Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

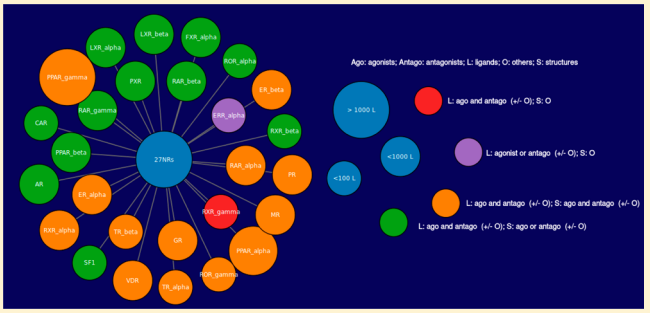
NRLiSt BDB, the Manually Curated Nuclear Receptors Ligands and Structures Benchmarking Database

Nathalie Lagarde, Nesrine Ben Nasr, Aurore Jérémie, Hélène Guillemain, Vincent Laville, Taoufik Labib, Jean-François Zagury, and Matthieu Montes*

Laboratoire Génomique, Bioinformatique et Applications, EA 4627, Conservatoire National des Arts et Métiers, 292 Rue Saint Martin, 75003 Paris, France

Supporting Information

ABSTRACT: Nuclear receptors (NRs) constitute an important class of drug targets. We created the most exhaustive NR-focused benchmarking database to date, the NRLiSt BDB (NRs ligands and structures benchmarking database). The 9905 compounds and 339 structures of the NRLiSt BDB are ready for structure-based and ligand-based virtual screening. In the present study, we detail the protocol used to generate the NRLiSt BDB and its features. We also give some examples of the errors that we found in ChEMBL that convinced us to manually review all original papers. Since extensive and manually curated experimental data about NR ligands and structures are provided in the NRLiSt BDB, it should become a powerful tool to assess the performance of virtual screening methods on NRs, to assist the understanding of NR's function and modulation, and to support the discovery of new drugs targeting NRs. NRLiSt BDB is freely available online at http://nrlist.drugdesign.fr.



INTRODUCTION

The nuclear receptors (NRs) are transcription factors naturally switched on and off by small-molecule hormones.[1] The NRs are involved in a wide range of physiological key functions and constitute an important class of drug targets. The evaluation of virtual screening methods, often achieved retrospectively, is necessary to ensure their reliability and relies on the availability of high quality benchmarking data sets. Among those developed over the years,[2−5] the directory of useful decoys (DUD)[6] and its enhanced version (DUD-E)[7] are considered to date as the current gold standard benchmarking database. In the DUD, 40 structurally different targets are available, including 8 NRs: AR, ER_alpha agonist and ER_alpha antagonist, GR, mineralocorticoid receptor (MR), peroxisome proliferative activated receptor γ (PPAR_gamma), progesterone receptor (PR), and retinoid X receptor α (RXR_alpha). Recently, an enhanced version of the DUD has been proposed. This new benchmarking data set has taken into account several "weaknesses" pointed out in the construction of the DUD[8−11] and consists of 102 targets including 11 NRs, the 7 previously available in the DUD (ER_alpha agonist and antagonist were merged into a unique data set), and 4 new NRs: estrogen receptor β (ER_beta), peroxisome proliferative activated receptor α (PPAR_alpha), peroxisome proliferative activated receptor δ also known as peroxisome proliferative activated receptor β (PPAR_beta), and thyroid hormone receptor β (TR_beta). To date, the most exhaustive benchmarking data set that includes NRs has been constructed in 2010 by Park et al[1] to perfom a study on NR ligands profiling and includes 3 to 33 active compounds on 13 nuclear receptors (NRs) with an ensemble of different pocket conformations. Benchmarking databases apart, several databases dedicated to NRs have been developed over the time. Some of them provide tools to elucidate NRs biology, for example, the NURSA (Nuclear Receptor Signaling Atlas)[12] which furnishes transcriptomic, proteomic, and metabolomic resources to study the expression, the organization, and the function of NRs.[13] Similarly, the IUPHAR-DB (International Union of Basic and Clinical Pharmacology database) is focused on the NRs pharmacology.[14] Other databases are more focused on NR sequences data, like the NucleaRDB which contains sequences, structures, and mutations of the NRs,[15] the NUREBASE[16,17] which provides the protein and DNA sequences, protein alignments and phylogenies, taxonomy and annotations of NRs, and the NRMD (nuclear receptor mutation database) which contains an atlas of NR mutations.[18] Finally, the NURBS (nuclear receptor binding site database)[19] includes data on experimental and predicted murine NRs. Unfortunately, none of them provides exhaustive data on ligands and structures of NRs, with annotated information about NR ligands activities and pharmacological profile.

We thus decided to create the most exhaustive NR-focused benchmarking database to date, the NRLiSt BDB (nuclear

Received: January 23, 2014
Published: March 17, 2014

   dx.doi.org/10.1021/jm500132p | J. Med. Chem. 2014, 57, 3117−3125

Table 1. Pharmacological Profile of the Ligands Bound in the 339 Structures and of the 9905 NR Ligands Presented in the NRLiSt BDB

	no. of ligands						no. of structures		
	agonists			antagonists					
	ligands	clusters	decoys	ligands	clusters	decoys	agonists-bound	antagonists-bound	others-bound
AR	179	11	8746	226	23	11586	29	0	7
CAR	33	16	1499	2	2	146	2	0	0
ER_alpha	434	26	21642	137	13	6555	11	4	17
ER_beta	392	28	18953	70	12	3547	20	2	3
ERR_alpha	13	2	1043	3	2	150	0	0	2
FXR_alpha	320	13	16559	28	4	1172	7	0	0
GR	295	12	15207	369	16	19664	6	1	0
LXR_alpha	259	18	14149	50	5	2485	4	0	0
LXR_beta	374	18	18743	38	4	1865	7	0	0
MR	9	3	495	146	18	7467	5	2	0
PPAR_alpha	1401	27	67058	7	3	420	11	1	0
PPAR_beta	906	17	44650	11	4	597	11	0	2
PPAR_gamma	1820	51	50760	9	6	494	62	1	17
PR	269	15	13803	531	26	26539	8	2	3
PXR	100	24	6272	7	6	327	7	0	0
RAR_alpha	133	11	6550	66	3	3292	2	1	1
RAR_beta	130	9	6446	31	3	1583	1	0	0
RAR_gamma	132	11	6683	57	3	2855	8	0	0
ROR_alpha	3	2	147	13	2	649	2	0	0
ROR_gamma	7	2	348	4	1	200	3	1	0
RXR_alpha	210	13	11790	135	3	7151	24	2	3
RXR_beta	65	6	3874	7	3	348	2	0	0
RXR_gamma	71	6	4124	6	2	300	0	0	1
SF1	19	1	837	20	2	991	2	0	0
TR_alpha	69	8	3974	17	7	841	6	0	0
TR_beta	78	8	3648	15	5	795	10	0	0
VDR	132	4	6626	47	3	2336	16	0	0
total	7853	362	354626	2052	181	104355	266	17	56

