Pharmacophore Modeling for Qualitative Prediction of Antiestrogenic Activity

Simone Brogi,[†] Maria Kladi,[‡] Constantinos Vagias,[‡] Panagiota Papazafiri,[§] Vassilios Roussis,[‡] and Andrea Tafi*,[†]

Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Siena, Via Aldo Moro, I-53100 Siena, Italy, Department of Pharmacognosy & Chemistry of Natural Products, School of Pharmacy, University of Athens, Panepistimiopolis Zografou, Athens 15771, Greece, and Department of Animal & Human Physiology, School of Sciences, University of Athens, Panepistimiopolis Zografou, Athens 157 84, Greece

Received July 20, 2009

A ligand-based pharmacophore approach for the prediction of antiestrogenic activity to be used as an in silico screening tool for bioactive compounds including natural products was developed using Catalyst HypoGen. The generated pharmacophore hypothesis (HYPO-7) consisted of five features, namely, one hydrophobic (HY1), two hydrophobic aromatic (HY2), one hydrogen-bond acceptor (HBA), and one hydrogen-bond donor (HBD). HYPO-7 successfully predicted the lack of cytotoxicity of a number of new metabolites isolated from the red alga *Laurencia glandulifera*. Furthermore, a screening of the Asinex Gold Collection database was performed by coupling HYPO-7 with a docking filtration, which resulted in a restricted set of 12 new scaffolds to be investigated as potential SERMs. The inhibitory activity of these compounds was evaluated in vitro using MCF7 human breast adenocarcinoma cell line. Ten out of the twelve compounds exhibited inhibitory activity with IC_{50} values between 26 and 188 μM. This result shows that application of HYPO-7 could assist in the selection of potentially active compounds, thus expediting the hit discovery process.

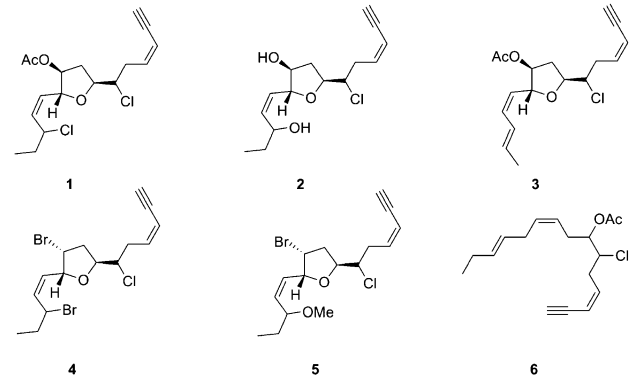
INTRODUCTION

In the course of our studies for the discovery of new bioactive secondary metabolites from marine organisms,[1,2] specimens of *Laurencia glandulifera* collected from the island of Crete, have been recently investigated, resulting in the isolation of five new tetrahydrofuran metabolites with an acetogeninic scaffold (**1**−**5**) and their linear biosynthethic precursor (**6**) (Chart 1).[3] The cytotoxicity of these compounds was assessed on HT29 (derived from colorectal adenocarcinoma), MCF7 (derived from a mammary adenocarcinoma), PC3 (derived from a prostate adenocarcinoma), HeLa (derived from cervix adenocarcinoma), and A431 (derived from epidermoid carcinoma) human tumor cell lines. After 48 h of incubation none of the metabolites showed any significant activity ($IC_{50} \geq 300$ μM) against the above-mentioned cells.

The inherent difficulty in the isolation of adequate quantities of new marine metabolites to enable reliable biological test data to be obtained in combination with the low inhibitory activity exhibited by compounds **1**−**6** prompted us to develop a computational bioactivity screening evaluation method. Pharmacophores, which are essentially drug−receptor interaction models, were considered as the most appropriate tool for this purpose.[3]

Pharmacophore modeling (PM) is a three-dimensional computational approach widely employed in drug design. In this framework, its usefulness covers two major domains:

Chart 1. Metabolites Isolated from *L. glandulifera*



first, the performance of "scaffold hopping", that is virtual screening of libraries of potentially active compounds in search for new leads, and, second, the rationalization of activity distribution within groups of molecules, exhibiting a similar pharmacological profile, that are recognized by the same target.[4]

In recent years, progressive applications of PM continue to appear in the literature. A novel approach, named parallel screening (PS), was introduced in 2006 by Thierry Langer and co-workers.[5] This is an in silico method to predict new potential biological activities of compounds by screening them with a multitude of pharmacophore models. The major aim of PS is the in silico determination of the biological activity profile of products in order to speed up the cost-intensive drug discovery development process and increase its efficiency. Moreover, in 2008, Langer and co-workers published a survey discussing the prospective impact of virtual screening techniques in the discovery of bioactive

* Corresponding author phone: (+39)-0577-234-313; fax: (+39)-0577-234-333; e-mail: tafi@unisi.it.
[†] Università degli Studi di Siena.
[‡] School of Pharmacy, University of Athens.
[§] School of Sciences, University of Athens.

10.1021/ci900254b CCC: $40.75
Published on Web 10/30/2009

natural products,[6] so as to support the applicability of these methods in natural products research. Finally, virtual screening techniques from the drug discovery field begin to be used for the bioactivity profiling of chemicals (especially those of potential environmental concern). Two of the major goals of such a purpose are (i) prioritizing compounds for further testing using more complex systems and (ii) reducing and ultimately replacing the use of animals in regulatory testing.[7] PM might be extremely helpful in both cases.

Our initial plan, related to the application of pharmacophore modeling to *Laurencia* metabolites, was to evaluate the potential of recursive application of such a methodology to several pharmacological targets as an in silico filtering tool for bioactive compounds including natural products. The achievement of such a goal, in fact, would reduce real pharmacological testing only to compounds and targets theoretically predicted to be complementary to each other.[3] Since a significant number of *Laurencia* metabolites had already been tested against MCF7 cells,[1,8] in combination with literature data, it was decided to test first our approach on this target.

In the framework summarized above our PM intended to work out a fast, ultimately qualitative, ligand-based tool to be applied independently from either the knowledge of the receptor three-dimensional structure or the availability of receptor affinity data that is ideally referring to the early definition by Paul Ehrlich, which described the pharmacophore as a molecular framework that carries the essential features responsible for a drug's biological activity.[9] On the other hand, since there is no reason to ignore protein data when such data are available, we preferred to first select an important and broadly explored target in order to substantiate our results by comparison with available crystallographic drug−receptor structures.

