

***In-silico* screening using flexible ligand binding pockets: a molecular dynamics-based approach**

Dakshanamurthy Sivanesan^a, Rajendram V Rajnarayanan^a, Jason Doherty^{a,c} & Nagarajan Pattabiraman^{a,b}

^aDepartment of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, USA; ^bDepartment of Biochemistry and Molecular Biology, Lombardi Comprehensive Cancer Center, Georgetown University, NRB, Room W417 3970, Reservoir Rd, NW, DC, 20005, USA; ^cDepartment of Pathology, Washington University in St. Louis, St. Louis, MO, 63130, USA

Received 9 November 2004; accepted 28 March 2005
© Springer 2005

Key words: docking, endocrine disrupting chemicals, estrogen receptor, *in-silico* screening, molecular dynamics, receptor conformation, receptor flexibility

Summary

In-silico screening of flexible ligands against flexible ligand binding pockets (LBP) is an emerging approach in structure-based drug discovery. Here, we describe a molecular dynamics (MD) based docking approach to investigate the influence on the high-throughput *in-silico* screening of small molecules against flexible ligand binding pockets. In our approach, an ensemble of 51 energetically favorable structures of the LBP of human estrogen receptor α (hER α) were collected from 3 ns MD simulations. *In-silico* screening of 3500 endocrine disrupting compounds against these flexible ligand binding pockets resulted in thousands of ER–ligand complexes of which 582 compounds were unique. Detailed analysis of MD generated structures showed that only 17 of the LBP residues significantly contribute to the overall binding pocket flexibility. Using the flexible LBP conformations generated, we have identified 32 compounds that bind better to the flexible ligand-binding pockets compared to the crystal structure. These compounds, though chemically divergent, are structurally similar to the natural hormone. Our MD-based approach in conjunction with grid-based distributed computing could be applied routinely for *in-silico* screening of large databases against any given target.

Abbreviations: MDGS – molecular dynamics generated structures; ER – estrogen receptor; EDC – endocrine disrupting chemicals; LBP – ligand binding pocket

Introduction

The molecular recognition between receptors and their cognate ligands play a crucial role in number of important biological processes. This recognition at the molecular level was explained by the ‘lock-and-key’ model, assuming a rigid receptor (lock) and a rigid ligand (key) [1] and later by an ‘induced-fit

model’ [2] based on flexible ligand and receptor complementarity to explain enzymatic catalysis [2]. The initial docking programs were developed based on shape (geometrical) complementarity, including parameters for physico-chemical properties to evaluate the docking results [3, 4]. Rigid receptor docking was greatly enhanced by visualization methods in tandem with real-time docking to calculate the total interaction energy between a drug and its cognate receptor using a 3D tabulation of potentials [5]. The docking programs such as

*To whom correspondence should be addressed. Fax +202-687-7505; E-mail: np47@georgetown.edu

AutoDock, DOCK and FlexX, consider only the ligand flexibility and allow limited side chain flexibility of the receptor [6–9].

Various studies have been reported by modeling the partial protein flexibility using different approximations [10–15]. These docking programs impart protein conformational changes only to a few residues and as a consequence the binding site still remains rigid. However, these models improve prediction to a certain extent as it reflects the conformational changes of key residues in the binding pocket during ligand binding in comparison with regular rigid receptor–ligand docking. The development of super computers catalyzed the progress of force-field based molecular simulation programs such as AMBER [16], CHARMM [17], GROMOS [18] for energy minimization and molecular dynamics. While these programs enable computation of the protein–ligand binding free energy [19, 20], it is computationally intensive and could only be applied to a small set of very closely related analogs.

Structure-based drug discovery has been successfully used not only to identify lead compounds, but also to optimize leads into drugs [21]. This led to the development of *in-silico* screening of large database of small molecules. Apart from practical concerns, the high-cost and time, it takes to perform high-throughput *in-silico* screening limits it to a few academic institutions and large pharmaceutical industries. Ignoring flexibility will inevitably reduce the number and diversity of successfully screened molecules. Hence, it is important to include various receptor conformations especially when screening against large database of organic molecules.

Till date, there is no computational approach to screen large database of flexible ligands against a flexible ligand binding pocket. The molecular dynamics (MD) simulations could be applied to include full flexibility of the LBP by generating an ensemble of protein conformations.

We propose MD simulations in tandem with *in-silico* screening as a viable approach to include both ligand and LBP flexibility [22]. The rationale behind this approach is that all ligand-binding pockets have some inherent flexibility, which allows them to accommodate structurally diverse ligands by resorting to variations in the displacement of backbone and/or in the orientation of side chains. Also, the MD-based approach presents a good

model of protein–ligand interactions in solution, as proteins are in dynamic equilibrium between low energy conformational substates. To validate our approach we chose estrogen receptor, one of the widely studied and validated biological targets. Natural hormone estrogen, bind to ER and induce conformational changes in the receptor architecture enabling the essential co-activator recruitment to trigger estrogenic activity. At the same time, antagonists like tamoxifen, disable the co-activator recruitment. Such ligand-triggered receptor agonism (antagonism incorporates the flexibility as a function of multiple ligand binding conformations. Apart from natural estrogens, a wide range of compounds diverse in their size, shape, and chemical properties [23] exert estrogenic effects by binding to hER α .

There exist few reports on the improvement of the prediction of the relative binding affinity of xeno and phytoestrogens using MD simulations of ER–Ligand complexes [24–27]. Gunsteren and co-workers [19] have reported free energies of binding of polychlorinated biphenyls (PCBs) to the ER from a single simulation. This study demonstrates that a ligand bound to its receptor has to be represented by an ensemble of different conformations and cannot be represented by a single conformation. A recent study by Welsh and co-workers used a knowledge-based multiple-conformation docking method to account for the receptor flexibility [28]. While reporting significant improvement in the results in comparison with single-conformation docking (rigid crystal structure), at the same time they observed a low correlation between docking scores and experimental values.

In this paper, we describe an efficient *in-silico* screening using an ensemble of flexible hER α ligand binding pockets collected at 50 ps intervals from 3 ns MD simulations. The MD generated flexible LBPs were screened against an EDC database, enriched with several tested (predicted ER agonists and antagonists). Using our method, we were able to identify chemically and structurally diverse set of ER ligands that bind to MDGS with higher affinity compared to the rigid crystal structure.

Methodology

In Figure 1, a flowchart of our MD-based flexible docking approach for the high throughput *in-silico*