receptors ligands and structures benchmarking database), optimized for the evaluation of structure-based and ligand-based virtual screening methods and for assisting the discovery of new NR ligands by providing comprehensive manually curated data on NR ligands and bound-structures pharmacological profiles. This new benchmarking database provides for the NRs having more than one agonist and one antagonist ligand and at least one experimental structure available, all available agonist and antagonist ligands identified in the literature. In order to assist more efficiently the discovery of new ligands of NRs, experimentally characterized agonist and antagonist ligands are provided in separated data sets, since evidence was reported that not only the ligand binding could induce conformational changes in the nuclear receptor structure but also these changes could vary according to the pharmacological profile of the ligand (agonist or antagonist).[20] Each data set of the NRLiSt BDB for a given NR comprises all the experimental human holo structures of the protein, the structures, and properties of all known active ligands (agonists or antagonists) and the structures of their corresponding decoys as provided by the DUD_E decoy generation tool.[7]

In the present study, we present the protocol used to generate the NRLiSt BDB. We also describe some of the errors found in ChEMBL that convinced us to manually review all original papers before including corresponding ligands into our database. NRLiSt BDB is the most comprehensive benchmarking data set on NR ligands and structures including their pharmacological profile and activity retrieved manually from the literature. Additionally to its natural usage for benchmarking virtual screening methods, this database could also be used to assist the understanding of NRs function and modulation and the discovery of new drugs targeting NRs.

RESULTS

Nuclear Receptors Ligands and Structures Benchmarking DataBase (NRLiSt BDB). The NRLiSt BDB is a manually curated benchmarking database dedicated to the NR ligands and structures pharmacological profiles. We focused the database on the NRs for which more than one agonist and one antagonist ligand and at least one experimental structure were available. Only 27 NRs out of the 48 identified to date satisfied these conditions. We retrieved from the PDB all experimentally resolved holo human structures of the 27 NRs except for RXR_gamma for which only one apo structure was available. Of note, GR and MR structures presented only 99% of identity with the human sequences since mutations were produced to enable the protein crystallization.[21−28] The holo structures were classified according to the pharmacological profile of the ligand that was cocrystallized in their binding site. An exhaustive review of the scientific literature using the ChEMBL database[29] as a starting point was conducted. We decided to focus only on the agonist and antagonist ligands and thus to ignore all ligands with alternative pharmacological profiles: inverse agonists, modulators, agonists/antagonists, and weak to partial agonists and antagonists. The structures of these ligands and their

dx.doi.org/10.1021/jm500132p | J. Med. Chem. 2014, 57, 3117−3125

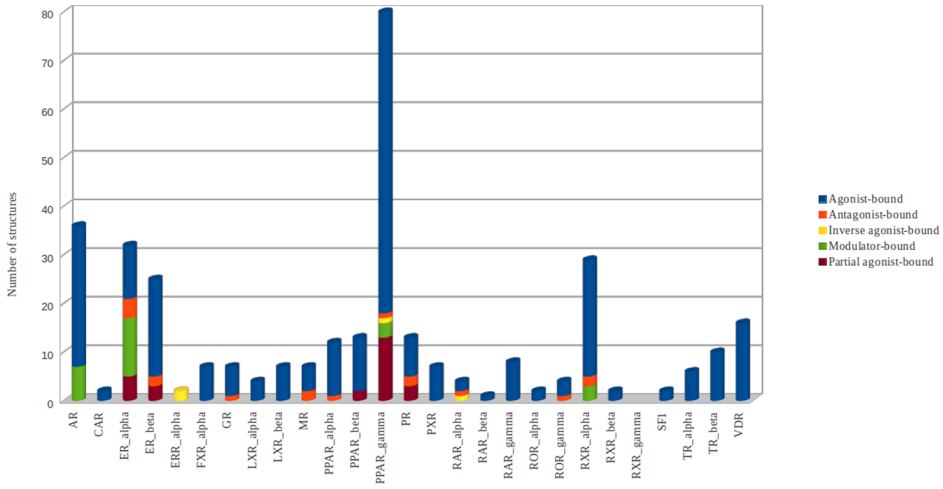


Figure 1. Classification of the holo structures included in the NRLiSt BDB according to the pharmacological profile of the bound ligand in the binding site. For RXR_gamma, only one apo structure was available.

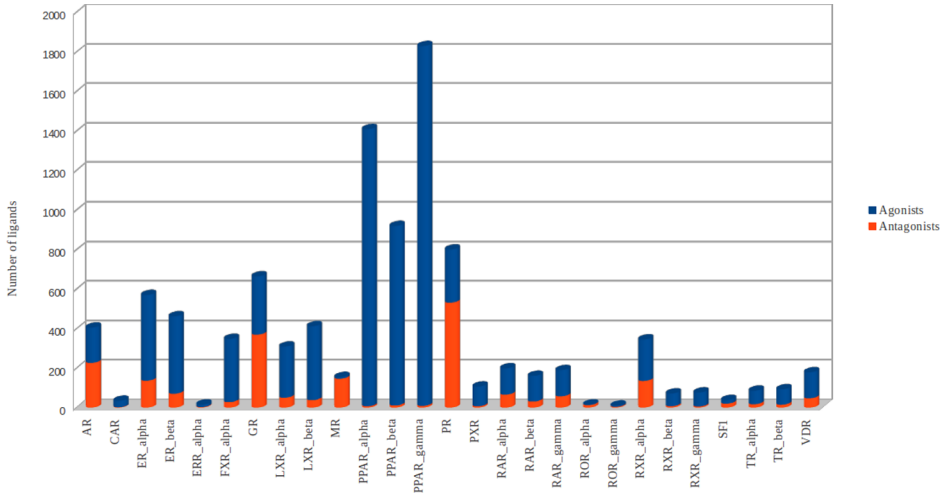


Figure 2. Classification of the ligands included in the NRLiSt BDB according to their pharmacological profiles.

corresponding activity and pharmacological profile are described in Supporting Information Table S1.