MCF7 is a well-charaterized estrogen receptor subtype α (ER-α) positive control cell line and therefore a useful in vitro model to study the activity of new metabolites against breast cancer. This neoplastic disease is a leading cause of morbidity and mortality in women in developed countries, and its incidence is increasing in developing countries as well. Approximately 75% of all patients with breast cancer have a tumor expressing the estrogen receptor (ER) and/or the progesterone receptor.[10] These tumors depend on estrogens for their survival, and endocrine therapy is used to deprive breast cancer cells of estrogens' stimuli either through competitive binding to ER or by estrogen deprivation. The antiestrogen tamoxifen, a nonsteroidal triphenylethylene derivate, and other related selective estrogen receptor modulators (SERMs) like raloxifene (arylbenzothiophene derivate), showing tissue-dependent agonistic or antagonistic behavior, have been used extensively in the clinical management of primary and advanced endocrine-sensitive breast cancer.[11] Considering the value of developing nonsteroidal estrogen analogs, the research on SERMs is still on to afford a variety of nonsteroidal ligands.[12]

Targeting the above-mentioned goals, we explored the pharmacophore features of a comprehensive set of SERMs derivatives, tested through their inhibitory activity toward MCF7 cells, by application of Catalyst[13] software. The ligand-based pharmacophore developed by this approach was first applied to verify its prediction ability on *Laurencia* metabolites **1**−**6**[3] and then used in a virtual screening within a chemical 3D database as a query for searching new scaffolds to be chemically developed as potential SERMs. Herein we report the results of our studies.

PHARMACOPHORE MODELING

The existence of a quantitative correlation between inhibition of MCF7 human breast adenocarcinoma cell line proliferation (measured by IC_{50} values) and ER-α receptor binding (relative binding affinity by competition with 17β-estradiol) has been reported for a series of SERM derivatives structurally related to raloxifene. Such a finding demonstrated that receptor binding is the first step in the pathway that leads to inhibition of tumor cell proliferation.[14] On that basis a pharmacophore has been recently developed for arylbenzothiophene derivatives active against MCF7 cells using IC_{50} values as a measure of the binding to ER-α receptor.[15]

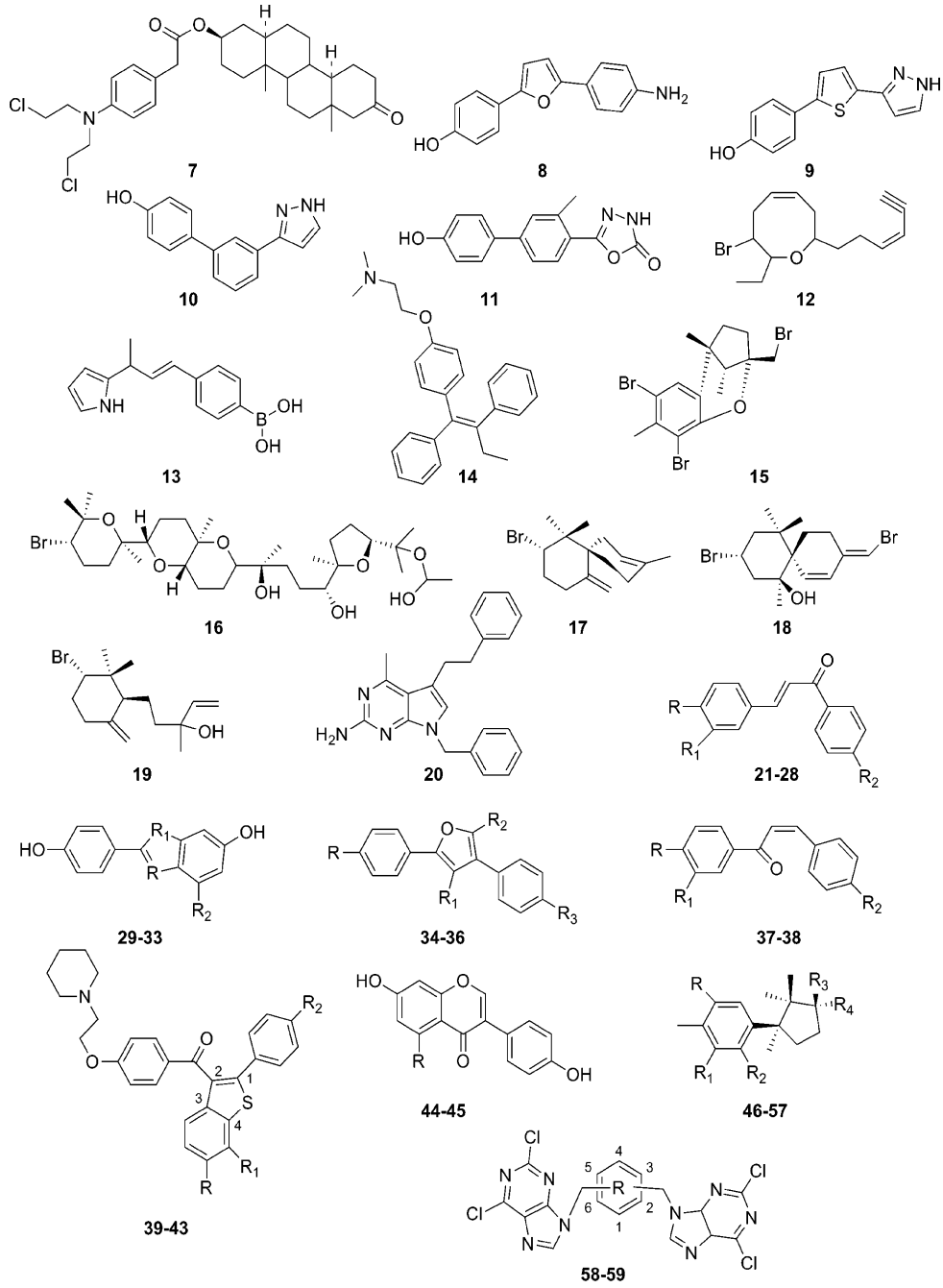
Taking into account the experience summarized above, we focused our PM approach on the qualitative prediction of the inhibitory activity of new chemical entities (including natural products **1**−**6**) toward MCF7 cells. Cytotoxicity against MCF7 cells (IC_{50} values) was used to approximate the ability of a compound to act as SERM by binding to ER-α receptor. Catalyst, one of the outstanding software packages enabling automated generation of pharmacophores, was selected for building our model. This software, in fact, should be applied to panels of compounds that have a wide spectrum of activity against a given biological target and possess structural diversity so as to produce reliable results,[16] and both requirements suited our purpose.

In order to endow the yet-to-be-built pharmacophore with comprehensiveness and to couple this property with the wish to give insight into the basic molecular functional requirements of compounds for ER-α binding attention was paid to maximize structural diversity within the set of estrogen receptor modulators selected to generate and validate the model. It was consequently decided that each IC_{50} order of magnitude had to be represented by at least three compounds belonging to different structural classes. This selection prevented us from taking into consideration the few most active SERMs reported in the literature in the low nanomolar range,[17] such as raloxifene.

Specifically, a data set of 53 SERMs belonging to several different structural classes (Chart 2) was compiled. These compounds covered a 4 orders of magnitude activity range (Table 1), centered at the IC_{50} median value of 13 μM and included active and less active derivatives reported in the literature[15,17−25] as well as metabolites isolated and tested by us.[26] The set was divided into the training and the test subsets according to Catalyst guidelines. Full details of the protocol applied to generate the pharmacophore can be found in the Experimental Section.

