

CoMFA and docking study of novel estrogen receptor subtype selective ligands

Peter Wolohan & David E. Reichert*

Mallinckrodt Institute of Radiology, Washington University School of Medicine, 510 South Kingshighway, Campus Box 8225, St. Louis, MO 63110

Received 31 December 2002; accepted in revised form 19 April 2003

Key words: CoMFA, docking, estrogen receptor, scoring function, subtype selective

Summary

We present the results from a Comparative Molecular Field Analysis (CoMFA) and docking study of a diverse set of 36 estrogen receptor ligands whose relative binding affinities (RBA) with respect to 17 β -Estradiol were available in both isoforms of the nuclear estrogen receptors (ER α , ER β). Initial CoMFA models exhibited a correlation between the experimental relative binding affinities and the molecular steric and electrostatic fields; ER α : $r^2 = 0.79$, $q^2 = 0.44$ ER β : $r^2 = 0.93$, $q^2 = 0.63$. Addition of the solvation energy of the isolated ligand improved the predictive nature of the ER β model initially; $r^2 = 0.96$, $q^2 = 0.70$ but upon rescrambling of the data-set and reselecting the training set at random, inclusion of the ligand solvation energy was found to have little effect on the predictive nature of the CoMFA models. The ligands were then docked inside the ligand binding domain (LBD) of both ER α and ER β utilizing the docking program Gold, after-which the program CScore was used to rank the resulting poses. Inclusion of both the Gold and CScore scoring parameters failed to improve the predictive ability of the original CoMFA models. The subtype selectivity expressed as RBA(ER α /ER β) of the test sets was predicted using the most predictive CoMFA models, as illustrated by the cross-validated r^2 . In each case the most selective ligands were ranked correctly illustrating the utility of this method as a prescreening tool in the development of novel estrogen receptor subtype selective ligands.

Introduction

Estrogens are predominantly formed in the reproductive organs of the human body, specifically the ovaries and testis, from which they infiltrate many cells in the body. They play a critical role in the growth, development and sustenance of a wide range of tissues. In particular they play a critical role in the physiology of the female reproductive system, the maintenance of bone density, and cardiovascular health. However stimulation of other tissues can increase the risk of cancer within these tissues, particularly the female breast. The estrogen receptor is found in two isoforms; the classical subtype alpha estrogen receptor (ER α) and the recently discovered subtype beta estrogen

receptor (ER β). [1, 2] Upon ligand binding, the receptor undergoes a conformational change, specifically in the activation function-2 helix (H12), allowing it to bind to chromatin resulting in the activation or repression of responsive genes. [3] As a result the gene transcription activity of the estrogen receptor is regulated by the type of ligand (agonist/antagonist) which is bound to the receptor. The exact mechanism by which the estrogen receptor affects gene transcription is poorly understood, as is the role of each subtype. The possibility that tissue and cell selectivity of certain estrogens is due to their specificity to the newly discovered ER β estrogen receptor has been validated by recent studies. [4, 5] Indeed it has been reported that the pharmacology of several classical estrogen receptor agonists and antagonists is reversed for ER β . [6, 7] As a result there exists the opportunity to develop novel subtype-selective ligands; to date the subtype-

*To whom correspondence should be addressed.
E-mail: reichertd@mir.wustl.edu

selectivity of the classical estrogens and antiestrogens has been modest [2].

As a result of the role estrogens play in the physiology of both humans and animal species the estrogen receptors represent a viable and important pharmaceutical target. In particular the estrogen receptors are a target for pharmaceutical agents for hormone replacement in menopausal women, reproductive cancers such as prostate cancer, uterine and breast cancer. Pharmaceuticals developed to date can be divided into three distinct categories. The first category acts solely as receptor agonists as with the estrogen receptor's natural ligand 17β -estradiol. The second category includes antiestrogens such as the compound ICI 164,384 and act as pure antagonists. The third category includes antiestrogens which have the ability to act as both agonists or antagonists and includes such compounds as tamoxifen and LY117018. As previously noted, to date these compounds have exhibited only modest ER isoform specificity.

However, recent research by Katzenellenbogen and co-workers has focused on several series of nonsteroidal compounds based on substituted furans, pyrazoles and tetrahydrochrysenes.[8–10] These compounds have been shown to exhibit unprecedented estrogen subtype selectivity compared to the classical steroidal compounds. In particular the compound 4-propyl-1,3,5-triphenolpyrazole (**4g**), Table 2, has been reported to have 400-fold affinity for the ER α isoform of the estrogen receptor. These new compounds provide an opportunity to understand the basis of ER subtype selectivity, and form the basis of the reported study. The development of models for predicting the isoform selectivity will allow for the development of more potent and selective compounds for these important pharmaceutical targets.

One approach to understanding these interactions is that of QSAR (Quantitative Structure-Activity Relationship) with two recent studies on estrogen ligands being found in the work of Gao and Hansch [11, 12]. Several approaches utilizing three-dimensional structural information have been developed with one of the most widely used being CoMFA (Comparative Molecular Field Analysis) [13]. Numerous CoMFA type studies of estrogen receptor ligands have appeared in the literature [14–18]. CoMFA is based on the premise that the pharmacophoric elements which are responsible for the biological activity of a compound, be it favorable or unfavorable, will be represented in the calculated steric and electrostatic fields of the compound. By studying a series of compounds, called

the training set, consisting of compounds with good, medium and poor bioactivity for a specific protein target it is possible to extrapolate a three-dimensional pharmacophoric model that explains the observed bioactivity. Indeed the model suggests how the steric and electrostatic fields might be manipulated to produce a novel compound with enhanced bioactivity.

One requirement of CoMFA is that the compounds in the training set be aligned against each other so that the overlap of the pharmacophoric elements responsible for producing a biological response is maximized. Often the lowest energy conformation of the each compound in the training set is used, but is the lowest energy conformation the bound conformation [19]? In cases where the crystal structure of the target protein in complexation with a ligand has been resolved the structure of the bound ligand can be used as a template. However, even in this advantageous case it is difficult to deal with compounds in the training set which might have multiple protein binding conformations while maintaining a high pharmacophoric overlap with the template compound. Several authors have developed enhanced 3D-QSAR routines where predictive models are constructed from descriptors which are independent of alignment or from the iterative sampling and statistical manipulation of multiple docking poses [15, 20].

An alternative approach to QSAR and CoMFA in evaluating potential ligands are receptor based docking methods, such as DOCK, FlexX, Gold and numerous other programs. The ligands are fit into the receptor site through a variety of techniques and the resultant poses evaluated through a variety of scoring functions. A relatively recent development has been the use of consensus scoring, whereby the poses are ranked according to a consensus of several scoring functions.[21, 22] Related to this method are the ranking of the docked poses by direct calculation of the interaction energies between the ligand and protein using molecular dynamics or molecular mechanics. Several authors have developed approaches which utilize a combination of receptor docking and enhanced 3D-QSAR routines. In several works it has been shown that the predictive power of the resultant models are improved over those developed by the individual methods [18, 20, 23, 24].

In this paper we present an examination of some of the most subtype selective ligands for the estrogen receptor. The docking program Gold has been utilized to predict the preferred pose for a diverse set of ligands, whose relative binding affinities have been ex-

perimentally determined within one laboratory. These poses were then utilized as the alignment of the ligands for CoMFA studies, resulting in predictive models for both ER α and ER β . Additionally the use of scoring functions as descriptors in CoMFA studies were evaluated.

Preparation of molecular structures

I. Preparation of nuclear receptor estrogen protein structures

The crystal structures of 17 β -estradiol complexed to ER α (1ERE) and genistein complexed to ER β (1QKM) were extracted from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB-PDB). These structures were read in and manipulated with the program Sybyl 6.7. In each case missing residues or atoms in the crystal structure were added in order to complete the protein chain. Hydrogen atoms were assigned to these crystal structures and then minimized using the Tripos force field, using a distance dependent dielectric constant of 1.0, while the rest of the structure was held fixed until the maximum derivative was <0.01 kcal/(mol Å). At this point Gasteiger-Huckel partial charges were added to each atom in the structures. The added or repaired residues were then minimized until the maximum derivative was <0.01 kcal/(mol Å) while all other atoms were fixed. Finally, for each protein-ligand the entire system was allowed to relax until the maximum derivative was <0.01 kcal/(mol Å), with the exception of the ligand and a single water molecule which is present in both protein-ligand complexes and is considered important for binding [25, 26].