Hence, the NRLiSt BDB is constituted of 27 NRs, divided into 54 data sets comprising 339 structures (266 agonist-bound, 17 antagonist-bound, 56 others-bound) and 9905 ligands (7853 agonists, 2052 antagonists) (Table 1 and Figures 1 and 2). The number of structures available per NR varies from 1 (RAR_beta, RXR_gamma) to 80 (PPAR_gamma). The smallest set is CAR_antagonist containing only 2 ligands, and the largest is PPAR_gamma_agonist containing 1820 ligands. The structural diversity of the ligands in each data set was investigated using a Tanimoto similarity distance threshold (Td < 0.5) (see Table 1). For each data set, 1 (SF1_agonist) to 51 (PPAR_gamma_agonist) clusters were obtained, with an average value of 10 clusters per data set. For each ligand, the DUD_E decoy generation tool[7] was used to generate appropriate decoys. 458 981 decoys were obtained, leading to an average rate of 51 decoys per ligand. We also reported the mean values of 6 structural descriptors of the ligands computed with Dragon 6[30] (see Table 2 and Figure 3): the molecular weight (MW), the number of rotable bonds (nrotB), the number of hydrogen bond donors (HBD) and acceptors (HBA), the Moriguchi octanol–water partition coefficient (mlogP), and the topological polar surface area using N, O, S, and P polar contributions (TPSA).

By analyzing the differences between the values of these six descriptors obtained with the two pharmacological classes of ligands for each NR (agonists and antagonists), we observed that the mean value of the MW was significantly different for 16 out of the 27 NRs studied according to a Wilcoxon test[31] (Supporting Information Table S2). Focusing on the NRs presenting at least 10 ligands in each class, this rate raised to 13 out of 17. Similarly, when considering the flexibility of ligands, for 13 out of the 27 NRs studied, the mean value of the nrotB was also significantly different between the two classes according to a Wilcoxon test[31] (Supporting Information Table S2). As observed with the MW descriptor, when focusing on the NRs presenting at least 10 ligands in each class, this rate raised to 11 out of 17. No significant difference in the mean value of HBD, HBA, mlogP, and TPSA was noticed between agonists and antagonists.

DISCUSSION

In the present study, we have presented an exhaustive NR-focused benchmarking database, the NRLiSt BDB, dedicated to a family of targets widely studied for their therapeutical

dx.doi.org/10.1021/jm500132p | J. Med. Chem. 2014, 57, 3117–3125

Table 2. Mean Values of Six Descriptors Computed with Dragon 6.0 for the 54 Data Sets of the NRLiST BDB[a]

	MW	nrotB	HBD	HBA	mlogP	TPSA
AR_agonist	337.323	1.724	1.149	6.326	2.883	60.728
AR_antagonist	386.177	3.659	0.889	6.478	3.088	67.529
CAR_agonist	347.318	5.212	0.758	3.667	3.504	62.520
CAR_antagonist	341.860	3.500	0.000	2.000	4.624	25.510
ER_alpha_agonist	323.653	2.765	2.012	3.825	3.169	60.660
ER_alpha_antagonist	458.390	8.110	1.971	5.360	3.451	65.916
ER_beta_agonist	303.690	2.362	1.906	3.778	2.961	60.246
ER_beta_antagonist	382.938	4.443	2.071	4.300	3.432	60.332
ERR_alpha_agonist	264.158	1.692	0.769	3.538	2.996	49.362
ERR_alpha_antagonist	238.293	2.000	0.667	0.667	4.965	13.487
FXR_alpha_agonist	506.109	6.959	0.966	6.066	4.221	82.750
FXR_alpha_antagonist	452.890	5.750	1.107	3.857	4.940	59.740
GR_agonist	426.141	4.414	2.227	5.973	3.593	73.831
GR_antagonist	490.209	5.144	1.277	5.448	4.347	71.244
LXR_alpha_agonist	466.515	6.436	0.884	7.301	4.450	62.959
LXR_alpha_antagonist	378.834	3.440	0.660	3.700	4.711	50.322
LXR_beta_agonist	475.726	6.283	0.735	6.794	4.591	66.174
LXR_beta_antagonist	416.367	3.763	0.737	4.711	4.712	58.997
MR_agonist	365.658	2.889	1.889	4.889	1.993	79.309
MR_antagonist	380.065	3.521	0.849	5.288	3.557	69.381
PPAR_alpha_agonist	449.990	9.592	0.673	6.453	3.766	84.637
PPAR_alpha_antagonist	465.703	8.286	1.000	5.429	3.530	122.473
PPAR_beta_agonist	476.844	9.413	0.522	7.134	4.006	89.735
PPAR_beta_antagonist	448.175	6.727	1.091	8.364	3.146	89.213
PPAR_gamma_agonist	455.864	9.878	0.777	6.331	3.845	84.219
PPAR_gamma_antagonist	336.536	6.667	1.000	5.444	2.832	85.980
PR_agonist	347.317	1.851	0.680	3.450	3.921	53.948
PR_antagonist	350.452	2.512	0.859	4.079	3.657	54.665
PXR_agonist	420.947	5.790	1.320	5.270	3.785	72.750
PXR_antagonist	443.186	3.143	1.714	6.571	2.865	88.894
RAR_alpha_agonist	364.844	3.812	0.504	3.308	4.739	56.010
RAR_alpha_antagonist	444.993	4.182	0.212	3.288	5.511	54.711
RAR_beta_agonist	354.842	4.277	0.277	3.062	4.948	52.585
RAR_beta_antagonist	422.596	3.219	0.469	3.281	5.679	51.220
RAR_gamma_agonist	367.609	3.556	0.331	3.128	4.913	54.436
RAR_gamma_antagonist	445.899	3.684	0.193	3.123	5.581	52.540
ROR_alpha_agonist	427.933	5.000	1.000	5.667	5.693	46.127
ROR_alpha_antagonist	416.732	7.462	1.231	5.615	3.473	71.082
ROR_gamma_agonist	404.523	4.857	1.857	3.286	5.506	38.837
ROR_gamma_antagonist	402.225	5.000	1.750	2.000	5.485	39.670
RXR_alpha_agonist	381.307	4.900	0.271	3.733	4.815	50.206
RXR_alpha_antagonist	514.234	6.200	1.022	6.096	5.169	78.311
RXR_beta_agonist	370.459	4.785	0.323	2.969	4.739	49.351
RXR_beta_antagonist	486.984	4.857	0.429	6.000	5.045	85.721
RXR_gamma_agonist	372.868	4.845	0.296	2.915	4.881	48.414
RXR_gamma_antagonist	495.543	4.333	0.333	6.333	5.022	89.408
SF1_agonist	367.253	8.211	0.579	0.789	6.312	9.717
SF1_antagonist	425.356	9.200	1.150	7.750	2.138	93.778
TR_alpha_agonist	487.943	5.841	1.986	5.652	3.371	85.939
TR_alpha_antagonist	491.622	7.824	1.882	5.824	3.979	106.041
TR_beta_agonist	482.584	5.846	1.949	5.679	3.376	86.554
TR_beta_antagonist	461.425	7.467	1.733	5.333	3.266	90.214
VDR_agonist	457.805	8.579	3.075	4.165	4.330	72.571
VDR_antagonist	505.089	6.702	2.553	4.596	4.641	75.961

