Nuclear Hormone Receptor Targeted Virtual Screening

Matthieu Schapira,* Ruben Abagyan, and Maxim Totrov

Molsoft LLC, 3366 North Torrey Pines Court, Suite 300, La Jolla, California 92037

Received January 9, 2003

Virtual library screening (VLS) is emerging as a valuable drug lead discovery tool. ICM-VLS implementation of this technology was evaluated on a benchmark set of nuclear hormone receptors (NRs), an important therapeutic target family. Over 5000 structurally diverse compounds, including 78 known NR ligands, were screened against 18 crystal structures and one computer model of 10 NR ligand binding domains in their active or inactive states. The results confirm the ability of the VLS method to generate highly focused subsets of the input chemical library, enriched 33- to 100-fold for all but one receptor studied. However, receptor flexibility remains to be fully addressed, and the choice of the specific conformation used for screening may determine the success of the exercise. We observe that for a particular ligand VLS can often identify the correct target within the receptor family, although the technology is unable to reliably discriminate between the closely related receptor isoforms. Additionally, our results suggest that VLS may be applied successfully without an experimental structure of the receptor by using a homology model. These data represent a realistic snapshot of the state-of-the-art of NR-targeted VLS and define the recent progress and the remaining limitations of the technology.

Introduction

Identification of the ligands that bind a specific target receptor is of paramount importance in drug design and in biochemistry in general. The rapidly increasing availability of structural data prompted the development of methods for in silico modeling of ligand−receptor association (docking) and virtual screening of compound libraries for putative ligands based on the structure of the receptor.[1]

Structure-based virtual ligand screening (VLS) algorithms typically involve (1) a docking procedure that generates a hypothetic bound structure(s) for each ligand and (2) a scoring procedure that evaluates the docked ligand structures and assigns a score used to rank the ligands according to the quality of their fit to the receptor, ideally reflecting the binding affinity. A growing number of docking methods have been proposed (DOCK, FlexX, Gold, Autodock, and ICM-Dock;[3−8] see refs 2 and 9 for review). Scoring functions are also evolving rapidly (Boehm/FlexX, PMF, ICM score, to name a few;[10−12] see ref 13 for review).

The ICM VLS algorithm used in the present study applies internal coordinates mechanics (ICM) methodology to rapidly dock flexible ligands into a grid representation of the receptor. The best-energy docked conformation is subsequently evaluated using the scoring function optimized for discrimination of the active ligands.[12] The scoring function includes a solvation electrostatic free energy term calculated by the fast REBEL algorithm.[13]

Successful applications of VLS algorithms to discover novel ligands for various receptor proteins and DNA were reported.[14−17] Doman and colleagues also compared the hit rates in random experimental high-throughput screening and VLS-based approach on the same ligand library and found a 1700-fold enrichment rate for the target enzyme PTP1B.[16] Interestingly, "inverse virtual screening" was recently proposed as a tool to identify in the protein structure database the receptors that bind a specific compound/drug.[18]

In a number of recent studies, the performance of various VLS methods was evaluated on benchmark sets of receptor structures. Charifson et al. applied DOCK and GAMBLER in combination with 13 different scoring functions to three receptors.[19] Bissantz et al. used three different docking programs and seven scoring functions in their assessment of the performance of VLS methods for two receptors.[20] Stahl and Rarey docked a 10 000 compound subset of the World Drug Index to seven receptors using FlexX and analyzed the performance of four scoring functions.[21] These studies involved one to seven receptors of diverse nature. Such tests are important to establish applicability of VLS techniques to various target proteins. However, many receptors of biomedical significance form large families such as kinases, G-protein coupled receptors (GPCRs), and nuclear receptors (NRs). To confirm the utility of VLS applications involving receptors from such families, it is essential to perform tests on a sufficiently large set of representative proteins and to evaluate the ability of the method to correctly identify specific interaction partners within a group of related receptors and ligands.

Nuclear receptors form an important class of transcriptional regulators involved in various signaling mechanisms controlling cell proliferation and differentiation as well as homeostasis and are implicated in pathologic conditions such as cancer, inflammatory diseases, and diabetes.[22] These transcription factors are naturally switched on and off by small-molecule hormones presenting physicochemical properties very similar to therapeutic chemical entities. NRs are therefore

* To whom correspondence should be addressed. Phone: 1-858-625-2000. Fax: 1-858-625-2888. E-mail: matthieu@molsoft.com.

10.1021/jm0300173 CCC: $25.00
Published on Web 05/31/2003

intrinsically good therapeutic targets. Several NRs are already targeted by known drugs, such as tamoxifen, an estrogen receptor (ER) partial antagonist against breast cancer;[23] troglitazone, a peroxisome proliferator activated receptor (PPAR) γ agonist against type II diabetes;[24] bicalutamide, an androgen receptor (AR) antagonist against prostate cancer;[25] and bexarotene, a retinoid X receptor (RXR) agonist against cutaneous T-cell lymphoma.[26] Additionally, many other members of the family are considered as good potential targets, such as the liver X-activated receptor and the farnesoid X-activated receptor for diabetes[27] and the vitamin D receptor (VDR) for cancer therapy.[28]

The crystal structures of the ligand-binding domain (LBD) of several NRs, both in their active and inactive states, have been solved and resulted in a detailed model for the structural mechanism of activation and inhibition of members of the NR family.[29−34] Binding of an agonist stabilizes a conformation of the receptor where the C-terminal H12 helix folds like a lid onto the ligand, thereby generating a hydrophobic cavity at the surface of the receptor used for the recruitment of coactivator proteins. Antagonists bind to the same site as agonists but destabilize the coactivator-recruitment state by preventing H12 from folding onto the ligand-binding pocket.[30−32]

Here, we analyze the performance of the ICM virtual screening approach on a benchmark set of 19 structures of active and inactive forms of 10 nuclear receptors, 78 known ligands of these receptors, and a library of 5000 randomly selected compounds from a diverse compound database.

Results

Receptor Preparation. Crystal structures of the LBD of the androgen receptor, the estrogen receptor, the pregnane X receptor (PXR), the progesterone receptor (PR), the peroxisome proliferator activated receptor, the retinoid X receptor, the retinoic acid receptor (RAR), the thyroid hormone receptor (TR), the vitamin D receptor, and a computer model of the glucocorticoid receptor (GR) were used to screen a benchmark library of 78 known NR ligands (Table 1). Structures of both active and inactive conformations of ER were tested, while other receptors were in the active state only. Different receptor isotypes were tested for PPAR (PPARα, PPARγ, PPARδ) to evaluate the ability of virtual screening to discriminate between isotype-specific ligands. When several crystal structures were available for a specific receptor, the highest resolution data were preferred. In the case of AR, active ERα, PR, PPARγ, and RXRα, two different crystal structures presenting alternative side chain conformations in the vicinity of the ligand binding pocket (LBP) were used for screening in order to measure to what extent variations of the binding pocket structure influence VLS efficiency.

In all cases but for GR, hydrogens were added to the crystal structure, and the system was converted into the internal coordinate representation according to the ICM method.[35] While the backbone and side chain conformations were left unchanged from the crystal structures, it was found in most cases that a limited number of polar atoms (mostly side-chain histidine nitrogens and serine oxygens) lining the LBP could act either as a hydrogen bond donor or as a hydrogen bond acceptor. The preferred orientation of the group was usually strongly suggested by the nature of neighboring atoms of the ligand in the crystallized complex. However, the favorable bias for the known ligands had to be avoided in order to reproduce a more realistic exercise where structure-based virtual screening is expected both to retrieve known binders and to identify active molecules presenting novel chemotypes. To reduce this bias, receptors with alternative tautomeric states or rotameric states of polar hydrogens along the LBP were systematically generated when possible (with the exception of PXR, which has too many such groups) and screened in parallel (see "Experimental Section" for details).

While it is clear that availability of the receptor crystal structure is preferable for VLS applications, the use of computer models of the receptor would open up a vast field of interesting targets to the technology that are otherwise only amenable to high-throughput screening (HTS). To test this option, we generated a theoretical model of the GR-LBD that was used for virtual screening. The model was built with ICM by homology to PR, the closest crystallized homologue (see Experimental Section).

Benchmark Library Preparation. A library of 78 known NR ligands was generated from the literature (Table 1). The molecules are agonists (such as dihydrotestosterone and calcipotriol) or antagonists (tamoxifen, mifepristone, BMS614), natural (cortisol) or synthetic (dexamethasone), isotype-specific (GW501516, THC) or universal (all-trans retinoic acid, triiodothyronine). Most ligands target at least one of the structures screened, but a few do not (five AR antagonists, three PR antagonists, and four RAR antagonists). For each receptor screened, between three and eight known ligands are present in the database. The initial 3D coordinates of the ligands were either extracted from a crystal structure or generated from their smiles string with ICM. The free molecule was then subjected to global energy optimization in the internal coordinate space by the ICM method (ICM2.8 manual), and the compounds were stored in the structure library.

Virtual Screening. The benchmark library was screened against each of the prepared receptor structures according to the ICM virtual screening algorithm:[8,13,15,35] rapidly, grid potentials were generated that account for the LBP shape, hydrophobicity, electrostatic potential, and hydrogen-bond profile, and each continuously flexible ligand was docked by a Monte Carlo simulation with grid receptor potentials in the internal coordinates space. Docking took an average of 1 min per molecule, depending on binding pocket size and the number of torsional degrees of freedom of the ligand.

We evaluated the precision of the docking procedure for the complexes with known X-ray structures. Results were encouraging, with 11 out of 16 structures reproduced to less than 1 Å heavy atom root-mean-square deviation (rmsd) and only one structure with rmsd worse than 2 Å (Table 4). Relatively poor rmsd in this case may be related to the quality of the X-ray structure (resolution 2.76 Å).