II. Preparation of ligand structures

The A-ring from the crystal structure of 17 β -estradiol in complexation with ER α was used a building block for the construction of molecular models of the other 36 inhibitors. Gasteiger-Huckel partial charges were added to each atom in the structures and they were all minimized using the Tripos force field, using a distance dependent dielectric constant of 1.0, until the maximum derivative was <0.01 kcal/(mol Å). The lowest energy conformation of each ligand was then located by using the Genetic Algorithm (GA) search implemented in Sybyl 6.7. In each case, all rotatable bonds were selected, 50000 conformations were generated by the GA with a population of 500 and unique

conformations within 2 kcal/mol of the lowest energy conformation were retained. The GA search was selected so as to be consistent with the program Flexidock, which is also implemented in Sybyl 6.7 and utilizes a GA to search the conformational space of a ligand once docked inside a protein.

Method and discussion

The program Gold has been used extensively to study the docking of ligands in proteins [27–29]. Gold uses a genetic algorithm to examine multiple protein-ligand binding configurations and, through the application of energy functions based on conformation and non-bonded contact information from the Cambridge Structural Database (CSD) and the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB-PDB), ranks these protein-ligand conformations in order of fitness. By utilizing such a program the human element of deciding which conformation is the most probable binding conformation of a particular set of compounds in a given ligand is removed.

Once a viable pose is obtained it is possible to calculate the theoretical binding affinity of the ligands. It has been reported by several authors that in many protein-ligand systems there exists a strong correlation between the calculated gas-phase binding affinity and the biological activity of the ligand [30–33]. Accurate prediction of the binding affinity is time consuming as even in the simplest cases the minimization of a large protein involves several thousand atoms. In many cases it is also necessary to further sample the conformational space of the ligand docked into the receptor adding an additional time cost to the process [34, 35]. As a consequence of the time consuming nature of calculating the theoretical binding affinities for a large series of protein-ligand complexes, many researchers have developed algorithms which are designed to very efficiently estimate the magnitude of the theoretical binding affinity of a protein-ligand complex for the purpose of screening large virtual databases of ligands. Several of these algorithms are available under the program CScore which is part of the Sybyl suite of programs [36]. Once a measure of the ligand binding affinity is calculated and a strong correlation with experimental biological activity is found, it is possible to utilize this method to evaluate new candidate ligands as viable protein inhibitors. Furthermore, it is possible to add the theoretical binding affinity to a CoMFA model in order to improve

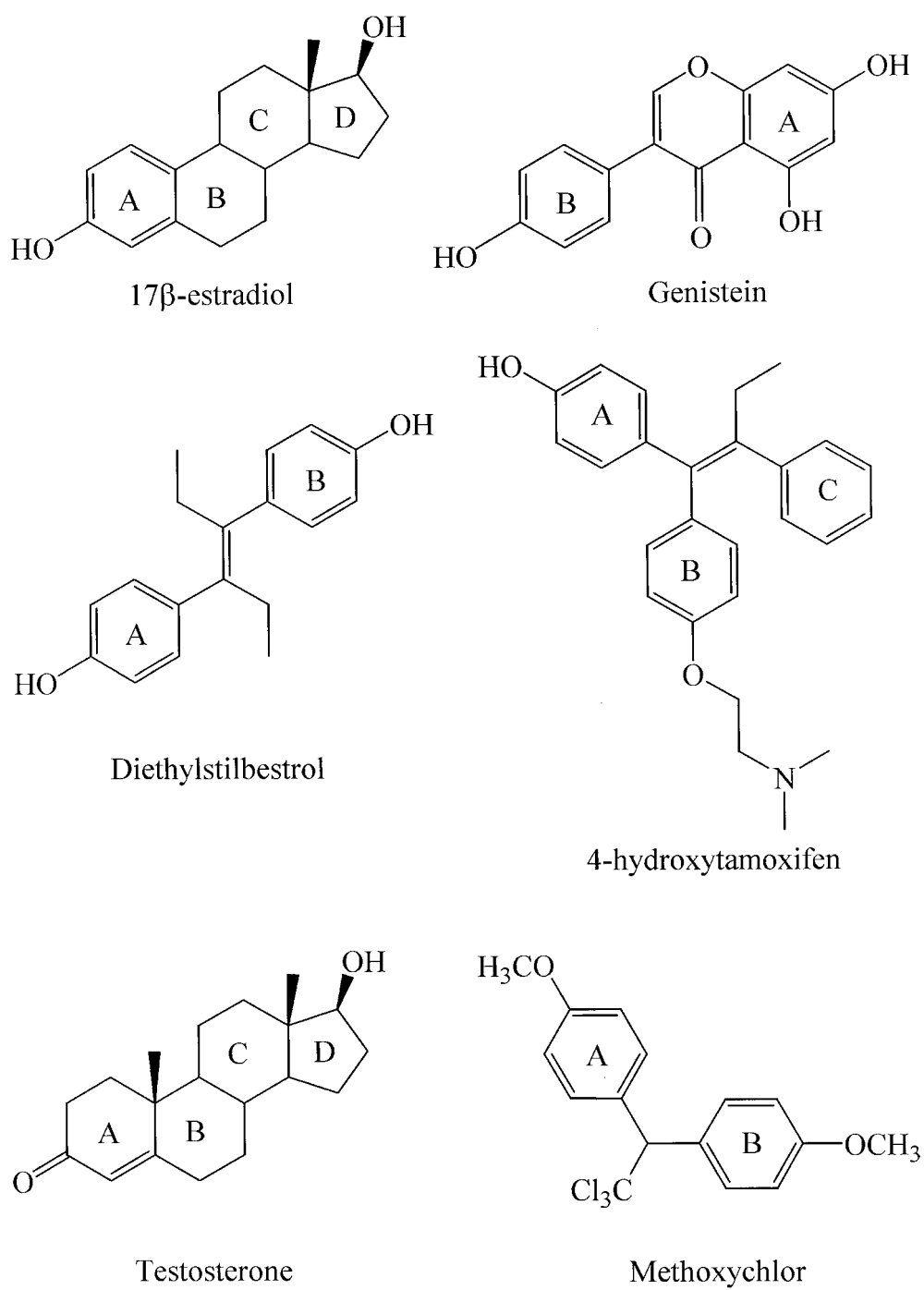


Figure 1. Common estrogenic ligands used in this study with their pharmacophoric elements highlighted.