[a](MW: Molecular Weight, nrotB: number of rotable bonds, HBD: number of hydrogen bond donors, HBA: number of hydrogen bond acceptors; MlogP: Moriguchi octanol-water partition coefficient, TPSA: topological polar surface area using N, O, S, P polar contributions).

potential, the NRs. This benchmarking database exhibits three important features: (1) we did not just look for putative NR ligands but for compounds for which the pharmacological profile on a given NR was determined experimentally; (2) all information about the compound activity was manually curated from the literature; (3) given (1) and (2), despite being a

dx.doi.org/10.1021/jm500132p | J. Med. Chem. 2014, 57, 3117−3125

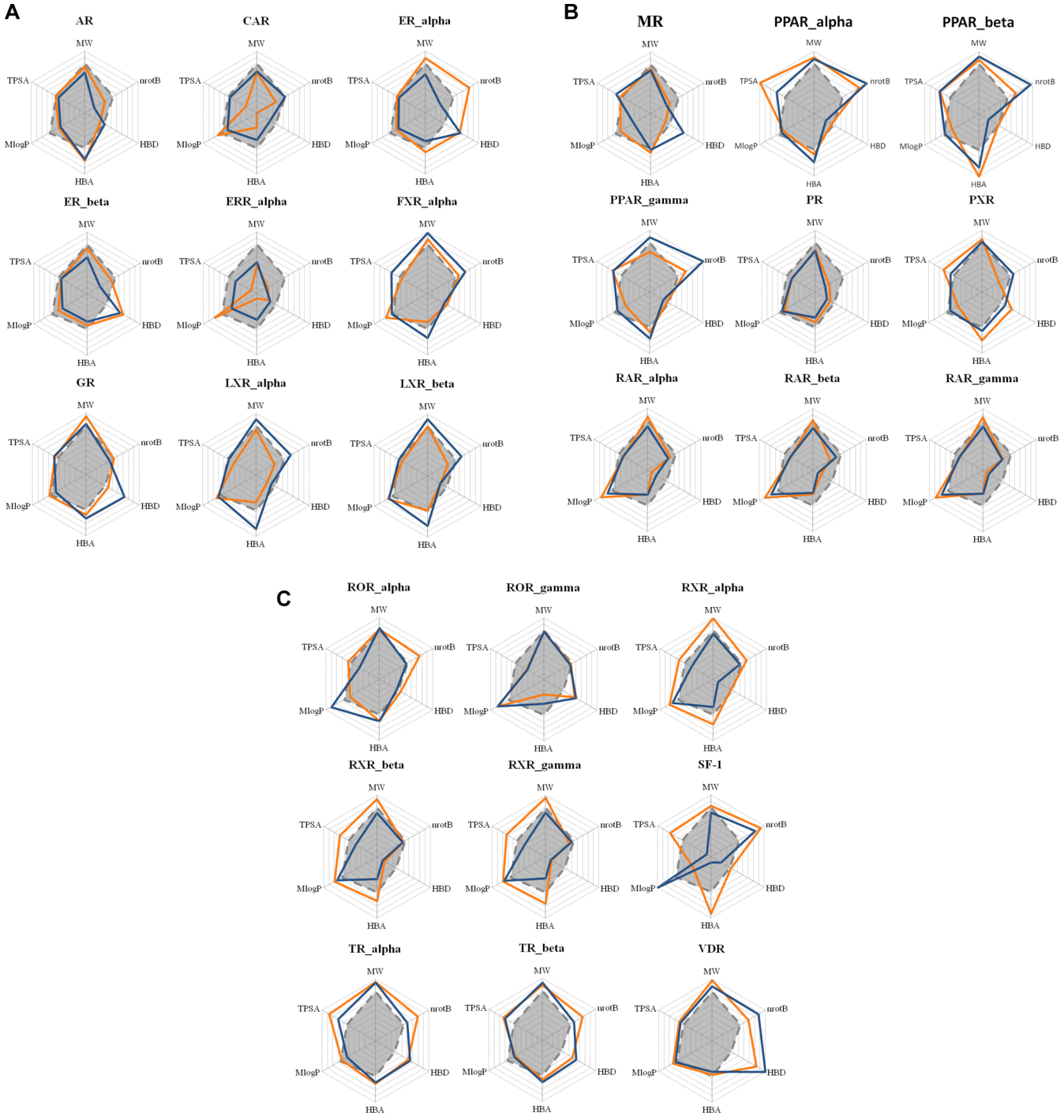


Figure 3. (A–C) Radiochart representation of mean values of the six descriptors computed with Dragon 6.0 for the agonist data sets (blue) and the antagonist data sets (orange) for each of the 27 NRs of the NRLiSt BDB: molecular weight (MW), number of rotable bonds (nrotB), number of hydrogen bond donors (HBD) and acceptors (HBA), Moriguchi octanol–water partition coefficient (MlogP), and topological polar surface area using N, O, S, and P polar contributions (TPSA).

benchmarking database, the NRList BDB can additionally be a precious source of information for assisting the discovery of new compounds targeting NRs.

The first benchmarking database was initiated by Bissantz et al.[2] more than 10 years ago. Several other benchmarking data sets have been proposed since then,[3−7,32] among which are the DUD[6] and DUD_E[7] databases that are considered as the gold standard to date. The initial DUD, contained 8 NRs and its enhanced version contained 11 NRs. Interestingly, for ER_alpha, the authors of the DUD decided to merge the agonist and antagonist data sets in a single one in the enhanced version of the DUD. Since we pointed out in a previous study[33] that this separation could be crucial for the quality of enrichment with structure-based virtual screening methods, we decided to create a benchmarking database specially focused on NRs with distinct data sets according to the pharmacological profile of the ligands. A similar distinction was previously

dx.doi.org/10.1021/jm500132p | J. Med. Chem. 2014, 57, 3117−3125

performed with GPCR ligands in the recently published GLL and GLD databases.[32]

To make our database suitable to benchmark structure-based virtual screening methods, we decided to focus on the 27 NRs for which at least one experimental structure was available. Contrary to the DUDs, we decided to provide, for each NR, all holo structures available in the PDB to avoid bias in future benchmarking studies by choosing a priori a "best" structure given its use in docking studies, resolution or feedback from previous use in the DUD.[7] In order to provide access to comprehensive data, for each holo structure, we have retrieved from the literature the information about the pharmacological profile of the bound ligand (agonist-bound, antagonist-bound, or other-bound). We decided to focus the NRLiSt BDB on agonist and antagonist ligands and thus eliminate all ligands with alternative pharmacological profiles: modulators, inverse agonist and partial agonists and antagonists.

