Journal of Pharmaceutical and Biomedical Analysis 48 (2008) 369–375

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Molecular structural characteristics as determinants of estrogen receptor selectivity

S. Agatonovic-Kustrin[a,*], J.V. Turner[b], B.D. Glass[a]

[a] School of Pharmacy and Molecular Sciences, James Cook University, Townsville, QLD 4811, Australia
[b] School of Medicine, University of Queensland, Brisbane, Australia

ARTICLE INFO

Article history:
Received 19 September 2007
Received in revised form 7 April 2008
Accepted 7 April 2008
Available online 15 April 2008

Keywords:
ER-alpha
ER-beta
Differential binding
Artificial neural network
QSAR
Theoretical descriptors

ABSTRACT

Recent reports that a wide variety of natural and man-made compounds are capable of competing with natural hormones for estrogen receptors serve as timely examples of the need to advance screening techniques to support human health and ascertain ecological risk. Quantitative structure–activity relationships (QSARs) can potentially serve as screening tools to identify and prioritize untested compounds for further empirical evaluations. Computer-based QSAR molecular models have been used to describe ligand–receptor interactions and to predict chemical structures that possess desired pharmacological characteristics. These have recently included combined and differential relative binding affinities of potential estrogenic compounds at estrogen receptors (ER) α and β. In the present study, artificial neural network (ANN) QSAR models were developed that were able to predict differential relative binding affinities of a series of structurally diverse compounds with estrogenic activity. The models were constructed with a dataset of 93 compounds and tested with an additional dataset of 30 independent compounds. High training correlations (r^2 = 0.83–0.91) were observed while validation results for the external compounds were encouraging (r^2 = 0.62–0.86). The models were used to identify structural features of phytoestrogens that are responsible for selective ligand binding to ERα and ERβ. Numerous structural characteristics are required for complexation with receptors. In particular, size, shape and polarity of ligands, heterocyclic rings, lipophilicity, hydrogen bonding, presence of quaternary carbon atom, presence, position, length and configuration of a bulky side chain, were identified as the most significant structural features responsible for selective binding to ERα and ERβ.

1. Introduction

Estrogens are steroid hormones integral to the human endocrine system. They have diverse physiological and pathophysiological functions in different tissues and cell types in the body. Estrogens regulate the growth and development of reproductive systems as well as homeostasis in a variety of tissues. They play important roles in bone maintenance, in the central nervous system, and also in the cardiovascular system where they display certain cardioprotective effects [1]. A large number of different pathological conditions are associated with changes in the production of estrogen and cellular responses to an estrogen stimulus. Thus, compounds with estrogenic activity have generated considerable interest as targets for the development of therapeutic agents.

The action of estrogen is mediated differentially through the estrogen receptors α and β (ERα and ERβ, respectively). At the discovery of the estrogen receptor (ER) it was thought that only one receptor existed. Then, in 1996, a second type of ER was cloned from rat [2], human [3–5] and mouse [6]. This receptor was named ERβ while the original ER is now referred to as ERα. The late discovery of a second ER is not surprising since physiological endogenous estrogens (estradiol, estrone, and estriol) bind equally well to both ER subtypes, whereas some anti-estrogens currently used, such as tamoxifen, raloxifen, block both receptor subtypes with little selectivity [7] (Fig. 1).

The discovery of a second type of ER demonstrated that the mechanisms behind the effects of estrogen are far more complex than previously assumed [8]. Studies have also shown that the two subtypes have different functions and distributions in certain tissues [9]. These differences have stimulated the search for subtype-specific ligands with tissue- or cell-specific estrogen activity. For example, ERα is dominant in the breast and uterus suggesting that ERβ selective ligands may be used as hormone

* Corresponding author. Tel.: +61 7 4781 6803; fax: +61 7 4781 5356.
E-mail address: nena.kustrin@jcu.edu.au (S. Agatonovic-Kustrin).