Table 1. NR Ligand Benchmark Database Composed of 78 Compounds of Agonists (+) and Antagonists (−) Taken from the Literature[a]

	ligands	activity		ligands	activity		ligands	activity
1	LG121071	AR+	27	GW0072	PPARg+	53	BMS270394	RARg+
2	dihydrotestosterone	AR+, ER+	28	GW501516	PPARd+	54	BMS961	RARg+
3	metribolone	AR+, PR+	29	GW7845	PPARg+	55	BMS181156	RARabg+
4	MPA	AR−	30	GI262570	PPARg+	56	BMS184394	RARg+
5	bicalutamide	AR−	31	L165041	PPARd+	57	CD564	RARbg+
6	flutamide	AR−	32	L783483	PPARg+	58	AGN193109	RAR−
7	**21b**	AR−	33	L796449	PPARg+	59	BMS614	RAR−
8	nilutamide	AR−	34	rosiglitazone	PPARg+	60	MX781	RAR−
9	coumestrol	ER+	35	troglitazone	PPARg+, PXR+	61	Ro415253	RAR−
10	DES	ER+	36	CP8481	PR+	62	ER35794	RXR+
11	genistein	ER+	37	promegestone	PR+	63	SR11237	RXR+
12	moxestrol	ER+	38	progesterone	PR+	64	**5**	RXR+
13	estradiol	ER+	39	trimegestone	PR+	65	**3**	RXR+
14	THC	ER+, ER−	40	mifepristone	PR−, PXR+	66	targretin	RXR+
15	EM343	ER−	41	**5g**	PR−	67	**5**	RXR+
16	ICI164384	ER−	42	onapristone	PR−	68	9cis-RA	RXR+, RAR+
17	LY326315	ER−	43	**12**	PR−	69	GC1	TR+
18	LY353381	ER−	44	SR12813	PXR+	70	T3	TR+
19	hydroxytamoxifen	ER−	45	clotrimazole	PXR+	71	TRIAC	TR+
20	nafoxidine	ER−	46	lovastatin	PXR+	72	19norD	VDR+
21	raloxifene	ER−	47	phenobarbital	PXR+	73	KH1060	VDR+
22	trioxifene	ER−	48	pregnenolone	PXR+	74	KS291	VDR+
23	cortisol	GR+	49	transnonachlor	PXR+	75	LG190178	VDR+
24	dexamethasone	GR+	50	all-trans RA	RARabg+	76	MC1288	VDR+
25	triamcinolone	GR+	51	Am580	RARa+	77	vitamin D3	VDR+
26	AZ242	PPARag+	52	BMS187949	RARabg+	78	calcipotriol	VDR+

[a] References for the entries (in brackets) are the following. [1] Hamann et al. *J. Med. Chem.* **1999**, *42*, 210−212. [2] Maggiolini et al. *Cancer Res.* **1999**, *59*, 4864−4869. [6, 10, 11, 13, 19, 21, 23−25, 38, 40, 50, 68, 77] *Goodman & Gilman's the Pharmacological Basis of Therapeutics*, 9th ed.; Hardman, J. G., Limbird, L. E., Eds.; McGraw-Hill, Health Professions Division: New York, 1996. [3] Matias et al. *J. Biol. Chem.* **2000**, *275*, 26164−26171. [4] Labrie et al. *J. Steroid Biochem.* *28*, 379−384. [5] McLeod *Cancer* **1993**, *71* (3, Suppl.), 1046−1049. [7] Hamann et al. *J. Med. Chem.* **1998**, *41*, 623−639. [8] Dole et al. *Ann. Pharmacother.* **1997**, *31*, 65−75. [9] Nelson et al. *Biochemistry* **1984**, *23*, 2565−2572. [12] Salmon et al. *J. Steroid Biochem.* **1983**, *18*, 565−573. [14] Shiau et al. *Nat. Struct. Biol.* **2002**, *9*, 359−364. [15] Luo. *Int. J. Cancer* **1997**, *73*, 735−759. [16] Poulin et al. *Breast Cancer Res. Treat.* **1989**, *14*, 65−76. [17] Mitchell et al. *Abstracts of Papers*, 219th National Meeting of the American Chemical Society, San Francisco, CA, 2000; American Chemical Society: Washington, DC, 2000; Paper 334. [18] Sato et al. *J. Pharmacol. Exp. Ther.* **1998**, *287*, 1−7. [20] Tagnon. *Cancer* **1977**, *39* (6, Suppl.), 2959−2964. [22] Arafah. *Eur. J. Cancer* **1980**, Suppl. 1, 281−285. [26] Cronet et al. *Structure* **2001**, *9*, 699−706. [27] Oberfield et al. *Proc. Natl. Acad. Sci U.S.A.* **1999**, *96*, 6102−6106. [28] Olivier et al. *Proc. Natl. Acad. Sci. U.S.A* **2001**, *98*, 5306−5311. [29] Suh et al. *Cancer Res.* **1999**, *59*, 5671−5673. [30] Gampe et al. *Mol Cell.* **2000**, *3* (March 5), 545−555. [31−33] Berger et al. *J. Biol. Chem.* **1999**, *274*, 6718−6725. [34, 35] Mudaliar et al. *Annu. Rev. Med.* **2001**, *52*, 239−257. Willson et al. *Nat. Rev. Drug Discovery* **2002**, *1*, 259−266. [36] Tabata et al. *Eur. J. Pharmacol.* **2001**, *430*, 159−165. [37] Bashirelahi et al. *J. Steroid Biochem.* **1986**, *3* (Sep 25), 367−374. [39] Wahab et al. *Expert Opin. Invest. Drugs* **2001**, *10* (9), 1737−1744. [41] Hamann et al. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2731−2736. [42] Wiechert et al. *J. Steroid Biochem.* **1987**, *27*, 851−858. [43] Zhi et al. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 415−418. [44−49] Moore et al. *J. Biol. Chem.* **2000**, *275*, 15122−15127. [51] Kagechika et al. *J. Med. Chem.* **1988**, *31*, 2182−2192. [52, 53] Klaholz et al. *Proc. Natl. Acad. Sci. U.S.A* **2000**, *97*, 6322−6327. [54] Klaholz et al. *Nat. Struct. Biol.* **1998**, *5*, 199−202. [55−57] Klaholz et al. *J. Mol. Biol.* **2000**, *302*, 155−170. [58] Sun et al. *Cancer Res.* **1997**, *57*, 4931−4939. [59] Bourguet et al. *Trends Pharmacol. Sci.* **2000**, *21*, 381−388. [60] Fanjul et al. *Cancer Res.* **1998**, *58*, 4607−4610. [61] Marth et al. *J. Steroid Biochem. Mol. Biol.* **1993**, *47*, 123−126. [62] Hibi et al. *J. Med. Chem.* **1998**, *41*, 3245−3252. [63] Lehmann et al. *Science* **1992**, *258*, 1944−1946. [64] Beard et al. *J. Med. Chem.* **1996**, *39*, 3556−3563. [65] Cannan-Koch et al. *J. Med. Chem.* **1999**, *42*, 742−750. [66] Boehh et al. *J. Med. Chem.* **1994**, *37*, 2930−2941. [67] Vuligonda et al. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 589−594. [69−71] Wagner et al. *Mol. Endocrinol.* **2001**, *15*, 398−410. [72, 78] Brown. *Am. J. Kidney Dis.* **1998**, *32* (2, Suppl. 2), S25−S39. [73, 76] Tocchini-Valentini et al. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5491−5496. [74] Verstuyf et al. *J. Bone Miner. Res.* **1998**, *13*, 549−558. [75] Boehm et al. *Chem. Biol.* **1999**, *6*, 265−275.

Once a docked conformation was found, the ICM scoring function that includes steric, hydrogen bonding, hydrophobicity, continuum electrostatics, and entropy terms (see Methods) was used to assign a score to the ligand reflecting the quality of the complex. The library was screened twice independently for each receptor structure used, and the best score was kept for each ligand (two additional screenings did not change the results). When alternative conformations were derived from a single-crystal structure, the best score for each ligand over all receptor conformations was retained.

A random library of 5000 compounds was screened following exactly the same protocol against each receptor. This library was generated by extracting every 34th molecule of the Chemdiv CombiLab database of 174 467 compounds (Chemical Diversity Inc., San Diego, CA). This source database was designed to represent wide chemical diversity, which is reflected in our sample library. The diversity and the overlap between the distributions of the basic properties for the benchmark and the random background library are essential for the meaningful evaluation of the VLS algorithm. Four properties were calculated: molecular weight (MW), number of flexible torsions (N_{flex}), and the numbers of hydrogen bond donors (N_{HBdo}) and acceptors (N_{HBacc}). The averages and standard deviations of these parameters for the random library and benchmark library were as follows: MW, 344.0 ± 75.0 and 393.0 ± 85.0; N_{flex}, 5.4 ± 2.2 and 4.3 ± 3.9; N_{HBdo}, 1.1 ± 1.0 and 1.2 ± 1.1; N_{HBacc}, 3.6 ± 1.5 and 3.5 ± 1.5 (see also histograms in the Supporting Information).

For each of the 18 receptor crystal structures used as well as for our model of GR, the best scores for the 78 compounds of the benchmark library are shown as histograms along with the score thresholds necessary to select 1% and 10% of the 5000 random molecules (Figure 1). In a real life exercise used in lead discovery programs, VLS would be applied to extract from large