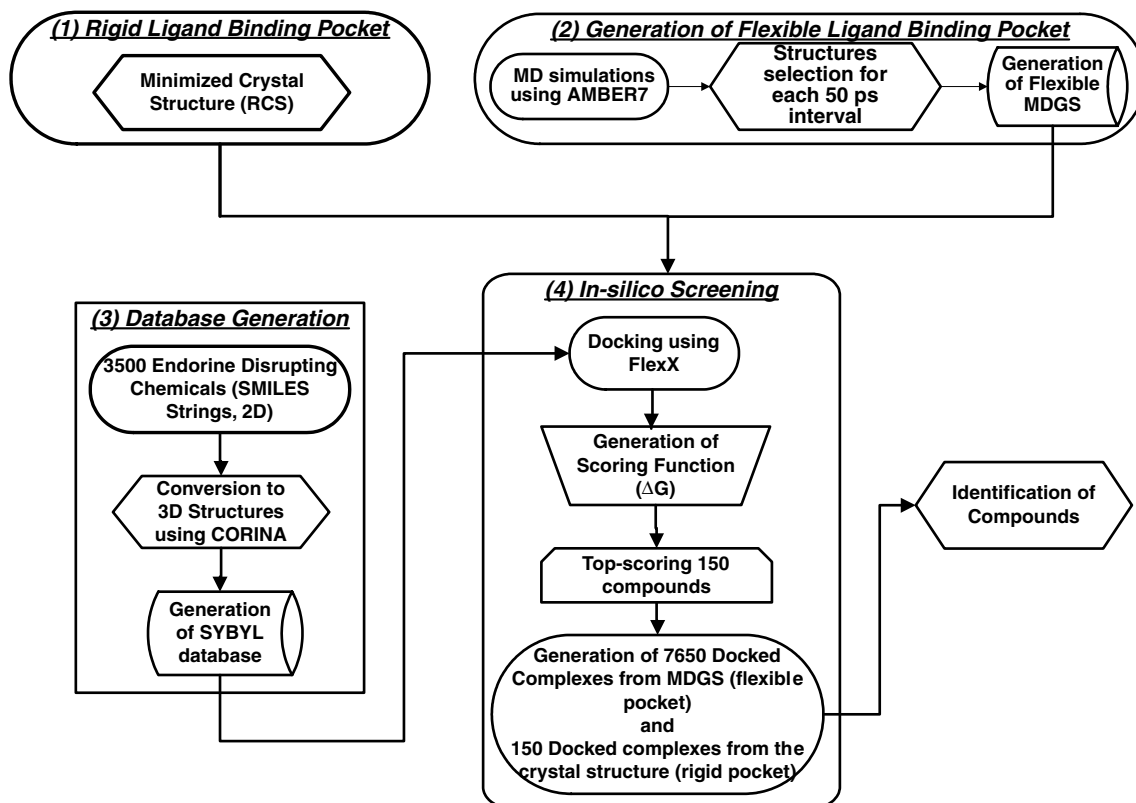


Figure 1. Schematic overview of the MD-based flexible docking approach. Step 1: the generation of rigid ligand binding pocket using the crystal structure (PDB: 3ERD); Step 2: the generation of flexible receptor conformations using MD simulations; Step 3: curation and generation of EDC database; and Step 4: the *in-silico* screening process.

screening of a small molecule database is shown. Our approach consists of four steps, which are described below:

Preparation of rigid ligand binding pocket (LBP)

The ligand binding domain (LBD) of hER α crystal structures was downloaded from the Protein Data bank [29]. Nine structures of hER α with bound agonist ligand were available in the PDB database. Out of these nine structures, 1L2I [30] and 3ERD [31] have resolution ~ 2 Å and both structures are similar with the root mean square deviation (RMSD) of 0.5 Å to each other. The chain-A (monomer) of the ER α crystal structure complexed with the potent synthetic estrogen diethylstilbestrol (PDB ID: 3ERD) was chosen for the present study. The bound ligand, co-activator and water molecules were removed from the PDB structure. The structure was energy minimized using the SANDER

module of AMBER7.0 [16] with the distance-dependent dielectric constant and PARM98 force-field parameters. This energy-minimized structure was used as the rigid binding pocket in our *in-silico* screening.

Generation of flexible ligand binding pockets

Using the energy minimized structure of hER α as the initial model, 3 ns MD simulations with distant-dependent dielectric constant were carried out by using AMBER7.0 with PARM98 force-field parameter [16]. The SHAKE algorithm [32] was employed to keep all bonds involving hydrogen atoms rigid, and weak coupling temperature and pressure coupling algorithms [33] were used to maintain constant temperature and pressure, respectively. MD Simulations were performed using 0.001 ps time steps with temperature set at 300 K, dielectric constant at $1R_{ij}$ and a non-bonded cut-off of 12 Å.

The energetically favorable receptor structures were collected from the low energy minima of potential energy surface within every 50 ps interval starting from 0.2 to 3 ns. A total of 51 MD generated structures were selected and energy minimized. In this paper, we refer 51 MD generated structures as MDGS. 3 ns MD simulations took 33 days of super computer time at the National Cancer Institute. Backbone structural analysis of MDGS was performed using the program NotesPlot unpublished results, developed in our laboratory that compares two backbone torsional angles, ϕ 's and ψ 's between two corresponding structures and identifies the flexible regions of the molecules. Unlike, the RMSD calculation, the two structures were not superimposed in the NotesPlot. A typical NotesPlot is divided into windows of 60 amino acids and for each window the amino acid sequence is shown along the top of the box. The stars '*' in the second line above the box represent amino acid identity. The predicted secondary structure elements, from the DSSP program [34] are displayed at the bottom of each box. The first line corresponds to the predicted secondary structure elements of the crystal structure and the second line that of MDGS at 3ns. In the DSSP output, the alphabet 'H' represents the α -Helix, 'B' represents extended strand participates in beta ladder, 'G' stands for 3_{10} Helix, 'T' stands for the hydrogen bonded turn, and 'S' represents bend. In the two bottom lines, each dash, '-' represents that the DSSP program could not assign secondary structure element, star, '*' in the bottom line represents the secondary structure which is same as predicted for the crystal structure. The binding site residues are marked by a dash '-' above the sequence. The y-axis represents the difference in backbone torsional angles (ϕ and ψ). For each residue, the $\Delta\phi$ and $\Delta\psi$ between the two structures is represented as a note (similar to a musical note). Each note is drawn starting from the $\Delta\phi$ (difference in the ' ψ ' angles between two structures) and ends at $\Delta\psi$ (difference in the ' ψ ' angles between two structures); the end of the note is represented by the letter 'o'. If the $\Delta\phi$ and $\Delta\psi$ between these two structures are zero, then the conformation of this residue is identical between these structures.

Database generation

A database comprising of more than 3500 structurally diverse endocrine disrupting chemicals was

generated. This database contains steroids, non-steroidal anti inflammatory drugs, poly aromatic hydrocarbons, phytoestrogens, phthalates and xenobiotics. Initial 3D structures of these compounds were generated using 2D to 3D conversion program CORINA [35].

In-silico screening

The ligand-binding pocket of hER α is a large steroid size lipophilic cavity with acceptor groups (in residues E353 and H524) at either end that can form hydrogen bonds with two hydroxyl groups of the ligand [31]. Based on the ligand binding pockets of all the available ER crystal structures, 32 residues that are in close contact with the bound ligands were selected as the binding site residues for our *in-silico* screening. The following residues (hydrophilic residues are shown in bold italics) were used as the binding site residues: I326, M343, L346, **T347**, L349, A350, **D351**, **E353**, W383, L384, I386, L387, M388, G390, L391, **R394**, L402, F404, V418, M421, I424, F425, L428, **K520**, G521, M522, **H524**, L525, **Y526**, M528, V534, L540. Molecular docking was performed using the FlexX module of SYBYL6.9 with standard parameters as implemented in the software. Ligand bound-MDGS complexes as obtained from the molecular-docking procedure were rank ordered based on the FlexX scoring function, since this scoring function has been shown to increase the performance of *in-silico* screening [36]. A list of top 150 compounds was generated based on FlexX binding free energy for each 51 MDGS including rigid crystal structure, which results in a total of 7650 complexes. From the list of 7650 complexes, we identified 582 unique compounds. Each molecular-docking took ~40 h of CPU time on the SGI Octane 2 workstation.

Results and discussions

Structural analysis of MD generated structures

The main focus of this study is to validate the potential of our MD based *in-silico* approach while exploring the flexibility of hER α ligand binding pocket. The analysis of the potential energy of the ER-LBD shows that it is stable during the entire course of MD simulations (Figure 2).

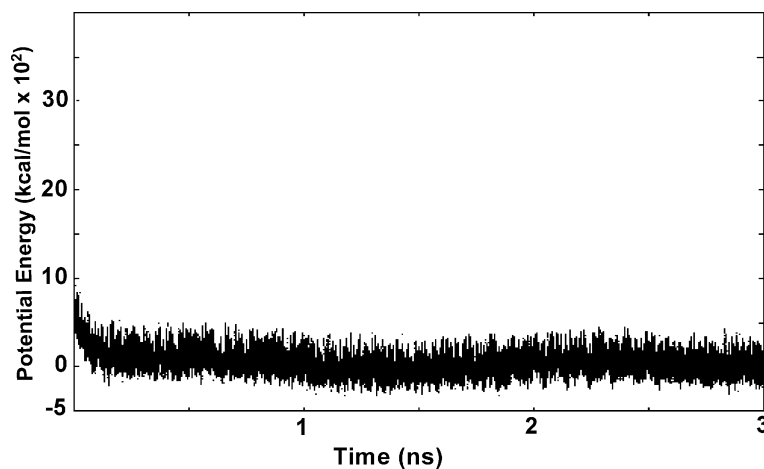


Figure 2. The Potential energy plot vs. simulation time in nanosecond.