0731-7085/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2008.04.008

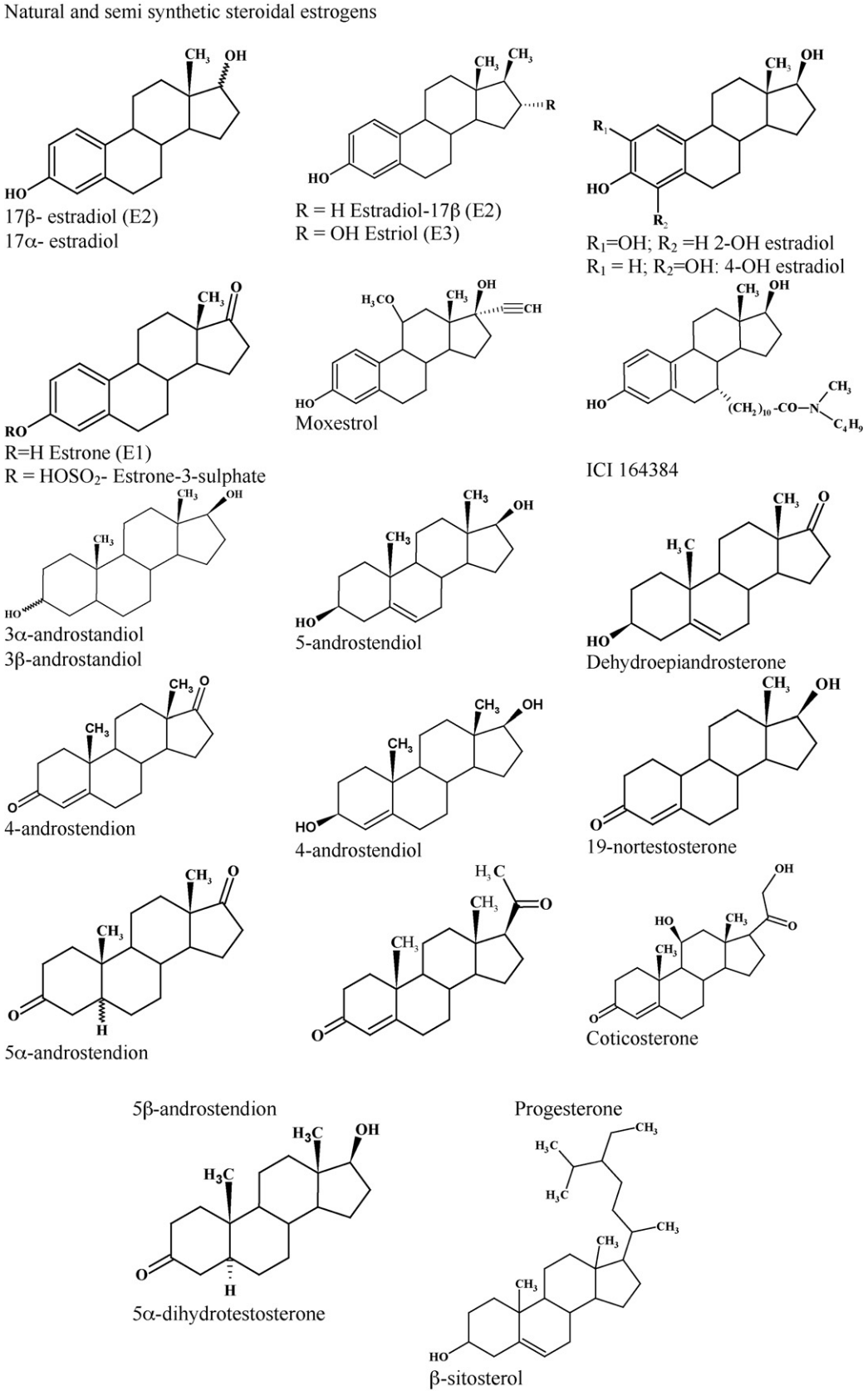

Fig. 1. Natural and semi-synthetic steroidal estrogens.

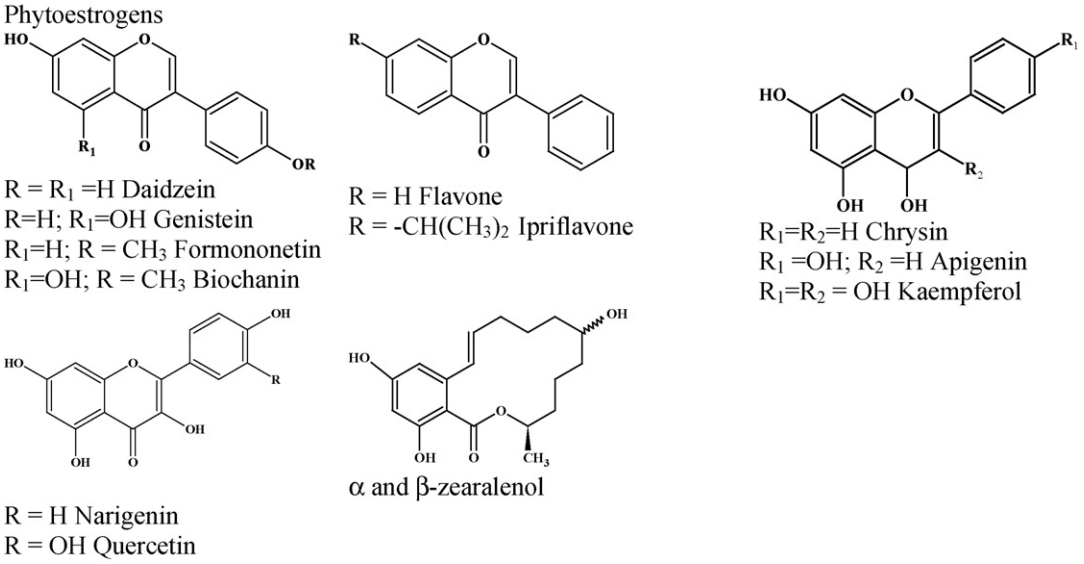

Fig. 2. Phytoestrogens.

replacement therapy without increasing the risk of breast or uterine cancer [10].

Estrogenic activity is unique in that it does not require a steroidal structure, as do the other sex hormones. Intensive research has revealed that some man-made and naturally occurring chemicals can directly interact with the estrogen receptors and disrupt the normal functioning of the endocrine systems of humans and wildlife. These chemicals, termed "endocrine-disrupting chemicals" (EDCs), are of such scientific and public concern that the screening and testing of 58,000 chemicals for EDC activity is now statutorily mandated in the USA [11]. It is believed that these chemicals can cause reproductive and child development disorders, mainly through ER mediated mechanisms of toxicity. There is thus a growing need for ER activity modeling in this field in order to reduce costs and enhance the screening and testing processes. One strategy lies in the development of quantitative structure–activity relationship (QSAR) models to screen chemicals based on their molecular structure for their ability to interact with the endocrine system. Structure–activity relationships (SARs) can serve as screening tools to prioritize untested compounds for more intensive and costly empirical evaluations at a later stage [12]. Numerous SARs and QSARs have been developed to predict hormone relative binding affinity (RBA) and to indicate potential estrogenicity [13].