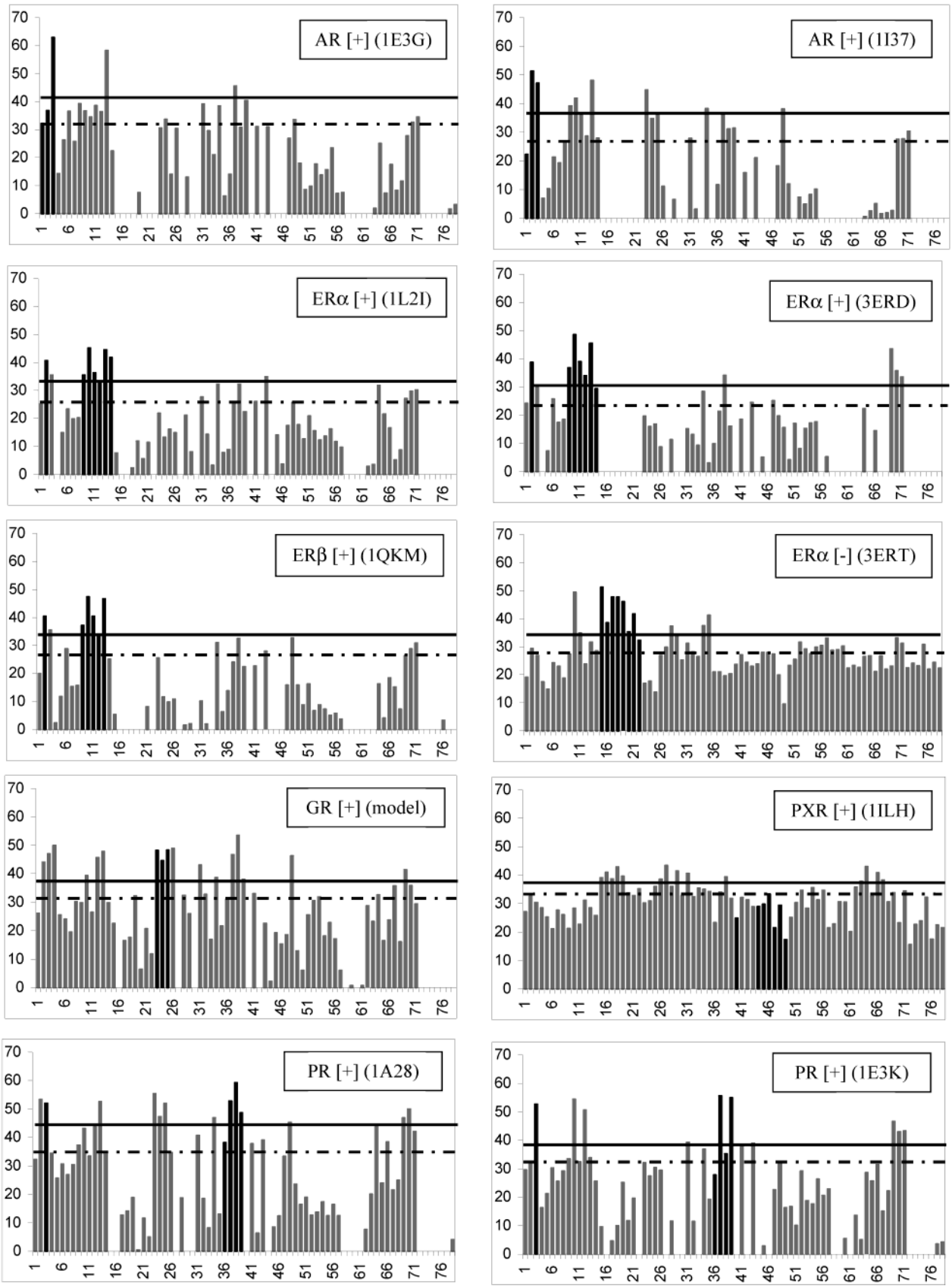

Figure 1 (continued on next page).

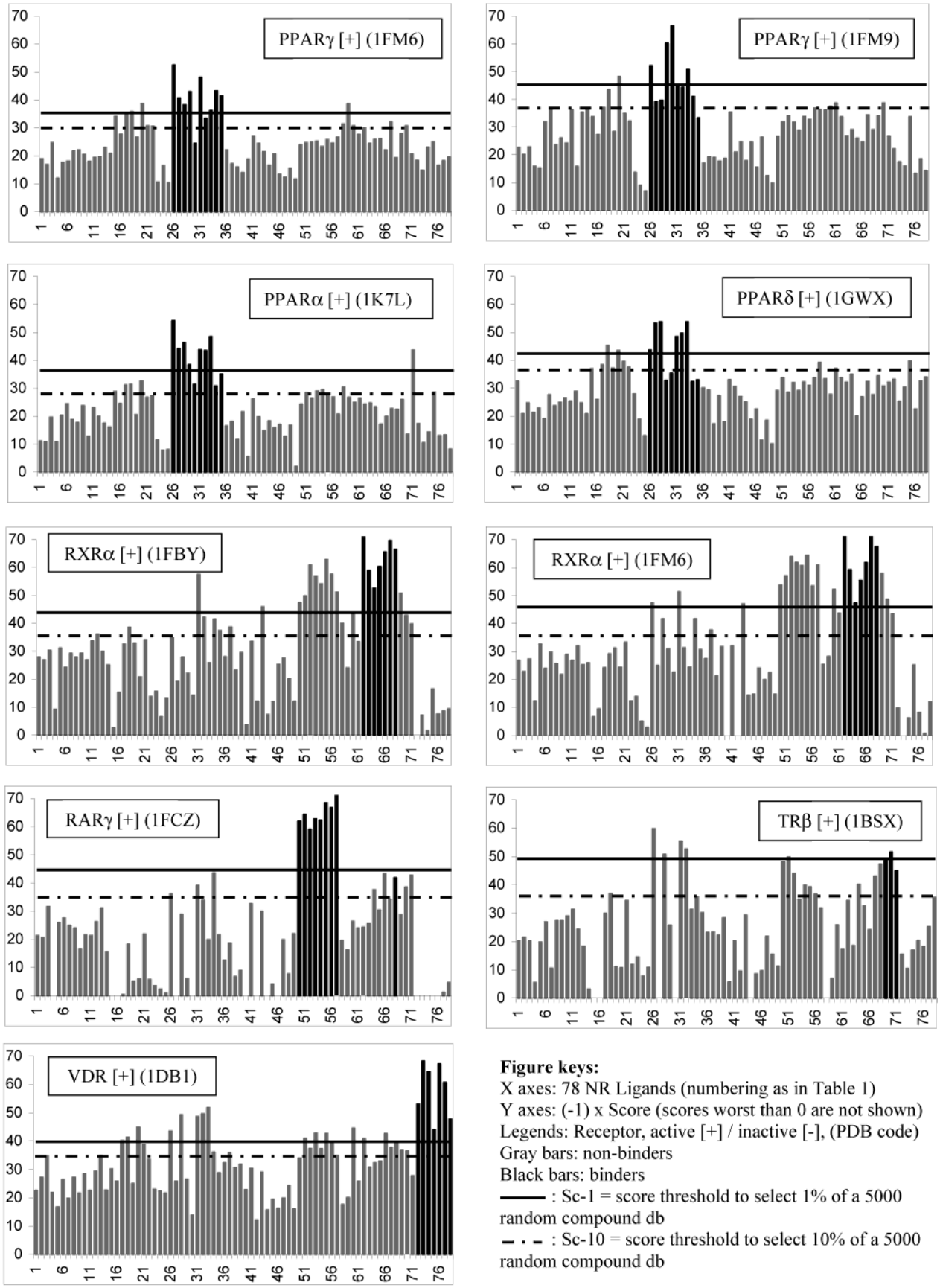

Figure 1. Virtual screening result. For each receptor conformation tested, the score of known binders (black) and presumed nonbinders (gray) from the benchmark library (Table 1) is shown. Scores were multiplied by −1 (best scoring compounds have the highest score), and scores worst than 0 are not displayed. The score thresholds necessary to select 1% (Sc-1) and 10% (Sc-10) of a diverse database of 5000 random compounds are also shown (see figure keys). The crystal structure 1FM6 is a heterodimer of PPARγ and RXRα and was used separately for both receptors.

Table 2. Screening Efficiency[a]

	score cutoff (top 1%)		score cutoff (top 10%)		enrichment (top 1%)		enrichment (top 10%)	
receptor	full scoring	steric scoring	full scoring	steric scoring	full scoring	steric scoring	full scoring	steric scoring
AR [+] (1E3G)	−41.	−43.	−32.	−39.	33	33	10	7
AR [+] (1I37)	−37.	−43.	−27.	−38.	67	50	7	5
ERa [+] (1L2I)	−33.	−46.	−27.	−42.	71	0	10	3
ERa [+] (3ERD)	−32.	−44.	−26.	−39.	71	14	10	6
ERb [+] (1QKM)	−34.	−44.	−28.	−39.	57	14	9	6
ERa [-] (3ERT)	−34.	−52.	−29.	−46.	87	87	10	10
GR [+] (model)	−38.	−51.	−31.	−46.	100	100	10	10
PXR [+] (1ILH)	−38.	−57.	−33.	−50.	0	14	2	4
PR [+] (1A28)	−44.	−51.	−35.	−46.	83	20	10	8
PR [+] (1E3K)	−39.	−50.	−32.	−45.	50	60	8	6
PPARg [+] (1FM6)	−36.	−54	−30.	−48.	80	10	9	3
PPARg [+] (1FM9)	−46.	−58.	−37.	−52.	40	30	9	5
PPARa [+] (1K7L)	−37.	−54.	−29.	−48.	70	20	10	7
PPARd [+] (1GWX)	−42.	−59.	−37.	−53.	60	20	6	4
RXRa [+] (1FBY)	−44.	−54.	−36.	−48.	100	71	10	10
RXRa [+] (1FM6)	−47.	−53.	−36.	−48.	100	29	10	10
RARg [+] (1FCZ)	−45.	−53.	−35.	−48.	88	89	10	9
TRb [+] (1BSX)	−50.	−55.	−36.	−49.	33	0	10	0
VDR [+] (1DB1)	−40.	−59.	−34.	−52.	100	71	10	10
average					68	39	9	6

[a] For each receptor conformation, the scores necessary to select 1% and 10% of a diverse database of 5000 compounds (score cutoff, also displayed as horizontal lines Figure 1) and enrichment factors of the corresponding focused libraries are shown. Results obtained with the full ICM scoring function and results obtained with the steric component (steric scoring) only of the scoring function are presented.

and diverse source virtual compound databases, smaller focused libraries. Thus, the goal is to retain a high quantity of hits while significantly decreasing the size of the database subset that needs to be actually tested in vitro. The corresponding hit enrichment factor (EF) of the focused library can be calculated as

$$EF = \frac{\text{hits}_{FL}}{\text{hits}_{SL}} \times \frac{\text{compounds}_{SL}}{\text{compounds}_{FL}}$$

where SL is the source library and FL is the focused library.

The enrichment factor of the NR-focused libraries (excluding libraries focused on PXR) varies from 33 to 100 times for the top 1% selections and from 6 to 10 times for the top 10% selections (Table 2). This illustrates significant differences in the VLS efficiency from one receptor to another, which were further analyzed on a case-by-case basis.

AR. Two crystal structures were used for AR screening: the 2.4 Å resolution structure of human AR-LBD complexed with the agonist metribolone (PDB code 1E3G)[36] and the 2.0 Å resolution structure of rat AR-LBD complexed with dihydrotestosterone (PDB code 1I37).[37] Throughout the text, PDB codes will be used to refer to specific structures of the receptor. Three known AR ligands were included in the benchmark database: dihydrotestosterone, metribolone, and LG121071 (references in Table 1). Not surprisingly, metribolone and dihydrotestosterone satisfied the top 1% score threshold against the structure of the receptor crystallized in their presence (1E3G and 1I37, respectively). Metribolone also passed the top 1% threshold against 1I37, while dihydrotestosterone was only in the top 10% selection of 1E3G due to the conformation of M745 in this complex, which projects further toward the binding pocket and sterically hinders correct positioning of the additional methyl group present at position 19 on the A ring of this compound. LG121071 did not satisfy the top 10% threshold condition against 1I37 and hardly met the threshold against 1E3G, which suggests that neither receptor conformation can accommodate this non-steroid ligand and which results in suboptimal enrichment of AR-based selections (Table 2).