Backbone flexibility

In this study, an ensemble of flexible LBP conformations was collected from the energetically favorable MDGS structures, approximately within every 50 ps time starting from 0.2 to 3 ns.¹ Figure 3 shows the NotesPlot comparison of the backbone torsion angles between the 3 ns MDGS and the crystal structure. Major backbone structural changes between the crystal structure and the structure generated at 3 ns were the residues from L327 to E339 that forms the loop between the helices H1 and H2, the residues from R394 to V412 that connects the helices H4/5 and H6, and the residues from M529 to L536 that forms the loop between helices H10/11 and H12. The residues from N455 to T465 that form a part of the dimer interface have significant structural changes between crystal structure and MDGS. These structural differences might be due to the fact that MD simulations were carried out using a monomer. In comparison with the α -helices in the binding pocket between crystal structure to that of MDGS, the Helix-12 which controls the agonist(antagonist conformation of hER α showed a displacement of ~ 4 Å of Helix-12 residues (Figure 4). Further, from the NotesPlot, we identified significant deviations in the backbone torsion angles of the binding site residues I326, R394, V418, M421, L428 and H524. Figure 5 shows the ribbon representation of the crystal structure superimposed on the 3 ns MDGS and the displaced residues identified by the NotesPlot were labeled.

Side chain flexibility

To identify the flexible side chain conformations, we calculated the deviation of the side chain torsion angles ($\Delta\chi$) of the binding site residues between the MDGS and crystal structure using ROTAMER [37]. The residues are set to be flexible, if one of the $\Delta\chi$'s of the residue is equal to or greater than a cut-off value of 25 degrees. Table 1 shows $\Delta\chi_1$, and $\Delta\chi_2$ for 14 binding pocket residues that are identified as flexible residues. From the comparison of ' χ ' angles between the MDGS and that of the corresponding crystal structure, the resulting $\Delta\chi_1$, varies from 10° to 38° , whereas $\Delta\chi_2$ varies from 15° to 79° . Figure 6 shows the stick model representation of the flexible binding pocket residues (for the crystal structure and the 3 ns MDGS) identified from the side chain torsional angle (χ_1 , χ_2) analysis. We found that both backbone and the side chains for residues I326, M421, H524 have maximum variations between MDGS and the crystal structure, which is in accordance with the earlier reported studies [38–40]. Superposition of all the MDGS reveal that the LBP corresponding to the A-ring of the estrogen have minimal structural differences when compared to that of the D-ring binding pocket. The residues I326, F404 (A-ring binding pocket) and M421, H524 and L525 (D-ring binding pocket) exhibit maximal structural variance. To understand the changes in size and shape of the LBP, the Accessible Surface Areas for the 32 LBP residues of all the structures were computed using ACCESS [41]. The peripheral residues I326, M343, M421, K520 and H524 that are in close contact with the D-ring of estradiol were more solvent accessible than the

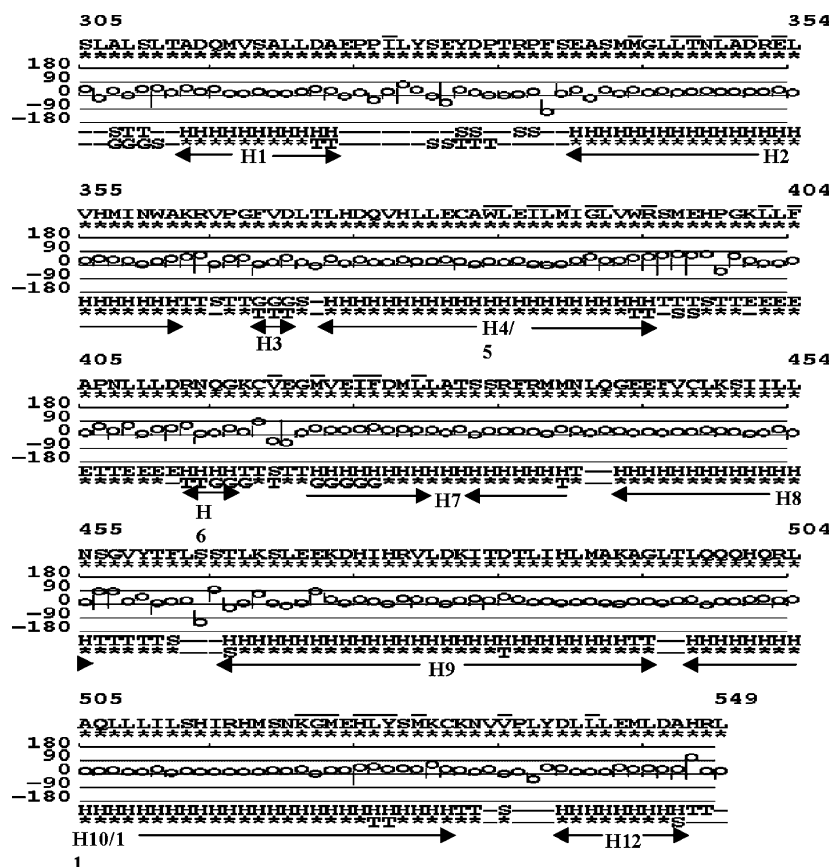


Figure 3. Notesplot representation of backbone torsional angles ($\Delta\phi$ and $\Delta\psi$) of the crystal structure vs. the flexible MDGS at 3 ns. The 32 binding site residues are marked by a dash '-' above the sequence. If the $\Delta\phi$ and $\Delta\psi$ between these two structures are zero (L308), then the conformation of this residue is identical between these structures.

residues that are in contact with the A-ring. The salt bridge interaction between the residues E353 and R394 further contributes to an asymmetric compression of about 3.3% of the solvent accessible area of the A-ring binding pocket. Consequently, the binding site residues forming the A-ring binding pocket are more rigid than the residues forming the D-ring side of the pocket. Only 17 out of 32 residues in the binding site affect the shape and size of the LBP, and H524 is the most flexible residue in the binding pocket. Altogether, these results provide a view of the estrogen-binding pocket having half rigid and half flexible pocket as illustrated by the dashed lines in Figure 8.

In-silico screening

In order to demonstrate that our MD-based approach clearly distinguishes active ER ligands from the inactive ones, we used the database

containing compounds randomly selected from ACD with similar physicochemical properties as that of the active compounds [42]. The database was used by Yang et al. [43] to evaluate the strengths and limitations of GEMDOCK and to compare with several widely used methods (DOCK, GOLD, and FlexX) against ER α . In order to know how the false positive rate might change with the inclusion of more conformations of the ER, we took 10 multiple conformations for every 500 ps interval from the 3 ns simulation. We docked the database with randomly selected compounds into these 10 MDGS conformations. We fix a binding free energy cut-off range of -32.9 to -14.2 kcal/mol to choose false positive rate. This cut-off was based on the binding free energy of range obtained from all the top 150 lists of each MDGS docked complexes using the Endocrine Disrupting Chemicals database. We found that an average of <2% of the random