SAR models based on different methods such as CoMFA [14], COMSIA [15], SOMFA [16], CoSA [16], kNN [17], COmmon REactivity PAttern (COREPA) [18], HQSAR [14], MTD-PLS [19], artificial neural networks (ANNs) [20,21] and Raptor [22], provide different perspectives on the interactions between the estrogen receptor and its ligands. Limitations of these QSAR models are associated with the size and chemical-structure diversity of the training set, experimental error, structure representation and correlation algorithms [23]. Compared with classical statistical optimization techniques, ANNs have demonstrated utility in managing data containing nonlinear relationships for modeling and predictive purposes. The absolute utility of a QSAR model depends on its ability to make accurate predictions for unknown chemicals. Regression models often demonstrate excellent fit to training data but fail to accurately predict chemicals that differ structurally from the training set of compounds.

In our recent QSAR model, ANNs were used to model selective binding of 48 phytoestrogens and structurally related compounds at ERα and ERβ [21]. The aim of the present study was to improve the previous model and overcome limitations associated with data size and diversity, and to further investigate the effects of structural parameters on ER subtype binding.

2. Methods

Relative binding affinities at ERα and ERβ of 93 compounds (Figs. 1–3) were taken from the literature [21,24–29]. From optimized molecular structures 73 theoretical descriptors were generated using Molecular Modeling Pro 5.1 [30] which described 2D and 3D structural information as well as molecular physical/chemical properties. These included constitutional, steric, electronic, topological and chemical descriptors described elsewhere [21]. Molecular descriptors were used as inputs and relative binding affinities as the outputs to build the ANN model. Compounds were randomly divided into training, testing and independent external validation subsets. Two models were built with different data division of 73/25/25 and 63/30/30 data subsets for training, internal testing and external validation.

2.1. ANN modeling

ANNs are learning systems based on a computational technique that can simulate the neurological processing ability of the human brain. They can be used to quantify a nonlinear relationship between causal factors and pharmaceutical responses by means of iterative training of experimental or theoretically derived data. Intelligent problem solver was used for the initial training to select the best ANN type. In contrast to traditional linear techniques in statistics, there is no method known that will automatically locate the optimal neural network to fit a particular data set. Neural network designers traditionally run training algorithms a number of times selecting the best network (or perhaps a few of the best). Therefore, a number of experiments with different designs are conducted, and a Generalized Regression Neural Network (GRNN) was selected based on its performance. Standard supervised network architectures (multilayer perceptrons and radial basis functions) infer a parameterized model from available training data, with the weights between neurons forming the parameters. The parameterized network generated is usually