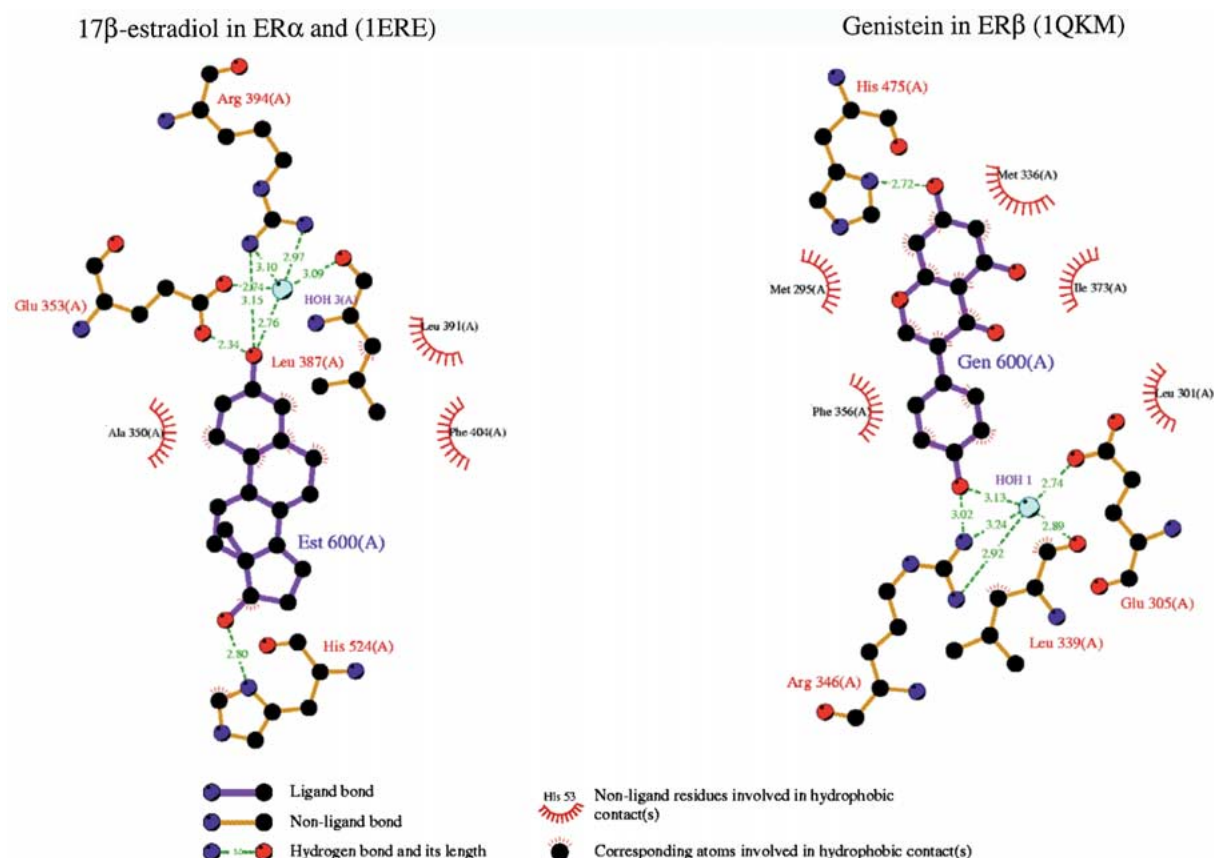


Figure 2. LIGPLOT schematic of hydrogen-bonding network of (a) 17β-estradiol in ERα and (1ERE) (b) Genistein in ERβ (1QKM).

the predictive nature of the model thus utilizing both theoretical techniques to aid in the process of the ligand-based design of new drug candidates [16–18, 24, 37].

Figure 1 illustrates the chemical structure of several of the common estrogenic and steroidal ligands used in this study with their pharmacophoric elements highlighted. Examining the ligand binding sites of both crystal structures, Figure 2, the natural estrogen 17β-estradiol interacts with ERα via a hydrogen-bonding network composed of the hydroxyl group of the A-ring with ARG394, GLU353 and a single water molecule, while the hydroxyl group of the D-ring of 17β-estradiol forms a hydrogen bond with HIS524. These hydrogen-bonding interactions form the basis of the favorable binding interaction of 17β-estradiol with ERα and the core elements for a pharmacophoric model of the ERα binding pocket. In the case of genistein the interaction with ERβ is virtually identical. Genistein interacts with ERβ via a hydrogen-bonding network with the hydroxyl group of the B-ring with

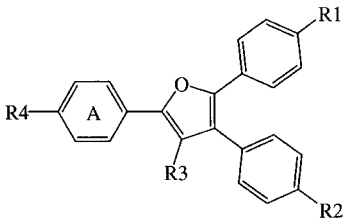
Table 1. Experimental relative binding affinities (RBA) of common estrogenic ligands.

Ligand	RBA ^a		Ratio RBA (ERα/ERβ)
	ERα	ERβ	
17β-Estradiol	100	100	1
17α-Estradiol	58	11	5.3
Genistein	5	36	0.1
Diethylstilbestrol	468	295	1.6
Dienestrol	223	404	0.6
4-OH-Tamoxifen	178	339	0.5
Tamoxifen	7	6	1.2
Methoxychlor	0.01	0.13	0.1
5-Androstenediol	6	17	0.4
Dihydrotestosterone	0.05	0.17	0.3
Norethindrone	0.07	0.01	7
Testosterone	<0.01	<0.01	1

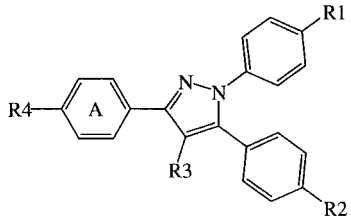
^aDetermined by competitive radiometric binding assay, where the RBA of 17β-Estradiol is arbitrarily set at 100. All experimental data was taken from Kuiper *et al.* [4].

Table 2. Experimental relative binding affinities (RBA) of substituted furans, pyrazoles and single THC species used in this study.

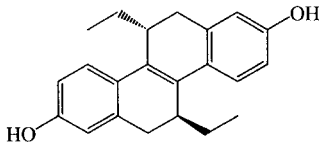
Ligand	R-Groups				RBA ^a		Ratio RBA (ER α /ER β)
	R1	R2	R3	R4	ER α	ER β	



Furans



Pyrazoles



<i>Furans</i>							
15a	OH	OH	CH ₃	OH	40 ± 6.5	0.62 ± 0.02	65
15b	OH	OH	C ₂ H ₅	OH	140 ± 38	2.9 ± 0.1	48
15c	OH	OH	C ₃ H ₇	OH	100 ± 14	1.8 ± 0.65	56
15d	OH	OH	C ₄ H ₉	OH	21 ± 0.6	3.9 ± 1.1	5.4
15e	OH	OH	<i>p</i> -HOC ₆ H ₄	OH	8.7 ± 1.1	0.25 ± 0.01	36
15f	H	OH	C ₂ H ₅	OH	82 ± 20	7.1 ± 1.2	12
15g	H	OH	C ₃ H ₇	OH	140 ± 13	15 ± 4.1	9.5
15h	OH	H	C ₂ H ₅	OH	16.5 ± 1.9	3.0 ± 0.6	5.5
15i	OH	OH	C ₂ H ₅	H	14.8 ± 3.1	4.5 ± 1.2	3.3
15j	H	H	C ₂ H ₅	OH	10.8 ± 2.6	3.4 ± 1.2	3.8
15k	H	OH	C ₂ H ₅	H	0.15 ± 0.01	0.07 ± 0.02	2.1
15l	OH	H	C ₂ H ₅	H	6.8 ± 0.4	2.0 ± 0.4	3.4
<i>Pyrazoles</i>							
4a	H	OH	CH ₃	OH	0.76 ± 0.18	0.28 ± 0.16	2.7
4b	H	OH	C ₂ H ₅	OH	31 ± 0.15	1.1 ± 0.2	28
4c	H	OH	<i>n</i> -C ₃ H ₇	OH	16.8 ± 0.3	0.52 ± 0.03	32
4d	H	OH	<i>i</i> -C ₄ H ₉	OH	56 ± 6	1.4 ± 0	40
4e	H	OH	<i>n</i> -C ₄ H ₉	OH	8.7 ± 2.0	0.47 ± 0.1	19
4f	OH	OH	C ₂ H ₅	OH	36 ± 6	0.15 ± 0.01	240
4g	OH	OH	<i>n</i> -C ₃ H ₇	OH	49 ± 12	0.12 ± 0.04	410
4h	OH	OH	<i>i</i> -C ₄ H ₉	OH	75 ± 6	0.89 ± 0.06	84
4i	OH	OH	<i>n</i> -C ₄ H ₉	OH	14 ± 4	0.18 ± 0.09	77
6	H	OH	<i>i</i> -C ₃ H ₇	OH	5.6 ± 2	0.86 ± 0.11	6.5
7c	H	OH	C ₂ H ₅	H	0.04 ± 0.11	0.06 ± 0.01	0.67
3b	THC				221 ± 42	432 ± 21	0.51

^aDetermined by competitive radiometric binding assay, where the RBA of 17 β -Estradiol is arbitrarily set at 100. Values represent the average (\pm SD or range) of multiple determinations. All experimental data for the furans was taken from Mortensen *et. al.* [8], for the pyrazoles from Stauffer *et. al.* [9], and for the THC (3b) from Meyers *et. al.* [10].

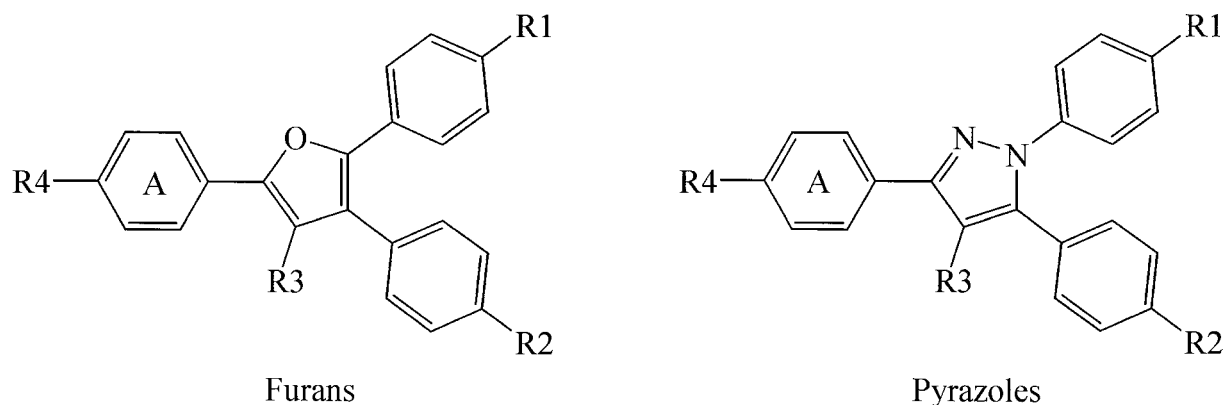


Figure 3. Novel estrogen subtype selective ligands with their A-ring estradiol mimics highlighted.

