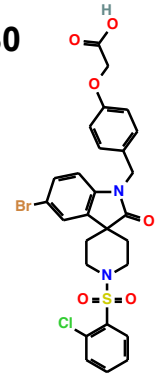
FXR_78



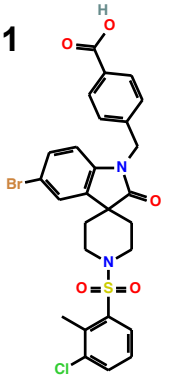
FXR_79



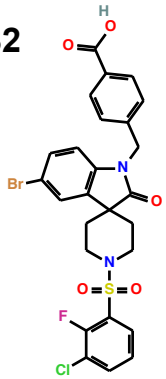
FXR_80



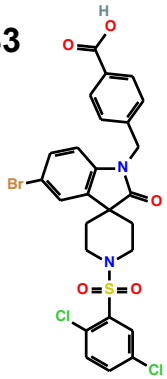
FXR_81



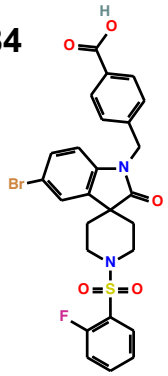
FXR_82



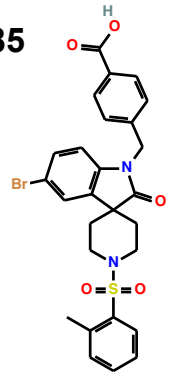
FXR_83



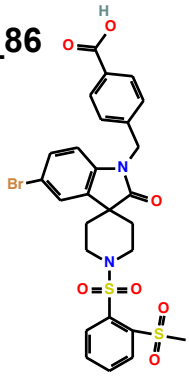
FXR_84



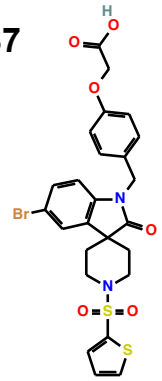
FXR_85



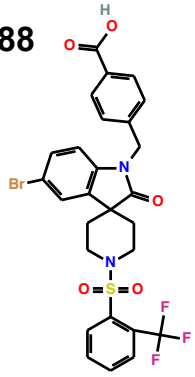
FXR_86



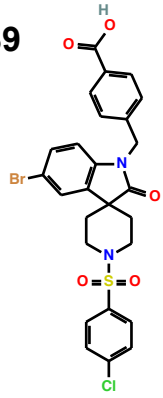
FXR_87



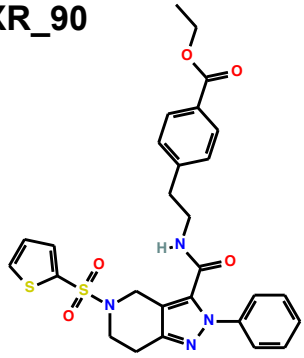
FXR_88



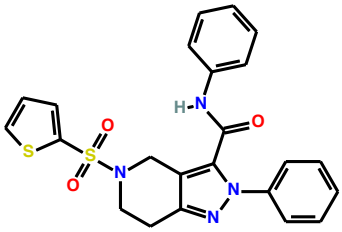
FXR_89