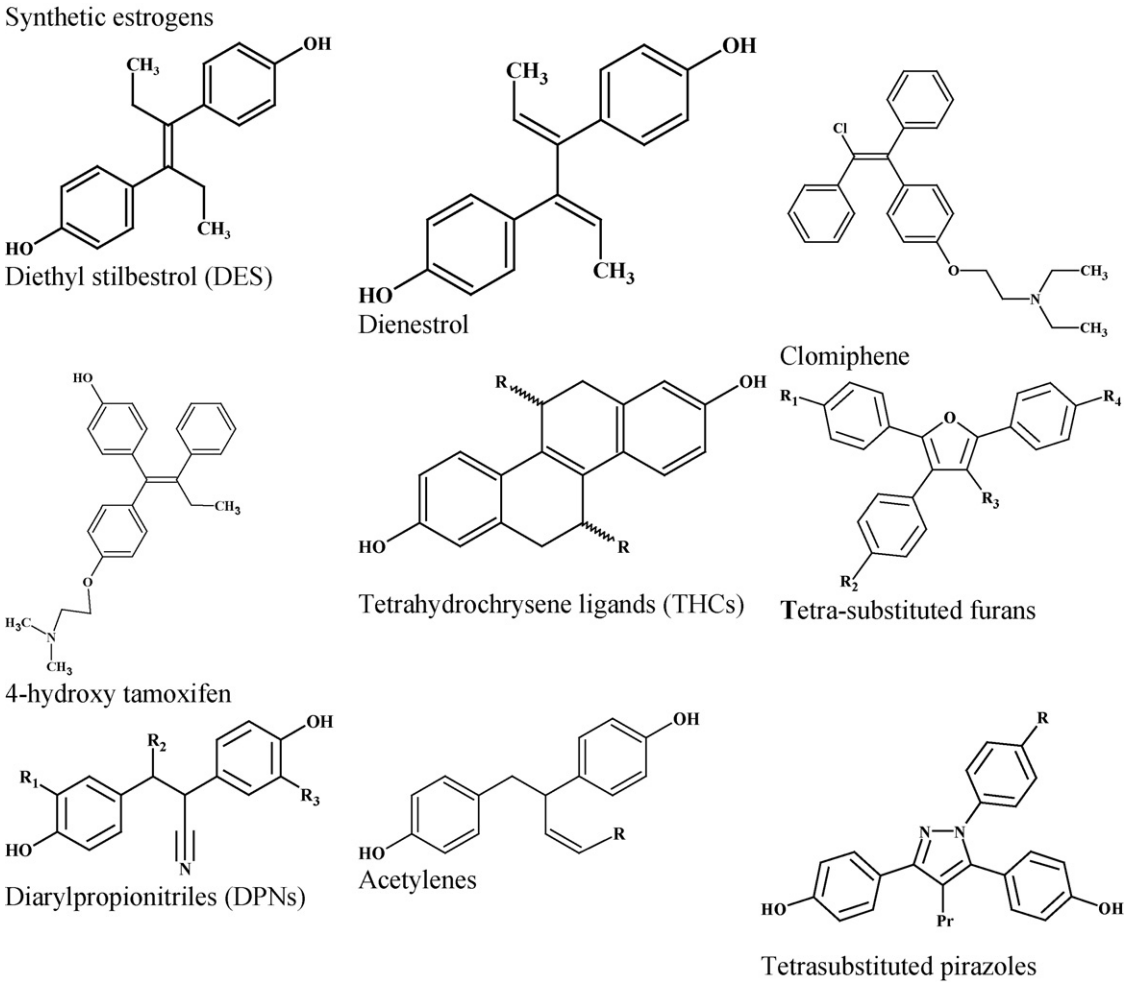

Fig. 3. Synthetic estrogens.

much smaller than the training data and can be executed quite quickly, although the time taken to train the model may be long. An alternative approach is to model the function more-or-less directly from the training data which substantially decreases training time. GRNNs, also known as Probabilistic Neural Network or Bayesian network GRNNs are such a method [31]. GRNNs have four layers: input, a layer of radial centres, a layer of regression neurons, and output. The radial layer neurons represent the centres of clusters of known training data. This layer must be trained by a clustering algorithm such as K-means, subsampling or Kohonen training. The radial layer is typically large but not necessarily as large as the number of training cases. The regression layer must have exactly one unit more than the output layer and contains linear neurons of two types. There is one type A neuron for each output unit and one type B neuron. Type A neurons calculate the "desired" regression outputs for the cases; the type B neuron calculates the probability density. The output layer executes a special post-synaptic division. Each unit simply divides the output of the type A unit by the output of the type B unit in the previous layer.

The primary advantage of GRNN is the speed at which the network can be trained. There are no training parameters such as learning rate and momentum in back-propagation network, but there is a smoothing factor that is applied after the network is trained. The smoothing factor allows the GRNN to interpolate between data in the training set.

2.2. GRNN optimization

The commercially available Statistical Neural Networks software package [32] was used throughout the study for model generation and optimization. Manually-determined parameters included: the number of neurons in the second radial layer, the smoothing factor controlling deviation of the Gaussian kernel function located at the radial centers (set at 0.1) and the clustering algorithm (K-means in the present study). A sum-squared error function was used in training the network whereby the error, calculated as the sum of the squared differences between the obtained and actual RBA value on each training prediction, determined training performance. Optimization of the GRNN model was achieved by monitoring predictions for the 30(25) internal testing compounds. Once the model was developed and optimized, sensitivity analysis was applied in order to select relevant molecular descriptors. Molecular descriptors were ranked according to sensitivity of the training subset and 21 most important descriptors were identified. Sensitivity analysis indicates which input variables are considered most important by neural network and identifies variables that can be safely ignored. Input variables are not, in general, independent and there are interdependencies between variables. Sensitivity analysis rates variables according to the deterioration in modeling performance that occurs if that variable is no longer available to the model and assigns a single rating value to each variable. The aim of descriptor subset selection was to balance minimum

Table 1
Optimum model characteristics

	Architecture	Data division	$r_{training}$ (r^2)		$r_{testing}$ (r^2)		$r_{validation}$ (r^2)	
			ERα	ERβ	ERα	ERβ	ERα	ERβ
Model 1	63/30/30	21-*64*-3-2	0.921	0.925	0.653	0.569	0.851	0.789
Model 2	73/25/25	21-*73*-3-2	0.914	0.956	0.414	0.654	0.872	0.925

model complexity with maximum predictive performance during internal testing. The optimum model was next validated with the external subset of compounds to assess true predictive ability. The final combination of descriptors could then be analysed.

3. Results and discussion

3.1. Optimum models

A diverse set of compounds were evaluated in the present study for various molecular attributes including size/shape parameters (molecular weight, surface area, volume, kappa shape and connectivity indices 0–3), electrostatic parameters (dipole moment, valence indices 0–3 and CIM indices), solubility parameters (log K_{ow}, solubility, polarity, percent hydrophilic surface and hydrophilic–lipophilic balance), hydrogen bonding potential, cyclic components and specific substitution.

Two optimum models with different ratios of training/internal testing/external validation subsets (63/30/30 and 73/25/25, respectively) were selected. High training variances ($r^2 > 0.835$) were achieved by both models for ERα and ERβ (Table 1) indicating that appropriate nonlinear relationships amongst the data were being modeled. As expected, true prediction correlations were not as high as those seen during training but were close to 0.8 or above. Model 1 accounted for more than 75% of the variance in the external prediction subset for both ERα and ERβ while Model 2 accounted for less than 75% of the variances in the external prediction subset.

Both optimum models had 21 input descriptors selected based on the sensitivity analysis. Ten of the final descriptors in these two models were identical (Table 1). The remaining 11 descriptors were both similar and highly correlated. Both models selected molecular length, κ2 shape index, hydrogen bond donor, water solubility, number of carbons in nonaromatic rings and number of carbons in six membered aromatic ring, substituted double bond, ether or oxygen bridge and connectivity index. The remaining of selected descriptors was related to the molecular size, polarity (CIM indices), solubility, lipophilicity, hydrogen bonding, presence of nitrogen and oxygen in nonaromatic ring, presence of a nitrile.

