Pharmaceutical Research, Vol. 26, No. 10, October 2009 (© 2009)
DOI: 10.1007/s11095-009-9941-z

Research Paper

Application of Screening Methods, Shape Signatures and Engineered Biosensors in Early Drug Discovery Process

Izabela Hartman,[1,2] Alison R. Gillies,[2] Sonia Arora,[3] Christina Andaya,[3] Nitya Royapet,[5] William J. Welsh,[3,4] David W. Wood,[2,6] and Randy J. Zauhar[1,6]

Received March 20, 2009; accepted July 6, 2009; published online July 22, 2009

Purpose. In this study, two unreported estrogen antagonists were identified using a combination of computational screening and a simple bacterial estrogen sensor.
Methods. Molecules here presented were initially part of a group obtained from a library of over a half million chemical compounds, using the Shape Signatures method. The structures within this group were then clustered and compared to known antagonists based on their physico-chemical parameters, and possible binding modes of the compounds to the Estrogen Receptor α (ERα) were analyzed. Finally, thirteen candidate compounds were purchased, and two of them were shown to behave as potential subtype-selective estrogen antagonists using a set of bacterial estrogen biosensors, which included sensors for ERα, ERβ, and a negative control thyroid hormone β biosensor. These activities were then analyzed using an ELISA assay against activated ERα in human MCF-7 cell extract.
Results. Two new estrogen receptor antagonists were detected using *in silico* Shape Signatures method with an engineered subtype-selective bacterial estrogen biosensor and commercially available ELISA assay. Additional thyroid biosensor control experiments confirmed no compounds interacted with human thyroid receptor β.
Conclusions. This work demonstrates an effective combination of computational analysis and simple bacterial screens for rapid identification of potential hormone-like therapeutics.

KEY WORDS: antagonist; drug discovery; estrogen receptor; estrogen receptor biosensor; shape signatures.

INTRODUCTION

Estrogen receptors (ERs) belong to the family of Nuclear Hormone Receptors (NHR). In humans, two ER subtypes, α and β, are known at this time (1). Their major function is regulation of transcription in response to binding of small hormones and hormone-like molecules. Similarly to other NHRs, both ERα and ERβ contain six domains (A-F). Homologous domains of the two ER subtypes have similar sequences, with sequence identity ranging from 18 to 97%. The Ligand Binding Domain (LBD) has 59% sequence identity between both ER subtypes. However, those residues in the vicinity of the ligand binding site are highly conserved, with substitutions only observed at two positions: residues

[1] Department of Chemistry & Biochemistry, University of the Sciences in Philadelphia, 600 S. 43rd Street, Philadelphia, Pennsylvania 19104, USA.
[2] Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544, USA.
[3] Department of Pharmacology, University of Medicine and Dentistry of New Jersey-Robert Wood Johnson Medical School (UMDNJ-RWJMS), New Brunswick, New Jersey, USA.
[4] UMDNJ Informatics Institute, 675 Hoes Lane, Piscataway, New Jersey 08854, USA.
[5] Mercer County Community College, 1200 Old Trenton Road, West Windsor, New Jersey 08550, USA.
[6] To whom correspondence should be addressed. (e-mail: r.zauhar@usip.edu, dwood@princeton.edu)

ABBREVIATIONS: ANOVA, Analysis of variance; BPA, Bisphenol A; COMFA, Comparative molecular field analysis; DMSO, Dimethylsulfoxide; ELISA, Enzyme-linked immunosorbent assay; ERs, Estrogen receptors; E2, 17-β-estradiol; GOLD, Genetic optimization for ligand docking algorithm; HBA, Number of hydrogen bonding acceptors; HBD, Number of hydrogen bonding donors; HRP, Horseradish peroxide; LB, Luria-Bertani; LBD, Ligand binding domain; logp, Octanol-water partition coefficient; MOE, Molecular operation environment; MEP, Molecular electrostatic potential; MW, Molecular weights; NCI, National cancer institute; NHR, Nuclear hormone receptors; NSC, National service center; OD, Optical density; pMIT::ER, **p**lasmid **M**BP (maltose-binding protein tag) -**I**ntein-**T**S (thymidylate synthase)::**E**strogen **R**eceptor; pMIT::TR, **p**lasmid **M**BP (maltose-binding protein tag) -**I**ntein-**T**S (thymidylate synthase)::Thyroid Receptor; PSA, Polar surface area (Å^2); QSAR, Quantitative structure-activity relationship; RB, Number of rotatable bonds; SERMs, Selective estrogen receptor modulators; TTM, Medium which contain -Thy medium supplemented with 10 μg/ml trimethoprim and 50 μg/ml thymine; UPGMA, Unweighted pair group method with arithmetic means; +THY, A **thy**mine-rich medium; used as a positive control; -THY, A **thy**mine-free medium; used as a negative control.

Leu384 and Met421 in ERα are replaced by Met336 and Ile373 in ERβ (2). This local modification is expected to alter the volume of the ligand-binding pocket from 490 $Å^3$ in ERα to 390 $Å^3$ in ERβ (3). These differences allow some compounds to bind selectively to one ER subtype over another, and these compounds are therefore known as Selective Estrogen Receptor Modulators (SERMs). In many cases, SERMs act like estrogen in some tissues but block the effect in others. They include a wide variety of structures with varying binding specificity, including benzopyran derivatives and benzothiophene derivatives (4–10). In addition, some compounds, such as Faslodex, and GW7604, cause degradation of the target receptor and are therefore known as Selective Estrogen Receptor Downregulators (SERDs) (5,11,12).

In this work, the LBD (known also as the E domain) of ERα was closely studied to determine possible interactions with molecules that resemble known antagonists. The Shape Signatures method was used to identify a novel group of compounds that could potentially bind selectively to the ERα-LBD. This is a computational screening tool, which allows for fast comparison of the shapes and electrostatic properties of molecules and receptor sites (13). Initially, both shape and molecular electrostatic potential of over a half million compounds derived from the National Cancer Institute (NCI), Ryan Scientific and Sigma-Aldrich databases, which are included in the ZINC database of commercially-available compounds, were compared to known SERMs and SERM-like compounds (14). The predictive accuracy of the Shape Signatures method has been statistically analyzed in previous work, where it was shown to be highly efficient in the identification of agonist molecules in a group of randomly selected compounds from the NCI database, as well as differentiation of agonist and antagonist molecules for the serotonin receptor (15). In order to identify novel compounds that might act as SERMs, the Shape Signatures method was trained using a set of known SERM molecules, as well as a known SERD (GW7604) (5).

The initial group of known SERM compounds is here referred to as queries. The molecules with the highest resemblance to queries were additionally examined by analyzing their physico-chemical properties to predict their solubility and permeability. The binding modes of ligand-ERα complexes were also predicted using the Genetic Optimization for Ligand Docking algorithm (GOLD; Cambridge Crystallographic Data Centre) (16,17). Two scoring schemes are used by the GOLD software to estimate the ligand-protein binding affinity, GOLD Score and Chem Score, and both were applied for each ligand considered. GOLD is a ligand-receptor docking method used for high-throughput library screening, which allows limited flexibility in the protein side chains and full flexibility of the ligand (18–21). GOLD Score is the sum of the calculated values: 1—hydrogen bond score, 2—van der Waals score, 3—internal hydrogen bond score and 4—sum of internal torsion and van der Waals energy terms. In the final GOLD Score calculation, the intermolecular van der Waals score is multiplied by 1.375, which is an empirical correction to maximize hydrophobic interaction between protein residues and ligand (22). Chem Score is also a fitness function. It uses multiple linear regression to predict free energy of binding of protein-ligand complex and additionally calculates clashes that occur between protein and ligand as well as the internal energy of the ligand. The sum of those three variables is described as the Chem Score (23). By comparing those results, several compounds were predicted to act as estrogen antagonists. These molecules were obtained and screened using a novel subtype-selective bacterial estrogen biosensor (24–26). Based on the observed results of the screen, two estrogen antagonists were subsequently tested and confirmed by an ELISA assay using nuclear extract from the human breast cancer cell line MCF-7 (NR Peptide ERα ELISAs, Active Motif, Carlsbad, CA).