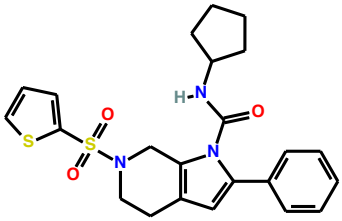
FXR_90



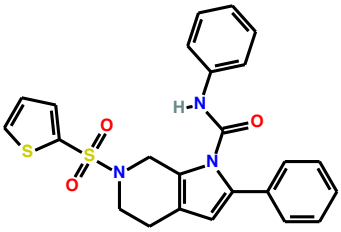
FXR_91



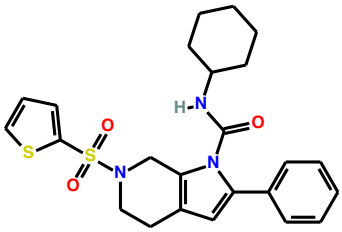
FXR_92



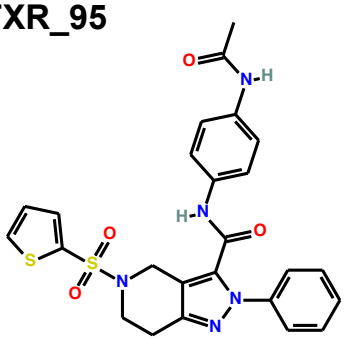
FXR_93



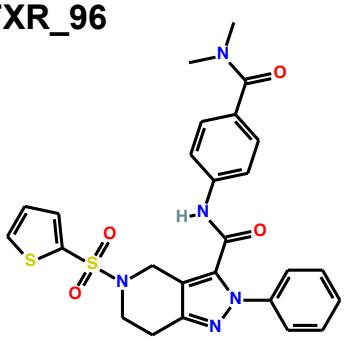
FXR_94



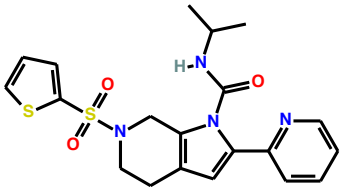
FXR_95



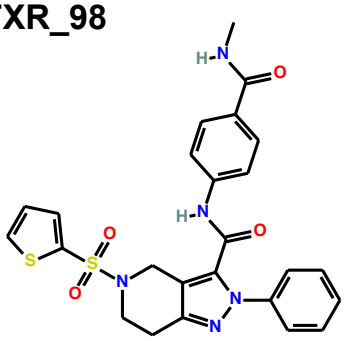
FXR_96



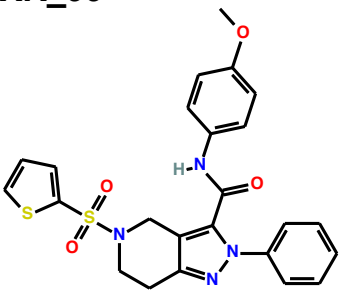
FXR_97



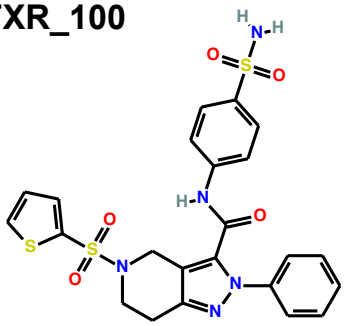
FXR_98



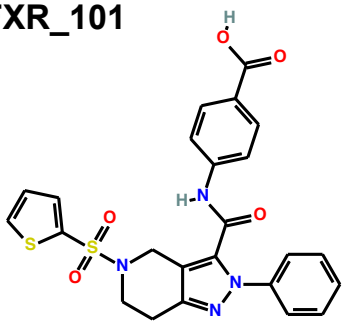
FXR_99



FXR_100



FXR_101



FXR_102

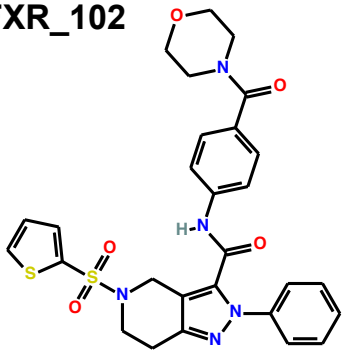
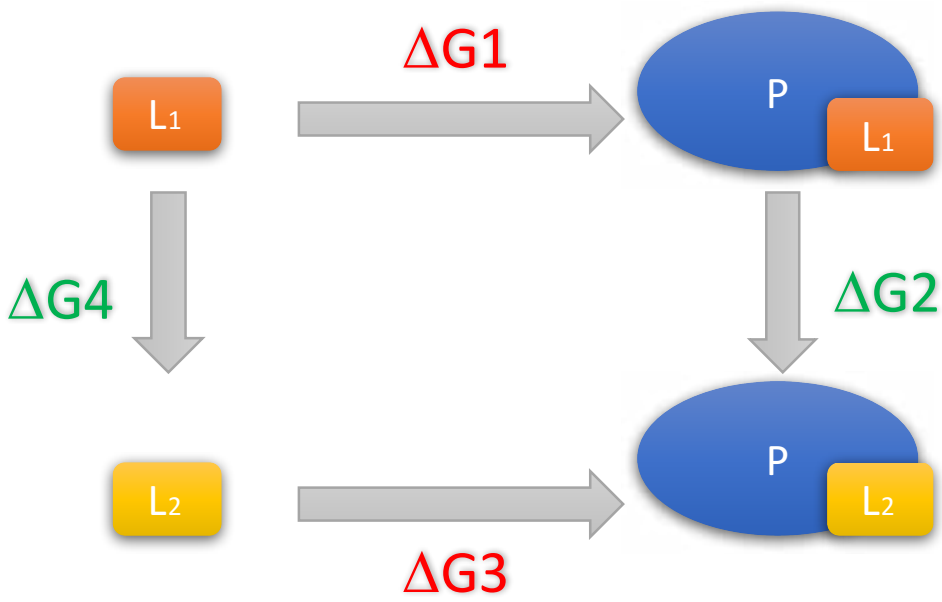