3.2. Descriptor analysis

While the structures of estrogen agonists vary widely, they can be classified as either steroids or nonsteroidal synthetic structures. The length and width of both the steroid skeleton and nonsteroidal synthetic estrogen diethylstilbestrol (DES) skeletons fit well into the receptor-binding pocket. Nonsteroidal molecules such as DES have activity similar to that of estradiol. The activity of DES analogues was explained in 1946 when it was proposed that the critical structural requirement for the receptor recognition is the distance between two oxygens that should be 12.1 Å [33]. Modern medicinal chemistry has shown that the actual distance is 12.1 Å in diethylstilbestrol and 10.9 Å in estradiol. However, in aqueous solution, estradiol has two water molecules hydrogen bonded to the 17-OH. If one water molecule is included in the distance measurements, there is a perfect fit to 12.1 Å [34].

The number of carbon atoms is related to the molecular size. The ligand-binding pockets of the alpha and beta subtypes are similar but not identical [35]. The ERβ ligand-binding pocket is smaller (390 Å versus 490 Å for ERα) and differs in two amino acids. Leu384 and Met421 in ERα are replaced by Met336 and Ile373 in ERβ [36]. In general, ERβ selective ligands seem to be smaller and more polar than ERα selective ligands. Furthermore, domain near the 17α position of estradiol is larger in ERα than in ERβ. This distinction suggests that increasing steric bulk in this region will enhance the binding affinity for ERα. Indeed, moxestrol (RBA = 43 and 5) and norethynodrel (RBA = 0.7 and 0.22), both with 17-ethynyl substituent, have higher binding affinities for ERα than ERβ.

Molecular lipophilicity was encoded in water solubility, selected by Model 1, and also in the log *P* and Hansen solubility descriptors that were selected by Model 2. Natural estrogens are lipophilic molecules with two polar groups attached at the ends of the molecule. Strong estrogens tend to be more hydrophobic. The volume available in the receptor pocket exceeds the size of natural ligand leaving a bit of empty space in the binding pocket around 7α and 11 position of the estradiol (B and C rings). This empty space provides an explanation for the high affinity of ligands with substituents at these positions. However, addition of a hydrophilic group at C-11 position (such as a hydroxyl or keto group) almost completely eliminates binding affinities for both receptors. Addition of a lipophilic group with even a bulkier size (acetate or methoxy group) does not affect the binding affinities of natural estrogens. This indicates that the large decrease in binding affinities is not due to steric hindrance caused by the C-11 position substitutions but is primarily due to alterations of the lipophilicity near the C-11 position.

Kappa (κ) indices encode attributes of molecular shape by quantifying the structure of a compound in terms of its relative star-likeness and straight chain-likeness shapes. The κ2 molecular shape index can distinguish among geometric *cis* and *trans* isomers. Presence of substitutions is essential for the nonsteroidal stilbene estrogens (diethylstilbesterol, dienestrol and hexestrol). Geometric isomers have different spatial arrangements of atoms and receptor interactions are also different. The *trans*-isomer of the diethylstilbestrol has 14 times greater estrogenic activity than the *cis*-isomer due to its overall structure and the interatomic distance between the two hydroxyls. The presence of a substituted double bond is essential for the nonsteroidal stilbene estrogens (diethylstilbesterol, dienestrol and hexestrol). It can be regarded as a measure of molecular flexibility. Recent studies indicate that the potency and agonist or antagonist activity of steroid hormone ligands are dependent, in part, on ligand–receptor binding affinity as well as the conformation of the ligand–receptor complex. The binding of ligands to hormone receptors is thought to involve interactions by which shapes of both the receptor and ligand are modified in the formation of the ligand–receptor complex. As a consequence, it is essential to explore the significance of ligand flexibility in the development of screening-level structure–activity relationships. The double bond contributes to rigidity of the molecule and its shape. As long as the OH-to-OH distance relationship is maintained, significant estrogenic activity is found. Without the central double bond and two ethyl or alkyl groups the molecule loses its rigidity and shape and the distance between two hydroxyls is not fixed.

Model 2 contained the quaternary carbon atom descriptor. Quaternary carbon atoms are found in many isoprenoic compounds. Enzymes aromatase catalyzes removal of C-19 from androgens, leading to the formation of estrogens. Aromatization of A ring in natural estrogens alters the overall shape of the molecule. The relative spatial orientation of the A-steroidal ring with respect to the B-ring, may be considered as important structural characteristics for ERα ligand recognition.

Estrogen receptors exhibit stereo-selective ligand binding and stereoselective recognition of several chiral compounds. Substi-

tuted tetrahydrochrysene (THC) ligands are potent agonists at ERα but also potent antagonists at ERβ. This characteristic is a function of substituent size and stereochemistry. SS enantiomers have similar agonist activity to ERα and ERβ. The difference in efficacy of *R,R*-THC on the two receptor subtypes arises from its optimal fit in the ERα ligand-binding pocket and its suboptimal fit in the smaller ERβ pocket. All THCs are agonists on ERα receptor, and THCs with small substituents are agonists on both ERα and ERβ. As substituent size increases, ERβ-selective antagonism is developed first in the R,R-cis enantiomer series and finally in the *trans* and *S,S-cis* enantiomer series.

The descriptor encoding the number of carbons in six-membered rings (C aromatic 6) was present in the final model, indicating the importance of the presence of an aromatic ring. Amongst the great variety of molecular structures found in the many classes of nonsteroidal ER ligands, it appears that a phenolic ring can mimic the steroid "A-ring" present in natural estrogens such that presence of a phenolic ring is often associated with estrogenic activity [37]. The phenolic ring is thought to develop significant H-bond at the ER binding domains with Glu 353 and Arg 394 (ERα) and Glu 305 and Arg 346 (Erβ) [38].