Several hypotheses were generated by Catalyst. The most reliable pharmacophore was selected among them by using the test set compounds' IC_{50} values prediction ($r^2_{\text{test-set}}$) as the selection standard. On this basis, a potentially predictive hypothesis, exhibiting a correlation coefficient $r^2_{\text{test-set}} = 0.91$ was chosen and established as HYPO-7. The five features of the pharmacophore, comprising one hydrogen-bond donor (HBD), one hydrogen-bond acceptor (HBA), two hydrophobic aromatic sites (HY1), and one hydrophobic (HY2) site

Chart 2. Structure of SERM Derivatives Used in This Study



Compd	R	R_1	R_2	R_3	Compd	R	R_1	R_2	R_3	R_4
21	Cl	Cl	$B(OH)_2$		**41**	OCH_3	H	CH_2OH		
22	H	I	OCH_2COOH		**42**	OH	H	CF_3		
23	Cl	Cl	OCH_2COOH		**43**	H	OH	OH		
24	H	I	OH		**44**	H				
25	F	Br	$B(OH)_2$		**45**	OH				
26	I	Br	$B(OH)_2$		**46**	H	H	H	H	Br
27	F	Cl	$B(OH)_2$		**47**	H	H	H	Br	H
28	F	F	$B(OH)_2$		**48**	Br	H	OH	H	CH_3
29	H	S	H		**49**	H	H	H	H	CH_3
30	O	N	H		**50**	Br	H	OH	CH_3	H
31	O	H	CH_2CH_2		**51**	H	H	H	H	CH_2
32	O	H	CH_2CN		**52**	H	H	H	H	H
33	O	N	CN		**53**	H	H	OH	H	CHBr
34	OH	CH_2CH_3	FSC	OH	**54**	Br	Br	OH	H	CH_3
35	H	OH	FSC	OH	**55**	Br	H	OH	H	CH_2
36	OH	Me	Me	OH	**56**	Br	H	OH	H	CHBr
37	Cl	Cl	CH_2COOH		**57**	Br	I	OH	H	CH_3
38	H	H	H		**58**	1,4				
39	H	H	H		**59**	1,3				
40	OH	H	1'-CH_2CH_3							

FSC (functional side chain): $(CH_2)_6NMe(CH_2)_3SC_5H_{11}$.

Table 1. Experimental and Predicted Activity IC_{50} (μM) of the Compounds Used in the Computational Study

cmpd	exp	calcd	error	cmpd	exp	calcd	error
7[a,18]	100	10	−10	**34**[a,23]	0.039	0.13	+3.3
8[a,19]	4.1	2.4	−1.7	**35**[b,23]	1.8	0.83	−2.2
9[a,19]	24	20	−1.2	**36**[b,23]	6.9	2.4	−2.9
10[b,19]	18	3.9	−4.6	**37**[b,20]	15	4.5	−3.3
11[b,19]	25	6.8	−3.7	**38**[b,26]	63	160	+2.6
12[a,26]	118.7	160	+1.3	**39**[a,15]	0.3	0.89	+3
13[a,19]	11	12	+1.1	**40**[a,15]	0.02	0.007	−2.8
14[a,17]	0.48	1.1	+2.4	**41**[b,15]	0.6	0.11	−5.5
15[b,26]	167.9	120	−1.4	**42**[b,15]	1	0.48	−2.1
16[b,26]	34.5	24	−1.4	**43**[b,15]	0.3	0.23	−1.3
17[b,26]	247.3	3600	+15	**44**[a,24]	2.3	1.5	−1.5
18[b,26]	74.8	130	+1.7	**45**[b,24]	3.4	12	+3.7
19[b,26]	148.4	130	−1.2	**46**[a,26]	177.6	160	−1.1
20[a,21]	10.7	8.4	−1.3	**47**[a,26]	265.4	160	−1.7
21[a,20]	5	7.8	+1.6	**48**[a,26]	103.7	160	+1.5
22[a,20]	9	6.5	−1.4	**49**[a,26]	234.8	170	−1.4
23[a,20]	13	5.8	−2.3	**50**[a,26]	67.2	160	+2.3
24[a,20]	7	8.2	+1.2	**51**[b,26]	173.8	130	−1.4
25[b,20]	8.5	5.8	−1.5	**52**[b,26]	274.6	130	−2
26[b,20]	9.5	5.6	−1.7	**53**[b,26]	95.5	20	−4.7
27[b,20]	6	6.4	+1.1	**54**[b,26]	104.1	130	+1.2
28[b,20]	7.8	7.2	−1.1	**55**[b,26]	167.3	120	−1.4
29[a,19]	0.34	0.46	+1.3	**56**[b,26]	78.3	31	−2.5
30[a,22]	1.11	0.42	−2.7	**57**[b,26]	86.3	120	+1.4
31[a,22]	0.079	0.3	+3.8	**58**[a,25]	61.66	140	+2.2
32[b,22]	1.06	0.28	−3.8	**59**[b,25]	100	24	−4.2
33[b,22]	0.39	0.31	−1.3				

[a] Compound included in the Training set. [b] Compound included in the Test set. In the error column "+" indicates that the estimated IC_{50} is higher than the experimental IC_{50}; "−" indicates that the estimated IC_{50} is lower than the experimental IC_{50}.

are shown in the left part of Figure 1 superimposed onto the most active compound (**40**) considered in this study.

The activity of the training and test set compounds was reasonably well estimated by HYPO-7. Experimental and predicted IC_{50} values of all compounds used in the computational study are shown in Figure 2 and Table 1. Notably, the ratio between experimental and predicted values (Table 1, error columns) was within 1 order of magnitude for almost all compounds and in most cases lower than ±3-fold. Only compound **17**, with an IC_{50} at the activity upper border, was mispredicted by HYPO-7, while the less active compounds were estimated at an IC_{50} value of about 140 ± 20 μM, reflecting the pharmacophore approach having some difficulty to quantitatively measure inactivity. Actually, in several cases, the absence of binding has been found not to be due to complete absence of pharmacophore elements, as to other problems such as steric hindrance,[27] cell permeability, pharmacokinetic issues, which are not taken into account by Catalyst.[28]

The pharmacophore model was subsequently used to predict the activity of the new *Laurencia* metabolites **1**−**6**. The in silico screening of all these compounds resulted in reliable agreement with their experimental activity, as shown in Table 2. The superposition of HYPO-7 onto **5** (Figure 3) assisted in clarifying the possible structural reasons for the inactivity demonstrated experimentally by these compounds since the best "function mapping" of all them covered only two of the pharmacophoric features.