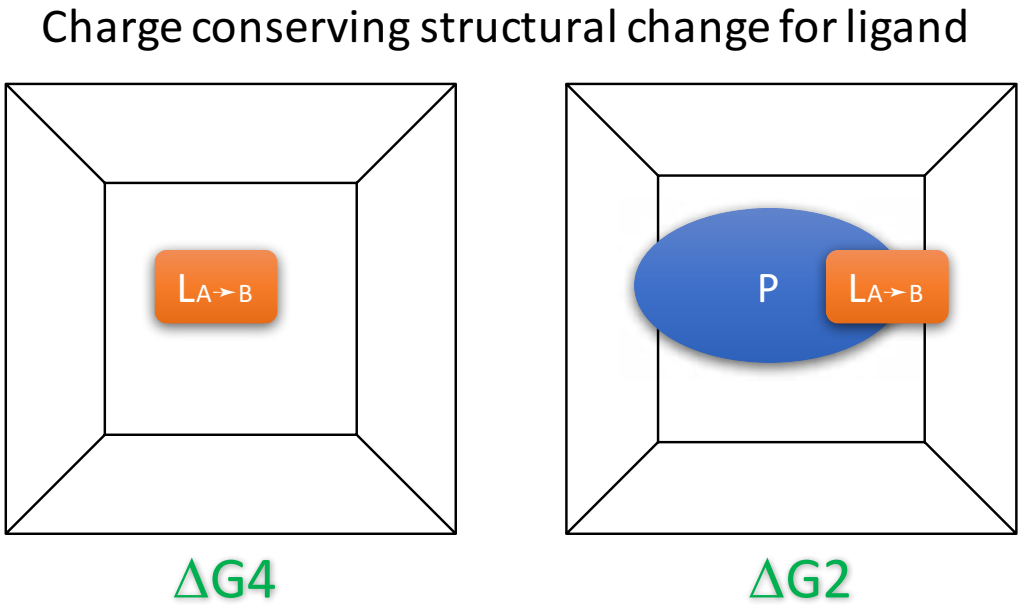
Figure S2. Thermodynamic cycle for the calculation of relative binding affinities of ligands L_1 and L_2 compared with the protein P.



$$\Delta G1 + \Delta G2 - \Delta G3 - \Delta G4 = 0$$

$$\Delta\Delta G = \Delta G1 - \Delta G3 = \Delta G4 - \Delta G2$$

Figure S3. Schematic representation of the system used for the calculation of relative binding affinities of ligands L_1 and L_2 for the protein P, in the case of a charge conserving structural change on the ligand.



$$\Delta\Delta G = \Delta G4 - \Delta G2$$

Figure S4. Schematic representation of the system used for the calculation of relative binding affinities of ligands L_1 and L_2 for the protein P, in the case of a charge modifying structural change on the ligand.