To retrieve the ligands of each NR, we first used the ChEMBL database.[29] For each target of interest, the ChEMBL database proposed compounds that were classified according to their activity available in the literature. However, the information on the pharmacological profile of the ligands was not necessarily available, and we thus decided to manually retrieve all information about ligand binding and pharmacological profiles. During this extensive literature screen, we pointed out numerous false positive results in the ChEMBL database, which confirmed the legitimacy of our decision to review manually all data. For example, there are two compounds (namely, CHEMBL1961797 and CHEMBL1961794) that are denoted as bioactive on all NRs. By manual checking of the associated references in the literature,[34] these compounds have been determined experimentally as active with an agonist profile on Rev-erb alpha and Rev-erb beta only. "Neither compound exhibited activity at the other NR."[34] For LXR_alpha, two compounds (CHEMBL209145 also known as DPPF-01 and CHEMBL210372 also known as DPHK-01) are proposed as active compounds. However, as claimed in the associated publication,[35] "DPFF-01 did not activate VDR, PPAR_alpha, PPAR_gamma, PPAR_delta, LXR_alpha, RAR_alpha, or RXR_alpha at 10 μM under these experimental conditions". Conversely, "DPKH-01 did not activate VDR, FXR, PPAR_gamma, PPAR_delta, LXR_alpha, RAR_alpha, or RXR_alpha at 10 μM under these experimental conditions".[35] Another example among numerous ones was about a factor Xa inhibitor classified as an ER_alpha ligand, but no trace of such activity could be found in the associated reference (ER_alpha is not even mentioned in the publication).[36]

To complete the benchmarking database, we added decoys for each ligand based on the procedure to generate the DUD_E benchmarking data sets.[7]

We ultimately obtained a large benchmarking database, comprising 27 NRS, divided into 54 data sets, with a total of 9905 ligands including 7853 agonists and 2052 antagonists, 458 981 decoys and 339 experimental structures including 266 agonist-bound, 17 antagonist-bound, and 56 others-bound. There is a large disequilibrium between and within data sets, in terms of number of ligands and number of structures since several NRs have been more widely explored. For example, the PPARs are the NRs for which the largest number of ligands and structures were available, underlying the pharmaceutical potential of targeting such receptors.[37] Indeed, the PPARs mediate the action of several commercially available drugs active for diabetes (thiazolidinediones like rosiglitazone, PPARγ agonist still used in the U.S.[38]), hypolipidemia (bezafibrate, PPARα agonist[39]), or inflammation.[40] Conversely, many NRs are still orphan receptors, and for some of them no experimental structure was even available.

Structural Diversity of the NRLiSt. As discussed by J. Irwin,[10] sampling a sufficient part of the chemical space is of high importance for a benchmarking database. The NRs form a family of highly conserved receptors, which means that the absolute part of the chemical space that will be covered by the NRLiSt BDB is necessarily small. However, the NRLiSt BDB comprises the most exhaustive list of NRs agonists and antagonists and thus covers the most extensive chemical space related to NR ligands. By analysis of the cluster diversity as defined with a Td cutoff of 0.5, the NRLiSt BDB contains an average rate of 10 clusters per data set. Only 10 out of the 54 data sets contain less than 3 clusters, illustrating the lack of sufficient data for some receptors. For example, there are only 19 SF1 agonists that constitute a single cluster, which were extracted from the only two publications available to date. On the opposite, there were some very diverse data sets, for instance, the PPAR_gamma agonists data set that contains 51 clusters with a large amount of ligands, illustrating the extensive work performed on this particular class of NR ligands.

Structural Features of NR Ligands, Leads for Their Profiling. Simple structural descriptors could be used in an attempt to profile NR ligands. In many data sets, agonists and antagonists appeared to be significantly different in terms of molecular weight (MW) and number of rotatable bonds (nrotB). This observation is strengthened with the "large and equilibrated" NR data sets in terms of number of compounds in the agonist and antagonist data sets (like AR, ER, GR, LXR, PR, RAR, RXR_alpha, and VDR). No significant signal was found with hydrogen bond donors and acceptors, log *P*, or TPSA. These findings can be rationalized with previous knowledge about differences in the binding of agonists and antagonists. Indeed, many antagonists can be perfectly superimposed with their corresponding agonists, sharing similar hydrogen bonds (explaining why there is no significant difference between agonists and antagonists HBD and HBA mean values). However, as it is known for TR[41] and RAR_alpha,[42] some antagonists present bulky extensions (explaining the difference in MW and nrotB) that could prevent the H12 positioning in its agonist-bound form. All these data, correlated with the biological data that were also provided whenever available, constitute a robust support for structure−activity relationship studies and/or structure−properties relationship studies on NR ligands and could be used to assist medicinal chemists in prioritizing the synthesis of new NR ligands. Indeed, in the literature, such studies already enabled the successful discovery of new NRs ligands.[43−45]

NRLiSt Web Site. We decided to provide freely all data collected during our bibliographic investigation and constituting the NRLiSt BDB on our Web site (http://nrlist.drugdesign.fr). All 9905 compounds identified to be NR agonists or antagonists with a confirmed biological activity in the literature are presented as tables, with their identification (ZINC, CHEMBL, or CID ID), their pharmacological profile (agonist or antagonist), their name (common name or name attributed in the associated reference in the literature), binding and/or activity data, the corresponding references from the literature, and finally their cross-reactivity data (cross-reactivity statistics are also presented in Supporting Information Table S3).

Similarly, all 339 holo human experimental structures of the NRs collected in the PDB are indicated, with their resolution, the name, and pharmacological profile of the bound ligand, and the organism from which the NR gene was extracted for expression. The 54 agonist and antagonist data sets corresponding to the 27 NRs selected for the NRLiSt BDB are available for download. The data sets are formed with three directories named "ligands", "decoys", and "targets". The "ligands" directory contains the ligands in MOL2 and SMILES format, and their corresponding six structural descriptors are computed with Dragon 6.0. The "decoys" directory contains the decoys identified in the ZINC database for each compound of the data set in MOL2 and SMILES format. The "targets" directory contains all experimental structures for a given NR. For each structure the data provided are the original PDB format (for example, 1E3G.pdb), the same structure without ligand (1E3G_WL.pdb), the structure prepared for docking using CHIMERA (1E3G_prot.mol2), and the ligand protonated with adequate charges using Gasteiger partial charges (1E3G_ligand_prot.mol2). We hope this Web site will be useful for chemoinformatics studies, since our downloadable database, the NRLiSt BDB, could be used for benchmarking virtual screening softwares but also for pharmaceutical chemistry studies, since the binding and activity data are provided together with the ligands and receptors structures.