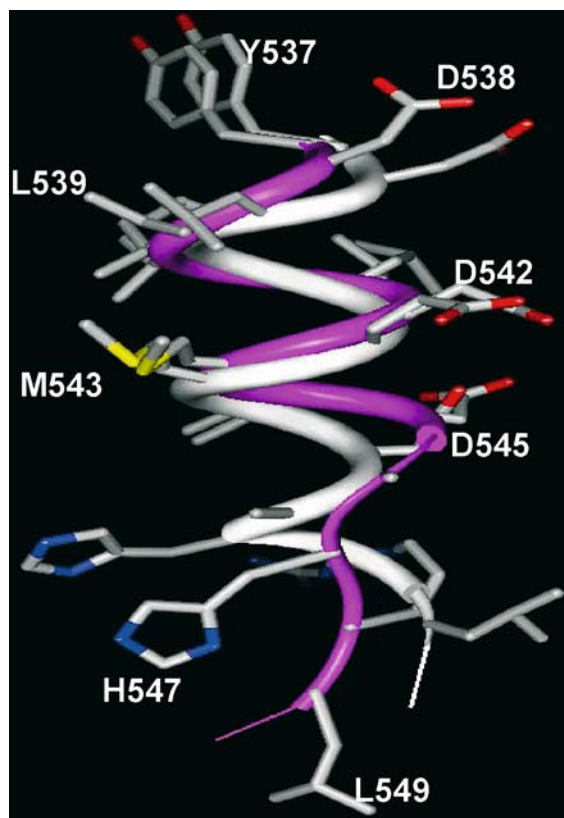


Figure 4. Graphical illustration of the flexibility of the Helix-12 during the MD simulations. Helix 12 from the 3 ns MDGS is colored in magenta and that of the crystal structure in white.

compounds were false positives as shown in Table 2. Hence, this lower false positive range predicted based on this approach does not change much with the inclusion of more ER conformations and therefore gives better performance in our MD-based *in-silico* screening approach. This clearly demonstrates our MD-based approach distinguishes active from the inactive ones.

To study the influence of LBP flexibility of hER α during the *in-silico* screening, 3500 EDCs were docked into the LBP of crystal structure and into all the 51 MDGS. This resulted in 7650 docked ER-ligand complexes comprising of 582 non-redundant compounds. To visualize the solutions from the *in-silico* screening in tandem with physico-chemical properties of the 582 unique compounds, plots of docked receptor-complexes verses physico-chemical properties of the compounds were generated (Figure 7a–c). In these plots, x -axes represents the crystal structure (first column) and 51 MDGS, y -axes represent the 582 unique compounds in the increas-

ing order of either the molecular weight (Figure 7a), or ClogP (Figure 7b) or PSA (Figure 7c). For a particular MDGS, a magenta dot was plotted for the top 150 docked-complexes rank ordered based on FlexX scoring function at the corresponding x_i, y_i coordinates; otherwise a yellow dot was plotted. Similarly, the colored dots were plotted for the crystal structure along the first column in Figures 7a–c. It could be observed that compounds with higher molecular weight cluster along the MDGS at the higher time steps. This is due to the flexibility of the D-ring binding pocket residues allowing MDGS at higher time steps to assume conformations similar to that of partial-antagonist conformation of the receptor. Only few low molecular weight hits were found at higher time steps due to lack of proper interaction with the active residues and poor overall docking scores. From Figure 7a, it is evident that medium molecular weight compounds (200–400 Da) have the strongest agonist interactions with the MDGS. We found that the 142 out of the 582 unique compounds have higher molecular weight (348.3 Da) of the compounds identified by *in-silico* screening against the crystal structure. Similarly, we observed compounds having higher lipophilicity (ClogP) docked preferentially in majority of the MDGS. This complements the lipophilic nature of the receptor cavity. In contrast, a reverse trend has been observed in the polar surface area (PSA) analysis on the unique compounds (Figure 7c). Compounds having lower PSA (low hydrogen bonding capacity) docked preferentially in majority of the MDGS. Since the side chain donor(acceptor groups like H524, has moved significantly from the ligand-binding pocket during the simulation and these groups were not available to form polar contacts. About 70% of the unique compounds having PSA value of (140, and 80% those with ClogP within the range 2–5 and molecular weight (505 Da. Hence, the compounds identified using our MD based *in-silico* screening approach was within the range defined for the drug-like properties. We found few compounds hit all the 51 MDGS. These compounds with medium molecular weight, higher ClogP and low PSA docked well into the flexible ligand-binding pocket of hER α . For example, compound ranked 273 (indicated by an arrow) in the molecular weight plot (Figure 7a); ranked 307 in the ClogP plot (Figure 7b); and ranked 214 in the PSA plot (Figure 7c). Out of 150 top ranking compounds

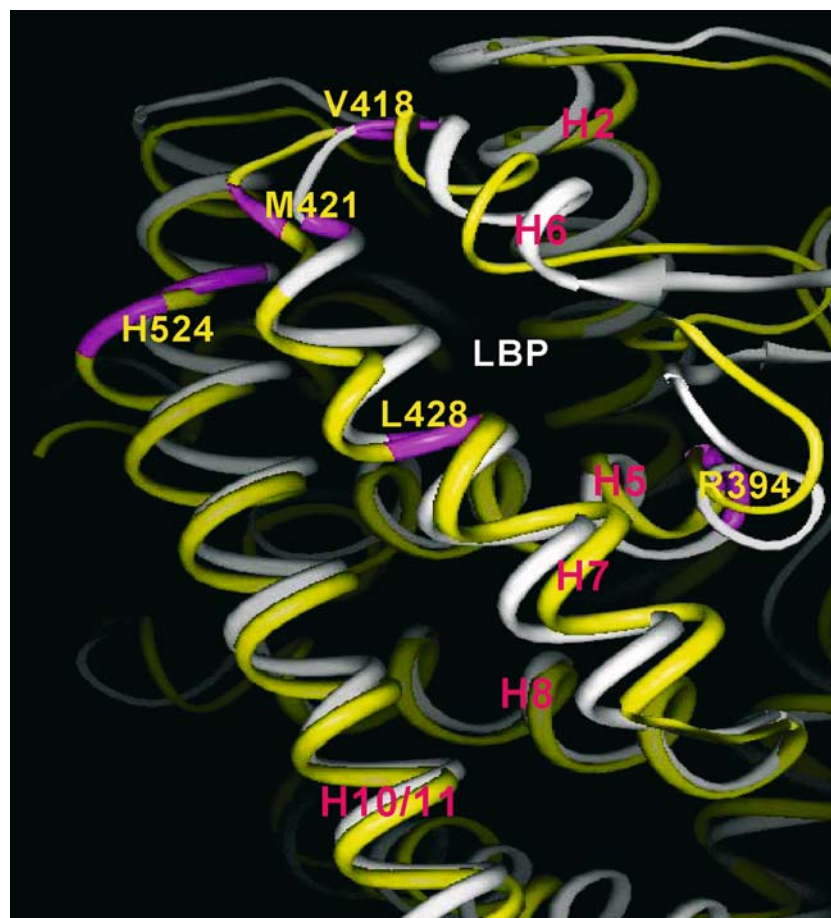


Figure 5. A ribbon representation of superimposed structures of 3 ns MDGS (colored in yellow) with that of the crystal structure (colored in white). Binding pocket residues that showed significant variations are highlighted in magenta color.

Table 1. Deviation in χ_1 , χ_2 angles for the identified flexible binding site residues of MDGS from the crystal structure.

Residues	$\Delta\chi_1^\circ$	$\Delta\chi_2^\circ$
I326	35	20
L349	10	30
D351	20	46
L384	10	32
L402	14	25
F404	36	25
M421	38	23
I424	35	17
F425	29	64
L540	22	36
H524	19	79
L525	10	26
M528	28	39
L540	11	32

that are docked into the crystal structure, only 141 were found in the list of 582 unique compounds that preferentially docked into at least one of the MDGS. Remaining nine compounds that do not appear in the unique list were of low molecular weight and lacked proper interactions with the binding site residues. In addition to these compounds, 441 compounds were identified from the MD-based *in-silico* screening. Clearly, our MD based *in-silico* screening approach has significantly higher hit rates than the rigid crystal structure.