MATERIALS AND METHODS

From in Silico Screening Using Shape Signatures to Multivariate Data Analysis

The Shape Signatures method has been described elsewhere (13,15,27). In brief, the starting point for the technique is to create a triangulated solvent accessible surface for each compound included in the study. Next, for each compound a reflecting ray is initiated from a random point on the surface of the molecule and internally propagated using the rules of optical reflection until 50,000 ray-trace segments are accumulated. The details of the Shape Signatures ray-tracing method, including diagrams to describe the geometry of the ray-trace methodology and optimization of ray density inside the molecule, have been previously shown and discussed in several references (13,15,27). Probability distributions are derived from the ray-trace and encode shape and property information. These distributions are stored as simple histograms and are referred to as Shape Signatures. They can be rapidly compared and can be used to score compounds for similarity on the basis of shape alone or in combination with other properties, such as molecular electrostatic potential (MEP) or lipophilicity, which can be measured on the molecular surface. Both one-dimensional (1D) signatures, which code only shape information, and two-dimensional (2D-MEP) signatures, which code both shape and electrostatic potential information, were computed for query and library compounds. The molecular electrostatic potential is computed using Columbic law at the ray-trace reflection points distributed on the surface of a molecule. In this work, query compounds presented in Fig. 1 were selected: 4-[(1Z)-1-[4-[2-(dimethylamino)ethoxy]phenyl]-2-phenyl-1-butenyl] phenol (4-hydroxytamoxifen), 3-[(1E)-1-[4-[2-(dimethylamino)ethoxy]phenyl]-2-phenyl-1-butenyl] phenol (droloxifene), 2-[4-[(1Z)-4-chloro-1,2-diphenyl-1-butenyl]phenoxy]-N,N-dimethyl ethanamine (toremifene), (S)-3-(4-hydroxyphenyl)-4-methyl-2-[4-[2-(1-piperidinyl) ethoxy]-phenyl]-2H-1-benzopyran-7-ol (EM652) and (2E)-3-[4-[(1E)-1-(4-hydroxyphenyl)-2-phenyl-1-butenyl]phenyl]-2-propenoic acid (GW7604) (2,11,12,28). All of these compounds are either known therapeutics for breast cancer or are currently in clinical trials. These queries were compared to compounds derived from the NCI, Ryan Scientific and Sigma-Aldrich databases and used to select library compounds based on their similar shape to a query or by similarity in shape and MEP. For simplicity, compounds that were derived from the NCI database are named based on their internal identification number (NSC; National Service Center).

Compounds from NCI database, which was our first analyzed set, contained only those stereoisomers which were available from the source. Partial charges were assigned using Sybyl 6.8 (Tripos Inc., St. Louis, MO). Compounds derived from ZINC are prepared as it is described in detail by Irwin and Shoiche and include proper protonating forms, a variety of SMILEs strings as well as partial atomic charges (14). However, ZINC database was extended by the generation of a wide variety of stereoisomers and conformations using STERGEN (Molecular Networks GmbH, Erlangen, Germany) and Molecular Operation Environment (MOE; Chemical Computing Group Inc.) (27). These increases allowed a significant expansion of the possible binding modes of the ZINC database compounds. The goal was to explore possible active forms of compounds, which were not detected previously. Structures of similar shape and MEP are likely to share similar biological activity. For the 1D and 2D search, Shape Signatures scores between 0.05–0.1 and 0.1–0.2, respectively, are an indication that structures strongly resemble the query, and these compounds were considered hits based on previously reported data (13). Commercial availability of the compounds, as well as Shape Signatures scores, were the primary criteria used for initial selection of compounds for higher evaluation. Computations were performed using a sixteen-processor Beowulf cluster running the Linux operating system.

The initial set of hit compounds was further filtered by clustering structures with similar properties. Both the Unweighted Pair Group Method with Arithmetic means (UPGMA) and K-means clustering methods (SPSS Inc., Chicago, IL) were applied for this purpose (29). With UPGMA the calculation of similarity of groups within compounds was based on the Squared Euclidean metric, which calculates the distance (D) between objects (compounds) based on standardized variable similarity. Longer distances indicate lesser similarity. In K-means analysis, the distances between the center of the cluster and the objects were calculated as the square of the Euclidean distance. Final cluster assignments obtained were evaluated by analysis of variance (ANOVA) within and between clusters, which provided a quantitative measure (F score) for each variable considered (30). Larger F values for a specific variable denoted its greater importance in defining the clusters.

Both clustering methods were used to find the most self-similar groups of hit compounds by comparing their molecular weights (MW), number of hydrogen bonding donors (HBD), number of hydrogen bonding acceptors (HBA), number of rotatable bonds (RB) and polar surface area (PSA; $Å^2$). In this particular step of the study, the octanol-water partition coefficient (logp) was omitted for compounds derived from the Ryan Scientific and Sigma-Aldrich databases due to the fact that some of the compounds were not neutral. However, the logp value was included in multivariate data analysis for structures derived from NCI database, and only the ones most similar to the query compounds were subjected to docking studies. For compounds derived from the Ryan Scientific and Sigma-Aldrich databases, multivariate data analysis and docking (described below) were performed simultaneously. The literature for the hits was reviewed when available.

Docking

The potential binding modes of the hits and queries in the active site of the ERα were identified using GOLD. 3ERT, a crystallographic structure of ERα bound to 4-hydroxytamoxifen, was obtained from the Protein Data Bank (31,32). The docking modes of the ligands in the active site of the receptor, as well as specific interactions between ligand functional groups and the receptor, were predicted and analyzed using the GOLD software based on this structure (18–21). Two fitness functions were calculated by GOLD for each ligand considered, GOLD Score and Chem Score (16,21,23). Both of them gave fitness scores for the predicted orientation of the ligand in the active site pocket of the receptor. The standard cutoff settings for the GOLD method were as follows: RMSD = 1.5 Å; maximal number of operations = 100000; population size = 100; selection pressure = 1.1; number of islands = 5; crossover = 95; mutate = 95; migrate = 10; niche size = 2; hydrogen bonding = 2.5 Å; and van der Waals interactions = 4.0 Å. The scores predicted by the GOLD method for the training and test sets were ranked. The orientation of molecules relative to the active pocket of the receptor and the predicted binding modes were then compared for each of the structures in the test and training sets. Compounds with high GOLD and Chem Scores that closely matched the expected orientation, based on the training set and their interactions with residues Glu353, Asp351 and Arg394, were chosen for further analysis. Among test compounds the minimum GOLD and Chem Score cutoffs were 40 and 29.9, respectively.

Some preparation of the protein structure was required before the docking process took place. Specifically, water was removed and hydrogens were added to the residues using Sybyl 6.8 (Tripos Inc., St. Louis, MO) (33). The molecules used in our Shape Signatures searches, both query compounds and hits, were docked to the target receptor structures using GOLD. The residues of the receptor's active site were extracted from the rest of the protein by including only those residues within 10 Å of the ligand (4-hydroxytamoxifen) included in the experimental structure 3ERT. Next, the ligand inside of the receptor was removed, and the coordinates of the active site center were calculated to define the active site of the receptor for GOLD studies. The quality of the predicted 4-hydroxytamoxifen binding mode was evaluated by comparison to the experimental structure, therefore this compound was used as a positive control for the experimental portions of this study.