ARG346, GLU305 and a single water molecule, while the hydroxyl group of the A-ring of genistein forms a hydrogen bond with HIS475. Again these hydrogen-bonding interactions form the basis of the favorable binding interaction of genistein with ER β and the core elements for a pharmacophoric model of the ER β binding pocket. Virtually all ligands designed to date exhibit at least an A-ring mimic. Figure 3 illustrates the generic structure of the novel estrogen subtype selective ligands with their A-ring estradiol mimics highlighted. Tables 1 and 2 list the experimental binding affinity of each of the ligands relative to that of the natural estrogen 17 β -estradiol that is arbitrarily set at 100. The subtype selectivity for each ligand is also listed in Tables 1 and 2.

As one can see from Table 1 diethylstilbestrol exhibits the highest RBA for ER α while from Table 2 the novel non-steroidal compound trans-5,11-diethyl-5,6,11,12-tetrahydrochrysene-2,8-diol exhibits the highest RBA for ER β . However, neither of these ligands exhibit high specificity for either ER α or ER β having a selectivity (RBA ER α /RBA ER β) of 1.6 and 0.5 respectively. As a result it is difficult to extrapolate whether one particular isoform in complex with the said ligand is responsible for the observed biological response. However, as one can see from Table 2 the novel non-steroidal substituted furans and pyrazoles exhibit unprecedented estrogen subtype selectivity compared to the classical ER ligands listed in Table 1. In particular the compound 4-propyl-1,3,5-triphenolpyrazole (**4g**) has been reported to have 400-fold affinity for ER α isoform of the estrogen receptor. It is essential to note that the origin of this enhanced specificity for ER α comes from the ligands poor affinity to bind in ER β (Table 2, RBA

ER β 0.12). Understanding the origin of the biological discrimination of **4g** in ER β is of great importance since it might lead to the development of more potent selective ER ligands.

I. CoMFA models

As previously mentioned the most important aspect of the CoMFA technique is the alignment of the ligand structures. It is essential that the ligands be aligned so as to maximize the superposition of the pharmacophoric elements responsible for producing the biological response of the ligands in the training set. For all of our ER α CoMFA models the ligands were aligned by atom fitting the atoms in the A-ring mimic of each ligand to the atoms in the A-ring of the minimized crystal structure of 17 β -estradiol from 1ERE (Figure 1). While for all of our ER β models the ligands were aligned by atom fitting the atoms in the A-ring mimic of each ligand to the atoms in the B-ring of the minimized crystal structure of genistein from 1QKM. The B-ring of genistein was used for the ER β model alignment because in the crystal structure 1QKM it is the B-ring of genistein that acts as the A-ring mimic. These alignments were chosen as the tetra-substituted furans and pyrazoles had extensive rotational freedom about the phenol rings, as well as multiple D-ring mimics; we concluded that utilizing an alignment involving all atoms in 17 β -estradiol or genistein would impose unnecessary positional restraints.

Once the compounds were aligned the electrostatic and steric fields of all of the ligands were calculated by the CoMFA technique. At this point a total of eight ligands, from the original thirty-six described in Tables 1 and 2, were selected at random and formed test set I used to validate our CoMFA models. The

ligands forming test set I were tamoxifen, methoxychlor, **15f**, **15k**, **15l**, **4b**, **4f**, **7c**. The remaining twenty-eight ligands formed training set I. A partial least squares analysis was performed to calculate the cross-validated correlation, utilizing the leave one out method, between the logbase10 of the experimental RBA and the calculated molecular fields of only the ligands in training set I.

Initial CoMFA models for both the ER α and ER β were constructed using the low energy conformation of the ligands as input. The solvation energy of the ligands, calculated using the GB/SA method and the OPLS all atom force field implemented in the molecular modeling program Macromodel, was added as an additional descriptor to the CoMFA model to examine whether it improved the predictive nature of the CoMFA [38, 39]. Tables 3 and 4 illustrate the calculated correlation's (r^2) and cross-validated correlation's (q^2) for these CoMFA models. The cross-validated q^2 is a measure of the external predictive nature of the CoMFA model with a q^2 greater than 0.50 said to represent a predictive model.

From Tables 3 and 4 the ER β model is extremely predictive exhibiting a q^2 of 0.63 using three principal components. The ER α model is less predictive exhibiting a q^2 of 0.44, which is a little less than the 0.50 representing a predictive model. Inclusion of the calculated ligand aqueous solvation energy actually decreases the predictive nature of the ER α model, contrary to reported research, but improves the predictive nature of the ER β model increasing the q^2 from 0.63 to 0.70 however using four principal components and contributing only 4% to the CoMFA model. Irregardless our initial CoMFA models for the ER α and ER β binding site exhibited promising predictive power based on the calculated q^2 considering the diverse nature of the ligands in training set I. These results are in good agreement with previously reported CoMFA and 3D-QSAR results, which are dominated by ER α models, in that the r^2 , q^2 and standard errors of estimate are of the same magnitude although it is difficult to do a clear comparison because each training set is different [12, 15, 16, 18].

Further CoMFA models were constructed by first utilizing the docking program Gold to dock the ligands in the ligand-binding domain (LBD) of both ER α and ER β respectively. Here a 16 Å cavity was defined around the carbon atom of the terminal methyl group of residue MET421 in the preprocessed crystal structure of ER α in complexation with 17 β -estradiol. In the case of ER β a 16 Å cavity was defined around carbon

number 4 of the phenyl ring in residue PHE356 in the preprocessed crystal structure of ER β in complexation with genistein. The result from a Gold run is a series of viable conformations of the ligand docked inside the LBD of the target protein together with an associated fitness function and other measures of the corresponding protein-ligand interaction energy. The program was set to allow early termination if the top 3 solutions are within 1.0 Angstroms root-mean squared deviation (rmsd) of each other which we feel would indicate that the poses have reached an optimum based on the fitness function. The 1st ranked pose was then selected for further analysis; in all but two cases the first three Gold solutions differed only in internal rotational freedom.

One measure from Gold for determining the fitness of a given pose is H_{ext}, which is a measure of the hydrogen-bonding interaction between the docked ligand and the protein target. As a validation of the accuracy of the docking program Gold the rmsd of the crystal structure of 17 β -estradiol from 1ERE was compared against the most favorably ranked conformation of 17 β -estradiol docked with Gold. Likewise the rmsd of the crystal structure of genistein from 1QKM was compared with the Gold docked genistein. The rmsd deviation between the experimental docked conformation and the calculated docked conformation for 17 β -estradiol in ER α was 0.26 and 0.39 for genistein in ER β . Given the low rmsd deviation between the experimental structures and the calculated docked structures it is reasonable to expect that the program Gold could produce reasonable docked conformations of the remaining ligands in our study. Indeed Gold was able to locate viable docking conformations, i.e. inside the LBD, of all of the ligands presented to it.

One surprising outcome from the docking study was that Gold found essentially a flipped conformation, relative to the same docked ligand in ER α , for the substituted furans and pyrazoles in ER β . Figure 4 illustrates the most favorable docking conformation for ligand **4g**, the most ER subtype selective ligand, in the LBD of ER α and ER β . Upon further analysis it appears that residue PHE356 in ER β which is identical to residue PHE404 in ER α protrudes into the α -face of the cavity of the LBD to a much greater extent in ER β than in ER α . As a result if the ligands were bound in the same conformation in ER α and ER β the functional groups added to the substituted furans/pyrazoles (R3 in Figure 3), which are considered to be the origin of the specificity of these ligands for ER α over ER β , would be in too close contact with

Table 3. Results from several CoMFA models for the ER α utilizing both the primary and secondary training sets.