Predictive ability of our MD-based in-silico screening approach

To assess and visualize the predictive ability of our approach, we subjected the top 150 solutions from *in-silico* screening against the rigid crystal structure

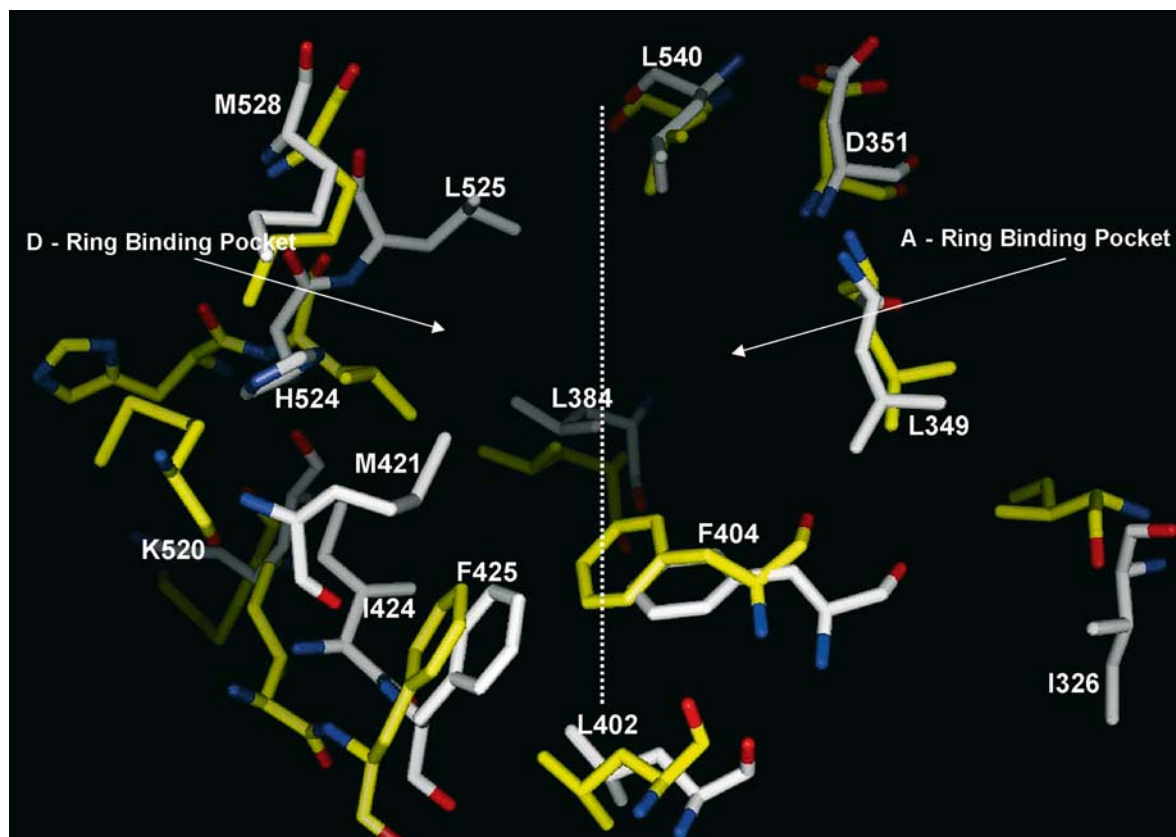


Figure 6. A closer look at the hER ligand-binding pocket. Superimposed structures of 3 ns MDGS (carbon atoms were colored yellow) with the crystal structure (carbon atoms were colored white).

Table 2. Percentage of false positives.

MDGS (ns)	Percentage of false positive ^a
3ERD (X-ray)	9.9
0.53	1.3
0.75	0.8
1.08	0.5
1.24	0.9
1.57	2.7
1.78	2.6
2.03	2.0
2.23	1.4
2.52	2.1
2.73	1.7
2.99	2.1

^aThe false positive compounds are those compounds have binding free energy with the range as observed for the active compounds docked into MDGS flexible binding pockets.

and the flexible LBP conformations to a series of graphical and mathematical analysis. The differences in the binding energies of the 141 compounds

identified in our *in-silico* screening using the crystal structure with that of the MDGS were calculated and plotted in Figure 8. Interestingly, we identified 32 compounds that bind better to the flexible MDGS compared to the corresponding compounds docked into the rigid crystal structure. Compounds with difference in the binding energy closer to zero were structurally similar to DES.

In order to understand the fate of the top 150 compounds docked using crystal structure during the *in-silico* screening against MDGS, we calculated the percentages of occurrence of these compounds in each of the MDGS and were plotted in Figure 9. The percentage of occurrence decreases from 60% to 35% as the number of MD time steps increases. Overall, an average of 45% of the compounds generated from the MD-based flexible *in-silico* screening was found in the top 150 compounds identified from the rigid crystal structure. Moreover, we observed that the compounds that hit first 20 MDGS (i.e. structures generated within 1.5 ns) resemble DES, whereas more divergent set of

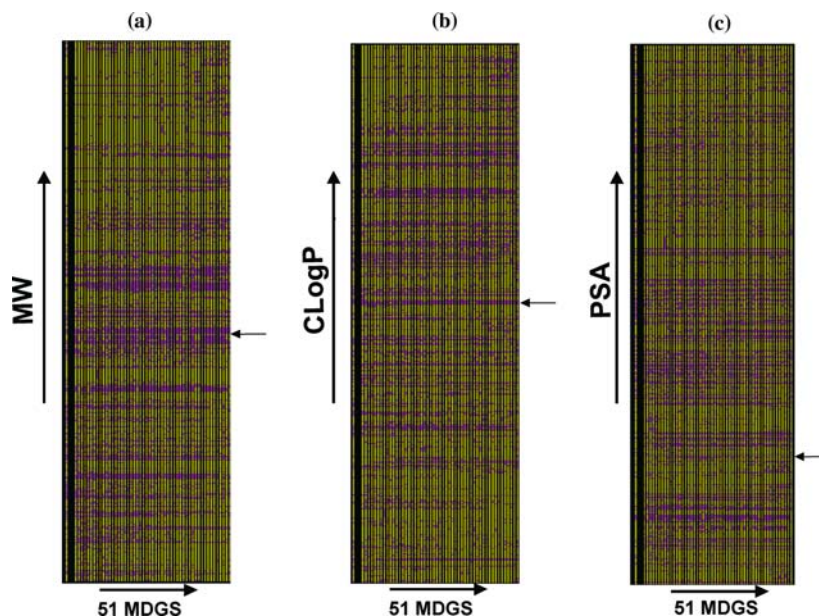


Figure 7a-c. Structure property plots of the small molecules identified using our MD-based flexible docking approach. The x-axes represent the 51 MDGS and the y-axes represent the increasing order of molecular weight, ClogP and polar surface area of the 582 unique compounds, respectively.

compounds appears in the structures generated above 1.5 ns. Hence, in the case of ER, for screening compounds that are closer to the cognate crystal structure ligand, 1.5 ns MD simulation is sufficient for the *in-silico* screening. On the other hand, to obtain more diverse group of compounds than the native ligand, it is adequate to extend the MD simulations at least up to 3 ns.

As discussed earlier, during the course of simulations, the A-ring pocket remains rigid and the

residues pertinent to the D-ring pocket tend to be more flexible. The MDGS structures up to 1 ns represent structure close to the crystal structure. This essentially means that the ensemble of LBP conformations generated via MD simulations contains a pool of rigid, partially flexible and flexible conformations. The 582 unique compounds identified in our *in-silico* screening, hit at least one of the MDGS and in many cases more than one MDGS (Figure 10). It should be noted that 464 compounds

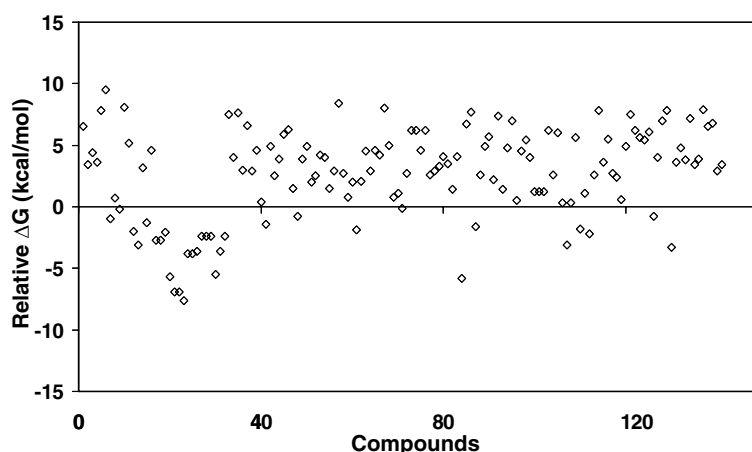


Figure 8. Relative binding free energy resulting from the difference between the ΔG of unique compounds and those docked into the crystal structure.