CONCLUSION

The NRLiSt BDB is, to date, the database that contains the most comprehensive experimental data about NR ligands and structures. Since its data have been entirely reviewed manually from the literature, it is also, to our knowledge, one of the most reliable benchmarking database and NR-focused database publicly available. Our choices regarding the construction of the NRLiSt BDB can provide new insights about the building of better benchmarking data sets in particular by taking into account the final phenotype (i.e pharmacological profile) resulting from the binding of ligands in addition to their sole affinity. NRLiSt BDB should become the database of reference to assess the performance of either structure or ligand-based virtual screening methods on NRs, to assist the understanding of NR's function and modulation, and to support the discovery of new drugs targeting NRs.

EXPERIMENTAL SECTION

NRLiSt BDB Generation. *Protein Target Selection and Ligand Collection.* The NRLiSt BDB was created by selecting from the 48 known NRs those for which more than one agonist, one antagonist ligand, and at least one experimental structure were available. For each of the 27 NRs corresponding to these criteria, all the human experimental holo structures available in the Protein Data Bank (PDB)[46] were downloaded. The pharmacological profile of the cocrystallized ligand was carefully monitored in the literature to define three classes of NR structures: agonist-bound, antagonist-bound, and others-bound (others corresponding to inverse agonists, modulators, agonists/antagonists, weak to partial agonists, and weak to partial antagonists). An extensive review of the scientific literature was achieved to collect all identified ligands for each NR, using the ChEMBL database[29] as a starting point. We manually curated all molecules proposed for a given NR to ascertain whether they were actually ligands or not and to classify them according to their pharmacological profile: agonist or antagonist. All inverse agonists, modulators, agonists/antagonists, weak to partial agonists, and weak to partial antagonists were eliminated. All ligands whose pharmacological profile was not informed were also eliminated. In addition to this qualitative activity data, quantitative data (K_d, K_i, EC_{50}, for example) for each compound for each NR were also retrieved. Of note, we also gathered the original reference from which the data were retrieved to allow rigorous activity comparisons between compounds tested in the same conditions.

Three data sets were created for each NR: "agonist", "antagonist", and a "total" data set including merged agonist and antagonist data sets. Each data set was formed of three elements: all the previously selected holo PDB structures (except for NRs for which only apo structures were available), all the ligands found to be agonists or antagonists, and their corresponding computed decoys using the DUD_E decoy generation tool.[7] The data set total is constituted, similar to the other data sets of all the previously selected PDB structures, of all ligands agonist and antagonist and the corresponding decoys.

Target Preparation. The structures were prepared to be usable for docking studies. The ligand bound in the active site was first removed from the protein. Then we used the DockPrep tool of UCSF CHIMERA[47] to delete the cocrystallized water molecules, repair truncated side chains, protonate the protein, and assign partial charges using the AMBER force field.

Ligands Preparation. All the ligands available in the ZINC database[48] were downloaded in MOL2 format, immediately usable for docking, and also in SMILES format. The remaining ligands were downloaded in SMILES format from the ChEMBL database[29] or the Pubchem database,[49] depending on their availability. The 3D structures of the ligands were generated with Corina online demo.[50] We used OpenBabel[51] to calculate Gasteiger's partial atomic charges.

Decoys Preparation. We used the DUD_E decoy generation tool[7] to select appropriate decoys for each ligand. These decoys were selected to have physical properties similar to those of their corresponding ligands (molecular weight, estimated water−octanol partition coefficient, rotatable bonds, hydrogen bond acceptors, hydrogen bond donors, and net charge) but to be topologically distant (computed with the Tanimoto coefficient, Tc). For each ligand and their alternative protonation states at pH 6−8 computed with Schrodinger's Epik, the DUD_E automated tool proposed, whenever possible, 50 decoys. The redundant decoys within a data set were removed, and the remaining decoys were then downloaded from the ZINC database in MOL2 and SMILES formats.

Ligand Cluster Definition. ICM, version 3.6, has been used to define ligand clusters in the NRLiSt BDB. Chemical descriptors fingerprints and Tanimoto similarity distance (Td) as implemented in ICM were used to classify the compounds in each data set. A Td cutoff was defined at 0.5 to obtain at least two equilibrated clusters in each data set, with the exception of SF1_agonist constituted by only one cluster.

Descriptors Calculation. Dragon 6. Dragon 6 enabled calculation of up to 4885 descriptors, from simple molecular descriptors to 2D and 3D descriptors, divided into 29 logical blocks.[30] We used the 3D structure of each compound to calculate 6 of their 4885 descriptors as proposed by Dragon 6.

Performance Metrics. All graphics were produced with the statistical and graphical tool R (http://www.r-project.org/).

ASSOCIATED CONTENT

Supporting Information

Details of the structure and inhibitory activity values for the 27 NR modulators, partial agonists and antagonists, and reverse agonists; *p*-values with a Wilcoxon test for comparing the mean values in four structural descriptors as computed by Dragon 6.0 for each data set; and cross-reactivity data about the NR ligands among the different NRs. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +33140272809. E-mail: matthieu.montes@cnam.fr.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Molsoft LLC for providing academic licenses for the ICM suite and Chemaxon for providing academic licenses for the Marvin suite. N.L. and V.L. are recipients of a CIFRE fellowship from ANRT. N.B.N. is recipient of a MNRT fellowship. H.G. is recipient of an ANSM fellowship.

■ ABBREVIATIONS USED

AR, androgen receptor; CAR, constitutive androstane receptor; DUD, directory of useful decoys; DUD-E, directory of useful decoys enhanced; EF, enrichment factor; ER_alpha, estrogen receptor α; ER_beta, estrogen receptor β; ERR_alpha, estrogen related receptor α; FXR_alpha, farnesoid X receptor α; GR, glucocorticoid receptor; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; LXR_alpha, liver X receptor α; LXR_beta, liver X receptor β; mlogP, Moriguchi octanol−water partition coefficient; MR, mineralocorticoid receptor; MW, molecular weight; NRLiSt BDB, nuclear receptors ligands and structures benchmarking database; nrotB, number of rotable bonds; NR, nuclear receptor; PDB, Protein Data Bank; PPAR_alpha, peroxisome proliferator activated receptor α; PPAR_beta, peroxisome proliferator activated receptor β; PPAR_gamma, peroxisome proliferator activated receptor γ; PR, progesterone receptor; PXR, pregnane X receptor; RAR_alpha, retinoic acid receptor α; RAR_beta, retinoic acid receptor β; RAR_gamma, retinoic acid receptor γ; ROR_alpha, retinoic acid receptor related orphan receptor α; ROR_gamma, retinoic acid receptor related orphan receptor γ; RXR_alpha, retinoid X receptor α; RXR_beta, retinoid X receptor β; RXR_gamma, retinoid X receptor γ; SF1, steroidogenic factor 1; Tc, Tanimoto coefficient; TPSA, topological polar surface area; TR_alpha, thyroid hormone receptor α; TR_beta, thyroid hormone receptor β; VDR, vitamin D receptor

■ REFERENCES

(1) Park, S. J.; Kufareva, I.; Abagyan, R. Improved docking, screening and selectivity prediction for small molecule nuclear receptor modulators using conformational ensembles. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 459−471.