Interestingly, a number of nonandrogenic NR ligands do score well against one or both structures (Figure 1). These are mostly estrogenic compounds such as estradiol, coumestrol, and DES; progestagens such as promegestone and trimegestone; and glucocorticoids such as cortisol. Such false positives do not constitute irrelevant background noise but actually reflect the fact that AR selectivity for androgenic steroids is very fragile. Indeed, it was shown in the past that a single residue mutation T877A in AR-LBP results in activation of the receptor by estrogen and progesterone receptor agonists.[38] Similarly, cortisol can activate the L701H AR mutant.[38] These illicit activations of mutated forms of AR can be associated with failed androgen ablation therapy in prostate cancer patients.[36] The nonperfect selectivity of the VLS algorithm, which is not as stringent as the biological selectivity of the receptor, fortuitously revealed the potential promiscuity of the AR-LBP for a diverse array of steroids.

ER. The crystal structures of ERα-LBD bound to the agonists (*R*,*R*)-5,11-*cis*-diethyl-5,6,11,12-tetrahydrocrysene-2,8-diol (THC) and diethylstilbestrol (DES) and to antagonist 4-hydroxytamoxifen (TAM) (PDB codes 1L2I, 3ERD, and 3ERT, respectively), as well as ER*β*-LBD complexed to the partial agonist genistein (PDB code 1QKM), were used.[40,41,33] Five ER agonists and eight ER antagonists were present in the benchmark library (Table 1). The ERα agonist/ER*β* antagonist THC was also included, and dihydrotestosterone was considered as a weak ER agonist.[42] On average, ER ligands scored better than other steroid receptor ligands against ER (Figure 1). Not surprisingly, ER agonists docked also reasonably well to the inactive form of the receptor (3ERT), since antagonists bind to the same LBP as agonists but generally present an additional bulky group that protrudes out of the agonist binding pocket.[41] Most ER*β* agonists also docked well to the structure of the LBD crystallized in the presence of the partial

antagonist genistein, which was expected, considering the structural similarity of the LBP with the active form of the receptor.[33]

Moxestrol scored rather poorly with all forms of the receptor compared to other ER agonists. Inspection of the docked complexes showed that the ethyne group of moxestrol constitutes a rather bulky hydrophobic entity that lightly clashes with M421 of ERα in 1L2I and 3ERD and with M295 of ERβ in 1QKM. Conformational rearrangement of some receptor side chains may be necessary to accommodate this agonist.

A few false positives scoring better than the top 1% threshold were present, most of which were ligands for other steroid hormone receptors such as metribolone and progesterone. More surprising was the top 1% selection of all three TR agonists against the 3ERD structure of ERα, where a hydrophobic cavity between M343, L346, and T347, too small in 1L2I or 1QKM, can accommodate the methyl group of the disubstituted phenolic ring of GC1 (compound **69**, Figure 1) or the corresponding T3 iodine. Whether this unexpected result simply represents an artifact of the docking simulations or points at possible cross-reactivity between thyroid hormone and estrogen receptors, as suggested elsewhere, remains to be tested.[43,44]

The 87-fold enrichment in ER antagonists of the top 1% selection against inactive ERα (Table 2) reflects the good feasibility of the VLS technology to identify ER antagonists. False-positive NR ligands are of two kinds (Figure 1). (1) ER agonists fit in the antagonist binding pocket but, as expected, do not present a bulky group that would destabilize the active state of helix H12.[30] A simple computerized filter could easily be set up that rejects docked compounds that are more than 2 Å away from active H12. (2) Three PPARγ agonists (rosiglitazone, troglitazone, and GW501516) also score better than the top 1% threshold (Figure 1). Inspection of the docked complexes reveals that these compounds actually occupy only half of the agonist binding cavity and mostly interact with residues that are out of the LBP. Such compounds could be avoided either by reducing the receptor domain available for docking during VLS or by automatically filtering out docked ligands that occupy only part of the agonist binding cavity. Such an exercise remains beyond the scope of this work, which is to compare the efficiency of one single VLS algorithm when applied to different receptors rather than customizing the algorithm or procedure to each receptor.

GR. A computer model of the GR-LBD built by homology to the crystal structure of PR was used for virtual screening (see above). Three known GR agonists were included in the benchmark library: cortisol, dexamethasone, and triamcinolone (Table 1). The observation that all three ligands are in the top 1% selection of a random database (Figure 1 and Table 2) strongly suggests that in some cases VLS can be applied successfully even when an experimental receptor structure is not available. A rather large number of false positives can be seen in Figure 1, which seems surprising considering the enrichment level observed (Table 2). This could be an expected outcome because a homology model was used. Interestingly, almost all false positives are ligands for AR, ER, or PR, all steroid hormone receptors structurally closer to GR than most enzyme active sites or receptor binding pockets. This underscores the fact that our 78-compound benchmark library is structurally biased toward NRs, and ligand discrimination within this library is more difficult than for random compound databases.

PXR. PXR, a xenobiotic receptor that was recently crystallized, is by nature able to bind to a variety of ligands presenting extreme structural diversity.[45] This promiscuity is based on a large spherical hydrophobic LBP lined by a discrete series of polar atoms that can be combined into a diverse array of hydrogen bond networks. Additionally, a flexible loop of the LBP can expand and contract, thereby accommodating ligands of different sizes and shapes.[45] The absence of salient features of the LBP, both in terms of shape and electrostatic potential, is extremely unfavorable to the fast and accurate docking required for VLS and translates into a lack of enrichment of focused libraries (Table 2). This clear result strongly suggests that high-throughput docking to PXR would not be an efficient alternative to binding or cell-based assays for rapid prediction of drug−drug interactions.[46]

PR. The crystal structures of human PR-LBD complexed to progesterone (PDB code 1A28) and metribolone (PDB code 1E3K) were used.[47,36] Five known PR agonists were present in the benchmark library: progesterone, metribolone, promegestone, trimegestone, and CP8481 (Table 1). A first observation is that both receptor structures yield to significant enrichment of focused libraries (Table 2), which indicates that VLS can efficiently identify PR agonists. The slightly higher enrichment of selections derived from 1A28 may be explained by the higher resolution of this structure (1.8 versus 2.8 Å), which translates into more accurate docking and more relevant scoring. Additionally, M759 adopts a more open conformation in 1A28, which allows placement of a methyl group that is present at position 19 of progesterone but absent from metribolone. The alternative conformation of M759 in 1E3K cannot accommodate this group or another methyl of CP8481, which explains the relatively poor score assigned to these compounds.

A number of false positives from the benchmark library satisfy the top 1% selection threshold (Figure 1). These false positives are mostly AR, ER, and GR agonists. The LBP of these steroid hormone receptors is structurally close to that of PR, and the promiscuity of these receptors is illustrated by the existence of ligands such as metribolone, which binds and activates both AR and PR[36] or cortisol, a GR agonist that also weakly binds PR.[48] This reflects the current limitations of the state-of-the-art in VLS at discriminating between very similar binding sites.

As observed above with one of the ERα structures, all three TR ligands present in the library are selected as PR agonist candidates (Figure 1, compounds **69**−**71**). The compounds dock reasonably and make both hydrophobic and electrostatic interactions; however, they seem to be lightly clashing with T894, which would destabilize the complex. The ICM VLS algorithm is designed to avoid penalizing strongly the limited steric clashes, based on the assumption that receptor flexibility may accommodate them. While this feature is useful to account for moderate side chain flexibility, it

can also result in the selection of false positives. More realistic implementation of receptor flexibility is a major, complex challenge of VLS and the subject of much effort to improve the technology.[13]

PPAR. Discriminating between ligands specific for different receptor isotypes seems so far to remain beyond the limits of virtual screening technology, especially when the LBPs of the receptor isoforms are almost identical.[13] To formally address this question, both nonspecific and isotype-specific PPAR agonists where included in our benchmark library and screened against the crystal structures of active PPARγ bound to rosiglitazone or GI262570 (PDB codes 1FM6 and 1FM9), PPARα complexed to GW409544 (PDB code 1K7L), and PPARδ bound to eicosapentanoic acid (PDB code 1GWX).[49−51]

A first observation is that PPAR focused libraries are significantly enriched in PPAR ligands: 80% of all PPAR ligands are in the top 1% VLS selection against the 1FM6 conformation of PPARγ, and 70% are in the top 1% VLS selection against PPARα (Table 2). The enrichment level decreases to 60% against PPARδ and 40% against the 1FM9 conformation of PPARγ. It is important to note here that PPAR ligands score equally well against receptor conformations yielding highly enriched or poor selections (Figure 1). The significant difference is rather the score threshold necessary for them to be selected. For instance, 1% of the 5000 random compounds scored better than −46 in the 1FM9 screen, while the threshold was only −36 against 1FM6. The much more open structure of the 1FM9 LBP underlies this difference: the conformation of F363 and F282 in 1FM9 creates an open hydrophobic channel, absent in 1FM6, that can accommodate a benzene ring of GI262570. As a result, numerous ligands that do not fit into the 1FM6 conformation dock well to and score well against 1FM9. GI262570 is one such ligand (Figure 1). Other compounds from the diverse library and not identified as PPAR ligands behave similarly. A corollary is that the actual enrichment of the 1FM9 selections may be higher than shown here because most true positives from our benchmark library as well as most known PPAR agonists are absent from the chemistry space represented by GI262570.

The score assigned to PPARγ specific ligands (compounds **27**, **29**, **30**, **32**−**35**), PPAR$\alpha\gamma$ dual agonists (compound **26**), and PPARδ selective ligands (compounds **28**, **31**) can be compared to address how VLS scoring correlates with receptor isotype selectivity. Figure 1 shows that the two PPARδ selective agonists are included in the top 1% selection focused on PPARα, PPARδ, and the 1FM6 conformation of PPARγ but are excluded from the 1FM9 PPARγ focused library. The PPAR$\alpha\gamma$ specific ligand passes the top 1% threshold against all PPAR structures. If only the top 0.5% top scoring compounds were retained, it would be included in the three PPARα/PPARγ selections while excluded from the PPARδ pool. At last, out of seven PPARγ selective agonists, five, four, and three are present in the top 1% PPARγ (1FM6), PPARα, and PPARδ focused libraries, respectively. These results show clearly that isotype selectivity is poorly represented by VLS exercises. This is particularly true when specificity relies on few residues, as it is the case here. For instance, structural and biochemical studies have shown that receptor selectivity between PPARα and PPARγ could be reverted by a single residue mutation.[50] This leaves room for the improvement of the virtual screening technology.

Among the few false-positive NR ligands, nafoxidine (compound **20**), an ER antagonist that has been used in advanced breast cancer,[52] passed the top 1% score threshold against both PPARγ conformations as well as PPARδ (Figure 1). This ligand also scored well against PPARα, as did the TR agonist TRIAC (compound **71**). Whether these results represent artifacts of the VLS approach or actually reflect PPAR agonist activities remains to be tested.