A subsequent comparison was done between the spatial disposition of HYPO-7 features (shown in Figure 1) and the interactions described within the active site of ER-α by Brzozowski and co-workers.[29] These authors observed "a combination of specific polar and nonpolar interactions" within ER-α receptor: specifically, hydrogen-bonding interactions at opposite ends of the ligands and a hydrophobic interaction system between the two ends due to the nonpolar character of the binding cavity. The crystal structure of ER-α in complex with raloxifene (pdb entry 1ERR)[29] was used for such a purpose. A valuable superposition was found by comparing the crystal structure of the bound ligand and its pharmacophore conformation, derived by the application to this SERM of a fast fit procedure. In Figure 4 the visual comparison between the two geometries is shown, highlighting the matching, confirmed by the experimental distance of 11.8 Å measured between the ligand oxygen atoms involved in the H-bond interactions, in good agreement with HYPO-7 (see Figure 1). The superpositions of the pharmacophore conformations of **32**[22] and **45**[30] (both members of the test set) with their respective X-ray structures (provided as Supporting Information) confirmed the result on raloxifene. Taking into account these achievements and considering that the low IC_{50} value exhibited by derivative **40** might imply that it possesses all of the crucial groups required to interact with the receptor, it could be assumed that HYPO-7 actually accounts for relevant interactions between inhibitors and ER-α, so that it is not arbitrary to state that matching the pharmacophore may indicate binding to this receptor.

VIRTUAL SCREENING AND MOLECULAR DOCKING

The data shown in Table 2 provided evidence that HYPO-7 would have been able to predict the inactivity of compounds **1**−**6** toward MCF7 cells. In our opinion, however, one more task would have to be fulfilled in order to rely on HYPO-7, that is to validate it by performing an in silico database screening. The favorable outcome of such a procedure would strengthen the confidence in the correlation shown by the model as well as allowing the discovery of new scaffolds to be developed as potential SERM derivatives.

Ligand-based virtual screening usually involves only the application of a pharmacophore to perform a fast database search. On the other hand, several recent studies have shown that coupling of pharmacophore-based virtual screening with docking filtration, that is docking of compounds into a protein binding site followed by scoring and ranking of compounds, can significantly improve the probability of identifying suitable hit candidates.[31] Even though the major aim of our studies was to set up a pharmacophoric filtering tool mainly intended for natural products, we decided to exploit the knowledge of the three-dimensional structure of ER-α to pragmatically combine pharmacophore- and structure-based virtual screening approaches. Such a choice allowed in fact combining the primary purpose of pharmacophore validation with that of hit identification. The steps of our protocol are summarized as follows. HYPO-7 was applied as the first filter to screen the whole Asinex Gold Collection, a database consisting of over 200 000 commercially available derivatives. All the compounds able to map all features in the pharmacophore were retained. Subsequently, in order to narrow down the number of hits, a stepwise postsearch filtering protocol was applied. First, the number of rotatable bonds and the Best Fit value were used to remove too flexible

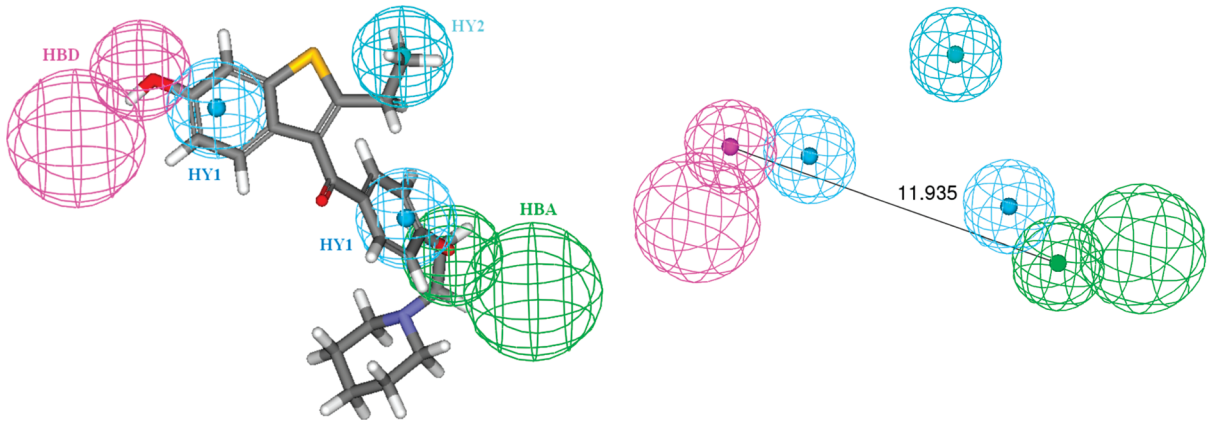


Figure 1. (Left) Superposition of HYPO-7 and **40** (the most active compound in the training set). Features are as follows: HY1, hydrophobic aromatic; HY2, hydrophobic; HBD, hydrogen-bond donor; HBA, hydrogen-bond acceptor. (Right) Distance between selected HYPO-7 features.

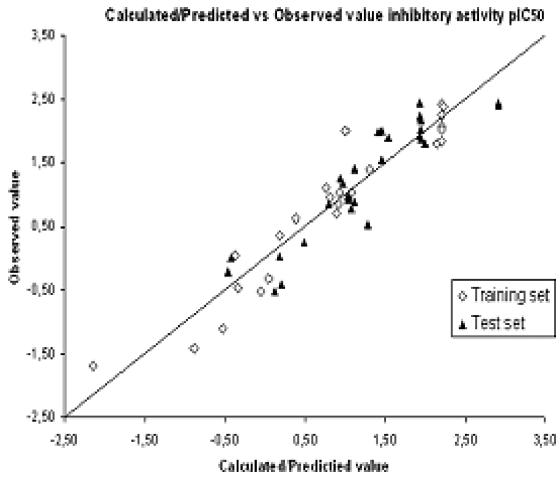


Figure 2. Correlation between the observed and the predicted activity (pIC_{50}) by HYPO-7 for the training and test set compounds.

Table 2. Experimental and Predicted Activity IC_{50} (μM) of Metabolites **1**−**6**

compd	exp	calcd	error[a]
1	>300	120	−2.4
2	>300	120	−2.4
3	N.M.[b]	120	//
4	N.M.[b]	120	//
5	>300	120	−2.4
6	>300	120	−2.4

[a] Calculated with respect to an arbitrarily chosen experimental value of 300. [b] Not measured. In the error column "−" indicates that the estimated IC_{50} is lower than the experimental IC_{50}.

(cutoff set to 10) and/or poorly fitting (cutoff set to 5) entries, respectively. Then an additional filter based on the Lipinski's rule of five[32] was applied. Finally, the resulting compounds were submitted to docking simulations.