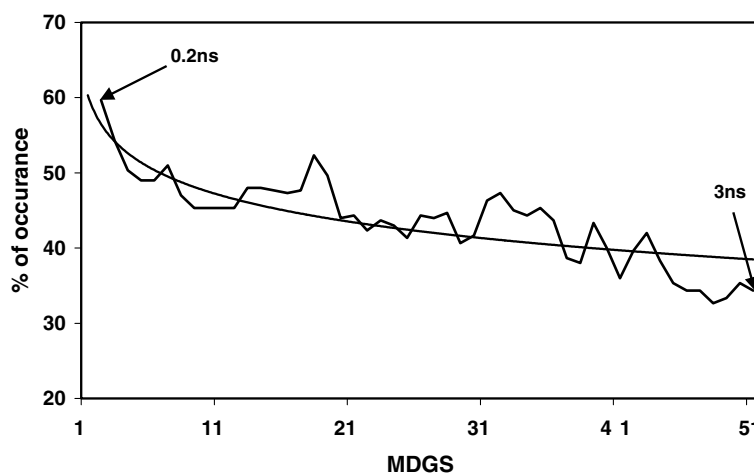


Figure 9. Comparative analysis of the frequency of occurrence of the top 150 hits obtained using 51 flexible MDGS with that of the rigid crystal structure.

hit at least two MDGS, essentially reflecting the fact that the number of sampled LBP conformations is still wide enough to accommodate many of the compounds chosen for the analysis. The compounds that hit all the structures in a particular ensemble tend to resemble the hits obtained by a combination of the smallest LBP subpockets serving as a volume constraint filter. Figure 10 shows that only six compounds hit all the 51 MDGS and seven hit all the 50 MDGS. These low molecular weight compounds are not affected by the flexibility of the binding pocket and thereby survived in both the rigid as well as flexible structures. Similarly, we carried out the analysis against the 51 MDGS for the top 150 compounds identified from the crystal structure. As expected, the compounds that hit at least 45 MDGS were agonists and were either structurally similar or identical for both of the lists analyzed. We observed that the compounds that hit all the 50 and 51 MDGS were identical in both the lists. These analyses provide an excellent example of fine tuning the solutions from *in-silico* screening to identify conformation specific and selective molecules by choosing the combination of energetically favorable ensemble of structures generated from MD simulations.

Identification of structurally diverse ligands

In order to identify structurally diverse compounds, we grouped all the compounds that hit MDGS obtained from 2.5 to 3 ns time steps. From

this list, a total of 32 unique compounds were generated as listed in Table 3. In most cases, the binding free energies of these compounds in complex with MDGS were more favorable than when compared with the crystal structure. As described earlier the rigid nature of the A-ring pocket, selectively restricted all the identified compounds to have at least one phenolic moiety. To assess the affinity of these identified compounds towards the MDGS (Table 3), we compared the ΔG of six compounds that hit all the 51 MDGS and thirteen compounds (with seven unique compounds) that hit at least 50 out of 51 MDGS with the corresponding compounds identified from the crystal structure (Figure 11). In majority of the cases, the calculated ΔG for the MDGS complexes were better than the crystal structure, owing to the conformational flexibility of the ligand-binding pocket of the MDGS.

Furthermore, we observed that more than one MDGS–ligand complexes have better binding energies when compared with the crystal structure–ligand complexes. To analyze the energy trend of these best binding compounds, we took one test case (compound number 2 in Figure 11) and the ΔG of that compound in complex with each 51 MDGS was plotted (Figure 12). From this plot, it was found that for the fifteen cases, it shows better binding free energy than the corresponding compounds docked into the crystal structure. Visual examination of the interactions made by the compounds with the highest ΔG ,

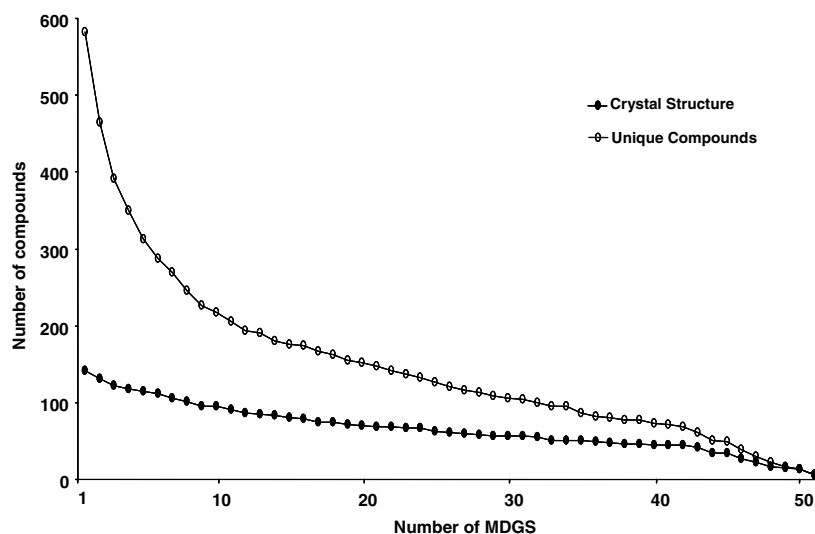


Figure 10. Plot represents the number of compounds from 582 unique lists (shown as unfilled circles) and number of compounds from top 150 lists identified from the crystal structure (shown as filled circles) that are present in any one of the top 150 compounds for all the MDGS. The y-axis denotes the number of compounds and x-axis denotes the number of MDGS.

revealed that they form hydrogen bonding interactions with the residues L346, E353, R394, and strong lipophilic interaction with L525, and stacking interactions with F404, F425. In the case of crystal structure, the same compounds do not form hydrogen bonding interaction with the residue L346; the corresponding lipophilic and stacking interactions were also not strong enough.

