Journal of Molecular Graphics and Modelling 29 (2011) 784–794

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

QSAR study of natural estrogen-like isoflavonoids and diphenolics from Thai medicinal plants

Wanchai De-Eknamkul[a,b,*], Kaoru Umehara[c,b], Orawan Monthakantirat[a,c,d], Radovan Toth[e], Vladimir Frecer[b,f,g], Lorena Knapic[h], Paolo Braiuca[h,b], Hiroshi Noguchi[c], Stanislav Miertus[b,**]

[a] Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok 10330, Thailand
[b] International Centre for Science and High Technology, UNIDO, AREA Science Park, Padriciano 99, Trieste 34012, Italy
[c] School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka 422-8526, Japan
[d] Faculty of Pharmaceutical Sciences, Khon Kaen University, Khon Kaen 40002, Thailand
[e] Department of Chemical Engineering, University of Trieste, Trieste 34127, Italy
[f] Department of Physical Chemistry of Drugs, Faculty of Pharmacy, Comenius University, Bratislava 83232, Slovakia
[g] Cancer Research Institute, Slovak Academy of Sciences, Bratislava 83391, Slovakia
[h] Department of Pharmaceutical Sciences, University of Trieste, Trieste 34127, Italy

ARTICLE INFO

Article history:
Received 3 August 2010
Received in revised form
24 December 2010
Accepted 5 January 2011
Available online 19 January 2011

Keywords:
QSAR
Molecular modeling
Isoflavonoids
Dalbergia parviflora
Belamcanda chinensis

ABSTRACT

Two species of Thai medicinal plants, *Dalbergia parviflora* R. (Leguminosae) and *Belamcanda chinensis* L. (Iridaceae), used traditionally for the regulation of menstrual disorders, have been found to contain a large number of potential estrogen-like compounds. A set of some 55 isolated isoflavonoids and diphenolics showed a wide range of estrogen activity as determined in breast cancer MCF-7 and T47D cell proliferation assays. This set of compounds was studied by means of computational techniques including quantitative structure–activity relationships (QSAR) and molecular modeling. It was found that the estrogenic potencies of the studied compounds depend mainly upon the presence/absence of hydroxyl groups attached to 3′ and 5′ positions of B ring of the isoflavone scaffold and the inter-atomic distance between the hydroxyl groups attached to the outer terminal positions 7 of A ring and 4′ of B ring. In a QSAR model employing ligand–receptor interaction energy descriptors, the LigScore scoring function of Cerius^2 virtual screening module, which describes the receptor affinities of simultaneous binding to estrogenic receptors α and β (ER_α and ER_β), led to the best correlation between the observed estrogenic activities and computed descriptors. Consideration of independent binding to ER_α and ER_β did not result in statistically significant QSAR models. It was thus concluded that simultaneous and possibly competitive interaction of the compounds with the ER_α and ER_β receptors, in which the presence of hydroxyl groups at the abovementioned positions of the isoflavonoids and diphenolics molecular scaffold plays a dominant role, may determine the estrogenic potency of the considered phytochemicals.

1. Introduction

Phytoestrogens are non-steroidal, diphenolic compounds found in fruits, vegetables, whole grains, legumes, and especially soy food products. This group of natural plant product has similar chemical and structural properties to those of estrogens [1]. Phytoestrogens have been known as protective agents against hormone-dependent cancers (e.g., breast and prostate cancer) and age-related diseases (e.g., osteoporosis and cardiovascular diseases) [2]. Currently, four different groups of plant phenolics are considered phytoestrogens: the isoflavonoids, stilbenes, lignans and coumestans [1]. Other less-investigated compounds include prenylflavonoids, flavones, flavans and phytosterol esters [3].

Like the female hormone estrogen, phytoestrogens have the ability to alter the growth, development, and function of estrogen-dependent target tissues, either through their estrogenic or anti-estrogenic activity [4]. These biological actions are mediated via two distinct nuclear estrogen receptor (ER) proteins, ER_α and ER_β, which belong to a large conserved superfamily of nuclear receptors [5]. Both receptors have specific tissue distribution and play a distinct role in human physiology [6]. ER_α and ER_β show 58% sequence identity in their ligand-binding domains and approximately 96% similarity in the DNA-binding domains [7]. Both ERs bind

* Corresponding author at: Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok 10330, Thailand. Tel.: +66 2 218 8393; fax: +66 2 218 8393.
** Corresponding author. Fax: +39 040 922 8115.
E-mail addresses: dwanchai@chula.ac.th (W. De-Eknamkul), stanislav.miertus@ics.trieste.it (S. Miertus).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.01.001