CoMFA Model	r ²	q ²	SEE	F	PC	% Contribution		
						Steric	Columbic	Other
Training Set I								
LowEnergyConformers								
Log RBA v CoMFA	0.79	0.44	0.54	48	2	45	55	*
Log RBA v CoMFA + SolvE ^a	0.68	0.40	0.67	27	2	39	57	4
Gold docked conformers								
Log RBA v CoMFA	0.99	0.70	0.14	300	6	36	64	*
Log RBA v CoMFA + GoldH_Ext ^b	0.95	0.64	0.27	111	4	33	53	14
Log RBA v CoMFA + FR ^c	0.99	0.57	0.16	240	6	35	57	8
Flexidock conformers								
Log RBA v CoMFA	0.87	0.67	0.87	86	2	32	68	*
Log RBA v CoMFA + FlexidockE ^d	0.89	0.65	0.40	65	3	31	62	7
Log RBA v CoMFA + F	0.95	0.55	0.27	118	4	33	62	5
Training Set II								
LowEnergyConformers								
Log RBA v CoMFA	0.98	0.19	0.23	150	6	36	64	*
Log RBA v CoMFA + SolvE	0.97	0.16	0.26	116	6	36	57	7
Gold docked conformers								
Log RBA v CoMFA	0.98	0.50	0.20	202	6	33	67	*
Log RBA v CoMFA + GoldH_Ext	0.98	0.59	0.21	179	6	29	57	14
Log RBA v CoMFA + FR	0.98	0.52	0.23	140	6	31	57	13
Flexidock conformers								
Log RBA v CoMFA	0.99	0.51	0.13	436	6	37	63	*
Log RBA v CoMFA + FlexidockE	0.99	0.53	0.16	300	6	34	60	6
Log RBA v CoMFA + F	0.99	0.51	0.18	287	5	35	57	7

^aSolvE refers to the addition of the calculated solvation energy of the ligand.^bGoldH_Ext refers to the addition of the hydrogen bonding score from Gold.^cFR and F refers to the addition of the relaxed and unrelaxed F score from CScore.^dFlexidock refers to the addition of the lowest energy score from Flexidock.

reside PHE356. To quantify this, if the docked **4g**-ER α complex is positioned so that its pyrazole ring overlaps the flipped pyrazole ring of **4g** in the docked **4g**-ER β complex the propyl functional group of **4g**-ER α would be within 2 Å of PHE356. Furthermore, as shown in figure 4, the propyl group of **4g** interacts with residues in the β -face of the cavity of the ER β LBD, particularly TRP335 and MET336, and with residues in the α -face of the cavity of the ER α LBD, particularly PHE404 and MET421. This is an extremely important finding, and contrary to reported research, since it suggests if true that the origin of the specificity of these novel compounds from this alternative docking conformation in ER β relative to ER α . Obviously if residues in the LBD of ER α are simply mutated to represent the LBD of ER β this configuration would not be observed since PHE404 in its ER α conformation would simply become PHE356. Of course it is possible that PHE356 could adopt an al-

ternate conformation in ER β to accommodate **4g** and ligands like it in the α -face of the cavity of the LBD however our theoretical studies do not suggest this to be the case as will be discussed below.

Using the program Flexidock implemented in the Sybyl suite of programs, we performed further analysis of the docked conformations found by Gold. As discussed, Flexidock is utilized to search for lower energy conformations of a protein-ligand complex. The output from a Flexidock run is a series of unique protein-ligand complexes with an associated Flexidock energy score. Flexidock models were constructed from the Gold docked protein-ligand complexes for both ER α and ER β .

In the Flexidock runs all rotatable bonds in the bound ligand and those in the protein within 6 Å of bound ligand were selected for conformational searching. In addition the hydrogen bonding networks between the ligands and residues ARG394, GLU353,

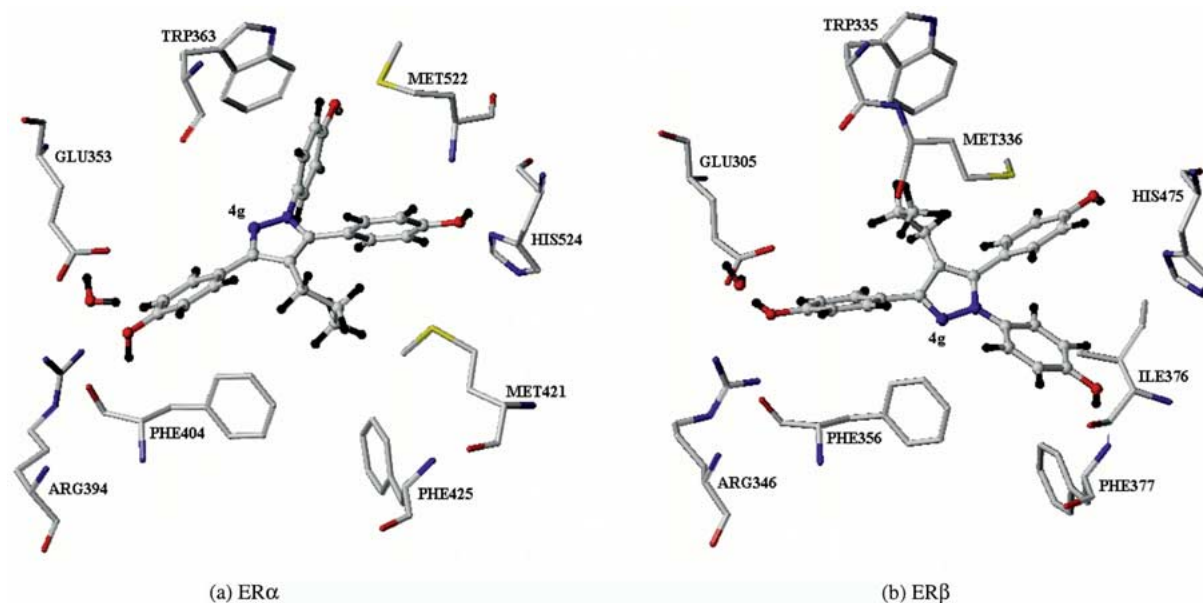


Figure 4. Unbiased best fit configuration of 4-propyl-1,3,5-triphenolpyrazole (**4g**) in (a) ER α and (b) ER β with the protein residues of interest highlighted.

the water molecule and HIS524 in ER α and ARG346, GLU305, the water molecule and HIS475 in ER β were defined. In each Flexidock run 100,000 generations for the GA were attempted in order to locate viable solutions for the docked ligands according to Flexidock. In addition to Flexidock runs for all of our ligands in both ER α and ER β , three additional complexes consisting of ligands **15a**, **15g** and **4d** bound in the α -face of the LBD of ER β , i.e. the same docking configuration as in ER α , were constructed in order to validate the flipped configurations in ER β found by Gold.

After Flexidock had completed its conformational searches the resulting **15a**, **15g** and **4d** low energy complexes, exhibiting both possible docking configurations in ER β , were subjected to full energy minimizations utilizing Gasteiger-Huckel partial charges and the Tripos force field until the maximum derivative was <0.01 kcal/(mol Å). In each case the flipped configuration was found to be the most favorable; **15a** by 1.21 kcal/mol, **15g** by 25.25 kcal/mol and in the case of **4d** by 24.63 kcal/mol. Obviously a difference of 1.21 kcal/mol is within the computational error of the technique however the two latter energies significantly favor the flipped configuration in ER β . It is important to note that in each case after conformational searching with Flexidock and subsequent full minimization PHE356 has had sufficient opportunity to adjust its

conformation in order to accommodate the ligand in the α -face of the cavity of the LBD yet the flipped configuration remains the most favorable.

The program CScore was applied to all of the ER α and ER β protein-ligand complexes obtained from Gold or subsequently from Flexidock, since each set of poses represent a unique binding conformation of the ligand in the protein. CScore produces estimates of the protein-ligand binding affinity in the form of a score as follows; F, G, D, PMF and ChemScore all of which can be applied within a rigid or flexible mode with respect to ability of the protein-ligand complex to relax. As estimates of the binding energy all of these functions were added to our CoMFA models. Tables 3 and 4 summarize the results from the addition of these additional descriptors. As in our initial models the ligands were aligned against the A-ring of 17 β -estradiol in ER α and the B-ring of genistein in ER β .