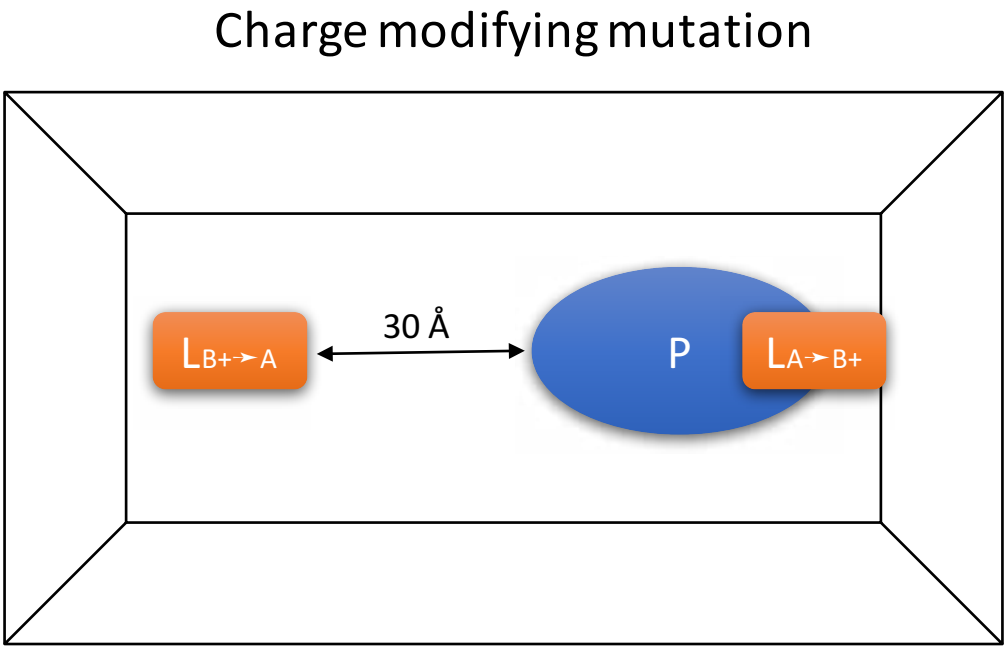
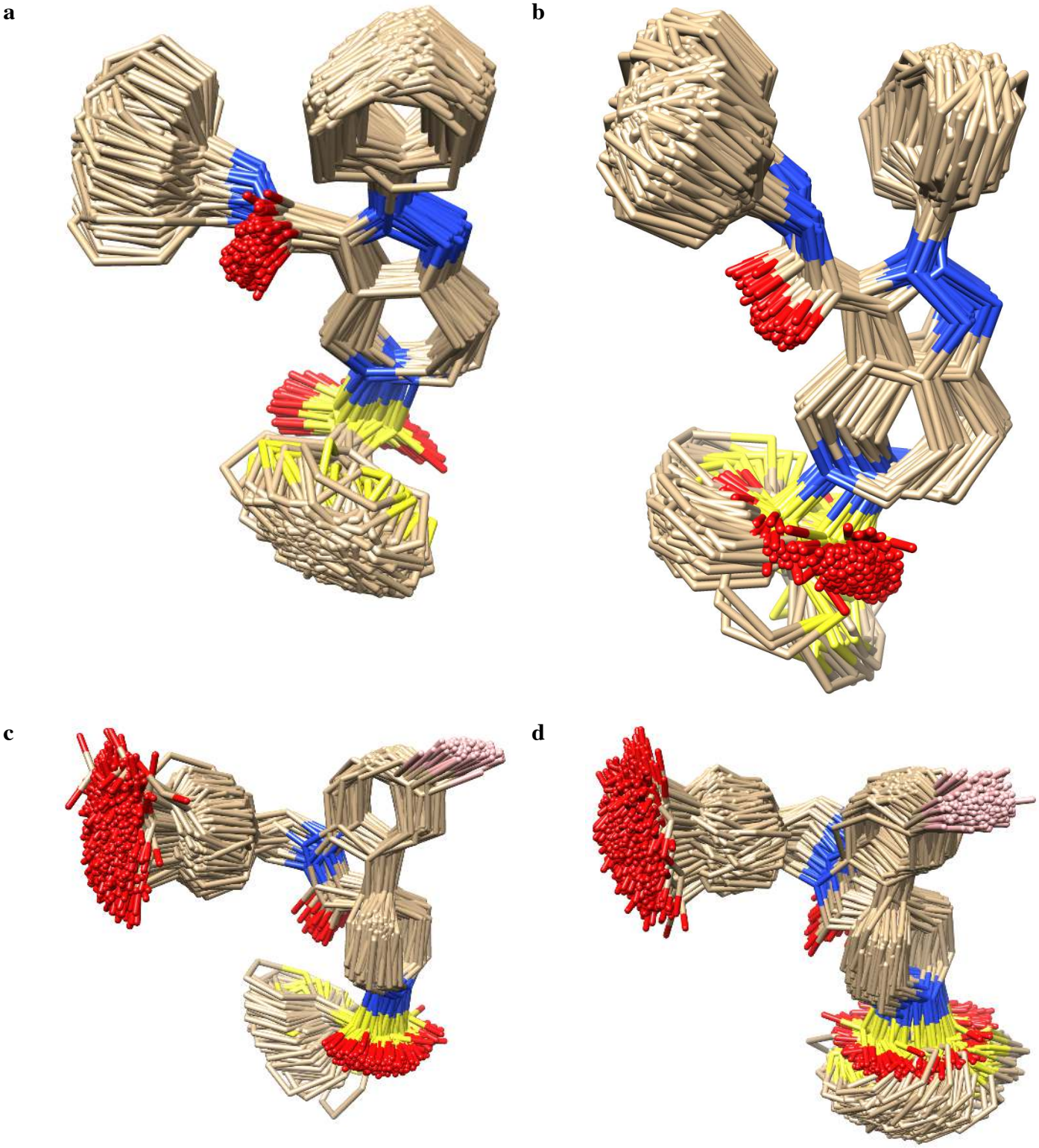