RXR. Two crystal structures of human RXRα bound to 9-*cis*-retinoic acid were used for screening (PDB codes 1FBY and 1FM6).[49,53] Seven RXR ligands were included in the benchmark library (Table 1). The different RXR isotype LBPs are almost identical, and all ligands are pan-agonists. Our results suggest that VLS can very efficiently enrich the RXR agonists random compound libraries; all known hits present in a source library would be selected if 99% of the library was filtered-out on the basis of VLS scoring (Table 1 and Figure 1). This result was repeated against both receptor conformations, which differ principally at the level of P264, V265, I324, F439, and L451, the LBP being slightly more compact in the 1FBY conformation.

Despite the observed efficiency of the RXR-based screening reflected by the enrichment of focused libraries (Table 2), the fact that all RAR agonists present in the benchmark library were also present in the RXR selections underscores the limitations of the in silico approach (Figure 1). The RAR LBP is only slightly more elongated than RXR but otherwise very similar in shape, and both display arginines on one end that interact with a carboxylate moiety present in all known retinoid receptor ligands. As mentioned above, the VLS algorithm used here is rather tolerant for light steric clashes, such as those present in most RAR ligand/RXR complexes (data not shown), to account for potential local rearrangement of side chain conformation. Once again, an improved representation of receptor flexibility would probably result in increased VLS efficiency. Alternatively, a more immediate approach that has proven successful in the past implements a second, more stringent filter with flexible receptor side chains against the limited number of compounds that have passed the VLS filter.[15]

The only other hits observed in both crystal structures are the PPARδ agonist L165041 and TR agonist GC1 (Figure 1). Inspection of the complex structures show that too short a distance between the positively charged side chain of R316 and the negatively charged carboxylate of L165041 compensates for the strained internal torsion angles of the PPAR ligand. This strongly suggests that this hit is a false positive. On the other hand, the RXR/GC1 docked complexes look reasonable, and in vitro experiments would be necessary to reject this unexpected hit.

RAR. The crystal structure of human RARγ complexed to the pan-agonist BMS181156 (PDB code 1FCZ) was used.[54] Of the nine RAR agonists included in the benchmark library, 89% satisfied the top 1% selection

threshold derived from the sample database of 5000 compounds. With a score slightly under the threshold, the natural agonist 9-*cis*-retinoic (9cis-RA) acid was the only binder filtered out of the selection (Figure 1). A conformational rearrangement of some receptor side chains may be necessary to properly dock this ligand. Along with 9cis-RA, all RXR agonists are excluded from the top 1% selection. This result could seem surprising considering the similarity between RAR and RXR LBPs and the observation made above that all RAR agonists satisfied the corresponding threshold in the RXR screening. Inspection of the docked complexes shows that most RXR ligands are actually not as elongated as RAR ligands (an observation already made in the past),[55] and while the carboxylate of RAR agonists interacts favorably with R278, the carboxylate of RXR agonists remains too distant from this residue. The score assigned by VLS to all RAR agonists except 9cis-RA is actually significantly better than that assigned to all other NR ligands from our benchmark library; they all score better than −50, which is the threshold necessary to retain only the best 10 hits (or 0.2%) of the 5000 random compound database. The corresponding enrichment value would be (8/10)/(9/5000) = 444 times.

These results are a good indication that VLS can efficiently accelerate the discovery of RAR agonist leads; however, the work presented here is only a test-case study and does not guarantee that selecting the top 0.2% of a large and diverse source library based on RAR VLS scoring would result in a focused library enriched over 400 times in RAR agonists. In this regard, two points should not be overlooked. First, Am580, an RARα selective agonist, was assigned a very good score (compound **51**, Figure 1) even though RARγ was the receptor isotype used for screening. Once again, this underscores the limitation of the method at discriminating between isotype specific ligands. Second, the structural diversity of known RAR agonists is relatively limited,[56] which could explain why all dock well to the same receptor conformation. Whether this receptor structure can allow identification of diverse RAR agonists presenting novel chemotypes remains an open question.

TR. The only TR-LBD crystal structure available is the 3.7 Å resolution structure of human TRβ bound to the natural hormone T3 (PDB code 1BSX).[57] As a rule of thumb, data worst than 2.5 Å resolution should not be considered reliable for VLS applications, and the poor resolution of this structure is probably in part responsible for the low screening efficiency observed (Table 2); the enrichment level of the top 1% selection is the second worst after PXR. The three known TR agonists screened ranked 3, 6, and 10 of the NR ligand benchmark library, while four PPAR agonists ranked in the top 10, including the best three compounds, as well as three RAR agonists (Figure 1).

The poor quality of the receptor structure is probably not the only factor responsible for the low enrichment of the top 1% selection. Indeed, the corresponding score threshold (−50) is the lowest observed for all receptors tested (Table 2). A cluster of three arginines (R99, R103, and R282) at one extremity of the TR LBP generates an electrostatic potential extremely favorable to carboxylate or nitro groups, which are rather common in compound libraries. Any molecule carrying such functional moiety and presenting structural features compatible with the TR LBP is likely to be assigned a good score.

This observation, however, does not explain why three PPAR agonists score better than all TR ligands. Three reasons can be advanced. (1) The receptor side chain conformation may not be optimal because of the above-mentioned low resolution of the crystal structure. (2) The torsion angles of several top scoring PPAR ligands, such as L165041 and GW501516, are very unfavorable, while their carboxylates lightly clash with R282, R103, and/or A234. The corresponding energetic penalty is compensated by too short a distance between positively charged R282 and/or R103 side chain amines and negatively charged carboxylates of the ligand, which translates into overly stabilizing electrostatic interactions. The same is true for RAR ligands. (3) The possibility that some PPAR ligands may bind to the TR-LBD should not be overlooked. This may reflect promiscuity between the PPAR and TR binding pockets. It is not impossible that a higher resolution structure or flexibility of the receptor side chains would allow a more reasonable, low-energy structure of the PPAR agonists/TR complex. The observation made above that TRIAC, a TR agonist, was selected among the top PPARα ligand candidates seems to reinforce this hypothesis (Figure 1). Whether such promiscuity would actually have a biological meaning (PPARs are therapeutic targets against diabetes and atherosclerosis, while TR activation has hypocholesterolemic effects)[58,59] remains beyond the scope of this work.

VDR. The high-resolution crystal structure of the human VDR-LBD complexed to vitamin D (PDB code 1DB1, 1.8 Å resolution)[60] was used for our VLS test-case study. All of the seven known VDR agonists screened scored better than the top 1% selection threshold derived from the 5000 random compounds (Figure 1, Table 2). Additionally, our results show that VDR-based VLS can significantly enrich compound libraries in structurally diverse hits. Indeed, compounds such as LG19078 and KS291, structurally unrelated to vitamin D, were selected by the algorithm. This result underscores a strength of the VLS technology, which is the ability to derive ligand candidates from the structure of the receptor, regardless of existing ligands, and to uncover novel chemotypes that may translate into improved pharmacological profiles. Indeed, while the use of VDR agonists against a number of diseases, such as cancer and psoriasis, has been impossible so far because of the hypercalcemic side effects of these compounds, newly discovered molecules, such as LG19078 and KS291, also identified by our screening, could uncouple the beneficial effect from the calcemic activity and are promising leads for cancer therapy.[61,62]

The main false positives observed among NR ligands are PPAR agonists (Figure 1). Some of these compounds, such as L796449 and GW501516, are structurally unrelated, and the docked complexes look reasonable. As always, in vitro experiments would be necessary to address the relevance of these unexpected hits.

Discussion

The efficiency of virtual screening was tested against a variety of NR-LBDs and different receptor conforma-

tions, with both random compound library and 78 known NR ligands. With the exception of PXR, the exercise could significantly increase the hit concentration of focused libraries. The average enrichment factor was 70 for the top 1% selections and was 9 for top 10% selections (Table 2). This result clearly shows that VLS can significantly enrich small libraries focused on crystallized NRs. Additionally, the 100 times enrichment rate of a top 1% selection derived from a computer model of GR suggests that, at least in some cases, VLS can be applied successfully in the absence of an experimental receptor structure. However, to evaluate the success rate of such an approach, VLS should be tested against a statistically representative number of structural models built by homology to more or less distant experimental template structures.

While the overall efficiency of NR-based VLS was demonstrated, results vary significantly from one receptor to another (Table 2). The observation that two different crystal structures of the same receptor can produce widely diverging VLS efficiency (see, for instance, 1FM6 and 1FM9 conformations of PPARγ) emphasizes that the conformation of the receptor, in addition to the nature of the target, is a critical parameter for VLS. The chemical diversity of hits present in the source library can also influence the results. For instance, the limited diversity of known RAR and GR ligands translated into high enrichment of corresponding selections. If one ligand docks well to the receptor's conformation, structurally related ligands have increased chances of also docking well. The apparently high VLS efficiency that results may be undermined by the low diversity of hits identified. As for HTS, diversity is an important variable of VLS technology because the purpose of a lead discovery program is to identify hits covering different regions of the chemistry space in order to increase the chances of developing a drug with acceptable pharmacology. In this regard, it is important to note that known ligands for most receptors that did produce quality selections, such as ERs, PPARγ, VDR, and RXR, are chemically diverse. This confirms that VLS is a relevant approach for NR-targeted lead discovery.

Selectivity is another important parameter to address the efficiency of VLS. Both receptor selectivity and selectivity between different isoforms of the same receptor can be considered. Comparing the scores assigned for each receptor to the different NR ligands clearly shows that, as a general trend, ligands for receptor X score better than ligands for receptor Y when screened against receptor X (Figure 1). While this distinction is very well defined for some receptors such as PPARs, VDR, RAR, and ER, the separation is sometimes not as clear for others, such as the steroid receptors AR, PR, and GR. As detailed above, in the latter case, nonspecific steroid receptor ligands constitute most of the false positives. While such miss-selected compounds result from a lack of VLS specificity, they can sometimes also reflect the presence of structurally similar features between the LBP of two receptors and the potential risk of cross-reactivity. For instance, the good scores assigned to several ER, PR, and GR ligands in the AR screenings (see Results) suggest possible cross-reactivity between these receptors. Indeed, some estrogens, progestagens, and glucocorticoids can activate variant forms of AR carrying a single mutation in their LBD.[38] A more surprising feature is the possible cross-reactivity between ER antagonists and active PPARs suggested by the high score assigned to some ER antagonists in the PPAR screens.

To systematically evaluate the predicted promiscuity between receptors used for screening in this work, a cross-reactivity matrix was generated on the basis of the scores assigned to all NR ligands against all receptors, which lists calculated distances between the LBPs (see Experimental Section for details). The distance between two receptors that may share common ligands is small while that between structurally remote receptors is large (Table 3). A striking feature of this matrix is the very small distance separating inactive ER (PDB code 3ERT) from all active PPARs. The distances are 0.07, 0.04, 0.04, and 0.05 against PPARα (1K7L), PPARδ (1GWX), PPARγ (1FM6), and PPARγ (1FM9), respectively. Testing this observation in vitro would be interesting but may prove difficult, since it is possible that only mutated forms of PPARs may be activated by ER antagonists.