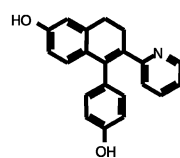
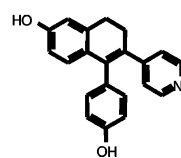
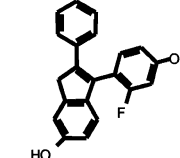
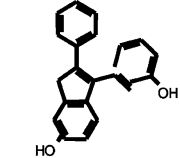
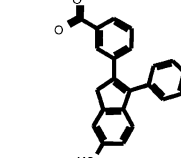
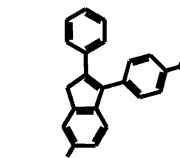
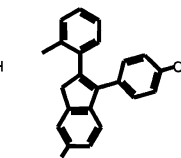
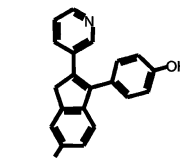
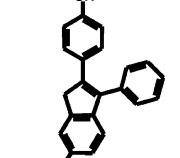
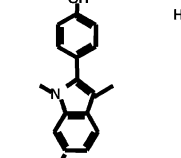
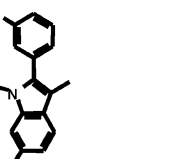
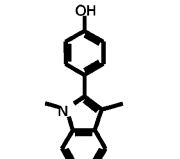
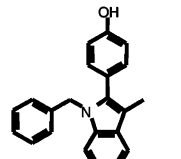
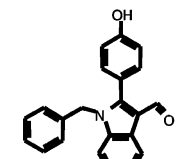
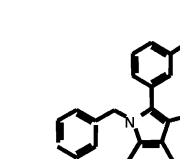
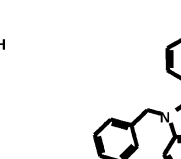
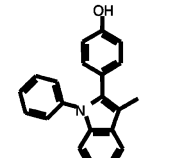
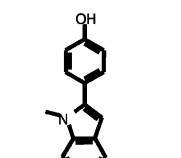
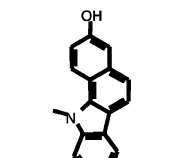
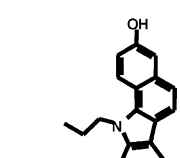
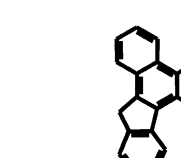
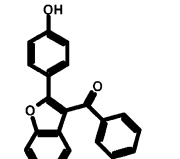
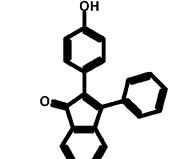
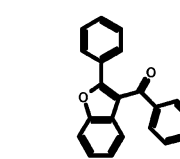
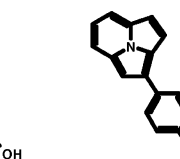
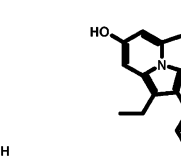
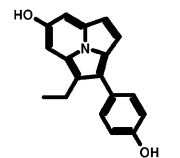
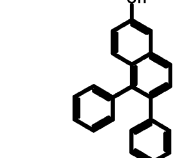
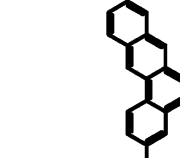
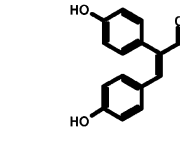
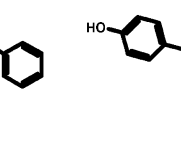
The 32 compounds identified in the present study (Table 3), though chemically diverse, possess the features as that of the natural hormone 17 β -estradiol. It is interesting to note that the MDGS based *in-silico* screening did not identify androgens and ligands of other nuclear receptors, which were enriched in the database used. Diols like 17 β -estradiol, and DES form hydrogen bond with E353 and H524. While examining all the MDGS complexes, it was found that the hydrogen bonding interaction with E353 is conserved. A closer look at the MDGS complexes generated above 0.5ns shows that H524 has moved sufficiently outside the ligand-binding pocket and could no longer make hydrogen-bonding interaction with some of the diol ligands. Though, H524 could no longer form hydrogen bonds with agonists, the calculated ΔG was better as the lipophilic interaction energies largely contribute to the binding energy. Moreover, H524 flexibility allows monohydroxy benzenes with steric substituents to be

accommodated. Additionally, as the number of time steps increases, the hER α structures tend to pick partial agonists. This is indicative of the role of H524 fluxionality in triggering receptor agonism/antagonism. These results were supported by the wealth of information in the literature about the importance of H524 and Helix 12 in altering estrogen receptor conformations [22, 44–47].

Implications for the structure-based drug design

ER is an intricate target for the structure-based drug design because of the flexibility of the ligand-binding pocket as discussed above. The standard structure-based drug design utilizing a rigid binding site yields comparatively high success rates in predicting near native binding modes and hence the resulting lead compounds structurally similar to the bound ligand. We have demonstrated the utility and discovery of 441 structurally unique compounds in addition to the 141 hits that bind to the rigid crystal structure using an ensemble of flexible LBP conformations from the MD simulations. Clearly, the hER α flexibility has a significant impact on the hit rates that is higher than the rigid crystal structure. Probably the concern however is whether the available ER crystal structures are sufficient to cover a wide range of

Table 3. The 2D chemical structures of identified compounds with their binding free energy of the corresponding MDGS complex and crystal structure complex. (shown in parenthesis)

					
-25.7 (-25.5)	-25.9 (-23.9)	-27.9 (-25.2)	-32 (-26.3)	-29.2 (*)	
					
-30.6 (-23.7)	-28.6 (*)	-31 (-23.4)	-28.3 (-24.5)	-24.6 (-27.1)	-25.7 (-24.9)
					
-24.2 (-27.7)	-27.3 (*)	-27.9 (*)	-29.2 (*)	-27.9 (*)	
					
-30.2 (*)	-24.2 (-28.7)	-27.6 (-25.7)	-24.2 (-26.3)	-26.6 (*)	
					
-27.1 (*)	-24.2 (*)	-28.2 (-24.9)	-24.7 (-25.3)	-26.3 (*)	
					
-24.9 (-29.8)	-29.5 (-31.5)	-26.2 (-27.3)	-26.5 (-31.7)	-27.9 (-25.7)	

* - Lead compounds were not present in the corresponding top ranked crystal structure compounds

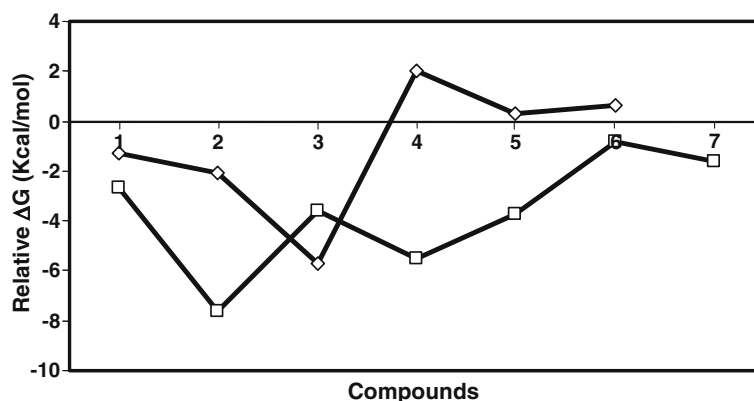


Figure 11. Comparison of ΔG of the six compounds that present in all 51 MDGS and seven that present in 50 of 51 MDGS, with the corresponding compounds of the crystal structure.

potentially binding-competent conformations. MD simulations improve this situation through the generation of ensembles of diverse conformational substates on a nanosecond time scale, thereby, covering a wide spectrum of accessible binding-site conformations. As expected in any protein with modular architecture, hER α structure is stable during the simulation period and has resulted in conformations that are sufficiently diverse, while maintaining the RMS deviations within reasonable boundaries during the simulation period. With respect to the binding site, it is compatible with the experimental data that different degrees of mobility are observed, for example, a relatively rigid site on the A-ring site and a highly flexible H524 on the D-ring site.

Conclusions

We have developed an efficient approach to perform *in-silico* screening of large database of flexible ligands against flexible LBPs. We have clearly documented the flexibility of the LBP of hER α from the analysis of backbone and side chain torsional angles. In particular, the residues D351, I326, F404, M421, I424, F425, H524, M528, and L540 were highly flexible and in turn influence the ligand binding. The nature of the binding pocket was characterized by a stable region surrounding the A-ring site and a highly mobile area around H524 in the D-ring site. This residue itself plays the dominant role in the binding-site adaptations; moreover, H524 flexibility allows monohydroxy

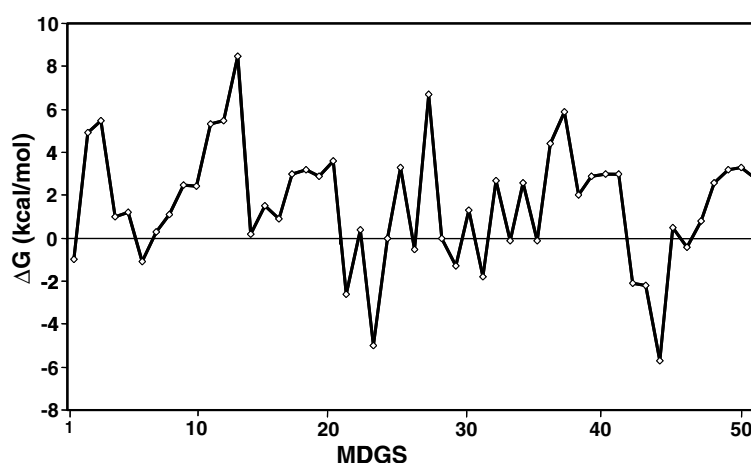


Figure 12. Difference in the binding free energies of the test compound in complex with each of the MDGS at different MD-time steps.

benzenes with steric substituents to be accommodated. Furthermore, the observed asymmetric compression of receptor binding pocket during the simulation period is representative of crucial molecular events that trigger ER-dependent signaling cascade. The flexibility of the binding pocket provides an explanation for the higher affinity of the compounds identified in the study with respect to DES. We have identified a diverse set of 441 unique compounds showing clear emphasis on hER α flexibility. Finally, in our *in-silico* screening, we have identified 32 compounds representative of seven chemically divergent pharmacophores and structurally similar to natural hormone. The present study could be extended to find hER α antagonists by continuing the MD simulations to generate partial agonists/antagonist ensembles of ER-LBD. Qualitatively, our results support a view of binding-site adaptations as resulting from preferential selection of binding-competent conformations from the ensemble. This approach in conjunction with grid based distributed computing [48], which largely reduces the computer time, has the likelihood of becoming an indispensable, and complimentary tool in structure based drug discovery.