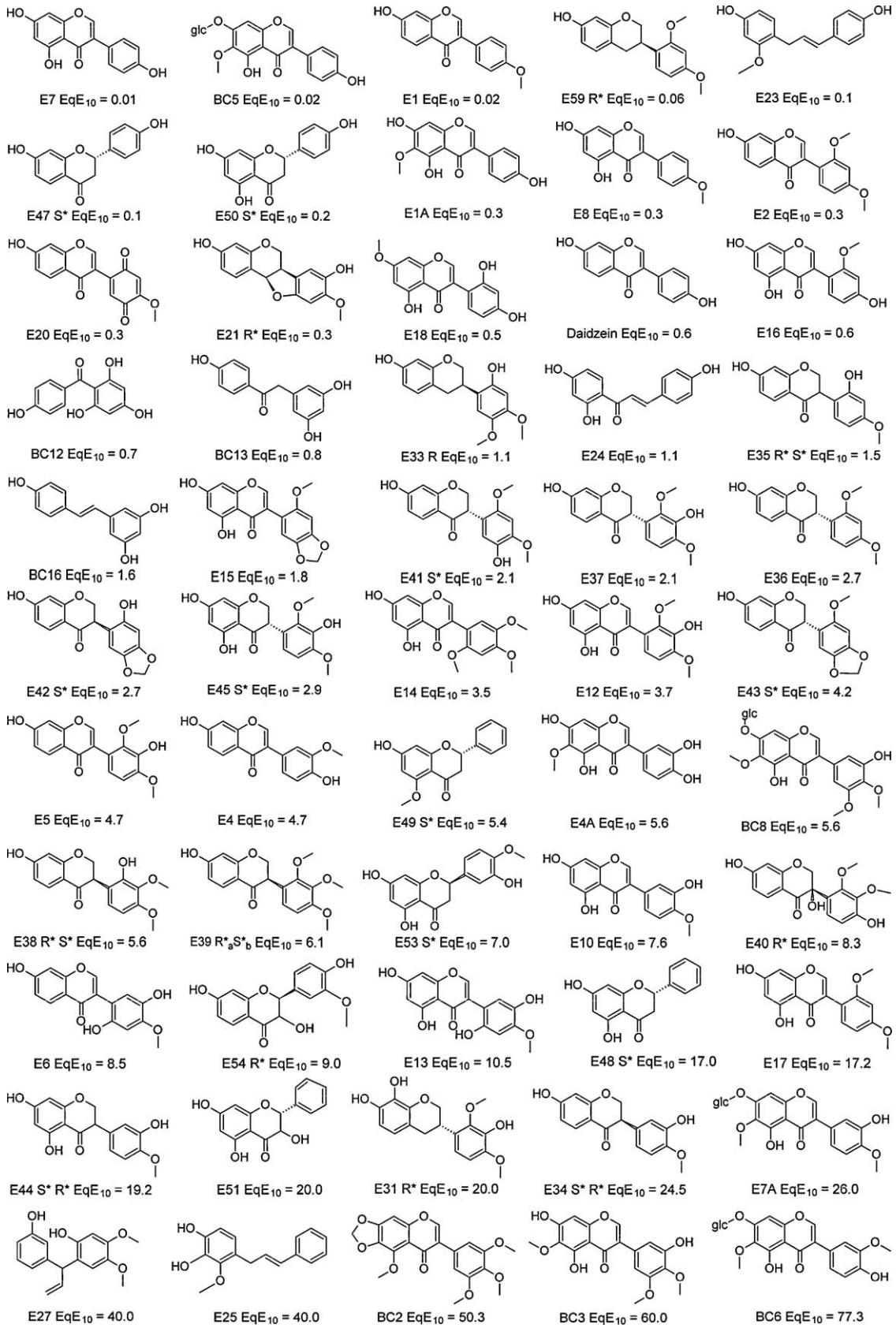

Fig. 1. Estrogenic compounds studied. Compound code and estrogenic biological activity measured in the MCF-7 cells proliferation assay expressed as equivalent concentrations stimulating the cell proliferation equivalent to 10 pM of estradiol (EqE_{10}) in [M] are shown. The EqE_{10} values were determined by linear regression analysis using at least five different concentrations and were done in quadruplicate [22–24]. In racemic mixtures, the preferred stereoisomers displaying the highest binding affinity towards the ER_α and ER_β receptors are indicated. For example, E_{59} R* means that E_{59} bound more strongly to both ER_α and ER_β as the *R*-enantiomer. On the other hand, E_{35} R* S* means that E_{35} bound more strongly as the *R*-enantiomer to ER_α (first index) and as the *S*-enantiomer to ER_β (second index). In structures in which both chiral forms are active, the wedged bond leading from the chiral carbon is replaced by a dashed bond (e.g., E_{35}).

estradiol with high affinity but vary in their ability to bind other natural and synthetic ligands and in the type of response elicited upon ligand binding [8,9]. Both receptors interact with the same conserved estrogen response element (ERE) as either homodimers or α/β heterodimers [10–12]. Tissue-specific expression and co-expression of receptor subtypes suggest that ER homodimers and heterodimers may mediate distinct hormone responses [13–16].