Evaluation of Hit Compounds Using Subtype-Selective ER Biosensors

For experimental evaluation of hit compounds, *Escherichia coli* D1210*ΔthyA*::Kan^R [F$^-$ *Δ*(*gpt-proA*)62 *leuB*6 *supE*44 *ara*-14 *galK*2 *lacY*1 *Δ*(*mcrC-mrr*) *rpsL*20 (*Str*^{r}) *xyl*-5 *mtl*-1 *recA*13 *lacI*^{q}] cells were transformed with either plasmid pMIT::ERα*, pMIT::ERβ* or pMIT::TRβ* as reported previously (24,25). The pMIT::ER* (or TR*) acronym is an abbreviation of Maltose binding protein-Intein- Thymidylate Synthase::Estrogen Receptor (or Thyroid Receptor). The pMIT::TRβ* biosensor strain is used as a negative control to verify that the growth effect of the biosensor is specific to the ER LBD and does not

arise from a more general impact on the *E. coli* host cell metabolism. Transformants were grown overnight in 5 ml cultures of Luria-Bertani (LB) medium [LB medium: 1% tryptone, 0.5% yeast extract, 1% NaCl, *w*/*v*] supplemented with 200 μg/ml ampicillin and 50 μg/ml thymine in a shaking water bath at 37°C. Cells carrying the pMIT::ERα*, pMIT::ERβ* or pMIT::TRβ* plasmids were grown to OD_{600} 0.9, 1.3–1.7 and 1.1, respectively. These cultures were then diluted 1:200 into 5 ml of thymineless medium [-Thy (per liter): 10 ml of 10% casamino acid, 10 ml of 20% glucose, 200 μL 1% thiamine HCl, 200 ml of 5×Minimal Davis Broth (MDB; 35 g dipotassium phosphate, 10 g monopotassium phosphate, 2.5 g sodium citrate, 0.5 g magnesium sulfate and 5 g ammonium sulfate), 10 ml of Thy Pool (2 mg/ml of each of following amino acids, L-Arg, L-His, L-Leu, L-Met, L-Pro and L-Thr), 1 ml of 0.1 M $CaCl_2$, pH 7.0] supplemented with 200 μg/ml ampicillin. Growth of the resulting cultures was then monitored by optical density measurements in the presence of specified concentration of control and test compounds.

Several control compounds known to bind ER or TR selectively were purchased: bisphenol A from ICN Biomedicals (Aurora, OH), raloxifene, tamoxifen, 17-β-estradiol as well as 3, 3', 5-triiodothyroacetic acid, 95% (triac) from Sigma (Saint Louis, MO), and 3, 3', 5-triiodo-L-thyronine sodium salt hydrate, 95% (T3) from Aldrich Chemical (Milwaukee, WI).

For simplicity, the library of thirteen test compounds was renamed **A**-**M** during the biosensor test. The structures of **A**-**M** molecules along with their original corresponding database designations are presented in Fig. 2. The stock concentrations of the test compounds were as follows: 10 mM in dimethylsulfoxide (DMSO) (compounds **B** and **D**-**L**), 9 mM in DMSO (compounds **A** and **M**) and 5 mM in a 50/50 mixture of DMSO and ethanol (compound **C**). The group of control compounds dissolved in 100% ethanol contained triac, T3, tamoxifen, raloxifene, bisphenol A (BPA) and 17-β-estradiol (E2). The stock concentrations of these compounds were 10 mM. The final solvent concentration in the cell growth tests was limited to 1% to minimize vehicle impact on growth curves. Optical densities were measured on a GENESYS™ 2 spectrophotometer at a wavelength of 600 nm.

To study antagonism, 5 ml -Thy cultures were supplemented with 500 nM and 10 μM 17-β-estradiol in the presence of ERβ and ERα, respectively. In these studies, the concentration of test compounds was set to 5 μM for ERβ and 10 μM for ERβ and ERα. For direct test of antagonism, TTM medium consisted of -Thy medium supplemented with 10 μg/ml trimethoprim and 50 μg/ml thymine. To determine potential toxicity of test compounds to the bacterial biosensor, non-selective medium was prepared, consisting of -Thy medium supplemented with 50 μg/ml thymine. The concentration of tested compounds for toxicity tests was 100 μM with incubation at 37°C. This concentration was chosen because it is close to the solubility limit of many of the compounds and is far above the concentration that would be relevant for a potential therapeutic. Further, the small amounts of test compounds that were available precluded experiments at higher concentrations.

Analysis of ERα Activation Using Enzyme-Linked ImmunoSorbent Assay (ELISA)

The NR Peptide ERα ELISA kit (Active Motif, Carlsbad, CA) allows detection of the ligand-activated ERα bound to co-activator. This assay was used to evaluate the predicted antagonistic action of two compounds identified by the bacterial subtype-selective ER binding biosensors. The test was repeated four times. The kit contained a 96-well plate coated with peptide containing an ERα co-activator binding motif (LXXLL, where L and X means a Leucine and any amino acid, respectively). The vehicle control solution contained the MCF-7 nuclear extract as well as diluent buffer and vehicle solvent. Blank solution contained only diluent buffer (data not shown). The antagonist control experiment contained 25 μM tamoxifen with 2.5% ethanol (final concentration), diluent buffer and MCF-7 nuclear extract. All of the reagents were included in the kit. The concentration of tamoxifen's stock solution was 1 mM. The two compounds **I** and **J** chosen for this test were dissolved in a mixture of 10% DMSO and 90% ethanol to form a 1 mM stock solution. Chemical names of compound **I** and **J** are 3-(2-aminophenyl) sulfanyl-1-(4-chlorophenyl)-3-phenyl-propan-1-one and 3-(2-aminophenyl)sulfanyl-3-(4-bromophenyl)-1-phenyl-propan-1-one, respectively (34,35). The stock solutions of compounds **I** and **J** were diluted to 25 μM for the ELISA assay. The final concentration of the solvents in the MCF-7 extract was kept below 2.5% in all cases. According to the manufacturer, the MCF-7 extract was diluted in 20 mM Hepes at pH 7.9, 100 mM KCl, 1 mM $MgCl_2$, 20% glycerol, 0.5 mM PMSF and 0.5 mM DTT.

The primary antibody was specific to recognize the activation function 1 (AF1) form of ERα, which was bound to peptide containing the LXXLL motif on the bottom of the well. The secondary antibody, a horseradish peroxide (HRP)-conjugated mouse antibody, bound to the primary antibody. The developing and stop solutions were added at the end of the procedure, allowing the amount of active co-activator-ERα complex to be quantified. The optical densities were measured on a GENios microplate reader (TECAN U.S. Inc., Research Triangle Park, NC). The tests were duplicated, the values were averaged and standard error bars were calculated. The final data were compared to the vehicle control.

RESULTS

Computational Prediction of Estrogen Antagonist's Activity

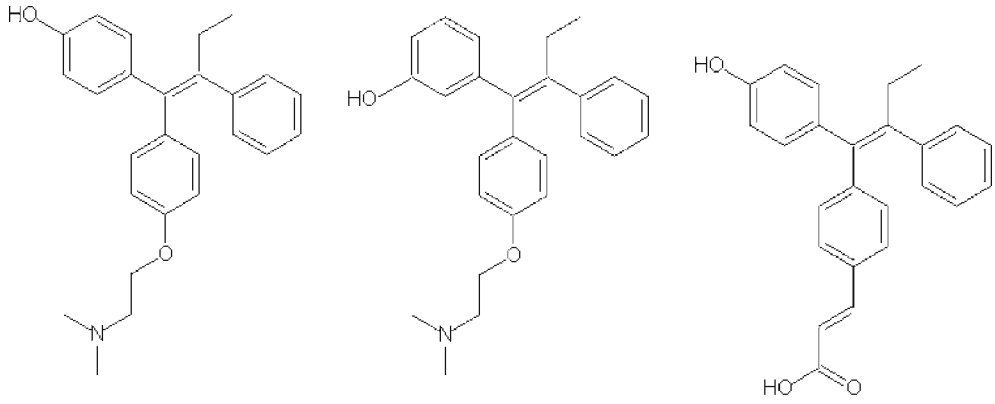
The Shape Signatures method allowed fast screening of approximately 604,000 compounds derived from the Ryan Scientific, Sigma-Aldrich and NCI Databases. However, only a few hundred structures, including 807 structures derived from ZINC database and eight commercially available structures from NCI Database, were found to be the most similar to the query by their shape or shape and MEP. Repeated structures were removed, and finally 533 molecules were tested from ZINC Database. Then, all of the compounds were clustered and docked to the LBD of ERα. Finally, thirteen compounds (denoted **A**-**M**, Fig. 2) were acquired for experimental evaluation in a simple bacterial