Selectivity between different isoforms of the same receptor can also represent an important parameter of lead discovery. As detailed above, the PPAR screenings show that VLS is so far unable to reliably discriminate between receptor isotypes (Figure 1). The observation that isoform specificity may rely on a single residue, as is the case here,[50] underscores the extreme selectivity required.

It may be hypothesized that steric fit dominates NR/ligand recognition. It is therefore instructive to evaluate the relative importance of the steric factors compared to more specific terms of the scoring function such as electrostatics and hydrogen bonding. Table 2 summarizes the enrichment factors of the compound selections obtained when the steric term (dEgrid) of the ICM scoring function was used alone without other components of the ICM score. For the top 1% selections, in 13 out of 19 cases the full score performs significantly better, especially for non-steroid binding receptors such as all PPARs and TRβ (the latter gets 0% enrichment with steric scoring). In five cases, the performance is similar. Only with PXR, the poorest performing receptor, does steric scoring work somewhat better for both the top 1% and 10% selections. Overall, compared to steric scoring, full scoring improves average enrichment by 75% (from 39- to 68-fold) for the top 1% selection.

The treatment of receptor flexibility in VLS simulations leaves room for improvement. In the present state of the algorithm used here, the ligand is continuously flexible during docking simulations while the receptor is represented by a series of superimposed potential maps that account for van der Waals, hydrogen bond, and hydrophobicity profiles, as well as electrostatic potential.[35] The boundaries of the LBPs are rather tolerant for small-range van der Waals clashes to simulate possible local rearrangement of receptor side chains, while tautomeric or rotameric side chain transformations are accounted for by screening independently different conformations of the receptor. Such screening strategy does result in a significant enrichment of compound libraries (Table 2) and can be further im-

Table 3. Predicted NR Cross-Reactivity[a]

	AR (1E3G)	AR (1I37)	ERα (1L2I)	ERα (3ERD)	ERα* (3ERT)	ERβ (1QKM)	GR (model)	PPARα (1K7L)	PPARδ (1GWX)	PPARγ (1FM6)	PPARγ (1FM9)	PR (1A28)	PR (1E3K)	PXR (1ILH)	RARγ (1FCZ)	RXR (1FBY)	RXR (1FM6)	TRβ (1BSX)	VDR (1DB1)
AR (1E3G)	0.	0.13	0.17	0.10	0.60	0.12	0.21	0.59	0.59	0.60	0.59	0.12	0.12	0.55	0.35	0.46	0.48	0.37	0.56
AR (1I37)		0.	0.17	0.09	0.85	0.10	0.32	0.89	0.90	0.89	0.88	0.20	0.24	0.80	0.60	0.74	0.77	0.65	0.86
ERα (1L2I)			0.	0.11	0.65	0.14	0.22	0.69	0.72	0.69	0.67	0.15	0.12	0.64	0.34	0.49	0.51	0.47	0.74
ERα (3ERD)				0.	0.72	0.08	0.31	0.75	0.77	0.76	0.73	0.19	0.19	0.70	0.42	0.59	0.60	0.50	0.73
ERα* (3ERT)					0.	0.66	0.28	0.07	0.04	0.04	0.05	0.40	0.44	0.03	0.39	0.15	0.16	0.19	0.08
ERβ (1QKM)						0.	0.28	0.72	0.73	0.71	0.70	0.18	0.18	0.65	0.45	0.56	0.59	0.51	0.71
GR (model)							0.	0.31	0.30	0.30	0.31	0.06	0.13	0.22	0.33	0.24	0.26	0.24	0.34
PPARα (1K7L)								0.	0.03	0.03	0.03	0.46	0.51	0.06	0.36	0.14	0.13	0.16	0.11
PPARδ (1GWX)									0.	0.02	0.04	0.43	0.48	0.03	0.39	0.13	0.14	0.16	0.06
PPARγ (1FM6)										0.	0.02	0.43	0.48	0.04	0.37	0.13	0.13	0.16	0.08
PPARγ (1FM9)											0.	0.45	0.50	0.06	0.36	0.13	0.13	0.18	0.11
PR (1A28)												0.	0.05	0.34	0.34	0.32	0.35	0.29	0.42
PR (1E3K)													0.	0.42	0.26	0.32	0.34	0.28	0.45
PXR (1ILH)														0.	0.39	0.11	0.13	0.17	0.08
RARγ (1FCZ)															0.	0.15	0.14	0.16	0.36
RXR (1FBY)																0.	0.01	0.10	0.16
RXR (1FM6)																	0.	0.10	0.17
TRβ (1BSX)																		0.	0.13
VDR (1DB1)																			0.

[a] A cross-reactivity matrix was computed where the distance between two receptors is lower when their binding profiles across the 78 NR ligands are similar. Very low distances suggest potential risk of cross-reactivity between the receptors' ligands.

Table 4. Docking Accuracy[a]

PDB	ligand	receptor	resolution X-ray structure (Å)	VLS score, X-ray ligand	VLS score, docked ligand	rmsd X-ray/docked (Å)
1a28	progesterone	PR	1.8	−55.2	−59.3	0.032
1bsx	T3	TRb	3.7	−35.7	−51.6	0.38
1db1	VD3	VDR	1.8	−52.3	−60.7	0.74
1e3g	metribolone	AR	2.4	−56.2	−62.9	0.22
1e3k	metribolone	PR	2.8	−52.4	−52.6	0.28
1fby	9cis-RA	RXR	2.25	−58.5	−66.4	0.37
1fcz	BMS181156	RARg	1.38	−72.7	−68.6	0.84
1fm6	rosiglitazone	PPARg	2.1	−41.	−43.3	1.39
1fm6	9cisRA	RXR	2.1	−51.4	−67.4	0.78
1fm9	GI262570	PPARg	2.1	−61.7	−66.5	1.72
1i37	dihydrotesto	AR	2.0	−48.6	−51.4	0.21
1ilh	SR12813	PXR	2.76	+28.6	−29.0	2.24
1l2i	THC	ERa	1.95	−31.2	−41.7	0.29
1qkm	genistein	ERb	1.8	−30.5	−40.4	0.79
3erd	DES	ERa	2.03	−43.0	−48.6	1.61
3ert	tamoxifen	ERa	1.9	−37.6	−46.2	1.41

[a] Docking and scoring performance for the ligand/receptor pairs for which complex X-ray structures were available. The rmsd of docked ligand versus X-ray structure as well as VLS scores for both structures are listed. X-ray structure resolution is provided as an indication of the quality of the experimental structures.

proved by adding a "medium-throughput" step, where compounds initially selected by VLS are docked again, this time against a flexible representation of the receptor, with higher stringency against van der Waals clashes.[15] While such fully flexible docking simulations still remain too slow for the screening of libraries of several hundred thousand molecules, it is reasonable for a few thousand compounds. Implementing receptor flexibility on all ligands of large source libraries would still be preferable because it may allow for the identification of interesting hits requiring unexpected side chain conformations and the exclusion of additional true negatives that may lightly clash with structurally locked residues.

The screenings performed against two different conformations of PPARγ (PDB codes 1FM6 and 1FM9) illustrate how critical receptor flexibility is for the identification of structurally diverse ligands. As detailed above (see Results), the LBP of the 1FM9 structure is larger than that of 1FM6. As a result, some real PPARγ agonists selected against 1FM9 cannot fit in the 1FM6 structure and are missed in the corresponding screening. Conversely, a few other true positives selected against 1FM6 fit in the 1FM9 structure but do not make optimal hydrophobic contacts with the LBP that is larger than necessary and are assigned a relatively poor score. This clearly illustrates that one receptor conformation does not match all ligands. If only PPARγ agonists of the glitazone family were known, the 1FM9 structure where the receptor is complexed to GI262570, a larger ligand, would not be available, and VLS against the 1FM6 conformation would not allow identification of GI262570-like hits. However, a flexible representation of PPARγ, starting from the 1FM6 structure, would account for more open conformations of the LBP and may allow the selection of GI262570 and related compounds as PPARγ agonist candidates. This example illustrates how implementation of receptor flexibility could open unexpected regions of the chemistry space to lead discovery efforts toward the development of active molecules presenting original chemotypes. Developing rapid and robust VLS algorithms that include such a feature is an important challenge of the present time (see ref 13 for a review).

Ongoing development efforts will undoubtedly result in improved enrichment levels of focused libraries toward an ideal case scenario where VLS would actually allow cherry picking of all true positives out of a large and diverse compound database. The present study, however, shows that significant differences in enrichment can exist between the top 10% selections and top 1% selections (Table 2). This result suggests that a conservative approach, where VLS is used to extract focused pools of a few thousand molecules from diverse or customized source libraries of hundreds of thousands of compounds, remains probably the most appropriate at this time.

Conclusion

The systematic virtual screening of 78 NR ligands as well as 5000 random compounds against 19 different NR structures provides a reliable assessment of the efficiency of the VLS technology against this important family of transcription factors. More generally, this study contributes to the evaluation of the state-of-the-art of VLS. It reveals encouraging progress, such as the efficient enrichment of a focused library in the absence of experimental structure of the targeted receptor, as well as important unmet challenges, such as the need for improvement of receptor isoform selectivity and implementation of side chain flexibility. While this study clearly confirms that VLS can be a useful tool to rapidly identify hits, it also defines the limitations of the approach and suggests that, to avoid missing active compounds, the technology may best be used to rapidly extract from large source libraries focused selections that are more easily amenable to high-throughput screening.