(2) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.

(3) Miteva, M. A.; Lee, W. H.; Montes, M. O.; Villoutreix, B. O. Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. *J. Med. Chem.* **2005**, *48*, 6012−6022.

(4) Montes, M.; Miteva, M. A.; Villoutreix, B. O. Structure-based virtual ligand screening with LigandFit: pose prediction and enrichment of compound collections. *Proteins* **2007**, *68*, 712−725.

(5) Pham, T. A.; Jain, A. N. Parameter estimation for scoring protein−ligand interactions using negative training data. *J. Med. Chem.* **2006**, *49*, 5856−5868.

(6) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.

(7) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **2012**, *55*, 6582−6594.

(8) Good, A. C.; Oprea, T. I. Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection? *J. Comput.-Aided Mol. Des.* **2008**, *22*, 169−178.

(9) Hawkins, P. C.; Warren, G. L.; Skillman, A. G.; Nicholls, A. How to do an evaluation: pitfalls and traps. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 179−190.

(10) Irwin, J. J. Community benchmarks for virtual screening. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 193−199.

(11) Mysinger, M. M.; Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* **2010**, *50*, 1561−1573.

(12) Margolis, R. N.; Evans, R. M.; O'Malley, B. W.; Consortium, N. A. The nuclear receptor signaling atlas: development of a functional atlas of nuclear receptors. *Mol. Endocrinol.* **2005**, *19*, 2433−2436.

(13) McKenna, N. J.; Cooney, A. J.; DeMayo, F. J.; Downes, M.; Glass, C. K.; Lanz, R. B.; Lazar, M. A.; Mangelsdorf, D. J.; Moore, D. D.; Qin, J.; Steffen, D. L.; Tsai, M. J.; Tsai, S. Y.; Yu, R.; Margolis, R. N.; Evans, R. M.; O'Malley, B. W. Minireview: Evolution of NURSA, the nuclear receptor signaling atlas. *Mol. Endocrinol.* **2009**, *23*, 740−746.

(14) Sharman, J. L.; Mpamhanga, C. P. IUPHAR-DB: an open-access, expert-curated resource for receptor and ion channel research. *ACS Chem. Neurosci.* **2011**, *2*, 232−235.

(15) Vroling, B.; Thorne, D.; McDermott, P.; Joosten, H. J.; Attwood, T. K.; Pettifer, S.; Vriend, G. NucleaRDB: information system for nuclear receptors. *Nucleic Acids Res.* **2012**, *40*, D377−380.

(16) Duarte, J.; Perriere, G.; Laudet, V.; Robinson-Rechavi, M. NUREBASE: database of nuclear hormone receptors. *Nucleic Acids Res.* **2002**, *30*, 364−368.

(17) Ruau, D.; Duarte, J.; Ourjdal, T.; Perriere, G.; Laudet, V.; Robinson-Rechavi, M. Update of NUREBASE: nuclear hormone receptor functional genomics. *Nucleic Acids Res.* **2004**, *32*, D165−D167.

(18) Van Durme, J. J.; Bettler, E.; Folkertsma, S.; Horn, F.; Vriend, G. NRMD: nuclear receptor mutation database. *Nucleic Acids Res.* **2003**, *31*, 331−333.

(19) Fang, Y.; Liu, H. X.; Zhang, N.; Guo, G. L.; Wan, Y. J.; Fang, J. NURBS: a database of experimental and predicted nuclear receptor binding sites of mouse. *Bioinformatics* **2013**, *29*, 295−297.

(20) Bourguet, W.; Germain, P.; Gronemeyer, H. Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications. *Trends Pharmacol. Sci.* **2000**, *21*, 381−388.

(21) Bledsoe, R. K.; Madauss, K. P.; Holt, J. A.; Apolito, C. J.; Lambert, M. H.; Pearce, K. H.; Stanley, T. B.; Stewart, E. L.; Trump, R. P.; Willson, T. M.; Williams, S. P. A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor. *J. Biol. Chem.* **2005**, *280*, 31283−31293.

(22) Hasui, T.; Matsunaga, N.; Ora, T.; Ohyabu, N.; Nishigaki, N.; Imura, Y.; Igata, Y.; Matsui, H.; Motoyaji, T.; Tanaka, T.; Habuka, N.; Sogabe, S.; Ono, M.; Siedem, C. S.; Tang, T. P.; Gauthier, C.; De Meese, L. A.; Boyd, S. A.; Fukumoto, S. Identification of benzoxazin-3-one derivatives as novel, potent, and selective nonsteroidal mineralocorticoid receptor antagonists. *J. Med. Chem.* **2011**, *54*, 8616−8631.

(23) Bledsoe, R. K.; Montana, V. G.; Stanley, T. B.; Delves, C. J.; Apolito, C. J.; McKee, D. D.; Consler, T. G.; Parks, D. J.; Stewart, E. L.; Willson, T. M.; Lambert, M. H.; Moore, J. T.; Pearce, K. H.; Xu, H. E. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **2002**, *110*, 93−105.

(24) Biggadike, K.; Bledsoe, R. K.; Coe, D. M.; Cooper, T. W.; House, D.; Iannone, M. A.; Macdonald, S. J.; Madauss, K. P.; McLay, I. M.; Shipley, T. J.; Taylor, S. J.; Tran, T. B.; Uings, I. J.; Weller, V.; Williams, S. P. Design and X-ray crystal structures of high-potency nonsteroidal glucocorticoid agonists exploiting a novel binding site on the receptor. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18114−18119.

(25) Biggadike, K.; Bledsoe, R. K.; Hassell, A. M.; Kirk, B. E.; McLay, I. M.; Shewchuk, L. M.; Stewart, E. L. X-ray crystal structure of the novel enhanced-affinity glucocorticoid agonist fluticasone furoate in the glucocorticoid receptor−ligand binding domain. *J. Med. Chem.* **2008**, *51*, 3349−3352.