biosensor growth assay (described below). Six out of the thirteen compounds were predicted to resemble various compounds, such as tamoxifen, droloxifen, GW7604 and EM652. This set included NSC326655, NSC60845, 1045337, 1042059, 675871 and 1033664. Seven other structures showed close similarity to known endocrine active compounds such as hexestrol (compound 402165), genistein (compounds 623563, 53962 and 118156), coumarin (compounds 639277 and 407890) as well as naphthol (compound 110356) (36). The 1D Shape Signatures scores (shape only) of molecules derived from NCI database were within the limits of 0 to 0.1, while the 2D scores (shape and electrostatic potential) ranged from 0.1 to 2.0. The structures from the ZINC database were approximately 0.15 for 1D and between approximately 0.25 and 0.3 for 2D.

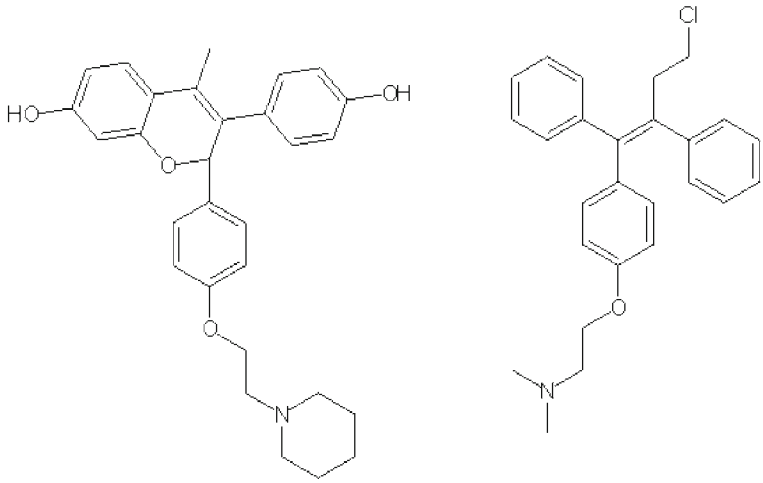
The GOLD method allowed prediction of the binding modes of the training and test sets. For the training set [4-hydroxytamoxifen, droloxifen, GW7604, EM652 and toremifene], the highest GOLD Score was 72.3 for EM652 and the lowest 55.5 for toremifene. The Chem Scores for this training set ranged from 54.7 (EM652) to 42.8 (GW7604). The GOLD Scores for test structures shown in Fig. 1 ranged from 67.5 (675871; compound **H**) to 41.5 (110356; compound **A**). However, the Chem Scores ranged from 41.34 (NSC326655, compound **C**) to 29.9 (110356; compound **A**). Active compounds **I** and **J** had the following GOLD and Chem Scores, respectively: 62.9 and 39.13 for compound **I** and 65.3 and 38.96 for compound **J**. Thus, both of the compounds **I** and **J** had scores within the range calculated for the training set compounds.

Detection of Estrogen Antagonist Activity in Bacterial Biosensor Strain

Test compounds were first screened for antagonism and agonism using the pMIT::ERβ* biosensor strain (Fig. 3) (26). This biosensor is designed such that growth of *E. coli* D1210*ΔthyA* cells in -Thy medium is enhanced in the presence of agonists and reduced in the presence of antagonists. In medium supplemented with trimethoprim and thymine, the phenotype is reversed, and growth of transformed *E. coli* D1210*ΔthyA* cells is inhibited in the presence of agonists and enhanced in the presence of antagonists. This biosensor strain has been characterized in previous work, where it has shown sensitivity to a wide range of estrogen agonists and antagonists. Further, the growth

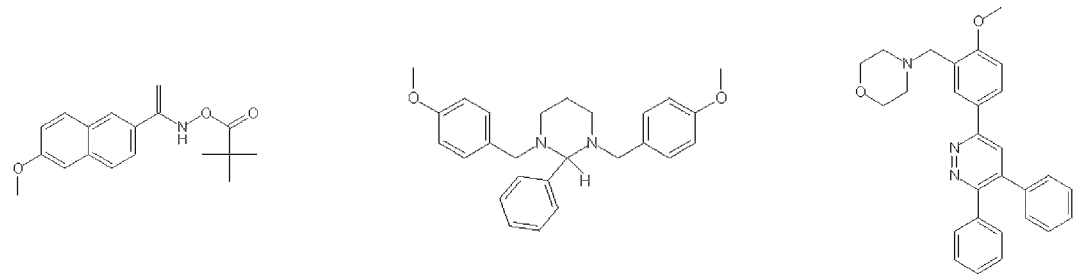


4-hydroxytamoxifen **Droloxifene** **GW7604**



EM652 **Toremifene**

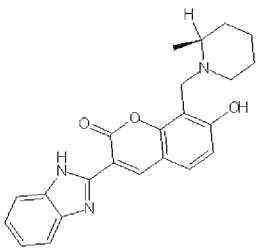
Fig. 1. Structures of queries used in the computational study.



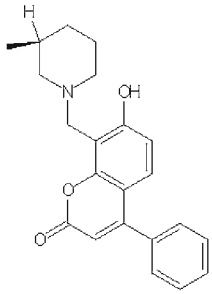
110356 (A)
RN: 258353-38-7

NSC60845 (B)
RN: 96764-88-4

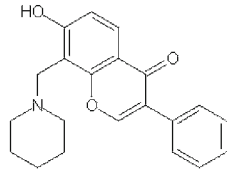
NSC326655 (C)
RN: 501655-45-4



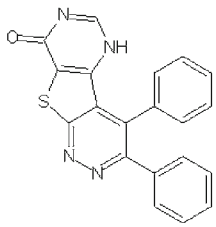
639277 (D)
RN: 307551-76-4



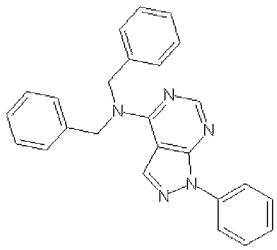
407890 (E)
RN: 344743-85-7



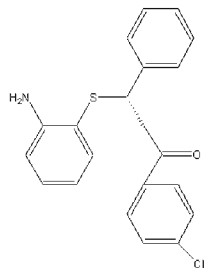
53962 (F)
RN: 111978-65-5



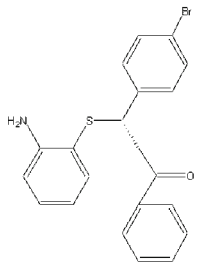
1033664 (G)
RN: 135127-79-6



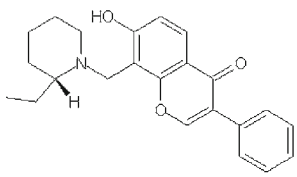
675871 (H)
RN: 331275-69-5



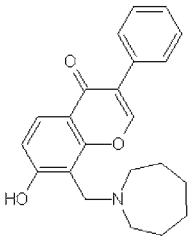
1042059 (I)
RN: 64820-37-7



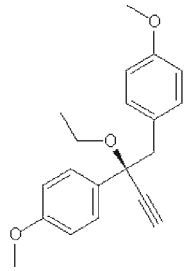
1045337 (J)
RN: 255714-26-2



623563 (K)
RN: 302575-30-0



118156 (L)
RN: 292170-06-0



402165 (M)
RN: 102553-37-7

Fig. 2. The final set of structures obtained by *in silico* screening. NSC compounds derived from National Cancer Database. The rest of the structures are commercially available and are a part of ZINC database. Alphabetical letters match the structures in Figs. 4, 5 and 6. Each of presented compounds is known, available commercially and their Chemical Abstracts Service Registration Numbers (CAS-RN) are included as well.