From Tables 3 and 4 despite every attempt to improve the predictive nature of our original CoMFA by the inclusion of addition terms such as the scoring functions from CScore the original CoMFA models utilizing the low energy conformations remained the most predictive. The only exception to this analysis is in the subsequent ER α CoMFA ($q^2 = 0.70$, $F = 300$) model using the Gold docked conformation of the ligands as the starting point for the alignment process. As a result it is clear that neither the Flexidock

Table 4. Results from several CoMFA models for the ER β utilizing both the primary and secondary training sets.

CoMFA Model	r ²	q ²	SEE	F	PC	% Contribution		
						Steric	Columbic	Other
Training Set I								
LowEnergyConformers								
Log RBA v CoMFA	0.93	0.63	0.36	104	3	38	62	*
Log RBA v CoMFA + SolvE ^a	0.96	0.70	0.28	131	4	33	59	8
Gold docked conformers								
Log RBA v CoMFA	0.94	0.23	0.34	117	3	36	64	*
Log RBA v CoMFA + GoldH_Ext ^b	0.94	0.22	0.35	84	4	30	61	9
Log RBA v CoMFA + D ^c	0.92	0.22	0.40	63	4	33	56	12
Flexidock conformers								
Log RBA v CoMFA	0.81	0.04	0.58	53	2	34	66	*
Log RBA v CoMFA + FlexidockE ^d	0.76	−0.03	0.65	40	2	36	54	10
Log RBA v CoMFA + DR	0.91	0.05	0.41	78	3	34	64	2
Training Set II								
LowEnergyConformers								
Log RBA v CoMFA	0.97	0.48	0.27	101	2	34	66	*
Log RBA v CoMFA + SolvE	0.96	0.43	0.30	81	3	33	65	2
Gold docked conformers								
Log RBA v CoMFA	0.85	0.13	0.52	72	2	36	64	*
Log RBA v CoMFA + GoldH_Ext	0.67	0.18	0.79	25	3	30	44	26
Log RBA v CoMFA + GR ^e	0.86	0.07	0.51	51	3	36	63	1
Flexidock conformers								
Log RBA v CoMFA	0.90	0.05	0.44	107	2	33	67	*
Log RBA v CoMFA + FlexidockE	0.91	0.39	0.41	85	3	30	55	15
Log RBA v CoMFA + DR	0.88	0.08	0.48	58	3	32	66	2

^aSolvE refers to the addition of the calculated solvation energy of the ligand.^bGoldH_Ext refers to the addition of the hydrogen-bonding score from Gold.^cDR and D refers to the addition of the relaxed and unrelaxed D score from CScore.^dFlexidock refers to the addition of the lowest energy score from Flexidock.^eGR refers to the addition of the relaxed g score from CScore.

energy score, H_ext from Gold nor any of the CScore scoring functions correlated particularly well with the experimental RBA data. Furthermore, variation in the q² of the various models illustrates the sensitivity of the CoMFA technique to the alignment process.

At this point the most predictive CoMFA models from training set I, were used to predict the experimental RBA data for test set I. Table 5 illustrates the results from the prediction of the experimental RBAs for test set I using the most predictive CoMFA models. Despite the relatively high q² for the most predictive models based on training set I 50% of the ligands from the test set are predicted with residuals (difference between predicted and experiment) greater than one log₁₀ and can be classified as outliers since they are so poorly predicted.

One of the specific goals of this research was to develop robust CoMFA models which could screen

a set of newly designed ligands for their ER subtype selectivity prior to synthesis and experimental determination of their corresponding RBAs. To test the ability of our CoMFA models to perform this goal we ranked the ligands in test set I in order of their RBA(ER α /ER β). We then calculated the predicted RBA(ER α /ER β) using our CoMFA models. As can be seen from Table 5 our CoMFA model correctly ranked 4f as the most subtype selective ligand in test set I. Indeed the top four compounds were correctly ranked although 4f was predicted ahead of 4b.

At this point in order to validate our CoMFA models we re-scrambled our ligands and repeated the entire process of constructing our CoMFA models. We achieved this by re-selecting eight ligands at random from the original set of thirty-six ligands to create a new test set consisting of the following eight ligands; 4g, 4h, 15c, 15e, 4b, 4e, 6 and 7c, the remaining

Table 5. Prediction of ER subtype selectivity ($ER\alpha/ER\beta$) of the test sets utilizing in each case the most predictive CoMFA model.

Ligand	$ER\alpha$			$ER\beta$			Ratio RBA ($ER\alpha/ER\beta$)			
	Pred. ^a	Exp. ^b	Resid. ^c	Pred.	Exp.	Resid.	Exp.	Rank	Pred.	Rank
Test Set I										
4f	1.07	1.56	-0.49	-0.23	-0.82	0.59	240	1	20.0	1
4b	0.70	1.49	-0.79	0.18	0.04	0.14	28	2	3.3	3
15f	1.33	1.91	-0.58	0.56	0.85	-0.29	12	3	5.9	2
15l	1.13	0.83	0.33	0.78	0.30	0.48	3.4	4	2.2	4
15k	1.13	-0.82	1.95	1.36	-1.15	2.51	2.1	5	0.6	6
Tamoxifen	1.59	0.85	0.74	2.48	0.78	1.70	1.2	6	0.1	8
7c	0.99	-1.40	2.39	1.00	-1.22	2.22	0.67	7	1.0	5
Methoxychlor	1.13	-2.00	3.13	1.42	-0.89	2.31	0.1	8	0.5	7
r^2		<0.01			0.02					
SEE _{cv}		2.17			1.32					
Test Set II										
4g	2.10	1.69	0.41	0.20	-0.92	1.12	408	1	79.4	1
4h	1.82	1.88	-0.06	-0.04	-0.05	0.01	84.3	2	72.4	2
15c	1.96	2.00	-0.04	0.63	0.26	0.37	56	3	21.4	5
15e	-0.01	0.94	-0.95	0.11	-0.60	0.71	36	4	0.8	8
4b	1.33	1.49	-0.16	-0.40	0.04	-0.44	28	5	53.7	3
4e	1.45	0.94	0.54	0.02	-0.33	0.35	19	6	26.9	4
6	0.85	0.75	0.10	-0.21	-0.07	-0.14	6.5	7	11.5	7
7c	0.52	-1.40	1.92	-0.68	-1.22	0.54	0.67	8	15.9	6
r^2		0.44			0.29					
SEE _{cv}		1.42			0.72					

^aPred. refers to predicted relative binding affinity (RBA).^bExp. refers to observed relative binding affinity (RBA).^cResid. refers to the corresponding difference between the predicted and observed.

twenty-eight ligands forming training set II. Again all ligands in the test set were excluded from the construction of the new CoMFA based on the ligands in training set II. CoMFA models were constructed in exactly the same way, using the same alignment procedure and adding the same descriptors at each stage in the model formation process. The results are shown in Table 4. In summary the most predictive CoMFA models from training set II for $ER\alpha$ was constructed from the alignment of the A-ring of the Gold docked ligands with the addition of the hydrogen bonding interaction function H_{ext} ($q^2 = 0.59$), for $ER\beta$ the best CoMFA model was constructed from the alignment of the A-ring of the low energy conformation of the ligands ($q^2 = 0.59$). These models were once again used to predict the experimental RBAs of the ligands in test set II and subsequent ER subtype selectivity (Table 5). With these new CoMFA models only two ligands from the test set were predicted with a residual greater than 1 \log_{10} unit. Unfortunately one of these poorly predicted ligands was **4g** which in this test set

exhibits the highest ER subtype selectivity. However, the error is in the $ER\beta$ model that predicts **4g** to be more active than is observed experimentally. Regardless our CoMFA model once again ranks at least the top two ligands correctly.