A total number of 33 compounds, mapping the pharmacophore as well as satisfying the postsearch filtering, was selected and submitted to our consensus docking protocol to be scored. The crystal structure of ER-α in complex with raloxifene (pdb entry 1ERR) was used to perform the docking experiments. An active site radius of 13.0 Å was chosen by *XYZ* coordinates of raloxifene in 1ERR. Compounds were docked employing the program GOLD 3.0.1 (The Cambridge Crystallographic Data Centre, Cambridge, U.K.)[33] and using

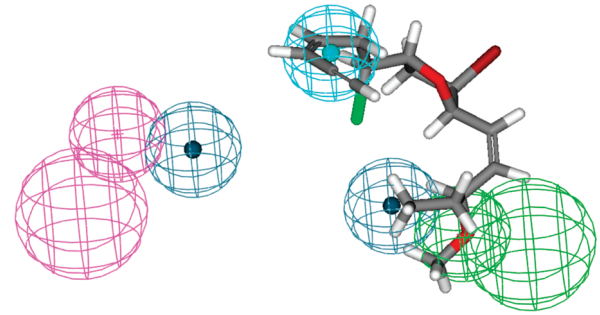


Figure 3. Superposition between HYPO-7 and *Laurencia* metabolite **5**.

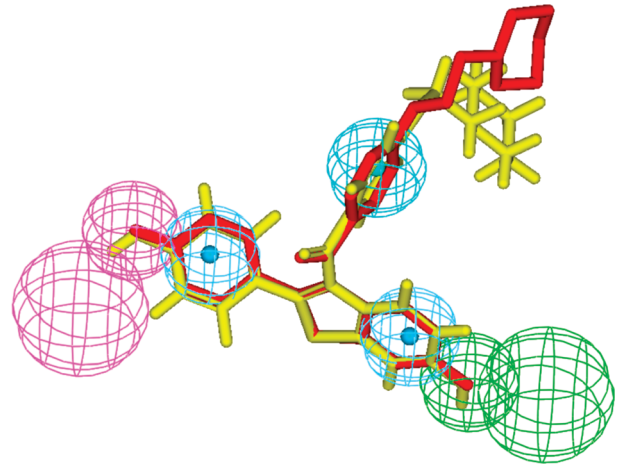


Figure 4. Superposition of HYPO7 with (yellow) the pharmacophore conformation of raloxifene (predicted IC_{50} = 0.014 μM) and (red) the crystal geometry of the ligand when complexed with ER-α. The 1-piperidyl moiety (up in the figure) was ignored in performing the superposition since it was not mapped by any pharmacophoric feature.

the scoring function Goldscore[34] to evaluate and rank binding affinities. Score values of 50 or higher (the choice of this lower limit was driven by the values found for raloxifene and other known potent ligands such as 17β-estradiol[29] and genistein[30]) and appropriate binding (in terms of interactions with the protein) were considered as further final cut off requirements. The combined ligand-based and structure-based calculations led to the final list of 12 entries to be purchased and tested. Different parameters were taken into account for this final selection, such as Catalyst predicted activity,

Table 3. Estimated Activity by HYPO-7, Results from Docking Protocol (Goldscore Value), and Biological Evaluation of Asinex's Compounds

compound	catalyst predicted activity	Goldscore	observed activity (IC_{50}, μM)
BAS1579754	0.010	52.65	60.1 ± 11.6
BAS1832998	0.032	51.26	170.6 ± 21.4
BAS3353570	0.062	50.35	>200
BAS2170314	0.066	50.2	188.4 ± 17.3
BAS3755311	0.07	53.31	67.4 ± 12.1
BAS0915895	0.079	52.72	148.5 ± 15.5
BAS8770034	0.14	51.97	26.4 ± 1.8
BAS0511524	0.490	50.97	31.8 ± 2.5
BAS1247250	0.5	51.15	124.3 ± 15.1
BAS0369551	0.67	52.03	58.5 ± 13.2
BAS1813090	1.0	51.84	40.9 ± 3.1
BAS2377597	3.6	51.49	>200

docking score, structural diversity, and commercial availability of the appropriate amount for biological tests.

The selected compounds were evaluated for their cytotoxicity toward MCF7 human tumor cell line. The outcome of our virtual screening protocol is summarized in Table 3. Activities predicted by Catalyst are reported in the second column of the table, fitness values calculated by GOLD in the third column, and measured IC_{50} values in the fourth column. As shown in Table 3, the virtual screening process was successful in identifying new scaffolds inhibiting MCF7 cell line as 10 out of 12 compounds were found to be active and exhibited IC_{50} values ranging from 26.4 (BAS8770034) to 188.4 μM (BAS2170314) (Chart 3). It is worth noticing that all compounds selected from the virtual screening contained at least two functionalized, with electronegative atoms, aromatic rings positioned at the ends of the molecules. Two of the three most active molecules (BAS1813090 and BAS0511524) are simple triphenyl compounds in which rotation of the aromatic rings is not restrained. The thiazol moiety present in the most active molecule (BAS8770034) is probably not important for expression of activity since a significant drop in activity is observed for compound BAS1247250 that differs only in the substituton pattern of the phenyl rings. The best docking pose of BAS8770034, BAS0511524, and BAS1813090 into ER-α ligand binding domain (see Supporting Information) suggested an interaction pattern of these derivatives comparable with that of raloxifene and of the SERMs used to derive HYPO-7.

In summary, our modeling resulted in a pharmacophore for the MCF7 cell line, which was useful in the estimation of inhibitory activity of SERMs with significantly high reliability. Application of HYPO-7 to new metabolites **1**−**6** would have been useful in predicting the lack of inhibitory activity of these compounds. A successful virtual screening was based on the pharmacophore, which provided a significant and cost-effective solution for identifying inhibitors to a desired target. Even if we cannot demonstrate that our ligand-based modeling provides information that a structure-based pharmacophore or even a docking study are not able to, however, the superpositions performed between pharmacophore conformations and crystallographic structures definitely support the soundness of our approach results. Accordingly, full development and application of such a fast computational tool could allow only potentially active compounds to be commited to pharmacological experiments, thereby effectively assisting the hit discovery process.

EXPERIMENTAL SECTION

Pharmacophore Modeling. Calculations and graphic manipulations were performed on a Silicon Graphic SGI Octane workstation by means of the Catalyst 4.8 software package. The entire set of 53 selected SERMs was divided into a training set of 24 compounds, with activity spanning 4 orders of magnitude (from 2.0×10^{-2} to 274.6 μM), and a complementary test set of 29 compounds (see Table 1). The training set was created according to Catalyst's guidelines.[35] Subsets of derivatives possessing the same structural scaffold were distributed equally between training and test set.