responses of the biosensor strains are consistent with independently measured binding affinities of know ER modulators (25,26). Based on observed growth phenotypes (Fig. 3a), compound **I** likely acts as an antagonist of ERβ. Further, the enhanced cell growth in TTM medium in the presence of compound **J** indicates that this compound may also be an ERβ antagonist and confirms that it is not toxic to the *E. coli* sensor strain. The low cell growth in -Thy in the presence of both compounds **I** and **J** relative to that of the negative control (Fig. 3b) strongly implies that these

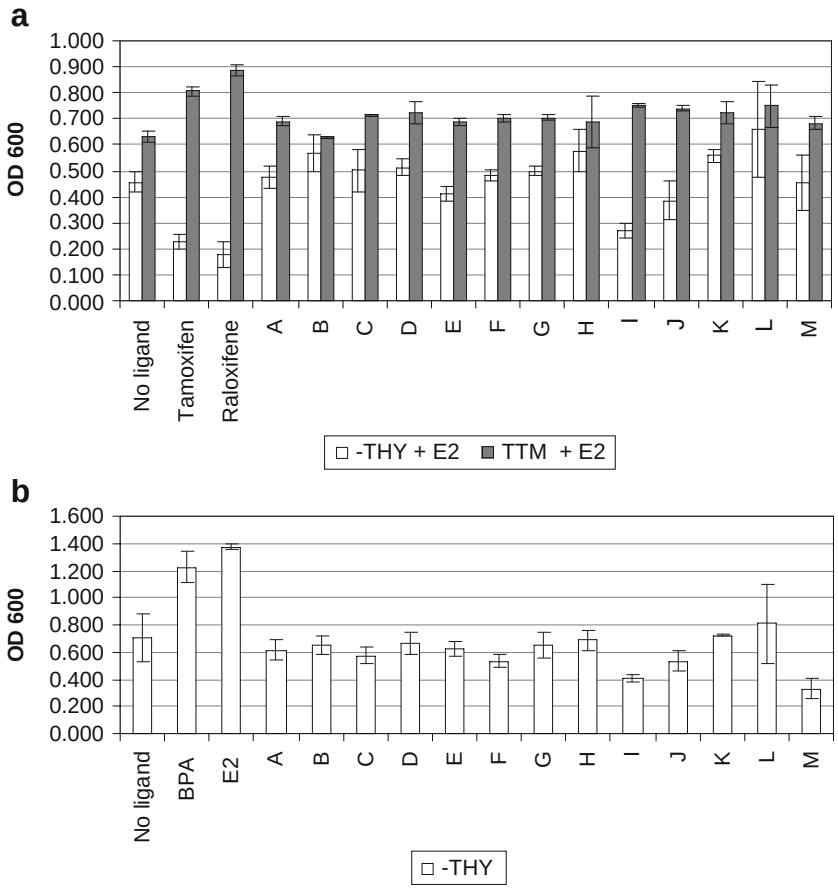


Fig. 3. (**a**) Cell growth as determined by optical density at 600 nm (OD 600) of *E. coli* D1210*ΔthyA* cells transformed with pMIT::ERβ* grown in -Thy medium + 500 nM E2 at 34°C (*clear bars*) and TTM medium + E2 media at 37°C for 16 h (*shaded bars*) in the presence of test ligands. The concentration of all of the tested compounds was 5 μM. Inhibition of ERβ leads to a decrease in growth relative to the no ligand control in -Thy medium, but increased growth in TTM medium. (**b**) Growth determined by OD 600 of *E. coli* D1210*ΔthyA* cells with pMIT::ERβ* in -Thy media at 34°C for 20 h. Agonist compounds increase cell growth under these conditions. The concentration of all compounds was 5 μM. Experiments were performed in duplicate where error bars represent one standard deviation. *BPA* bisphenol A.

compounds act as antagonists of ERβ. Phenotypes also suggest that compound **E** may be a very weak antagonist, but this was not confirmed in further tests. None of the other compounds showed consistent agonistic or antagonistic behavior in subsequent tests, and they were not examined further.

Ligands were also tested for agonism and antagonism towards ERα using the pMIT::ERα* biosensor (Fig. 4), permitting the identification of potential subtype-specific ER modulators (24,25). Similarly to the pMIT::ERβ* biosensor, growth of *E. coli* D1210*ΔthyA* cells containing the pMIT::ERα* plasmid in -Thy medium is enhanced in the presence of agonists (E2 control compound). Antagonist activity can be detected with the pMIT::ERα* sensor in the presence of E2, where addition of antagonist tends to decrease growth. The low cell growth of *E. coli* D1210*ΔthyA* cells in -Thy supplemented with compound **I** and 17-β-estradiol (Fig. 4a) indicates that this compound may act as an antagonist for ERα, while compound **J** does not appear to modulate this receptor (Fig. 4a and b). These results suggest that compound **I** is a general ER antagonist, while compound **J** may be an ERβ-specific antagonist. Toxicity of the test ligands at the concentrations used in the biosensor was ruled out by confirming positive *E. coli* D1210*ΔthyA* growth phenotypes in non-selective thymine-rich medium in the presence of the ligands (data not shown). This further supports our hypothesis that the low cell growth observed in the biosensor screens results from modulation of the ERs by the test ligands. Compounds **D**, **H**, **J** and **L** presented in Fig. 4b were not confirmed to show agonistic behavior in further tests. None of the molecules (**A**-**M**) had an effect on the TRβ sensor strain control (data not shown).

Analysis of Estrogen Antagonism of Test Compounds by ELISA

Additional ELISA analysis based on MCF-7 cell extracts (Fig. 5) showed that compound **J** slightly decreases the