Arguably the most relevant published work for comparison is that of Sippl in which enhanced 3D-QSAR models for $ER\alpha$ are constructed from different alignment procedures using a training set of thirty steroidal estrogen receptor ligands [18]. Utilizing a mathematical approach called the Smart Region Definition/Fractional Factorial Design Sippl was able to improve on the predictive nature of Sadler *et al.*'s [15] $ER\alpha$ CoMFA model increasing the q^2 from 0.80 to 0.92 with a conventional standard error of estimate (SEE) of 0.11. From Table 3 the SEE from our most predictive $ER\alpha$ CoMFA models are 0.14 and 0.21 respectively, of comparable accuracy based on this criteria (SEE). However it is difficult to make a real comparison between the accuracy of our models based on our test set predictions, Table 5, using the

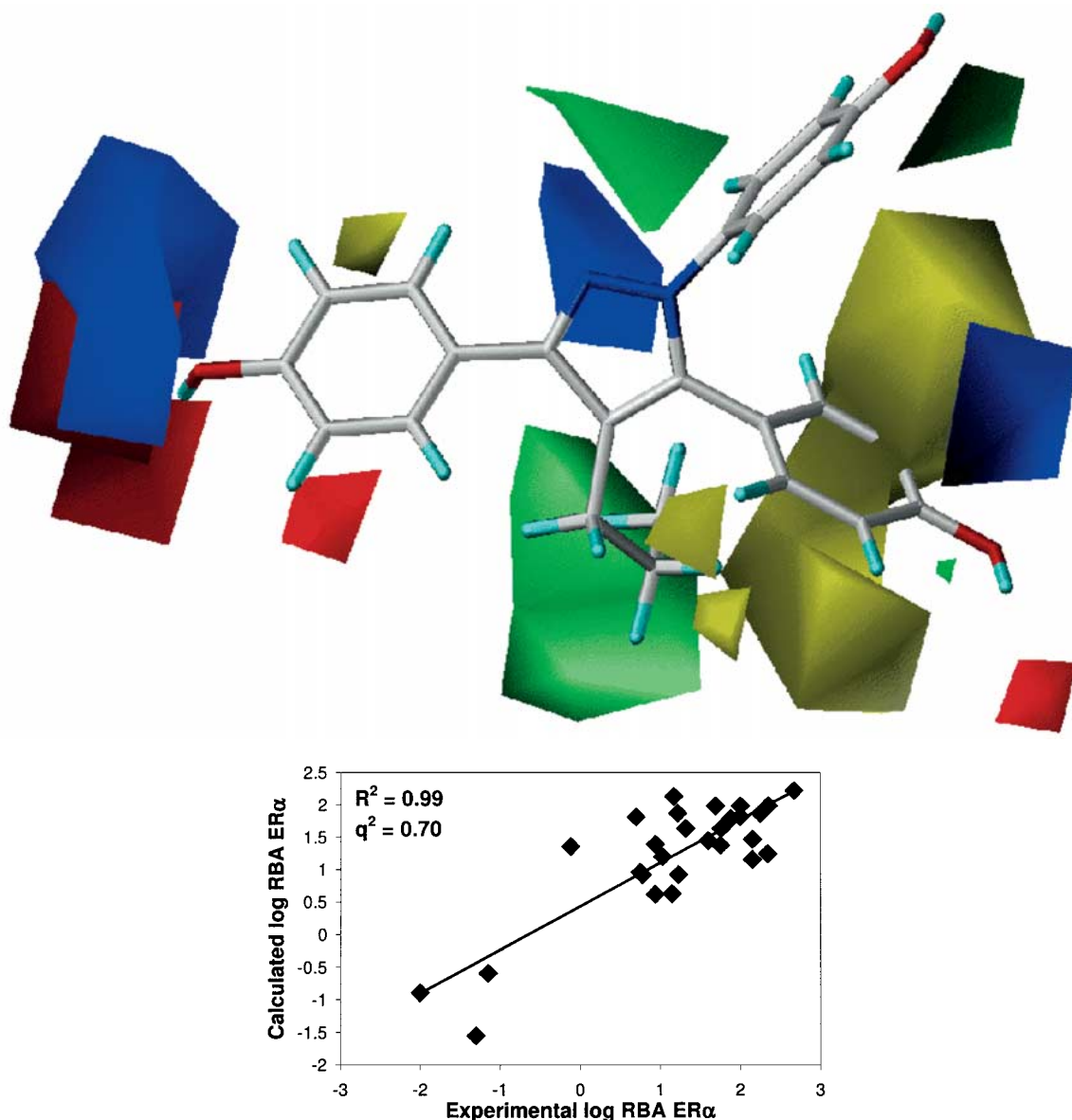


Figure 5. CoMFA derived molecular field maps from the most predictive ER α CoMFA model with the highly subtype ER selective ligand **4g** illustrated. Contours of the steric map are shown in yellow and green, while those of the electrostatic map are shown in red and blue. Increased biological activity is correlated with: more bulk near green; less bulk near yellow; more positive charge near blue, and more negative charge near red.

criteria of r^2 and cross-validated SEE (SEECv) since despite being ER ligands the test sets are not identical.

Figures 5 and 6 represent illustrations of the standard deviation of the calculated three-dimensional molecular fields from our most predictive ER α and ER β CoMFA models (Tables 3 and 4). Within these molecular fields the ligand **4g** has been superimposed in the same orientation for easy comparison. In each case contours of the steric map are shown in yellow

and green, while those of the electrostatic map are shown in red and blue. Increased biological activity is correlated with: more bulk near green; less bulk near yellow; more positive charge near blue, and more negative charge near red. As a result these fields could be manipulated in order to increase the subtype selectivity of ligand **4g**. For example, one might want to increase the biological activity of **4g** in ER α using the ER α molecular fields for guidance. In this case one

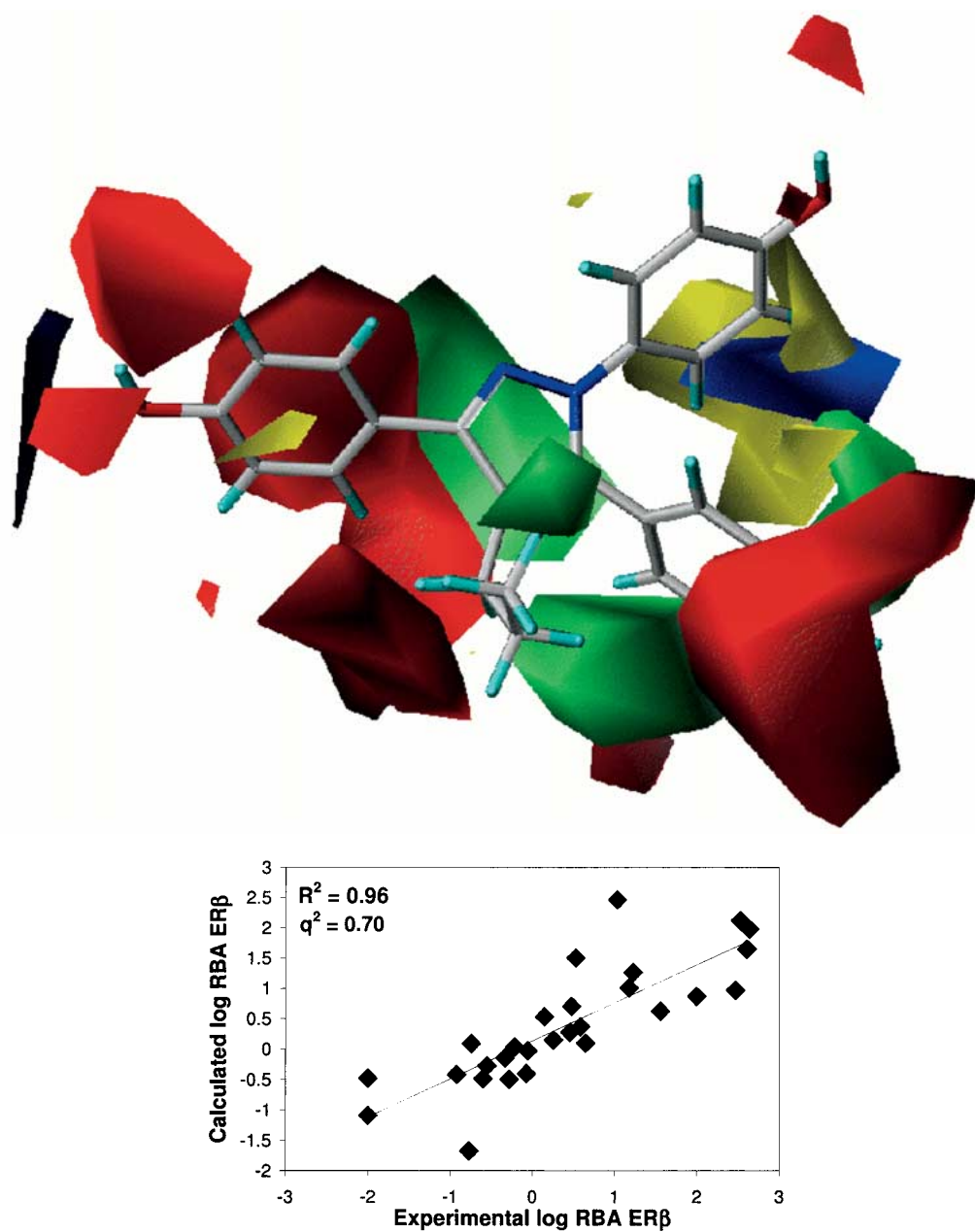


Figure 6. CoMFA derived molecular field maps from the most predictive ER β CoMFA model with the highly subtype ER selective ligand **4g** illustrated. Contours of the steric map are shown in yellow and green, while those of the electrostatic map are shown in red and blue. Increased biological activity is correlated with: more bulk near green; less bulk near yellow; more positive charge near blue, and more negative charge near red.

might decrease the steric bulk of the phenol ring on the 5-position where the steric contour map is densely yellow and or increase the positive charge of the N of the pyrazole ring where the electrostatic contour map is blue, while at the same time decreasing the biological activity of **4g** in ER β , using in turn the ER β molecular fields for guidance, by adding functional groups, which exhibit an overall positive charge, to the phenol A-ring mimic where the electrostatic map is densely red.