A three-dimensional geometry of each compound was built using Catalyst's View Compound workbench, and the stereochemistry of the chiral centers was specified during this operation. Although experimental data on the biologically relevant conformations of a few ER-α inhibitors were available (atomic coordinates derived from X-ray crystallographic studies of the complexes of 4-hydroxytamoxifen,[36] **32**,[22] and genistein[30] with the receptor), a molecular mechanics approach was applied for the construction of the conformational models of all derivatives, as required in pharmacophore generation, in order to use a homogeneous set of input data. Following Catalyst guidelines, training and test set compounds were submitted to both fast and best conformational searches using the poling algorithm and the CHARMm force field as implemented in Catalyst with the aim of collecting a representative set of conformers chosen within a range of 20 kcal/mol with respect to the global minimum.[16]

Ten hypotheses were generated for the training set after selection of the following features: hydrophobic aromatic (HY1), hydrophobic (HY2), ring aromatic (RA), hydrogen-bond acceptor (HBA), hydrogen-bond donor (HBD). All control parameters were set to their default values except for the uncertainly value, which was set to a value of 1.5 instead of 3. Details concerning the generated hypotheses are given in Table 4 and in the Catalyst output file provided as Supporting Information. A brief assessment of the statistical parameters of all 10 hypotheses was done through the cost analysis. According to Catalyst guidelines,[35] a 39.8 bit cost range gained over the models, suggesting the existence of a strong signal generated by the training set; moreover, a difference much greater than 100 bits between the cost of each hypothesis and the null cost was indicative of more than 90% of true correlation. Emphasis was given to test the models for arbitrary correlation. The pharmacophores were evaluated for statistical significance using a randomization trial procedure derived from the Fischer method.

Test set molecules were fitted to the generated hypothesis using the Catalyst fast fit feature. Fast fit refers to the method of finding the optimum fit of the substrate to the hypothesis among all the conformers of the molecule without performing an energy minimization. This method was also used to fit the training set molecules to the pharmacophore. The prediction of the test set (r^2 test set in Table 4) was used as the factor to select the most reliable pharmacophore among

Chart 3. Structure of Asinex Compounds Selected by the Virtual Screening

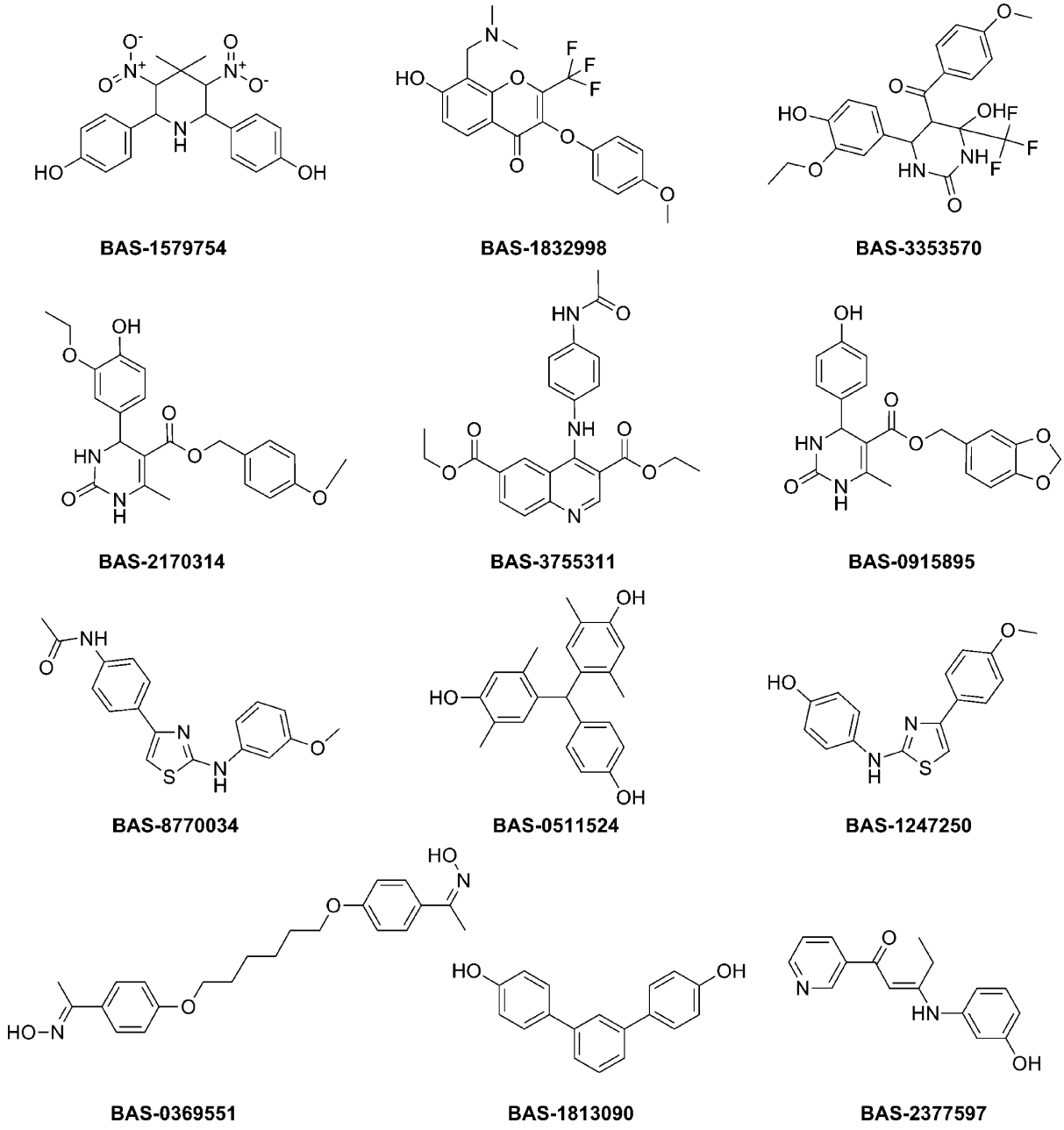


Table 4. Total Cost, Statistical Parameters (RMSD, $r^2_{tr.-set}$ and $r^{v2}_{test-set}$), and Composition Features Associated with the 10 Pharmacophoric Hypotheses (Hypo) Generated by Catalyst

	statistical parameter				composition feature[b]				
Hypo	total cost[a]	rmsd	$r^2_{tr.set}$	$r^2_{testset}$	1	2	3	4	5
1	110.7	1.53	0.97	0.77	HBD	HY1	HY2	RA	
2	122.5	1.80	0.97	0.78	HBD	HY1	HY2	RA	
3	148.2	2.31	0.94	0.78	HBD	HY1	HY2	RA	
4	162.1	2.49	0.93	0.81	HBD	HY1	HY2	RA	
5	166.7	2.61	0.92	0.74	HBA	HY1	HY1	HY2	RA
6	169.8	2.64	0.92	0.79	HBD	HY1	HY2	RA	
7	171.1	2.65	0.96	0.91	HBA	HBD	HY1	HY1	HY2
8	174.4	2.73	0.92	0.69	HBA	HY1	HY1	HY2	RA
9	174.9	2.71	0.94	0.88	HBA	HBD	HY1	HY1	HY2
10	176.9	2.76	0.94	0.75	HBA	HY1	HY2	RA	

[a] All costs are reported in bits. Fixed cost = 81.2 is the cost of a theoretical ideal hypothesis which is able to perfectly predict activities. Null cost = 636.6 is the cost of a hypothesis that gives no correlation between experimental and predicted activities. [b] HBA: hydrogen-bond acceptor. HBD: hydrogen-bond donor. HY1: hydrophobic aromatic. HY2: hydrophobic. RA: ring aromatic.