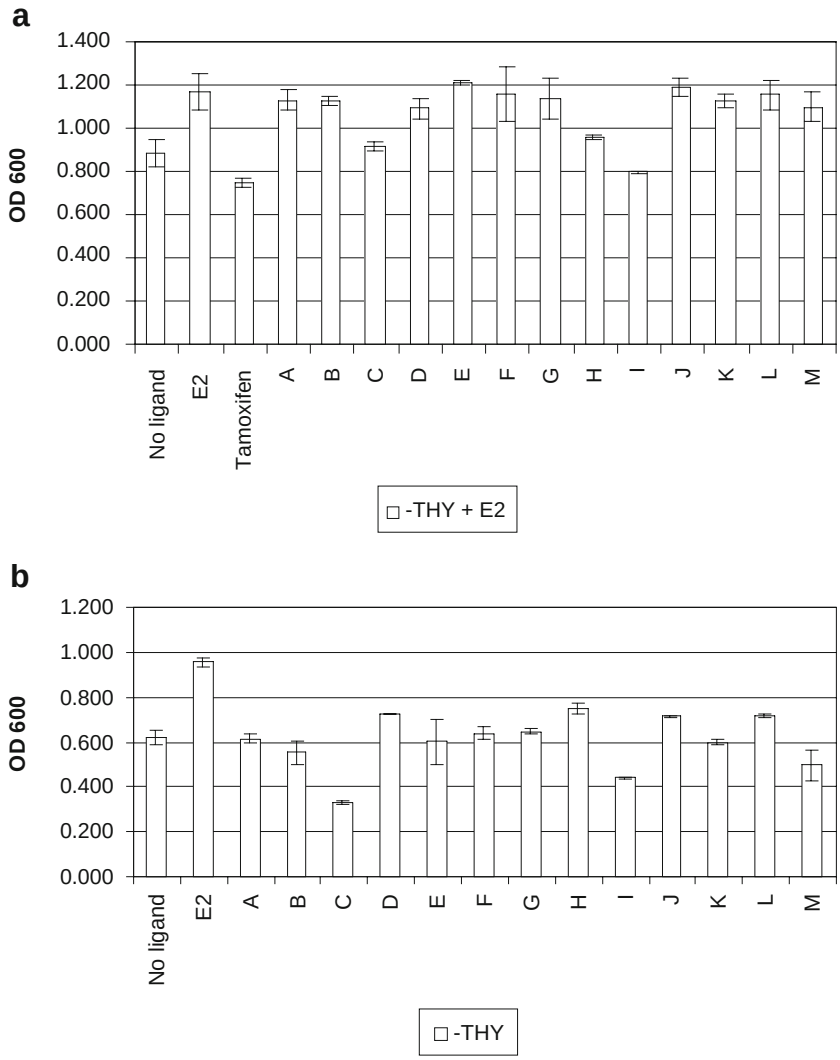


Fig. 4. (**a**) Cell growth as determined by OD 600 of *E. coli* D1210*ΔthyA* cells transformed with pMIT::ERα* and grown in -Thy medium with 10 μM E2 at 34°C for 13 h in the presence of the test ligands. The concentration of all of the ligands was 10 μM except of tamoxifen, which had a concentration of 2 μM. (**b**) Cell growth as determined by optical density [OD 600] of *E. coli* D1210*ΔthyA* cells transformed with pMIT::ERα* and grown in -Thy medium at 34°C for 13 h. The concentration of all compounds was 10 μM. Experiments were performed in triplicate and error bars represent one standard deviation.

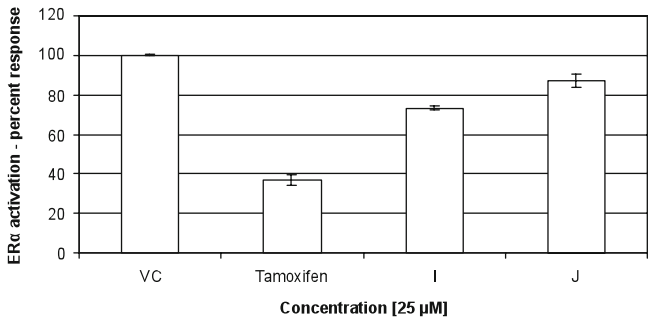
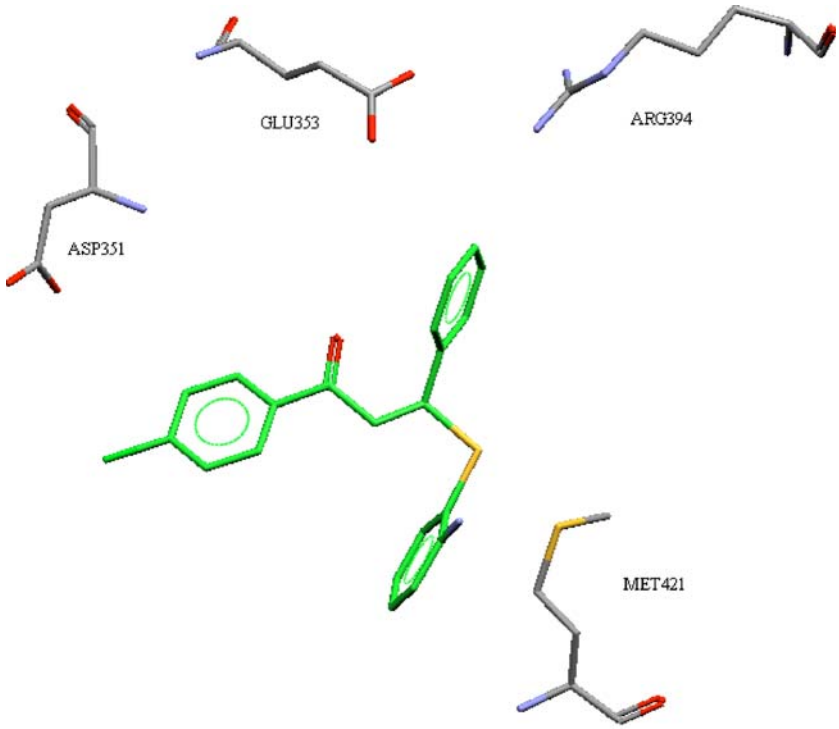
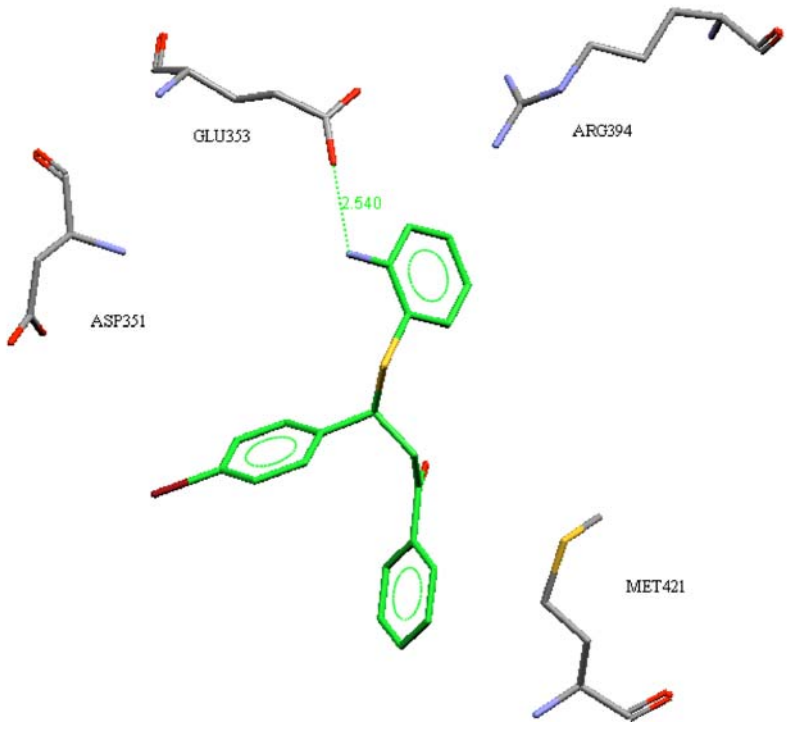


Fig. 5. The results of ER-α activation using ELISA (NR Peptide ERα ELISAs, Active Motif, Carlsbad, CA). The final concentration of tamoxifen and the compounds **I** and **J** was 25 μM. VC refers to vehicle control.