Drugs that target the ER can exhibit a variety of effects in different tissues. For example, tamoxifen is an ER antagonist in breast tissue [17] but an ER agonist in bone [18] and uterine tissue [19]. Raloxifene is also an ER antagonist in breast tissue and an ER agonist in bone, but it does not exert agonist activity in uterine tissue [20]. Indeed, one of the greatest challenges in understanding the pharmacology of the ER is determining how different ER ligands produce such diverse biological effects [21]. Pharmaceutical companies have therefore been interested in the search for compounds suitable for hormone replacement therapy as selective estrogen receptor modulators (SERMs) [4]. So far, the group of isoflavonic phytoestrogens and isoflavones with properties similar to those of SERMs has attracted a substantial degree of attention.

Very recently, we reported the isolation and structural elucidation of estrogen-like compounds from two Thai medicinal plants, namely *Dalbergia parviflora* R. (Leguminosae) [22,23] and *Belamcanda chinensis* L. (Iridaceae) [24]. These two plant species were chosen based on their traditional uses related to their estrogenic activities: the former as a blood tonic to normalize menstruation and the latter for the regulation of menstrual disorders. As expected, a number of estrogen-like constituents were isolated and identified, including seventy compounds from the heartwood of *D. parviflora* [22,23], and sixteen compounds from the rhizome of *B. chinensis* [24]. Interestingly, among these, as many as 22 isoflavones, 12 isoflavanones, 10 isoflavans, 6 flavanones, 4 flavanonols, and 6 pterocarpans were found in *D. parviflora*, and 6 isoflavones, 2 phenones and one stilbene in *B. chinensis*. This has readily provided a library of natural phytoestrogens with ample structural diversity. All of these isolates have been subjected to estrogen-like activity assays using breast cancer cell lines of MCF-7 and T47D and appear to have a wide range of activities [22–24].

In an effort to explore the relationship between the structure and estrogen-like activity with respect to the specific role of individual substituents, the set of phytoestrogenic molecules (Fig. 1) has been studied by exploring quantitative structure–activity relationships (QSAR) and molecular modeling of estrogen receptor binding. We report here the resulting QSAR models, which can assist the investigation of the possible mechanisms of action of the bioactive compounds and can be used to predict estrogenic activities of similar naturally occurring compounds or synthetically decorated analogs.

2. Methodology

2.1. Experimental estrogenic activities

The estrogenic activities of isolated phytoestrogens were obtained from a proliferation assay using breast cancer cell lines of MCF-7 and T47D and expressed by Umehara et al. [22–24] as concentrations of the compound that stimulated cell proliferation equivalent to 10 pM of estradiol (EqE_{10}).

2.2. 2D-QSAR analysis by GFA approach

Atomic descriptors of the phytoestrogens were computed to determine gas-phase molecular structures with geometries optimized by an *ab initio* quantum mechanical method at the HF/6-31G (d,p) level of theory using the Gaussian 03 program

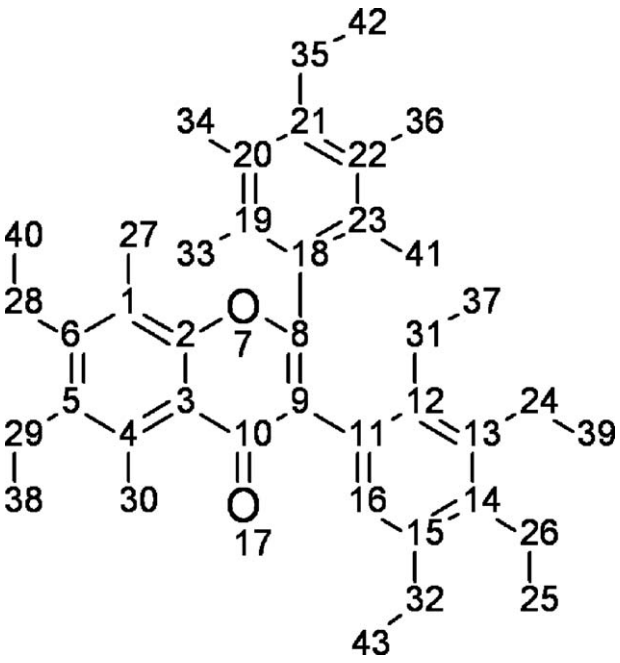

Fig. 2. Structural formula of a hypothetical super-molecule with an estrogen-like core obtained by superposition of chemical structures of the estrogenic compounds. Atomic position numbering is shown.