the generated hypotheses. On the basis of this consideration, HYPO-7 was chosen. The ligand-based pharmacophore model exhibited a correlation coefficient $r^2 = 0.96$ for the training set and $r^2 = 0.91$ for the test set. Fast fit of raloxifene

onto HYPO-7 and other graphic manipulations were performed with the Discovery Studio 2.0 software package.[37]

Molecular Docking. Docking was carried out using GOLD 3.0.1 (Genetic Optimization for Ligand Docking) software from Cambridge Crystallographic Data Center, U.K., that uses the Genetic algorithm (GA),[33] running under Linux OS. This method allows a partial flexibility of protein and full flexibility of ligand. All water molecules and ligand were removed from the protein to evaluate the scoring function Goldscore in GOLD software. The three-dimensional structure of the ER-α target was taken from PDB entry 1ERR.[29] For each of the 100 independent GA runs, a maximum number of 100 000 GA operations were performed on a set of five groups with a population size of 100 individuals. Operator weights for crossover, mutation, and migration were set to 95, 95, and 10, respectively. The active site radius of 13 Å was chosen by *XYZ* coordinates of the ligand (raloxifene) in PDB entry 1ERR. Default cut off values of 2.5 Å (dH−X) for hydrogen bonds and 4.0 Å for the van der Waals distance were employed. When the top three solutions attained rmsd values within 1.5 Å, GA docking was terminated. The rmsd values for the docking calculations are based on the rmsd matrix of the ranked solutions.

Database Searching. Catalyst-generated best pharmacophore model (HYPO-7) comprising of five chemical features was used as a query for searching in the chemical 3D databases Asinex (Asinex Ltd., 5 Gabrichevskogo, St. Building 8, Moscow 125367, Russia), implemented in the Catalyst 4.8 software package. This chemical 3D databases consists of about 200 000 structurally diversified small molecules. Virtual screening of such databases can serve two main purposes: first, validating the quality of the generated pharmacophore models by selective detection of compounds with known inhibitory activity, and, second, finding novel, potential leads suitable for further development. The best flexible search method was used for database searching to retrieve new lead molecules.

Determination of Cytotoxicity. Cells were routinely cultured in Dulbecco's minimal essential medium supplemented with penicillin (100 U/mL), streptomycin (100 μg/mL), and 10% fetal bovine serum in an environment of 5% CO_2, 85% humidity, and 37 °C, and they were subcultured using a trypsin 0.25%−EDTA 0.02% solution. Cells were plated in 96-well flat-bottomed microplates at a density of 1×10^5 cells/mL (100 μL/well), and 24 h later the test compounds were added, appropriately diluted with DMSO. Cells were exposed to various concentrations of the compounds for 48 h. The cytotoxicity was determined with the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) dye reduction assay[38] with minor modifications.[39] Briefly, after incubation with the test compounds, MTT solution (5 mg/mL in PBS) was added (20 μL/well). Plates were further incubated for 4 h at 37 °C, and the formazan crystals formed were dissolved by adding 100 μL/well of 0.1 N HCl in 2-propanol. Absorption was measured by an enzyme-linked immunosorbant assay (ELISA) reader at 545 nm with a reference filter at 690 nm. For each concentration at least nine wells were used from three separate experiments. One hundred microliters of culture medium supplemented with the same amount of MTT solution and solvent was used as blank solution. Data obtained were presented as IC_{50} (μM), which is the concentration of the compound where $100 \times (A_0 - A)/A_0 = 50$. In this formula, A is the optical density of the wells after 48 h of exposure to test compound and A_0 is the optical density of the control wells. All data are expressed as mean ± SD.

Culture media and antibiotics were from Biochrom KG (Berlin, Germany). All other chemicals were from Sigma-Aldrich.

ACKNOWLEDGMENT

This study was partially supported by the 01ED146 PENED program of the Greek Secretariat for Research and Technology. A.T. gratefully acknowledges project FIRB 2003, protocol number RBNE034XSW_005, of the Ministero Italiano della Università e della Ricerca.

Supporting Information Available: Superpositions of the pharmacophore conformations of **32** and **45** with their respective X-ray structures; Best docking pose of BAS8770034, BAS0511524, BAS1813090, and raloxifene into ER-α ligand binding domain; Catalyst output file; Aligned structures of all the studied derivatives (three-dimensional .SD file). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Kladi, M.; Vagias, C.; Papazafiri, P.; Furnari, G.; Serio, D.; Roussis, V. New sesquiterpenes from the red alga Laurencia microcladia. *Tetrahedron* **2007**, *63*, 7606–7611.
(2) Kontiza, I.; Stavri, M.; Zloh, M.; Vagias, C.; Gibbons, S.; Roussis, V. New metabolites with antibacterial activity from the marine angiosperm Cymodocea nodosa. *Tetrahedron* **2008**, *64*, 1696–1702.
(3) Kladi, M.; Vagias, C.; Papazafiri, P.; Brogi, S.; Tafi, A.; Roussis, V. Tetrahydrofuran acetogenins from Laurencia glandulifera. *J. Nat. Prod.* **2009**, *72*, 190–193.
(4) Wermuth, C. G. Pharmacophores: historical perspective and viewpoint from a medicinal chemist. In *Pharmacophores and Pharmacophore Searches*; Langer, T., Hoffman, R. D., Eds.; Wiley-VCH: Weinheim, Germany, 2006; Vol. 32, pp 3−11.
(5) Steindl, T. M.; Schuster, D.; Laggner, C.; Langer, T. Parallel screening: a novel concept in pharmacophore modeling and virtual screening. *J. Chem. Inf. Model.* **2006**, *46*, 2146–2157.
(6) Rollinger, J. M.; Stuppner, H.; Langer, T. Virtual screening for the discovery of bioactive natural products. In *Natural Compounds as Drugs Volume I*; Petersen, F., Amstuz, R., Eds.; Birkhäuser: Basel, Switzerland, 2008; Vol. 65, pp 211−249.
(7) International Science Forum on Computational Toxicology; http://www.epa.gov/comptox/forum/ (accessed February 14, 2007).
(8) Kontiza, I.; Abatis, D.; Malakate, K.; Vagias, C.; Roussis, V. 3-Keto steroids from the marine organisms Dendrophyllia cornigera and Cymodocea nodosa. *Steroids* **2006**, *71*, 177–181.
(9) Ehrlich, P. Über den jetzigen Stand der Chemotherapie. *Dtsch. Chem. Ges.* **1909**, *42*, 17–47.
(10) Perera, N. M.; Gui, G. P. Multi-ethnic differences in breast cancer: current concepts and future directions. *Int. J. Cancer* **2003**, *106*, 463–467.
(11) Shao, W.; Brown, M. Advances in estrogen receptor biology: prospects for improvements in targeted breast cancer therapy. *Breast Cancer Res.* **2004**, *6*, 39–52.
(12) Dunn, B. K.; Ryan, A. Phase 3 trials of aromatase inhibitors for breast cancer prevention: following in the path of the selective estrogen receptor modulators. *Ann. N.Y. Acad. Sci.* **2009**, *1155*, 141–161.
(13) *Catalyst*, version 4.8; Accelrys, Inc.: San Diego, CA, 2003.
(14) Grese, T. A.; Cho, S.; Finley, D. A.; Godfrey, A. G.; Jones, C. D.; Lugar, C. W., III; Martin, M. J.; Matsumoto, K.; Pennington, L. D.; Winter, M. A.; Adrian, M. D.; Cole, H. W.; Magee, D. E.; Phillips, D. L.; Rowley, E. R.; Short, L. L.; Glasebrook, A. L.; Bryant, H. U. Structure-activity relationships of selective estrogen receptor modulators: modifications to the 2-arylbenzothiophene core of raloxifene. *J. Med. Chem.* **1997**, *40*, 146–167.
(15) Mukherjee, S.; Nagar, S.; Mullick, S.; Mukherjee, A.; Saha, A. Pharmacophore mapping of arylbenzothiophene derivatives for MCF cell inhibition using classical and 3D space modeling approaches. *J. Mol. Graph. Model.* **2008**, *26*, 884–892.