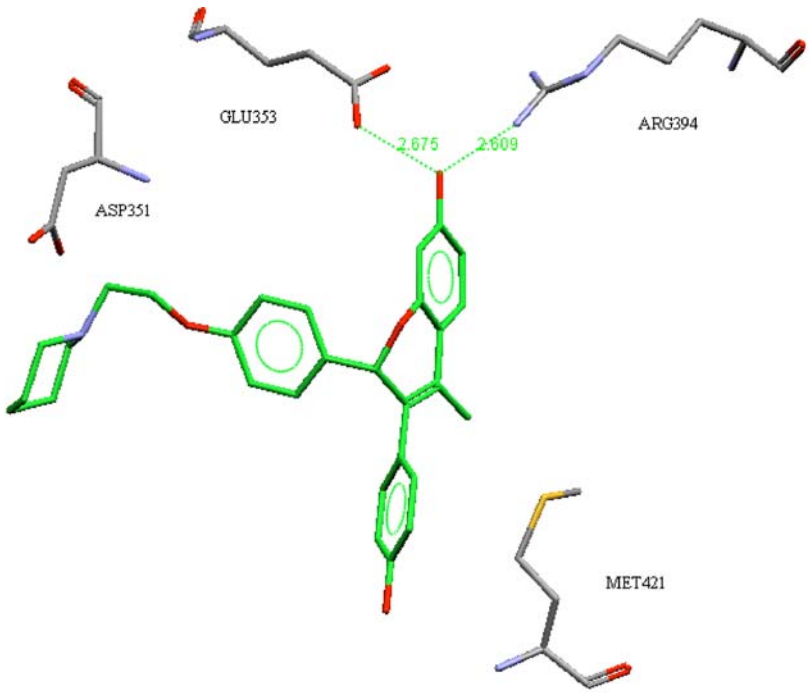


Compound I

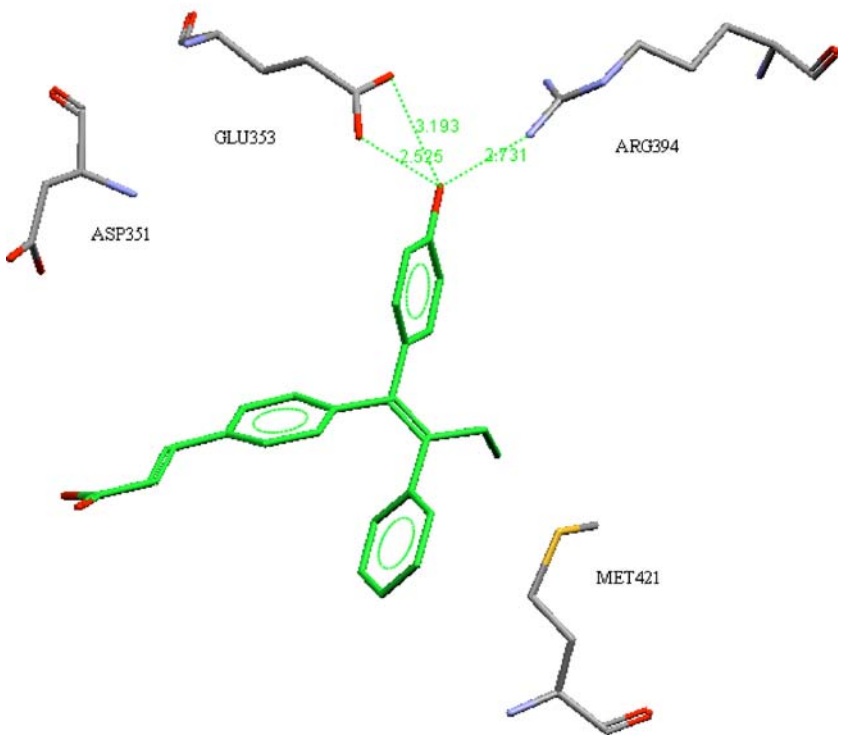


Compound J

Fig. 6. Structures of compound **I**, **J,** EM652, GW7604 and their binding modes predicted by GOLD. Highlighted amino acids are [clockwise from left to right] Asp351, Glu353, Arg394 and Met421. Atom colors represent carbon (*gray*), oxygen (*red*), nitrogen (*blue*), sulfur (*yellow*), chlorine (*green*) and bromine (*bordeaux*). Figures generated by SILVER ver. 1.0 (17).



EM652



GW7604

Fig. 6. (continued)

amount of activated ERα, thus indicating weak antagonistic activity. In the same test, compound **I** exhibited stronger antagonistic activity than compound **J**. The vehicle control was normalized to 100% to represent baseline of ERα activation by endogenous estrogens. Therefore, an antagonistic effect was assumed when measured ELISA values were below 100%. If any of the test compounds stimulated ERα activation, their values would exceed 100%. Both compounds **I** and **J** were found to be antagonists.

Computational Analysis of Compounds *I* and *J* Binding Models

The predicted binding mode, using the GOLD Score fitness function, indicated that receptor residues would likely not bind by hydrogen bonds to compound **I** (Fig. 6). Instead, similar to tamoxifen, this compound may form weak van der Waals contacts with Glu353 (Helix 3) and Arg394 (Helix 5). One of the binding modes predicted by GOLD showed that the amine group on compound **J** may possibly form a hydrogen bond with the γ-carboxylate group of Glu353 (O···N distance of 2.486 Å). However, despite the close similarities of both structures, compound **J** was not confirmed by biosensors to bind to ERα but only to ERβ (Figs. 4 and 5). Interaction with His 524 (Helix 11) was not predicted by GOLD for either of the compounds. Based on these parameters and Lipinski's Rule of Five, both of the compounds were predicted to be orally bioavailable.

DISCUSSION

The Shape Signatures method facilitates fast screening of chemical libraries and is a powerful tool for identifying lead compounds when used in conjunction with a quick and sensitive biosensor assay. The Shape Signatures method does not require conformational alignments of query compounds and is independent of their orientation in 3D space. The electrostatic fields are also not sampled at intersections of grids, as is required in Comparative Molecular Field Analysis (CoMFA). Instead, the molecular electrostatic potential is calculated over the surface of the molecule. Thus, the presented techniques complement QSAR methods and might be used for identifying structurally similar compounds with differing partial antagonist potencies. This is relevant with respect to receptor interactions that can modulate hormone balance with potential therapeutic consequences. The results obtained from Shape Signatures and docking methods could be verified by QSAR methods or utilized independently.

Our method is simply based on comparisons of molecular shape and polarity for compounds from large databases. These characteristics can easily be determined and compared to known biologically-active compounds with shape complementarity to the active site of the receptor. The assumption is that similar compounds are likely to also be biologically-active. The central processing unit time for the creation of a ray-trace image for one structure depends on the computer specification. For a 3.5 GHz Intel Pentium 4 with 2 GB RAM, each structure required about 10 s. The ray-tracing process is required only once and can be significantly improved with faster computers or a Beowulf cluster. Further, the number of analyzed structures is not limited by their type; organometalic, organic, inorganic, polymers or ions can be included, and this method also accounts for stereoisomerism. Since this method does not require experimental data, the structures do not have to be synthesized for analysis, thus greatly simplifying the identification of hits. Finally, once the ray traces are calculated, this technique is able to compare test to known compounds at a rate of 1.4 billion molecules per day, which is a thousand times more structures than the relatively simple ROCS overlay method (27).

Once the Shape Signatures analysis has provided possible hit compounds, the bacterial biosensor can quickly provide rough experimental information on their biological activity (25,26). This biosensor provides 1—detection of differential binding to subtypes of ER, including very weak binding, and 2—classification of the compounds into two groups, agonists and antagonists. This method is non-radioactive and economical and can be easily automated using a variety of platforms. It is easy to use and requires only a simple evaluation of bacterial growth phenotype in fully defined medium. This aspect further eliminates many confounding variables associated with more complex assays. Further, simplicity in the construction of the biosensor allows for exchanging the LBD of ER to other known receptors such as TRβ. By simply forming the new plasmid pMIT::TRβ*, it becomes possible to readily detect compounds active against this alternative active site (25).

The pMIT::ERβ* biosensor has already successfully detected several previously unidentified ERβ ligands and has consistently identified known modulators (26). However, the pMIT::ERα* biosensor could be improved. For example, the long functional group on Faslodex appeared to disturb the stability of the protein, and it was not detected correctly in the ERα system (25). This could explain the failure of the pMIT::ERα* biosensor to detect antagonism of ERα by compound **J**, resulting in some inconsistency with the ELISA method.

Several projects have already successfully identified compounds that are active against a specific receptor using the Shape Signatures technique. It was shown to be efficient in classifying agonists and antagonists from over 11,000 molecules, as well as their specificity of binding to serotonin receptor subtypes (15). Additional ERα antagonists have also been identified in previous studies (37). Shape Signatures can also be verified using other scoring approaches, such as docking scoring functions (GOLD and Chem Scores), and can be combined with other computer-based drug design methods, such as 3D Quantitative Structure-Activity Relationship (3D-QSAR) or Molecular Superposition (38). This method is sensitive to stereoisomerism, but it is not sensitive to small changes of conformation (27).