The most common types of phytoestrogens are flavones, isoflavones and lignans. These compounds belong to the large group of substituted phenolic compounds known as flavonoids which contain the essentially planar benzopyrane moiety. Of note was inclusion of the oxygen bridge or ether (ORR) descriptor in the final model which encodes the presence of an ether oxygen such as the one found in the flavanoids. Recently, the prenylated flavanone, 8-prenylnaringenin (8-PN) was identified as a potent estrogen demonstrating the highest in vitro estrogenic activity among all phytoestrogens known to date [39].

Several nitrogen descriptors were included in the final models: N in aromatic rings, double and triple bonded N, and N adjacent to another N. Several triaryl-substituted five-membered heterocycles show exceptionally large potency and efficacy preferences for ERα [40]. They function as agonists on both ERα and ERβ but in cell-based assays of gene transcription they activate ERα at much lower concentrations [41]. The best of these are triaryl–alkyl-substituted pyrazoles and furans that function as complete ERα agonists, but are almost completely inactive on ERβ. Propyl pyrazole triol (PPT) is approximately 10,000-fold more potent on ERα than on ERβ and it shows ERα-selective effects in vivo [42]. Other larger ring heterocycles, such as tetrasubstituted pyrimidines and pyrazines, also retain greater potency and efficacy on ERα than on ERβ [43].

Triple-bonded C and N (C≡N) are clearly nitrile functional groups. New synthetic bis-benzylnitriles synthetic and related compounds have up to 170-fold potency selectivity on ERβ [44]. Recently a number of diarylpropionitriles, diarylsuccinonitriles as well as acetylene and polar analogues of these nitriles were also found to be ERβ-selective agonists. These ligands have been shown to have high receptor selectivity and considerable ER binding affinity, some almost as much as that of the estradiol. It has been found that ERα has a lesser ability to tolerate the polar nature of the nitrile functionality, while the ERβ is less affected by the polar nature of the nitrile function than by the geometric requirement of the sp hybridization. As a result, ligands with linear groups show high selectivity for ERβ, and the increased polarity of the nitrile group reduces the affinity of the ligand for ERα, resulting in higher ERβ binding selectivities [44].

Hydrogen bonding is a key interaction in estradiol binding at the ER. The A ring alcohol group is known to form H-bond contacts with Glu 353 and Arg 394 while the D ring hydroxyl binds with His 524 residues in the receptor pocket [37]. X-ray crystallography has also confirmed that the phenol alcohol group of 17β-estradiol acts as a hydrogen donor while the 17β-hydroxyl group is a better hydrogen bond acceptor than donor [38]. In other studies on the structural requirements for ER binding and associated estrogenic responses, it was inferred that hydroxylation at specific sites of the estratrien-17β-ol aromatic A ring is a critical requirement. Hydroxylation at the 2 or 3 positions promoted high affinity of a ligand for the ER, while hydroxylation at the 1 or 4 positions attenuate binding affinity. It has been hypothesized that the hydroxyl groups at positions 2 and 3 may share, via hydrogen bonding, a common H acceptor/donor site in the receptor cavity [45].

Chemically intuitive molecular (CIM) indices are noted for their utility in describing chemical diversity [46]. In contrast to the classic topological indices that depend mostly on the size and bulk of the molecule, CIM indices accounts for the interaction between atoms in molecule and molecule with environment. These interactions give rise to electron distribution about the atoms as well as defining the molecular shape and giving rise to polarity, both along individual bond and for the molecule as a whole [47,48]. In general, ERβ selective ligands seem to be smaller and more polar than ERα selective ligands.

The descriptor R=C encodes the number of double bonded carbons with at least one nonhydrogen attachments while the descriptor R=CR encodes the number of double bonded carbons with at least two nonhydrogen attachments (both inclusive of aromatic carbon). Substituted tetrahydrochrysene (THC) ligands are potent agonists on ERα but also potent antagonists on ERβ [24]. This characteristic is a function of substituent size and stereochemistry. THCs can be regarded as ring-fused derivatives of diethylstilbestrol, containing an electron-donating hydroxyl group at C8 and a rigid four-ring structure reminiscent of steroidal estrogens. RR and SS enantiomers of THC have differing activities at ERα and ERβ, for example, ER selective antagonists reside completely in the RR enantiomer. The difference in efficacy of *R,R*-THC on the two ER subtypes appears to arise from its optimal fit in the ERα ligand-binding pocket and its suboptimal fit in the slightly smaller ERβ pocket [29]. In contrast, SS enantiomers have similar agonist activity at ERα and ERβ.