(16) Accelrys, Inc. Accelrys: Catalyst; http://www.accelrys.com/products/catalyst (accessed June 26, 2007)

(17) Dowers, T. S.; Qin, Z.; Thatcher, G. R. J.; Bolton, J. Bioactivation of Selective Estrogen Receptor Modulators (SERMs). *Chem. Res. Toxicol.* **2006**, *19*, 1125–1137.

(18) Trafalis, D. T. P.; Geromichalos, G. D.; Koukoulitsa, C.; Papageorgiou, A.; Karamanakos, P.; Camoutsis, C. Lactandrate: a D-homo-aza-androsterone alkylator in the treatment of breast cancer. *Brest Cancer Res. Treat.* **2006**, *97*, 17–31.

(19) Firth-Clarck, S.; Willems, H. M.; Williams, A.; Harris, W. Generation and selection of novel estrogen receptor ligands using de novo structure-based design tool, SkelGen. *J. Chem. Inf. Model.* **2006**, *46*, 642–647.

(20) Kumar, S. K.; Hager, E.; Pettit, C.; Gurulingappa, H.; Davidson, N. E.; Khan, S. R. Design, synthesis and evaluation of novel boronic-chalcone derivatives as antitumor agents. *J. Med. Chem.* **2003**, *46*, 2813–2815.

(21) Gangjee, A.; Yu, J.; Copper, J. E.; Smith, C. D. Discovery of novel antitumor antimicotic agents that also reserve tumor resistance. *J. Med. Chem.* **2007**, *50*, 3290–3301.

(22) Manas, E. S.; Unwalla, R. J.; Xu, Z. B.; Malamas, M. S.; Miller, C. P; Harris, H. A.; Hsiao, C.; Akopian, T.; Hum, W. T.; Malakian, K.; Wolfrom, S.; Bapat, A.; Bhat, R. A.; Stahl, M. L.; Somers, W. S.; Alvarez, J. C. Structure-based design of estrogen receptor-β selective ligands. *J. Am. Chem. Soc.* **2004**, *126*, 15106–15119.

(23) Zimmermann, J.; von Angerer, E. Estrogenic and antiestrogenic activities of 2,4-diphenylfuran-based ligands of estrogen receptors α and β. *J. Steroid Biochem. Mol. Biol.* **2007**, *104*, 259–268.

(24) Larrosa, M.; Gonzàlez-Sarrìas, A.; Garcia-Conesa, M. T.; Tomàs-Barberan, F. A.; Espìn, J. C. Urolithins, Elagic, Acid-derived metabolites produced by human colonic microflora exhibit estrogenic and antiestrogenic activities. *J. Agric. Food Chem.* **2006**, *54*, 1611–1620.

(25) Kode, N.; Chen, L.; Murthy, D.; Adewumi, D.; Phadtare, S. New bis-N9-(methylphenylmethyl) purine derivatives: synthesis and antitumor activity. *Eur. J. Med. Chem.* **2007**, *42*, 327–333.

(26) Kladi, M. Isolation and pharmacological evaluation of bioactive metabolites from the genus Laurencia, Asparagopsis and Falkenbergia. Ph.D. Thesis; University of Athens, Greece, 2006; pp 377−384.

(27) La Regina, G.; Diodata D'Auria, F.; Tafi, A.; Piscitelli, F.; Olla, S.; Caporuscio, F.; Nencioni, L.; Cirilli, R.; La Torre, F.; Rodrigues De Melo, N.; Kelly, S. L.; Lamb, D. C.; Artico, M.; Botta, M.; Palmara, A. T.; Silvestri, R. 1-[(3-Aryloxy-3-aryl)propyl]-1H-imidazoles, new imidazoles with potent activity against Candida albicans and dermatophytes. Synthesis, structure-activity relationship, and molecular modeling studies. *J. Med. Chem.* **2008**, *51*, 3841–3855.

(28) Chang, C.; Ekin, S.; Bahadduri, P.; Swaan, P. W. Pharmacophore-based discovery of ligands for drug transporters. *Adv. Drug Delivery Rev.* **2006**, *58*, 1431–1450.

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

(30) Manas, E. S.; Xu, Z. B.; Unwalla, R. J.; Somers, W. S. Understanding the selectivity of genistein for human estrogen receptor-beta using X-ray crystallography and computational methods. *Structure* **2004**, *12*, 2197–2207.

(31) Olla, S.; Manetti, F.; Crespan, E.; Maga, G.; Angelucci, A.; Schenone, S.; Bologna, M.; Botta, M. Indolyl-pyrrolone as a new scaffold for Pim1 inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 1512–1516.

(32) Lipinsky, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.

(33) Jones, G.; Willet, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53.

(34) Jones, G.; Willet, P.; Glen, R. C.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

(35) Accelrys, Inc. Automated Training Set Selection used for Developing a 3D QSAR Pharmacophore Model in CATALYST; http://accelrys.com/references/case-studies/training-set-selection.html (accessed September 29, 2009)

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

(37) *Discovery Studio*, version 2.0; Accelrys: San Diego, CA, 2007.

(38) Mosmann, T. Rapid colorimetric assay for cellular assay for cellular growth and survival application to proliferation and cytotoxicity assay. *J. Immunol. Methods* **1983**, *65*, 55–63.

(39) Alley, M. C.; Scudiero, A. D.; Monks, A.; Hursey, M. L.; Czerwinski, M. J.; Fine, D. L.; Abbott, B. J.; Mayo, J. G.; Shoemaker, R. H.; Boyd, M. R. Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay. *Cancer Res.* **1988**, *48*, 589–601.

CI900254B