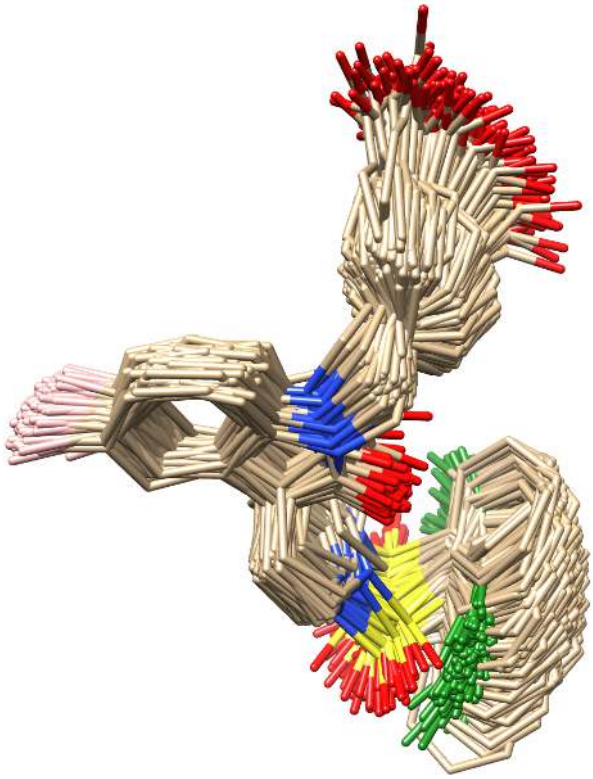