Experimental Section

Receptor Preparation. For all receptors but GR, the conformations targeted for virtual screening were directly derived from crystal structures. Several receptor conformations were used when multiple tautomeric or rotameric states could match the experimental electron density of the structure. For each of the 1E3G and 1I37 structures of AR,[36,37] alternative conformations were used where the χ2 torsion of N705 points either the side chain oxygen or side chain nitrogen toward the LBP and where the hydroxyl hydrogen of T877 points either

toward or away from the LBP. For the 1L2I, 3ERD, and 1QKM structures of ER,[33,40,41] the tautomeric state of ERα-H524 or ERβ-H475 was set so that the residue could either accept a hydrogen from or give a hydrogen to potential ligands. In the case of the 1A28 and 1E3K PR structures,[47,36] the χ3 torsion angle of Q725 was set either to 0° or to 180°, while the hydroxyl oxygen of T894 was either pointing away from or toward the receptor's LBP. Representations of PPARs with alternative hydroxyl rotameric states and histidine tautomeric states of S289, Y473, H449, and H473 for PPARγ, S280, Y314, and Y464 for PPARα, and Y473, H323, and H449 for PPARδ were used for screening. The two possible tautomeric states of TRβ-H435 that preserve the rotamer observed in the crystal structure were used for this receptor. Additionally, while the experimental conformation of N223 was retained for one screening, the conformational space available to this residue was also sampled in the presence of bound T3 and the lowest energy conformation found was used for screening. This exceptional, local diversion from the experimental electronic density was only justified by the poor 3.7 Å resolution of the TR data currently available.[57] In the case of VDR, varied rotameric states of S237 and S278 hydroxyls and tautomeric states of H305 were applied for screening the 1DB1 structure.[60] Finally, a single receptor conformation was derived from the RAR and RXR experimental structures, since no variation seemed necessary or justified, and a single conformational state of PXR was used out of the too numerous possible conformations produced by a series of hydrogen-bond donors and acceptors lining the LBP of this receptor.

The GR-LBD 3D model used for screening was built by homology to the 1A28 crystal structure of the AR-LBD, a template sharing 53.8% sequence identity.[47] The target sequence was aligned on the 3D template, and the energy of the system was minimized by a Monte Carlo simulation through a series of random global moves and gradient local minimizations in the internal coordinates space.[35] A single rotameric state was used for screening.

Virtual Screening. The ICM virtual library screening (VLS) module was applied (Molsoft LLC, La Jolla, CA). A series of five grid potential representations of the receptor were automatically generated, which accounted for the hydrophobic interactions, heavy-atom and hydrogen van der Waals interactions, hydrogen-bonding, and electrostatic potential of the predefined ligand binding site. Grid calculation took an average of 3 min per receptor conformation. Each flexible compound was first energy-minimized in the absence of the receptor. The lowest energy conformations identified were then used as a starting point for docking simulations to the receptor grids by the ICM method,[8,63] and the ligand was assigned a score that reflects the quality of the complex. Docking took an average of 1 min per processor and per ligand. Since ICM docking is a stochastic optimization procedure, to ensure convergence, the whole process was conducted two times in parallel for each receptor structure (an additional two screenings tested on the benchmark library of NR ligands did not change the results) and the lowest score assigned to each ligand was retained.

The ICM scoring function[12] used consisted of the following terms:

$$\Delta E_{\mathrm{IntFF}} + T\Delta S_{\mathrm{Tor}} + \alpha_1\Delta E_{\mathrm{HBond}} + \alpha_2\Delta E_{\mathrm{HBDesol}} + \alpha_3\Delta E_{\mathrm{SolEl}} + \alpha_4\Delta E_{\mathrm{HPhob}} + \alpha_5 Q_{\mathrm{Size}}$$

where $\Delta E_{\mathrm{IntFF}}$ includes the van der Waals interaction of the ligand with the receptor as well as internal force-field energy of the ligand. Total van der Waals repulsion for any ligand atom was truncated to 4 kcal/mol. This term accounts for the quality of the steric fit, including any conformational strain induced in the ligand by receptor binding.

$T\Delta S_{\mathrm{Tor}}$ is the change in free energy due to the conformational entropy loss for the ligand upon binding, which is assumed to be proportional to the number of free torsions (N_{Tor}) in the ligand and calculated as $0.6N_{\mathrm{Tor}}$.

$\Delta E_{\mathrm{HBond}}$ is the hydrogen bonding term. For a hydrogen bond acceptor atom i and a hydrogen atom j located at $\mathbf{r}_j$, the hydrogen bonding interaction was estimated as $F_{\mathrm{ang}}(\varphi)F_{\mathrm{dist}}(\mathbf{r}_{\mathrm{LP}i}-\mathbf{r}_j)$, where φ is the angle formed by hydrogen bond acceptor atom, hydrogen, and the hydrogen bond donor, and $\mathbf{r}_{\mathrm{LP}i}$ is the radius vector of the center of the lone electron pair (LP) closest to the hydrogen. The angular function used was defined as $F_{\mathrm{ang}}(\varphi) = 1 - \cos(\varphi)$. Distance function $F_{\mathrm{dist}}(\mathbf{r}_{\mathrm{LP}i}-\mathbf{r}_j)$ was constant (1.0) within $\lambda_{\mathrm{HB}}/2$ from the lone pair center and dropped as

$$\exp\left(\frac{-(|\mathbf{r}_{\mathrm{LP}i} - \mathbf{r}_j| - \lambda_{\mathrm{HB}}/2)^2}{{\lambda_{\mathrm{HB}}}^2}\right)$$

beyond that distance, where λ_{HB} is the characteristic range of hydrogen bonding interaction (value of 1.6 Å was used). Lone pair centers were placed at 1 Å from the hydrogen bond acceptor atom, assuming symmetrical planar trigonal configuration for sp^2 atoms and tetrahedral configuration for sp^3 atoms. The resulting functional dependence reflects (at least qualitatively) the physical nature and observed statistics of the hydrogen bond interactions. The interaction is maximized when the hydrogen atom is pointing directly to the acceptor atom along a lone pair axis and drops quickly as the hydrogen is moved farther away. It declines more gradually as the hydrogen moves out of the LP axis or, as hydrogen bond donor, hydrogen atom, or hydrogen bond acceptor, moves out of alignment. Appropriate strength of the interaction was achieved through the weighting factor α_1.

$\Delta E_{\mathrm{HBDesol}}$ is the term that accounts for the disruption of hydrogen bonds with solvent (desolvation of hydrogen bond donors and acceptors). Appropriate strength was achieved through the weighting factor α_2.

$\Delta E_{\mathrm{SolEl}}$ is the solvation electrostatic energy change upon binding. It was calculated using the boundary element Poisson equation solver as implemented in the REBEL[64] module of ICM. Calculations were performed using dielectric constants of 4 and 80 for the protein interior and exterior, respectively. Owing to the large uncertainty of the current estimates for internal dielectric constant and partial charges, the weighting factor α_3 was applied.

$\Delta E_{\mathrm{Hphobm}}$ is the hydrophobic free energy gain. Following the popular accessible surface area (ASA) model, it was assumed to be proportional to the ASA of the hydrophobic atoms of the receptor and ligand, buried upon binding. Initial estimate of the surface tension was set at 0.012 kcal/Å^2. Again, a wide variety of values for this parameter can be found in the literature, prompting us to apply an adjustable weighting factor α_4.

Q_{Size} is a size correction term. It is often observed that scoring functions have a bias toward larger ligands. Addition of a size factor proportional to the number of atoms in the ligand allowed us to largely eliminate this bias. While this term has no direct physical meaning, it may account for otherwise omitted interactions of the ligand with the solvent, primarily the van der Waals dispersion interaction. Q_{Size} was evaluated simply as the number of atoms in the ligand, with the weighting factor α_5.

The weights α_1–α_5 were optimized on a diverse benchmark of ligands and receptors as previously described.[12]

Cross-Reactivity Matrix. The predicted binding profiles across the 78 NR ligand benchmark library were used to scale the LBP structural similarity between different receptors. Two structures presenting similar scoring profiles across the library are separated by a small LBP-structure similarity distance and have higher risks of cross-reactivity. The distances listed in the matrix are calculated as scalar products between the normalized sets of scores assigned for each receptor to all ligands.

Acknowledgment. This work was supported in part by an SBIR/STTR grant awarded by the National Institutes of Health to Molsoft, Grant No. 1R41-DK-59041-01.

Supporting Information Available: Figure 2 showing 19 charts of the details of the steric scores (dEgrid term of the ICM scoring function) of all NR ligands for all receptors and Figure 3 showing the chemical diversity of the 5000-compound library used in this work (evaluated using the molecular weight, number of torsional degrees of freedom, number of hydrogen bond donors, and number of hydrogen bond acceptors as molecular descriptors), and the benchmark library of 78 NR ligands. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Walters, P.; Stahl, M. T.; Murcko, M. A. Virtual screening an overview. *Drug Discovery Today* **1998**, *3*, 160−178.

(2) Muegge, I.; Rarey, M. Small Molecule Docking and Scoring. *Rev. Comput. Chem*. **2001**, *17*, 1−60.

(3) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule−ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.

(4) Ewing, T. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated docking and database screening. *J. Comput. Chem*. **1997**, *18*, 1175−1189.

(5) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(6) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol*. **1997**, *267*, 727−748.

(7) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem*. **1998**, *18*, 1639−1662.

(8) Totrov, M.; Abagyan, R. Protein−ligand docking as an energy optimization problem. *Drug−Receptor Thermodynamics: Introduction and Applications*; Raffa, R. B., Ed.; John Wiley & Sons: New York, 2001; pp 603−624.

(9) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **2002**, *47*, 409−443.

(10) Boehm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein−ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.

(11) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein−ligand interactions: a simplified potential approach. *J. Med. Chem*. **1999**, *42*, 791−804.

(12) Totrov, M.; Abagyan, R. Derivation of Sensitive Discrimination Potential for Virtual Ligand Screening. In *RECOMB '99: Proceedings of the Third Annual International Conference on Computational Molecular Biology*, Lyon, France, April 11−14, 1999; Istrail, S., Pevzner, P., Waterman, M., Eds.; Association for Computing Machinery: New York, 1999.

(13) Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375−382.

(14) Perola, E.; Xu, K.; Kollmeyer, T. M.; Kaufmann, S. H.; Prendergast, F. G.; Pang, Y. P. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J. Med. Chem*. **2000**, *43*, 401−408.

(15) 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.

(16) Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem*. **2002**, *45*, 2213−2221.

(17) Filikov, A. V.; Mohan, V.; Vickers, T. A.; Griffey, R. H.; Cook, P. D.; Abagyan, R. A.; James, T. L. Identification of Ligands for HIV-1 TAR RNA via Structure-Based Virtual Screening *J. Comput.-Aided Drug Des.* **2000**, *14*, 593−610.

(18) Chen, Y. Z.; Zhi, D. G. Ligand−protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* **2001**, *43*, 217−226.

(19) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem*. **1999**, *42*, 5100−5109.

(20) 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.

(21) Stahl, M.; Rarey, M. Detailed Analysis of Scoring Functions for Virtual Screening. *J. Med. Chem*. **2001**, *44*, 1035−1042.

(22) Ribeiro, R. C.; Kushner, P. J.; Baxter, J. D. The nuclear hormone receptor gene superfamily. *Annu. Rev. Med*. **1995**, *46*, 443−453.

(23) Dees, E. C.; Kennedym, M. J. Recent advances in systemic therapy for breast cancer. *Curr. Opin. Oncol.* **1998**, *10*, 517−522.