CONCLUSION

Virtual study and functional screening identified two compounds (**I** and **J**) with predicted binding modes that overlap that of 4-hydroxytamoxifen in the active site of the pocket of ERα. Both compounds were also drug-like, based on Lipinski's Rule of Five. Additional analysis using engineered biosensors confirmed that both of the structures are in fact estrogen antagonists. Compound **I** bound to both subtypes of the ER, but compound **J** appeared to be ERβ selective. An ELISA assay based on human MCF-7 extract confirmed that compound **I** somewhat suppresses ERα activation while compound **J** demonstrated a similar but much weaker effect. Thus the system has demonstrated the capability to rapidly screen large chemical libraries and quickly identified drug-like compounds that can modulate a given NHR target.

ACKNOWLEDGMENT

This work was partially supported by NIH grant 1R21ES16630.

REFERENCES

1. Kuiper GG, Enmark E, Pelto-Huikko M, Nilsson S, Gustafsson JA. Cloning of a novel receptor expressed in rat prostate and ovary. Proc Natl Acad Sci U S A. 1996;93:5925–30.
2. Manas ES, Unwalla RJ, Xu ZB, Malamas MS, Miller CP, Harris HA, *et al*. Structure-based design of estrogen receptor-beta selective ligands. J Am Chem Soc. 2004;126:15106–19.
3. Pike AC, Brzozowski AM, Hubbard RE, Bonn T, Thorsell AG, Engstrom O, *et al*. Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. EMBO J. 1999;18:4608–18.
4. Tremblay A, Tremblay GB, Labrie C, Labrie F, Giguere V. EM-800, a novel antiestrogen, acts as a pure antagonist of the transcriptional functions of estrogen receptors alpha and beta. Endocrinology. 1998;139:111–8.
5. McDonnell DP. The molecular pharmacology of estrogen receptor modulators: implications for the treatment of breast cancer. Clin Cancer Res. 2005;11:871s–7s.
6. Jensen EV, Jordan VC. The estrogen receptor: a model for molecular medicine. Clin Cancer Res. 2003;9:1980–9.
7. Shiau AK, Barstad D, Radek JT, Meyers MJ, Nettles KW, Katzenellenbogen BS, *et al*. Structural characterization of a subtype-selective ligand reveals a novel mode of estrogen receptor antagonism. Nat Struct Biol. 2002;9:359–64.
8. Liu H, Lee ES, Deb Los Reyes A, Zapf JW, Jordan VC. Silencing and reactivation of the selective estrogen receptor modulator-estrogen receptor alpha complex. Cancer Res. 2001; 61:3632–9.
9. Grese TA, Sluka JP, Bryant HU, Cullinan GJ, Glasebrook AL, Jones CD, *et al*. Molecular determinants of tissue selectivity in estrogen receptor modulators. Proc Natl Acad Sci U S A. 1997;94:14105–10.
10. Gauthier S, Caron B, Cloutier J, Dory YL, Favre A, Larouche D, *et al*. (S)-(+)-4-[7-(2, 2-dimethyl-1-oxopropoxy)-4-methyl-2-[4-[2-(1-piperidinyl)- ethoxy]phenyl]-2H–1-benzopyran-3-yl]-phenyl 2, 2-dimethylpropanoate (EM-800): a highly potent, specific, and orally active nonsteroidal antiestrogen. J Med Chem. 1997;40:2117–22.
11. Bentrem D, Dardes R, Liu H, MacGregor-Schafer J, Zapf J, Jordan V. Molecular mechanism of action at estrogen receptor alpha of a new clinically relevant antiestrogen (GW7604) related to tamoxifen. Endocrinology. 2001;142: 838–46.
12. Fan M, Rickert EL, Chen L, Aftab SA, Nephew KP, Weatherman RV. Characterization of molecular and structural determinants of selective estrogen receptor downregulators. Breast Cancer Res Treat. 2007;103:37–44.
13. Zauhar RJ, Moyna G, Tian L, Li Z, Welsh WJ. Shape signatures: a new approach to computer-aided ligand- and receptor-based drug design. J Med Chem. 2003;46:5674–90.
14. Irwin JJ, Shoichet BK. ZINC—a free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005;45:177–82.
15. Nagarajan K, Zauhar R, Welsh WJ. Enrichment of ligands for the serotonin receptor using the Shape Signatures approach. J Chem Inf Model. 2005;45:49–57.
16. GOLD. Cambridge, UK: Cambridge crystallographic data centre; 2004.
17. SILVER. Cambridge, UK: Cambridge crystallographic data centre; 2004.
18. Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J Mol Biol. 1995;245:43–53.
19. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol. 1997;267:727–48.
20. Nissink JW, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R. A new test set for validating predictions of protein-ligand interaction. Proteins. 2002;49:457–71.
21. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins. 2003;52:609–23.
22. Annamala MK, Inampudi KK, Guruprasad L. Docking of phosphonate and trehalose analog inhibitors into M. tuberculosis mycolyltransferase Ag85C: Comparison of the two scoring fitness functions GoldScore and ChemScore, in the GOLD software. Bioinformation. 2007;1:339–50.
23. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins. 1998;33:367–82.
24. Skretas G, Wood DW. Rapid detection of subtype-selective nuclear hormone receptor binding with bacterial genetic selection. Appl Environ Microbiol. 2005;71:8995–7.
25. Skretas G, Wood DW. A bacterial biosensor of endocrine modulators. J Mol Biol. 2005;349:464–74.
26. Skretas G, Meligova AK, Villalonga-Barber C, Mitsiou DJ, Alexis MN, Micha-Screttas M, *et al*. Engineered chimeric enzymes as tools for drug discovery: generating reliable bacterial screens for the detection, discovery, and assessment of estrogen receptor modulators. J Am Chem Soc. 2007;129:8443–57.
27. Meek PJ, Liu Z, Tian L, Wang CY, Welsh WJ, Zauhar RJ. Shape Signatures: speeding up computer aided drug discovery. Drug Discov Today. 2006;11:895–904.
28. Labrie F, Labrie C, Belanger A, Simard J, Gauthier S, Luu-The V, *et al*. EM-652 (SCH 57068), a third generation SERM acting as pure antiestrogen in the mammary gland and endometrium. J Steroid Biochem Mol Biol. 1999;69:51–84.
29. SPSS. Chicago, IL: SPSS; 2005.
30. Grimm LG, Yarnold PR. Reading and understanding more multivariate statistics. Washington, DC: American Psychological Association; 2000.
31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al*. The protein data bank. Nucleic Acids Res. 2000;28:235–42.
32. Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, *et al*. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. Cell. 1998;95:927–37.
33. SYBYL. St. Louis, MO: Tripos Inc; 2001.
34. Levai A, Bognár R. Oxazepines and thiazepines, III. Synthesis of 2,3-dihydro-2,4-diphenyl-1,5-benzothiazepines by the reaction of 2-aminothiophenol with chalcones substituted in ring A. Acta Chim Acad Sci Hung. 1977;92:415–9.
35. Sharma G, Raj K, Chakraborti AK. Fluoroboric acid adsorbed on silica-gel (HBF4-SiO2) as a new, highly efficient and reusable heterogeneous catalyst for thia-Michael addition to α, β-unsaturated carbonyl compounds. Tetrahedron Lett. 2008;49:4272–5.
36. Fang H, Tong W, Shi LM, Blair R, Perkins R, Branham W, *et al*. Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens. Chem Res Toxicol. 2001;14:280–94.
37. Wang CY, Ai N, Arora S, Erenrich E, Nagarajan K, Zauhar R, *et al*. Identification of previously unrecognized antiestrogenic chemicals using a novel virtual screening approach. Chem Res Toxicol. 2006;19:1595–601.
38. Ambrose Amin E, Welsh WJ. Three-dimensional quantitative structure-activity relationship (3D-QSAR) models for a novel class of piperazine-based stromelysin-1 (MMP-3) inhibitors: applying a "divide and conquer" strategy. J Med Chem. 2001;44:3849–55.