Figure S5. Conformational distribution of ligands FXR_17 (**a**, **b**), FXR_10 (**c**, **d**) and FXR_12 (**e**, **f**), as representative structures for set1 and set2. In each case, all 501 conformers extracted from the 10 ns molecular dynamics simulation of the ligand alone in water, using the OPLS-AA (**a**, **c**, **e**) and AMBER/GAFF (**b**, **d**, **f**) force fields, are represented. Hydrogen atoms are not shown for more clarity.



e



f

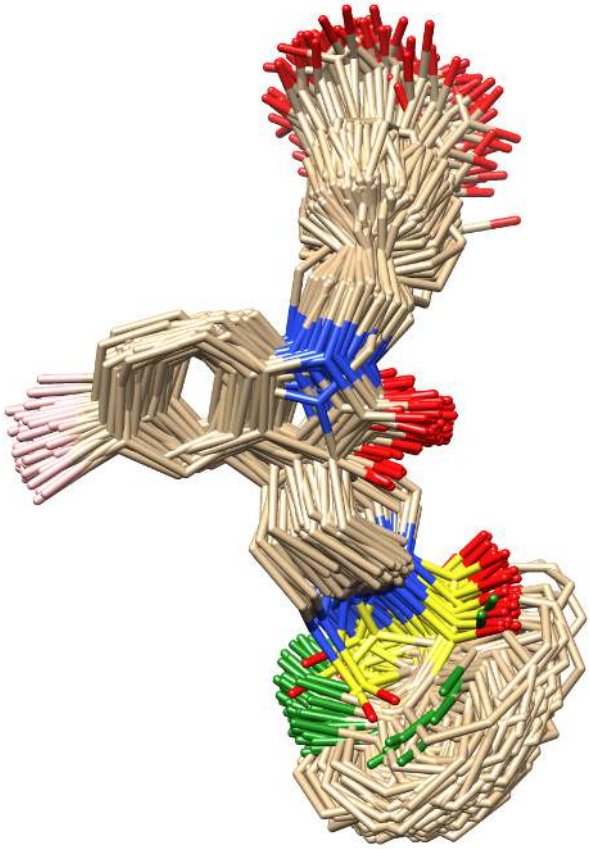


Table S1. Rank of the best RMSD poses. When no reference structure was available, the score of the first ranked pose was reported for the two submissions, therefore no rank of best RMSD pose is considered.

Ligand	Score of first ranked pose	Score of best RMSD pose	Rank of best RMSD pose (out of 10 poses)
FXR_1	57.25	57.25	–
FXR_2	57.27	57.27	–
FXR_3	59.03	59.03	–
FXR_4	63.35	64.13	6
FXR_5	58.53	62.45	7
FXR_6	49.10	52.44	4
FXR_7	48.07	49.36	4
FXR_8	53.28	54.13	3
FXR_9	43.55	43.55	1
FXR_10	51.19	51.19	–
FXR_11	44.97	44.97	–
FXR_12	47.66	47.66	–
FXR_13	41.27	41.27	1
FXR_14	45.85	46.23	2
FXR_15	49.65	49.65	–
FXR_16	56.47	56.47	–
FXR_17	42.05	42.05	–
FXR_18	54.34	54.34	–
FXR_19	49.24	52.23	9
FXR_20	48.05	49.73	2
FXR_21	46.07	46.26	2
FXR_22	51.95	51.95	1
FXR_23	53.66	57.20	10
FXR_24	44.42	45.46	7
FXR_25	46.74	47.83	7
FXR_26	39.43	39.43	1
FXR_27	39.50	41.40	5
FXR_28	40.42	42.27	7
FXR_29	41.16	43.78	6
FXR_30	41.54	44.58	7
FXR_31	44.30	45.09	4
FXR_32	52.50	53.19	2
FXR_33	40.94	42.95	5
FXR_34	52.87	59.51	10
FXR_35	40.64	40.64	1
FXR_36	32.84	35.97	7
FXR_37	55.97	58.21	10
FXR_38	50.16	50.16	–
FXR_39	51.58	51.66	4
FXR_40	48.89	51.89	5
FXR_41	45.89	45.89	–
FXR_42	45.92	50.25	7
FXR_43	46.24	46.24	–
FXR_44	48.13	48.13	–
FXR_45	39.38	39.38	–
FXR_46	42.54	42.54	–
FXR_47	51.05	51.05	–
FXR_48	43.75	43.75	–
FXR_49	43.65	43.65	–
FXR_50	52.93	54.39	6
FXR_51	48.25	48.74	2
FXR_52	45.76	54.79	3
FXR_53	39.81	41.12	7