Table 2
Descriptors and combined sensitivity ranking in the two optimum models

Model 1 (21-*63*-3-2)	Rank	Model 2 (21-*73*-3-2)	Rank
Molecular length	18	Molecular length	14
H bond donor	1	H bond donor	3
Kappa 2	11	Kappa 2	15
Water solubility	15	Water solubility	13
ORR	6	ORR	4
CIM 5	14	CIM 5	8
C in nonaromatic rings	9	C in nonaromatic rings	9
R=C	12	R=C	11
R=CR	12	R=CR	12
C in six membered aromatic rings	19	C in six membered aromatic rings	1
CIM 6	17	CIM 4	16
CIM 7	12	Hydrophilic surface area	7
CIM 8	7	CHRRR (chiral C)	10
CIM 9	3	Log *P*	2
CIM 10	5	H bond acceptor	5
C in molecule	11	O in nonaromatic ring	17
R=CRR	13	Singly bonded N in aromatic ring	18
Triple bonded C	10	N in five membered aromatic rings	19
Triple bonded C with no H	15	Double bonded N	20
N≡C	13	N next to another N	21
C in five membered aromatic rings	8	Solubility parameter	6

3.3. GRNN modeling

In silico methods of prediction are gaining increasing popularity in drug discovery due to their speed and relatively low cost. A structure–activity study can indicate which features of a given molecule correlate with its activity, thus making it possible to synthesize new and more potent compounds with enhanced biological activities. QSAR analysis is based on the assumption that the behaviour of compounds is correlated to the characteristics of their structure. Several approaches have been previously proposed for the development of QSAR models. Linear regression has been one of the most common techniques used to construct QSAR models. However, even with moderate numbers of features this technique can result in over-fitting [49]. In order to avoid over-fitting, linear regression is often used in combination with principal component analysis (PCA) [50]. Recently, neural networks and genetic algorithms were found to be efficient in constructing QSAR models [51]. The advantage of using a nonlinear method compared to a linear method such as linear regression is that more complex and nonlinear QSAR models can be derived, which in turn can better reflect the possible relationship between the features of the molecule and its activity (Table 2).

4. Conclusion

The two QSAR models developed revealed the importance of simple molecular characteristics for differential ER binding. Both models selected 10 identical descriptors. These molecular descriptors encode molecular characteristics that are responsible for nonselective binding. The remaining 11 descriptors were characteristic for ERα selectivity (Model 1) and ERβ selectivity (Model 2). Mutual descriptors included molecular size and shape, cyclic structures, solubility parameters, hydrogen bonding/donating potential, electrostatic parameters and the number of ether oxygens. The model has confirmed that five distinguishing criteria are essential for nonselective ER activity of phytoestrogens: H-bonding ability of the phenolic ring mimicking the 3-OH, H-bond donor mimicking the 17β-OH, oxygen–oxygen distance between 3- and 17β-OH, precise steric hydrophobic centers at 7α- and 11β-substituents, hydrophobicity and ring structure. Furthermore, predominant molecular characteristics important for estrogen receptor subtype selectivity for ERα are 17β substituents and substituted heterocyclic structure. Molecular size, polarity and electronic affect, lipophilic substituents are on the other hand important for ERb selective binding. ANNs have proven to be a useful tool in predictive QSAR modeling and further utility may be found by using larger numbers of compounds for model development.