(24) Olefsky, J. M.; Saltiel, A. R. PPAR gamma and the treatment of insulin resistance. *Trends Endocrinol. Metab.* **2000**, *11*, 362−368.

(25) Kolvenbag, G. J.; Iversen, P.; Newling, D. W. Antiandrogen monotherapy: a new form of treatment for patients with prostate cancer. *Urology* **2001**, *58* (2, Suppl. 1), 16−23.

(26) Wong, S. F. Oral bexarotene in the treatment of cutaneous T-cell lymphoma. *Ann. Pharmacother.* **2001**, *35*, 1056−1065.

(27) Repa, J. J.; Mangelsdorf, D. J. The role of orphan nuclear receptors in the regulation of cholesterol homeostasis. *Annu. Rev. Cell Dev. Biol.* **2000**, *16*, 459−481.

(28) Brown, A. J. Mechanisms for the selective actions of vitamin D analogues. *Curr. Pharm. Des.* **2000**, *6*, 701−716.

(29) Brzozowski, A. M.; Pike, A. C.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engstrom, O.; Ohman, L.; Greene, G. L.; Gustafsson, J. A.; Carlquist, M. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **1997**, *389*, 753−758.

(30) Moras, D.; Gronemeyer, H. The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* **1998**, *10*, 384−391.

(31) Weatherman, R. V.; Fletterick, R. J.; Scanlanm, T. S. Nuclear-receptor ligands and ligand-binding domains. *Annu. Rev. Biochem*. **1999**, *68*, 559−581.

(32) 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.

(33) Pike, A. C.; Brzozowski, A. M.; Hubbard, R. E.; Bonn, T.; Thorsell, A.-G.; Engstrom, O.; Ljunggren, J.; Gustaffson, J.-A.; Carlquist, M. Structure of the Ligand-Binding Domain of Oestrogen Receptor Beta in the Presence of a Partial Agonist and a Full Antagonist. *EMBO J.* **1999**, *18*, 4608−4618.

(34) Xu, H. E.; Stanley, T. B.; Montana, V. G.; Lambert, M. H.; Shearer, B. G.; Cobb, J. E,; McKee, D. D.; Galardi, C. M.; Plunket, K. D.; Nolte, R. T.; Parks, D. J.; Moore, J. T.; Kliewer, S. A.; Willson, T. M.; Stimmel, J. B. Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha. *Nature* **2002**, *415*, 813−817.

(35) *ICM 2.8 Manual*, Molsoft LLC: San Diego, CA; www.molsoft.com.

(36) Matias, P. M.; Donner, P.; Coelho, R.; Thomaz, M.; Peixoto, C.; Macedo, S.; Otto, N.; Joschko, S.; Scholz, P.; Wegg, A.; Basler, S.; Schafer, M.; Egner, U.; Carrondo, M. A. Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J. Biol. Chem*. **2000**, *275*, 26164−26171.

(37) Sack, J. S.; Kish, K. F.; Wang, C.; Attar, R. M.; Kiefer, S. E.; An, Y.; Wu, G. Y.; Scheffler, J. E.; Salvati, M. E.; Krystek, S. R., Jr.; Weinmann, R.; Einspahr, H. M. Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4904−4909.

(38) Brinkmann, A. O.; Trapman, J. Prostate cancer schemes for androgen escape. *Nat. Med.* **2000**, *6*, 628−629.

(39) Zhao, X. Y.; Malloy, P. J.; Krishnan, A. V.; Swami, S.; Navone, N. M.; Peehl, D. M.; Feldman, D. Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. *Nat. Med*. **2000**, *6*, 703−706.

(40) Shiau, A. K.; Barstad, D.; Radek, J. T.; Meyers, M. J.; Nettles, K. W.; Katzenellenbogen, B. S.; Katzenellenbogen, J. A.; Agard, D. A.; Greene, G. L. Structural Characterization of a Subtype-Selective Ligand Reveals a Novel Mode of Estrogen Receptor Antagonism. *Nat. Struct. Biol.* **2002**, *9*, 359−364.

(41) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998**, *95*, 927−937.

(42) Maggiolini, M.; Donze, O.; Jeanninm, E.; Andom, S.; Picard, D. Adrenal androgens stimulate the proliferation of breast cancer cells as direct activators of estrogen receptor alpha. *Cancer Res.* **1999**, *59*, 4864−4869.

(43) Nogueira, C. R.; Brentanim, M. M. Triiodothyronine mimics the effects of estrogen in breast cancer cell lines. *J. Steroid Biochem. Mol. Biol.* **1996**, *59*, 271−279.

(44) Gregoraszczuk, E. L. Is thyroid hormone a modulator of estrogen receptor in porcine follicular cells? *Endocr. Regul.* **2000**, *34*, 151−155.

(45) Watkins, R. E.; Wisely, G. B.; Moore, L. B.; Collins, J. L.; Lambert, M. H.; Williams, S. P.; Willson, T. M.; Kliewer, S. A.; Redinbo, M. R. The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science* **2001**, *292*, 2329−2333.

(46) Willson, T. M.; Kliewer, S. A. PXR, CAR and drug metabolism. *Nat. Rev. Drug Discovery* **2002**, *1*, 259−66.

(47) Williams, S. P.; Sigler, P. B.; Atomic structure of progesterone complexed with its receptor. *Nature* **1998**, *393*, 392−396.
(48) Selman, P. J.; Wolfswinkel, J.; Mol, J. A. Binding specificity of medroxyprogesterone acetate and proligestone for the progesterone and glucocorticoid receptor in the dog. *Steroids* **1996**, *61*, 133−137.
(49) Gampe, R. T., Jr; Montana, V. G.; Lambert, M. H.; Miller, A. B.; Bledsoe, R. K.; Milburn, M. V.; Kliewer, S. A.; Willson, T. M.; Xu, H. E. Asymmetry in the PPARgamma/RXRalpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. *Mol. Cell* **2000**, *5*, 545−555.
(50) Xu, H. E.; Lambert, M. H.; Montana, V. G.; Plunket, K. D.; Moore, L. B.; Collins, J. L.; Oplinger, J. A.; Kliewer, S. A.; Gampe, R. T.; Mckee, D. D.; Moore, J. T.; Willson, T. M. Structural Determinants of Ligand Binding Selectivity between the Peroxisome Proliferator-Activated Receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13919−13924.
(51) Xu, H. E.; Lambert, M. H.; Montana, V. G.; Parks, D. J.; Blanchard, S. G.; Brown, P. J.; Sternbach, D. D.; Lehmann, J. M.; Wisely, G. B.; Willson, T. M.; Kliewer, S. A.; Milburn, M. V. Molecular recognition of fatty acids by peroxisome proliferator-activated receptors. *Mol. Cell* **1999**, *3*, 397−403.
(52) Tagnon, H. J. Antiestrogens in treatment of breast cancer *Cancer* **1977**, *39*, 2959−2964.
(53) Egea, P. F.; Mitschler, A.; Rochel, N.; Ruff, M.; Chambon, P.; Moras, D. Crystal Structure of the Human Rxr Alpha Ligand-Binding Domain Bound to Its Natural Ligand: 9-Cis Retinoic Acid. *EMBO J.* **2000**, *19*, 2592−2601.
(54) Klaholz, B. P.; Mitschler, A.; Moras, D. Structural Basis for Isotype Selectivity of the Human Retinoic Acid Nuclear Receptor. *J. Mol. Biol.* **2000**, *302*, 155−170.
(55) Dawson, M. I.; Jong, L.; Hobbs, P. D.; Cameron, J. F.; Chao, W. R.; Pfahl, M.; Lee, M. O.; Shroot, B.; Pfahl, M. *J. Med. Chem.* **1995**, *38*, 3368−3383.
(56) Sun, S. Y.; Yue, P.; Dawson, M. I.; Shroot, B.; Miche, I. S.; Lamph, W. W.; Heyman, R. A.; Teng, M.; Chandraratna, R. A.; Shudo, K.; Hong, W. K.; Lotan, R. Differential effects of synthetic nuclear retinoid receptor-selective retinoids on the growth of human non-small cell lung carcinoma cells. *Cancer Res.* **1997**, *57*, 4931−4939.
(57) Darimont, B. D.; Wagner, R. L.; Apriletti, J. W.; Stallcup, M. R.; Kushner, P. J.; Baxter, J. D.; Fletterick, R. J.; Yamamoto, K. R. Structure and Specificity of Nuclear Receptor: Coactivator Interactions. *Genes Dev.* **1998**, *12*, 3343−3356.
(58) Schoonjans, K.; Martin, G.; Staels, B.; Auwerx, J. Peroxisome proliferator-activated receptors, orphans with ligands and functions *Curr. Opin. Lipidol.* **1997**, *8*, 159−166.
(59) Ness, G. C. Thyroid hormone. Basis for its hypocholesterolemic effect. *J. Fla. Med. Assoc.* **1991**, *78*, 383−385.
(60) Rochel, N.; Wurtz, J. M.; Mitschler, A.; Klaholz, B.; Moras, D. The Crystal Structure of the Nuclear Receptor for Vitamin D Bound to Its Natural Ligand. *Mol. Cell* **2000**, *5*, 173−179.
(61) Boehm, M. F.; Fitzgerald, P.; Zou, A.; Elgort, M. G.; Bischoff, E. D.; Mere, L.; Mais, D. E.; Bissonnette, R. P.; Heyman, R. A.; Nadzan, A. M.; Reichman, M.; Allegretto, E. A. Novel non-secosteroidal vitamin D mimics exert VDR-modulating activities with less calcium mobilization than 1,25-dihydroxyvitamin D3. *Chem. Biol.* **1999**, *6*, 265−267.
(62) Verstuyf, A.; Verlinden, L.; van Etten, E.; Shi, L.; Wu, Y.; D'Halleweyn, C.; Van Haver, D.; Zhu, G. D.; Chen, Y. J.; Zhou, X.; Haussler, M. R.; De Clercq, P.; Vandewalle, M.; Van Baelen, H.; Mathieu, C.; Bouillon, R. Biological activity of CD-ring modified 1alpha,25-dihydroxyvitamin D analogues: C-ring and five-membered D-ring analogues *J. Bone Miner. Res.* **2000**, *15*, 237−252.
(63) Totrov, M.; Abagyan, R. Flexible protein−ligand docking by global energy optimization in internal coordinates. *Proteins*, **1997**, Suppl. 1, 215−220.
(64) Totrov, M.; Abagyan, R. Rapid boundary element solvation electrostatics calculations in folding simulations: successful folding of a 23-residue peptide. *Biopolymers* **2001**, *60*, 124−133.

JM0300173