FXR_54	46.19	48.11	4
FXR_55	45.06	45.37	4
FXR_56	43.69	43.69	1
FXR_57	47.05	48.29	8
FXR_58	38.19	38.19	1
FXR_59	39.35	41.05	4
FXR_60	41.27	43.93	2
FXR_61	40.38	43.31	5
FXR_62	38.54	38.63	7
FXR_63	40.07	40.55	9
FXR_64	39.42	41.64	4
FXR_65	38.25	40.65	7
FXR_66	41.96	43.83	5
FXR_67	36.46	37.87	2
FXR_68	41.87	44.72	3
FXR_69	33.62	33.66	2
FXR_70	39.28	45.34	6
FXR_71	36.91	38.68	7
FXR_72	38.51	38.77	2
FXR_73	49.38	49.38	–
FXR_74	48.03	48.03	–
FXR_75	50.52	50.52	–
FXR_76	48.20	48.20	–
FXR_77	44.43	44.43	–
FXR_78	46.49	46.49	–
FXR_79	48.64	48.64	–
FXR_80	37.81	37.81	–
FXR_81	44.02	44.02	–
FXR_82	45.46	45.46	–
FXR_83	45.72	45.72	–
FXR_84	47.15	47.15	–
FXR_85	46.19	46.19	–
FXR_86	44.59	44.59	–
FXR_87	41.90	41.90	–
FXR_88	41.68	41.68	–
FXR_89	47.94	47.94	–
FXR_90	41.55	41.55	–
FXR_91	48.44	48.44	–
FXR_92	55.18	55.18	–
FXR_93	49.39	49.39	–
FXR_94	53.68	53.68	–
FXR_95	41.40	41.40	–
FXR_96	42.61	42.61	–
FXR_97	55.88	55.88	–
FXR_98	42.50	42.50	–
FXR_99	45.73	45.73	–
FXR_100	40.67	40.67	–
FXR_101	45.12	45.12	–
FXR_102	40.80	40.80	–
Mean rank of best RMSD pose (from 53 values)			**4.68**

Table S2. Tanimoto similarity matrix for the compounds belonging to set1.

Tanimoto	FXR_17	FXR_45	FXR_46	FXR_47	FXR_48	FXR_49	FXR_91	FXR_93	FXR_95	FXR_96	FXR_98	FXR_99	FXR_100	FXR_101	FXR_102	Mean value per compound
FXR_17	1.00	0.91	0.88	0.95	0.93	0.87	0.84	0.68	0.85	0.88	0.88	0.86	0.84	0.94	0.87	0.88
FXR_45	0.91	1.00	0.81	0.88	0.85	0.81	0.78	0.63	0.78	0.81	0.81	0.91	0.77	0.86	0.80	0.83
FXR_46	0.88	0.81	1.00	0.86	0.91	0.97	0.95	0.76	0.94	0.98	0.98	0.83	0.95	0.90	0.94	0.91
FXR_47	0.95	0.88	0.86	1.00	0.93	0.86	0.86	0.69	0.86	0.86	0.86	0.88	0.86	0.90	0.85	0.87
FXR_48	0.93	0.85	0.91	0.93	1.00	0.91	0.90	0.73	0.90	0.90	0.90	0.85	0.89	0.90	0.89	0.89
FXR_49	0.87	0.81	0.97	0.86	0.91	1.00	0.93	0.76	0.93	0.97	0.97	0.83	0.93	0.88	0.92	0.90
FXR_91	0.84	0.78	0.95	0.86	0.90	0.93	1.00	0.78	0.97	0.94	0.94	0.85	0.98	0.86	0.90	0.90
FXR_93	0.68	0.63	0.76	0.69	0.73	0.76	0.78	1.00	0.77	0.76	0.75	0.68	0.77	0.69	0.73	0.75
FXR_95	0.85	0.78	0.94	0.86	0.90	0.93	0.97	0.77	1.00	0.94	0.94	0.85	0.97	0.85	0.89	0.90
FXR_96	0.88	0.81	0.98	0.86	0.90	0.97	0.94	0.76	0.94	1.00	0.98	0.82	0.93	0.89	0.94	0.91
FXR_98	0.88	0.81	0.98	0.86	0.90	0.97	0.94	0.75	0.94	0.98	1.00	0.82	0.93	0.89	0.93	0.91
FXR_99	0.86	0.91	0.83	0.88	0.85	0.83	0.85	0.68	0.85	0.82	0.82	1.00	0.85	0.83	0.80	0.84
FXR_100	0.84	0.77	0.95	0.86	0.89	0.93	0.98	0.77	0.97	0.93	0.93	0.85	1.00	0.86	0.89	0.89
FXR_101	0.94	0.86	0.90	0.90	0.90	0.88	0.86	0.69	0.85	0.89	0.89	0.83	0.86	1.00	0.85	0.87
FXR_102	0.87	0.80	0.94	0.85	0.89	0.92	0.90	0.73	0.89	0.94	0.93	0.80	0.89	0.85	1.00	0.88
Global mean value																**0.88**

Table S3. Tanimoto similarity matrix for the compounds belonging to set2.