References

[1] S. Nilsson, J.A. Gustafsson, Crit. Rev. Biochem. Mol. Biol. 37 (2002) 1–28.
[2] G.G. Kuiper, E. Enmark, M. Pelto-Huikko, S. Nilsson, J.A. Gustafsson, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 5925–5930.
[3] S. Mosselman, J. Polman, R. Dijkema, FEBS Lett. 392 (1996) 49–53.
[4] E. Enmark, M. Pelto-Huikko, K. Grandien, S. Lagercrantz, J. Lagercrantz, G. Fried, M. Nordenskjold, J.A. Gustafsson, J. Clin. Endocrinol. Metab. 82 (1997) 4258–4265.
[5] S. Ogawa, S. Inoue, T. Watanabe, H. Hiroi, A. Orimo, T. Hosoi, Y. Ouchi, M. Muramatsu, Biochem. Biophys. Res. Commun. 243 (1998) 122–126.
[6] G.B. Tremblay, A. Tremblay, N.G. Copeland, D.J. Gilbert, N.A. Jenkins, F. Labrie, V. Giguere, Mol. Endocrinol. 11 (1997) 353–365.
[7] B.T. Zhu, G.Z. Han, J.Y. Shim, Y. Wen, X.R. Jiang, Endocrinology 147 (2006) 4132–4150.
[8] B.S. Katzenellenbogen, K.S. Korach, Endocrinology 138 (1997) 861–862.
[9] K.E. Carlson, I. Choi, A. Gee, B.S. Katzenellenbogen, J.A. Katzenellenbogen, Biochemistry 36 (1997) 14897–14905.
[10] J. Balasch, Hum. Reprod. Update 9 (2003) 207–222.
[11] W. Tong, R. Perkins, H. Fang, H. Hong, Q. Xie, W. Branham, D.M. Sheehan, J. Anson, Regul. Res. Perspect. 1 (2002) 1–16.
[12] G.T. Ankley, R.D. Johnson, N.E. Detenbeck, S.P. Bradbury, G. Toth, L.C. Folmar, Rev. Toxicol. Environ. Toxicol. 1 (1997) 71–106.
[13] L.E. Gray Jr., Toxicol. Lett. 102–103 (1998) 677–680.
[14] L.M. Shi, H. Fang, W. Tong, J. Wu, R. Perkins, R.M. Blair, W.S. Branham, S.L. Dial, C.L. Moland, D.M. Sheehan, J. Chem. Inf. Comput. Sci. 41 (2001) 186–195.
[15] K.P. Coleman, W.A. Toscano, T.E. Wiese, QSAR Comb. Sci. 22 (2003) 78–88.
[16] A. Asikainen, J. Ruuskanen, K. Tuppurainen, J. Chem. Inf. Model. 43 (2003) 1974–1981.
[17] A.H. Asikainen, J. Ruuskanen, K.A. Tuppurainen, Environ. Sci. Technol. 38 (2004) 6724–6729.
[18] O. Mekenyan, V. Kamenska, R. Serafimova, L. Poellinger, A. Brouwer, J. Walker, SAR QSAR Environ. Res. 13 (2002) 579–595.
[19] L. Kurunczi, E. Seclaman, T.I. Oprea, L. Crisan, Z. Simon, J. Chem. Inf. Model. 45 (2005) 1275–1281.
[20] F. Marini, A. Roncaglioni, M. Novic, J. Chem. Inf. Model. 45 (2005) 1507–1519.
[21] S. Agatonovic-Kustrin, J.V. Turner, Lett. Drug Des. Discov. 7 (2006) 436–442.
[22] M.A. Lill, A. Vedani, M. Dobler, J. Med. Chem. 47 (2004) 6174–6186.
[23] H. Hong, W. Tong, H. Fang, L. Shi, Q. Xie, J. Wu, R. Perkins, J. Walker, W. Branham, D. Sheehan, Environ. Health Perspect. 110 (2002) 29–36.
[24] M.J. Meyers, J. Sun, K.E. Carlson, B.S. Katzenellenbogen, J.A. Katzenellenbogen, J. Med. Chem. 42 (1999) 2456–2468.
[25] J. Sun, Y.R. Huang, W.R. Harrington, S. Sheng, J.A. Katzenellenbogen, B.S. Katzenellenbogen, Endocrinology 143 (2002) 941–947.
[26] W. Sippl, J. Comput. Aided Mol. Des. 14 (2000) 559–572.
[27] D.S. Mortensen, A.L. Rodriguez, K.E. Carlson, J. Sun, B.S. Katzenellenbogen, J.A. Katzenellenbogen, J. Med. Chem. 44 (2001) 3838–3848.
[28] F. Naoum, K.M. Kasiotis, P. Magiatis, S.A. Haroutounian, Molecules 12 (2007) 1259–1273.
[29] A.K. Shiau, D. Barstad, J.L. Radek, M.J. Meyers, K.W. Nettles, B.S. Katzenellenbogen, J.A. Katzenellenbogen, D.A. Agard, G.L. Greene, Nat. Struct. Biol. 9 (2002) 359–364.
[30] ChemSW, Fairfield, USA, 2003.
[31] D.F. Specht, IEEE Trans. Neural Netw. 2 (1991) 568–576.
[32] StatSoft Inc., Tulsa, USA, 2005.
[33] F.W. Schueler, Science 103 (1946) 213–240.
[34] G. Prêcigoux, P. Marsau, F. Leroy, B. Busetta, Acta Cryst. B36 (1980) 749–751.
[35] J.A. Katzenellenbogen, R. Muthyala, B.S. Katzenellenbogen, Pure Appl. Chem. 75 (2003) 2397–2403.
[36] A.C.W. Pike, A.M. Brozozowski, R.E. Hubbard, T. Bonn, A.-G. Thorsell, O. Engstrom, L. Jan, J.-A. Gustafsson, C. Mats, EMBO J. 18 (1999) 4608–4618.
[37] G.M. Anstead, K.E. Carlson, J.A. Katzenellenbogen, Steroids 62 (1997) 268–303.
[38] A.M. Brzozowski, A.C. Pike, Z. Dauter, R.E. Hubbard, T. Bonn, O. Engstrom, L. Ohman, G.L. Greene, J.A. Gustafsson, M. Carlquist, Nature 389 (1997) 753–758.
[39] N. Fokialakis, G. Lambrinidis, D.J. Mitsiou, N. Aligiannis, S. Mitakou, A.L. Skaltsounis, H. Pratsinis, E. Mikros, M.N. Alexis, Chem. Biol. 11 (2004) 397–406.
[40] S.R. Stauffer, Y. Huang, C.J. Coletta, R. Tedesco, J.A. Katzenellenbogen, Bioorg. Med. Chem. 9 (2001) 141–150.
[41] S.R. Stauffer, J. Sun, B.S. Katzenellenbogen, J.A. Katzenellenbogen, Bioorg. Med. Chem. 8 (2000) 1293–1316.
[42] H.A. Harris, J.A. Katzenellenbogen, B.S. Katzenellenbogen, Endocrinology 143 (2002) 4172–4177.
[43] U. Ghosh, D. Ganessunker, V.J. Sattigeri, K.E. Carlson, D.J. Mortensen, B.S. Katzenellenbogen, J.A. Katzenellenbogen, Bioorg. Med. Chem. 11 (2003) 629–657.
[44] M.J. Meyers, J. Sun, K.E. Carlson, G.A. Marriner, B.S. Katzenellenbogen, J.A. Katzenellenbogen, J. Med. Chem. 44 (2001) 4230–4251.
[45] S. Bradbury, O. Mekenyan, G. Ankley, Environ. Toxicol. Chem. 15 (1996) 1945–1954.
[46] R.S. Pearlman, K.M. Smith, J. Chem. Inf. Model. 39 (1999) 28–35.
[47] F. Burden, Quant. Struct. Act. Relat. 16 (1997) 309–314.
[48] F.R. Burden, J. Chem. Inf. Comput. Sci. 29 (1989) 225–227.
[49] D. Rogers, A.J. Hopfinger, J. Chem. Inf. Comput. Sci. 34 (1994) 854–866.
[50] W.D. Glen, W.J. Dunn, R.D. Scott, Tetrahedron Comput. Methodol. 2 (1989) 349–376.
[51] J. Wikel, E. Dow, Bioorg. Med. Chem. Soc. 110 (1988) 5959–5967.