Conclusions

In conclusion we have utilized the computational techniques of CoMFA, the unbiased docking of ligands utilizing Gold, prediction of subsequent binding affinity via CScore and location of low energy docked protein-ligand complexes using Flexidock, in order to study the origin of the subtype selectivity of a new class of ER agonists. Our unbiased docking study has highlighted a possible unconsidered binding configuration for these novel compounds that may well prove to be the origin of their enhanced specificity for ER α . Robust CoMFA models, consisting of several classes of ER ligands, have been developed and validated extensively within the framework of our original set of ligands. Indeed, we have shown how these predictive CoMFA models can be used to focus and prescreen new ligands for their ER subtype selectivity prior to experimental determination.

Ongoing work is being performed to improve our CoMFA models and to understand the origin of the subtype selectivity of the substituted pyrazoles in ER by completing a complete binding energy study of the ER and sex hormone-binding globulin (SHBG) a protein which also plays an important role in the regulation of estrogens in the human body. The goal being to utilize this understanding to aid in the ligand based design of new more selective and active non-steroidal ER agonists or antagonists.

Acknowledgements

We wish to thank the National Institute of Biomedical Imaging and Bioengineering, EB00340, for funding this research.

References

- Kuiper, G.G.J.M., Enmark, E., Peltö-Huikko, M., Nilsson, S., Gustafsson, J.-A. *Proc. Natl. Acad. Sci. USA* 93 (1996) 5925.
- Mosselman, S., Polman, J., Dijkema, R. *FEBS Letters* 392 (1996) 49.
- Fawell, S.E., Lees, J.A., White, R., Parker, M.G. *Cell* 60 (1990) 953.
- Kuiper, G.G.J.M., Carlsson, B., Grandien, K., Enmark, E., Haggblad, J., Nilsson, S., Gustafsson, J.-A. *Endocrinology* 138 (1997) 863.
- Register, T.C., Adams, M.R. *The Journal of Steroid Biochemistry and Molecular Biology* 64 (1998) 187.
- Paech, K., Webb, P., Kuiper, G.G.J.M., Nilsson, S., Gustafsson, J.-A., Kushner, P.J., Scanlan, T.S. *Science* 277 (1997) 1508.
- Montano, M.M., Jaiswal, A.K., Katzenellenbogen, B.S. *J. Biol. Chem.* 273 (1998) 25443.
- Mortensen, D.S., Rodriguez, A.L., Sun, J., Katzenellenbogen, B.S., Katzenellenbogen, J.A. *Bioorganic & Medicinal Chemistry Letters* 11 (2001) 2521.
- Stauffer, S.R., Coletta, C.J., Tedesco, R., Nishiguchi, G., Carlson, K., Sun, J., Katzenellenbogen, B.S., Katzenellenbogen, J.A. *J. Med. Chem.* 43 (2000) 4934.
- Meyers, M.J., Sun, J., Carlson, K.E., Katzenellenbogen, B.S., Katzenellenbogen, J.A. *J. Med. Chem.* 42 (1999) 2456.
- Gao, H., Williams, C., Labute, P., Bajorath, J. *Journal of Chemical Information & Computer Science* 39 (1999) 164.
- Gao, H., Katzenellenbogen, J. A., Garg, R., Hansch, C. *Chem. Rev.* 99 (1999) 723.
- Cramer, R.D., III, Patterson, D.E., Bunce, J.D. *J. Am. Chem. Soc.* 110 (1988) 5959.
- Tong, W., Perkins, R., Xing, L., Welsh, W.J., Sheehan, D.M. *Endocrinology* 138 (1997) 4022.
- Sadler, B.R., Cho, S.J., Ishaq, K.S., Chae, K., Korach, K.S. *J. Med. Chem.* 41 (1998) 2261.
- Sippl, W. *J. Comp.-Aided Mol. Des.* 14 (2000) 559.
- Sippl, W., Holtje, H.D. *Theochem* 503 (2000) 31.
- Sippl, W. *Bioorg. & Med. Chem.* 10 (2002) 3741.
- Hopfinger, A.J. *J. Am. Chem. Soc.* 102 (1980) 7196.
- Pastor, M., Cruciani, G., McLay, I., Pickett, S., Clementi, S. *J. Med. Chem.* 43 (2000) 3233.
- Charifson, P.S., Corkery, J.J., Murcko, M.A., Walters, W.P. *J. Med. Chem.* 42 (1999) 5100.
- Clark, R.D., Strizhev, A., Leonard, J.M., Blake, J.F., Matthew, J.B. *Journal of Molecular Graphics and Modelling* 20 (2002) 281.
- Vieth, M., Cummins, D.J. *J. Med. Chem.* 43 (2000) 3020.
- Lozano, J.J., Pastor, M., Cruciani, G., Gaedt, K., Centeno, N.B., Gago, F., Sanz, F. *Journal of Computer-Aided Molecular Design* 14 (2000) 341.
- Brzozowski, A.M., Pike, A.C.W., Dauter, Z., Hubbard, R.E., Bonn, T., Engström, O., Öhman, L., Greene, G.L., Gustafsson, J.-Å., Carlquist, M. *Nature* 389 (1997) 753.
- Pike, A.C.W., Brzozowski, A.M., Hubbard, R.E., Bonn, T., Thorsell, A.-G., Engstrom, O., Ljunggren, J., Gustafsson, J.-Å., Carlquist, M. *EMBO J.* 18 (1999) 4608.
- Jones, G., Willett, P., Glen, R.C. *Journal of Molecular Biology* 245 (1995) 43.
- Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R. *Journal of Molecular Biology* 267 (1997) 727.
- Bissantz, C., Folkers, G., Rognan, D. *J. Med. Chem.* 43 (2000) 4759.

30. Åqvist, J., Medina, C., Samuelsson, J.-E. *Protein Engineering* 7 (1994) 385.
31. Holloway, M.K., Wai, J.M., Halgren, T.A., Fitzgerald, P.M.D., Vacca, J.P., Dorsey, B.D., Levin, R.B., Thompson, W. J., Chen, L. J., et al. *J. Med. Chem.* 38 (1995) 305.
32. Ortiz, A.R., Pisabarro, M.T., Gago, F., Wade, R.C. *J. Med. Chem.* 38 (1995) 2681.
33. Pérez, C., Pastor, M., Ortiz, A.R., Gago, F. *J. Med. Chem.* 41 (1998) 836.
34. Dhar, A., Liu, S., Klucik, J., Berlin, K.D., Madler, M.M., Lu, S., Ivey, R.T., Zacheis, D., Brown, C.W., Nelson, E.C., Birckbichler, P.J., Benbrook, D. M. *Journal of Medicinal Chemistry*. 42 (1999) 3602.
35. Bertelli, M., El-Bastawissy, E., Knaggs, M.H., Barrett, M.P., Hanau, S., Gilbert, I.H. *Journal of Computer-Aided Molecular Design*. 15 (2001) 465.
36. Tripos SYBYL; 6.7 ed.; Tripos Inc.: St Louis, 1995.
37. Jayatilke, P.R.N., Nair, A.C., Zauhar, R., Welsh, W. J. *J. Med. Chem.* 43 (2000) 4446.
38. Qiu, D., Shenkin, P.S., Hollinger, F.P., Still, W.C. *J. Phys. Chem. A* 101 (1997) 3005.
39. Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caulfield, C., Chang, G., Hendrickson, T., Still, W.C. *J. Comput. Chem.* 11 (1990) 440.