Tanimoto	FXR_10	FXR_12	FXR_38	FXR_41	FXR_73	FXR_74	FXR_75	FXR_76	FXR_77	FXR_78	FXR_79	FXR_81	FXR_82	FXR_83	FXR_84	FXR_85	FXR_88	FXR_89	Mean value per compound
FXR_10	1.00	0.85	0.94	0.81	0.87	0.89	0.75	0.90	0.82	0.83	0.95	0.83	0.84	0.83	0.89	0.89	0.87	0.85	0.87
FXR_12	0.85	1.00	0.81	0.94	0.75	0.92	0.64	0.94	0.96	0.97	0.82	0.95	0.96	0.97	0.92	0.92	0.91	0.97	0.90
FXR_38	0.94	0.81	1.00	0.85	0.83	0.84	0.74	0.85	0.78	0.79	0.91	0.80	0.80	0.79	0.84	0.85	0.83	0.81	0.84
FXR_41	0.81	0.94	0.85	1.00	0.71	0.87	0.63	0.89	0.91	0.92	0.78	0.91	0.91	0.92	0.87	0.88	0.86	0.92	0.87
FXR_73	0.87	0.75	0.83	0.71	1.00	0.77	0.75	0.79	0.72	0.73	0.88	0.73	0.73	0.73	0.77	0.77	0.76	0.75	0.78
FXR_74	0.89	0.92	0.84	0.87	0.77	1.00	0.66	0.98	0.89	0.90	0.85	0.90	0.91	0.90	0.96	0.96	0.95	0.92	0.89
FXR_75	0.75	0.64	0.74	0.63	0.75	0.66	1.00	0.67	0.62	0.62	0.74	0.62	0.63	0.62	0.66	0.66	0.65	0.64	0.68
FXR_76	0.90	0.94	0.85	0.89	0.79	0.98	0.67	1.00	0.90	0.92	0.86	0.91	0.92	0.91	0.98	0.98	0.96	0.94	0.91
FXR_77	0.82	0.96	0.78	0.91	0.72	0.89	0.62	0.90	1.00	0.95	0.79	0.93	0.94	0.96	0.89	0.89	0.88	0.94	0.88
FXR_78	0.83	0.97	0.79	0.92	0.73	0.90	0.62	0.92	0.95	1.00	0.80	0.93	0.94	0.96	0.90	0.90	0.89	0.95	0.88
FXR_79	0.95	0.82	0.91	0.78	0.88	0.85	0.74	0.86	0.79	0.80	1.00	0.80	0.81	0.80	0.85	0.85	0.84	0.82	0.84
FXR_81	0.83	0.95	0.80	0.91	0.73	0.90	0.62	0.91	0.93	0.93	0.80	1.00	0.95	0.94	0.90	0.93	0.90	0.95	0.88
FXR_82	0.84	0.96	0.80	0.91	0.73	0.91	0.63	0.92	0.94	0.94	0.81	0.95	1.00	0.95	0.94	0.91	0.91	0.96	0.89
FXR_83	0.83	0.97	0.79	0.92	0.73	0.90	0.62	0.91	0.96	0.96	0.80	0.94	0.95	1.00	0.90	0.90	0.88	0.95	0.88
FXR_84	0.89	0.92	0.84	0.87	0.77	0.96	0.66	0.98	0.89	0.90	0.85	0.90	0.94	0.90	1.00	0.96	0.97	0.92	0.90
FXR_85	0.89	0.92	0.85	0.88	0.77	0.96	0.66	0.98	0.89	0.90	0.85	0.93	0.91	0.90	0.96	1.00	0.97	0.92	0.90
FXR_88	0.87	0.91	0.83	0.86	0.76	0.95	0.65	0.96	0.88	0.89	0.84	0.90	0.91	0.88	0.97	0.97	1.00	0.91	0.89
FXR_89	0.85	0.97	0.81	0.92	0.75	0.92	0.64	0.94	0.94	0.95	0.82	0.95	0.96	0.95	0.92	0.92	0.91	1.00	0.90
Global mean value																			**0.86**