(26) Kauppi, B.; Jakob, C.; Farnegardh, M.; Yang, J.; Ahola, H.; Alarcon, M.; Calles, K.; Engstrom, O.; Harlan, J.; Muchmore, S.; Ramqvist, A. K.; Thorell, S.; Ohman, L.; Greer, J.; Gustafsson, J. A.;

Carlstedt-Duke, J.; Carlquist, M. The three-dimensional structures of antagonistic and agonistic forms of the glucocorticoid receptor ligand-binding domain: RU-486 induces a transconformation that leads to active antagonism. *J. Biol. Chem.* **2003**, *278*, 22748−22754.
(27) Suino-Powell, K.; Xu, Y.; Zhang, C.; Tao, Y. G.; Tolbert, W. D.; Simons, S. S., Jr.; Xu, H. E. Doubling the size of the glucocorticoid receptor ligand binding pocket by deacylcortivazol. *Mol. Cell. Biol.* **2008**, *28*, 1915−1923.
(28) Madauss, K. P.; Bledsoe, R. K.; McLay, I.; Stewart, E. L.; Uings, I. J.; Weingarten, G.; Williams, S. P. The first X-ray crystal structure of the glucocorticoid receptor bound to a non-steroidal agonist. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6097−6099.
(29) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.
(30) Talete SRL. Dragon 6. http://www.talete.mi.it/products/dragon_description.htm.
(31) Wilcoxon, F. Individual comparisons by ranking methods. *Biom. Bull.* **1945**, *1*, 80−83.
(32) Gatica, E. A.; Cavasotto, C. N. Ligand and decoy sets for docking to G protein-coupled receptors. *J. Chem. Inf. Model.* **2012**, *52*, 1−6.
(33) Ben Nasr, N.; Guillemain, H.; Lagarde, N.; Zagury, J. F.; Montes, M. Multiple structures for virtual ligand screening: defining binding site properties-based criteria to optimize the selection of the query. *J. Chem. Inf. Model.* **2013**, *53*, 293−311.
(34) Solt, L. A.; Wang, Y.; Banerjee, S.; Hughes, T.; Kojetin, D. J.; Lundasen, T.; Shin, Y.; Liu, J.; Cameron, M. D.; Noel, R.; Yoo, S. H.; Takahashi, J. S.; Butler, A. A.; Kamenecka, T. M.; Burris, T. P. Regulation of circadian behaviour and metabolism by synthetic REV-ERB agonists. *Nature* **2012**, *485*, 62−68.
(35) Kainuma, M.; Kasuga, J.; Hosoda, S.; Wakabayashi, K.; Tanatani, A.; Nagasawa, K.; Miyachi, H.; Makishima, M.; Hashimoto, Y. Diphenylmethane skeleton as a multi-template for nuclear receptor ligands: preparation of FXR and PPAR ligands. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3213−3218.
(36) Shi, Y.; Zhang, J.; Shi, M.; O'Connor, S. P.; Bisaha, S. N.; Li, C.; Sitkoff, D.; Pudzianowski, A. T.; Chong, S.; Klei, H. E.; Kish, K.; Yanchunas, J., Jr.; Liu, E. C.; Hartl, K. S.; Seiler, S. M.; Steinbacher, T. E.; Schumacher, W. A.; Atwal, K. S.; Stein, P. D. Cyanoguanidine-based lactam derivatives as a novel class of orally bioavailable factor Xa inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4034−4041.
(37) Desvergne, B.; Wahli, W. Peroxisome proliferator-activated receptors: nuclear control of metabolism. *Endocr. Rev.* **1999**, *20*, 649−688.
(38) Vázquez, M.; Silvestre, J. S.; Prous, J. R. Experimental approaches to study PPAR gamma agonists as antidiabetic drugs. *Methods Find. Exp. Clin. Pharmacol.* **2002**, *24*, 515−523.
(39) Krey, G.; Braissant, O.; L'Horset, F.; Kalkhoven, E.; Perroud, M.; Parker, M. G.; Wahli, W. Fatty acids, eicosanoids, and hypolipidemic agents identified as ligands of peroxisome proliferator-activated receptors by coactivator-dependent receptor ligand assay. *Mol. Endocrinol.* **1997**, 779−791.
(40) Straus, D. S.; Glass, C. K. Anti-inflammatory actions of PPAR ligands: new insights on cellular and molecular mechanisms. *Trends Immunol.* **2007**, *28*, 551−558.
(41) Togashi, M.; Borngraeber, S.; Sandler, B.; Fletterick, R. J.; Webb, P.; Baxter, J. D. Conformational adaptation of nuclear receptor ligand binding domains to agonists: potential for novel approaches to ligand design. *J. Steroid Biochem. Mol. Biol.* **2005**, *93*, 127−137.
(42) Schapira, M.; Raaka, B. M.; Samuels, H. H.; Abagyan, R. Rational discovery of novel nuclear hormone receptor antagonists. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1008−1013.
(43) Stauffer, S. R.; Coletta, C. J.; Tedesco, R.; Nishiguchi, G.; Carlson, K.; Sun, J.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. Pyrazole ligands: structure−affinity/activity relationships and estrogen receptor-alpha-selective agonists. *J. Med. Chem.* **2000**, *43*, 4934−4947.
(44) Meyers, M. J.; Sun, J.; Carlson, K. E.; Marriner, G. A.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A. Estrogen receptor-beta potency-selective ligands: structure−activity relationship studies of diarylpropionitriles and their acetylene and polar analogues. *J. Med. Chem.* **2001**, *44*, 4230−4251.
(45) Greschik, H.; Moras, D. Structure−activity relationship of nuclear receptor−ligand interactions. *Curr. Top. Med. Chem.* **2003**, *3*, 1573−1599.
(46) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(47) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(48) Irwin, J. J.; Shoichet, B. K. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−182.
(49) Sayers, E. W.; Barrett, T.; Benson, D. A.; Bolton, E.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; Dicuccio, M.; Federhen, S.; Feolo, M.; Geer, L. Y.; Helmberg, W.; Kapustin, Y.; Landsman, D.; Lipman, D. J.; Lu, Z.; Madden, T. L.; Madej, T.; Maglott, D. R.; Marchler-Bauer, A.; Miller, V.; Mizrachi, I.; Ostell, J.; Panchenko, A.; Pruitt, K. D.; Schuler, G. D.; Sequeira, E.; Sherry, S. T.; Shumway, M.; Sirotkin, K.; Slotta, D.; Souvorov, A.; Starchenko, G.; Tatusova, T. A.; Wagner, L.; Wang, Y.; John Wilbur, W.; Yaschenko, E.; Ye, J. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2010**, *38*, D5−16.
(50) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537−547.
(51) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: an open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.