Note

The 51 MD generated structures with their corresponding time steps ordered (in ns) as below: 0.2, 0.53, 0.57, 0.62, 0.67, 0.75, 0.79, 0.82, 0.89, 0.94, 0.99, 1.03, 1.08, 1.12, 1.19, 1.24, 1.29, 1.31, 1.37, 1.41, 1.47, 1.52, 1.57, 1.61, 1.67, 1.73, 1.78, 1.83, 1.88, 1.93, 1.99, 2.03, 2.08, 2.14, 2.18, 2.23, 2.28, 2.33, 2.39, 2.44, 2.48, 2.52, 2.58, 2.63, 2.68, 2.73, 2.79, 2.84, 2.89, 2.94, 2.99

Acknowledgement

We acknowledge the National Cancer Institute (NCI) for allocation of computing time and staff support at the Advanced Biomedical Computing Center of the National Cancer Institute, Frederick.

References

- Fischer, E., Ber.DtschChem. Ges., 27 (1894) 2985.
- Koshland, G.E., Proc Natl. Acad. Sci (USA), 44 (1958) 98.
- Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comp. Chem., 13 (1992) 505.
- Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
- Pattabiraman, N., Levitt, M., Ferrin, T.E. and Langridge, R., J. Comp. Chem., 6 (1985) 432.
- Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., J. Comp. Chem., 19 (1998) 1639.
- Ewing, T.J.A. and Kuntz, I.D., J. Comp. Chem., 18 (1997) 1175.
- Kramer, B., Metz, G., Rarey, M. and Lengauer, T., Med. Chem. Res., 8 (1999) 463.
- Ronald, M.A. and Knegetel, M.W., proteins Struct. Funct. Genet., 37 (1999) 334.
- Jones, G., J. Mol. Biol., 267 (1997) 727.
- Jiang, F. and Kim, S.H., J. Mol. Biol., 219 (1991) 79.
- Sandak, B., Nussinov, R. and Wolfson, H., J. Comput. Appl. Biosci., 1: (1995) 87.
- Vakser, I.A.P.d.f.l.-r.s., Protein Eng., 8 (1995) 371.
- Walls, P.H. and Sternberg, M.J., J. Mol. Biol., 228 (1992) 277.
- Gschwend, D.A., Good, A.C. and Kuntz, I.D., J. Mol. Recog., 9 (1996) 175.
- Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham III, T.E., Wang, J., Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., Vincent, J.J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R.J., Duan, Y., Pitera, J., Massova, I., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A., AMBER 7, University of California, San Francisco, 2002.
- Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., J. Comp. Chem., 4 (1983) 187.
- van Gunsteren, W.F. and Berendsen, H.J.C., Angew.Chem. Int. Ed. (Engl.), 29 (1990), 992.
- Oostenbrink, C. and van Gunsteren, W.F., Proteins Struct. Funct. Genet., 54 (2004) 237.
- Junmei, W., Paul, M., Wei, W. and Kollman, P.A., J. Am. Chem. Soc., 123 (2001) 5221.
- Meegan, M.J. and Lloyd, D.G.W., Advances in the science of estrogen receptor. Curr. Med. Chem., 10 (2003) 181.
- Sivanesan, D., Rajnarayanan, R.V., Doherty, J. and Pattabiraman, N., Pro. Am. Assoc. Can. Res., 44 (2004) LB-185.
- Blair, R.M., Fang, H., Branham, W.S., Hass, B.S., Dial, S.L., Moland, C.L., Tong, W., Shi, L., Perkins, R. and Sheehan, D.M., Toxicol. Sci., 54 (2000) 138.
- van Lipzig, M.M., Ter Laak, A.M., Jongejan, A., Vermeulen, N.P., Wamelink, M., Geerke, D. and Meerman, J.H., J. Med. Chem., 47 (2004) 1018.
- Oostenbrink, C., Pitera, J.W., Marola, M.H., van Lipzig, M.M., John, H., Meerman, N. and van Gunsteren, W.F., J. Med. Chem., 43 (2000) 4594.
- Nam, K., Marshall, P., Wolf, R.M. and Cornell, W., Biopolymers, 68 (2003) 130.
- Laila, A., Abou-Zeid, and Abdalla, M.E., J. Mol. Struct. (Theochem.), 593 (2002) 39.
- Yoon, S. and Welsh, W.J., J. Chem. Inf. Comput. Sci., 44 (2004) 88.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., Nucleic Acids Res., 28 (2000) 235.
- Shiau, A.K., Barstad, D., Radek, J.T., Meyers, M.J., Nettles, K.W., Katzenellenbogen, B.S., Katzenellenbogen,

- J.A., Agard, D.A. and Greene, G.L., *Nat. Struct. Biol.*, 9 (2002) 359.
31. Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A. and Greene, G.L., *Cell*, 95 (1998) 927.
32. Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C., *J. Comput. Phys.*, 23 (1977) 327.
33. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R., *J. Chem. Phys.*, 81 (1984) 3684.
34. Kabsch, W. and Sander, C., *Biopolymers*, 22 (1983) 2577.
35. Sadowski, J. and Gasteiger, J., *Chem. Rev.*, 93 (1993) 2567.
36. Stahl, M. and Rarey, M., *J. Med. Chem.*, 44 (2001) 1035.
37. Lovell, S.C., Word, J.M., Richardson, J.S. and Richardson, D.C., *Proc. Natl. Acad. Sci. (USA)*, 96 (1999) 400.
38. Hanson, R.N., Lee, C.Y., Friel, C.J., Dilis, R., Hughes, A. and DeSombre, E.R., *J. Med. Chem.*, 462 (2003) 865.
39. Nicolas, F., Elie, S., Anne, V., Emmanuel, S., Jan-Martin, H. and Michel, H., *J. Chem. Bio. Chem.*, 4 (2003) 494.
40. Anstead, G.M., Carlson, K.E. and Katzenellenbogen, J.A., *Steroids*, 62 (1997) 268–303.
41. Lee, B. and Richards, F.M., *J. Mol. Biol.*, 55 (1971) 379.
42. Bissantz, C.F. and Rognan, G. D., *J. Med. Chem.*, 43 (2000) 4759.
43. Yang, J. M.S. and , T.W., *Proteins*, 25 (2005) 777.
44. Mattras, H., Aliau, S., Richard, E., Bonnafous, J.C., Jouin, P. and Borgna, J.L., *Biochem.*, 41 (2002) 15713.
45. Sumbayev, V.V., Bonefeld-Jorgensen, E.C., Wind, T. and Andreassen, P.A., *FEBS Lett.*, 579 (2005) 541.
46. Aliau, S M.H., Richard, E., Bonnafous, J.C. and Borgna, J.L., *Biochemistry*, 41 (2002) 7979.
47. Borgna, J.L. S.J., *Eur. J. Biochem.*, 199 (1991) 575 .
48. Richards, W.G., *Nat. Rev. Drug Discov.*, 7 (2002) 551.