[25] (see Supplementary Information for the atomic coordinates of the considered compounds). Atomic descriptors such as Mertz–Kollman net atomic charges [26] were combined with a selection of molecular descriptors. Interatomic distances between the hydroxyl group oxygens at positions 6 and 14 (Fig. 2), total, polar and hydrophobic molecular surfaces, volume, molecular mass, total charge, dipole moment, sum of atomic polarizabilities, number of hydrogen bond donors and acceptors, water/octanol partitioning coefficient $A\log P_{o/w}$, number of rotatable bonds, principal moment of inertia, radius of gyration, selected interatomic distances, desolvation free energy in water and octanol, molar polarizability and refractivity, and the topological indices of Balaban, Wiener and Zagreb were computed by the Cerius^2 program [27]. Atomic charge equal to zero was assigned to vacant positions of the hypothetical super-molecule scaffold (Fig. 2) not occupied by atoms in the individual molecules (Fig. 1).

The space of possible activity–descriptors correlations was sampled by the GFA algorithm [28] of the Cerius^2 package [27] using the squared correlation coefficient (R^2) to rank the fitness of each model. In the GFA approach, progeny populations of equations are created by evolving a random initial population of equations constructed from the set of available descriptors by standard regression techniques, using the genetic algorithm. The genetic algorithm automatically selects the type and number of descriptors to be used in the models by combining (*crossing-over*) equations with the best fitness scores, as *good combinations of genes* (i.e., descriptors) are expected to lead to improved QSAR models in the resulting offspring population of equations. For the activity data, the analyses were begun by building a population of 400 randomly constructed linear equations containing combinations of 4 descriptor terms. The initial population was then evolved for 10^5 generations with a variable equation length and mutation probabilities set at: (i) adding new term 50%, (ii) reducing number of terms 50% and (iii) extending number of terms 50%. The smoothness parameter was set to 2.0 to bias the equation scoring and evolution towards a smaller number of terms, i.e., more compact and easier-to-interpret QSAR models. The least-squares regression method was used to generate the QSAR equations for the sampled combinations of descriptors.

Table 1
QSAR models of the MCF-7 activity derived by GFA approach from atomic and molecular descriptors.

Descriptor	Compound[a]	R^2	R^2_{xv}	*F*-test	Regression equation
Q_{13}[b] Q_{32} D_{OO}[c]	41	0.85	0.82	70.9	$\log_{10}\mathrm{EqE}_{10} = -1.1052 - 2.2404Q_{13} + 0.8079Q_{32} + 0.0659D_{OO}$
Q_{13} Q_{32}	38	0.83	0.80	84.3	$\log_{10}\mathrm{EqE}_{10} = -0.3556 - 2.2559Q_{13} + 0.7698Q_{32}$
Q_{13} M_w [d]	41	0.81	0.78	80.8	$\log_{10}\mathrm{EqE}_{10} = 3.6418 - 1.3488Q_{13} - 0.0128M_w$
Q_{24} HB_{acc}[e]	39	0.81	0.78	76.2	$\log_{10}\mathrm{EqE}_{10} = 2.6812 - 0.8086Q_{24} - 0.4909HB_{acc}$
Q_{13} HB_{acc}	40	0.80	0.77	73.4	$\log_{10}\mathrm{EqE}_{10} = 2.8177 - 1.3931Q_{13} - 0.5272HB_{acc}$
Q_{13} D_{OO}	39	0.80	0.77	70.8	$\log_{10}\mathrm{EqE}_{10} = -1.0892 - 2.6961Q_{13} - 0.0571D_{OO}$

[a] Number of compounds in the correlation after removal of outliers, squared correlation coefficient of the regression (R^2), predictive leave-one-out cross-validated squared correlation coefficient (R^2_{xv}), statistical significance of the correlation expressed as Fischer-test (*F*-test).
[b] Q_{C13} – net Mertz–Kollman atomic charge [26] on atom C13 (see Fig. 2 for atom numbering) computed using Gaussian 03 [25].
[c] D_{OO} – interatomic distance between hydroxyl groups atoms O_{28} and O_{26} attached to carbons C_{14} and C_6 (Fig. 2) [Å].
[d] M_w – molecular weight [Da].
[e] HB_{acc} – number of hydrogen bond (proton) acceptors in the molecule.

In the end, the final population of equations was ranked based on the R^2 value, and the highest-scoring QSAR models were subjected to the removal of outliers[1] and leave-one-out cross-validation [29].

2.3. Molecular modeling of estrogen receptor binding

The crystal structure of ER_α complexed with estradiol (PDB entry code 1A52 [30]) and ER_β complexed with genistein (PDB entry code 1X7J [31]) were used as the target 3D receptor models for docking of the phytoestrogenic compounds. The shape and size of the ligand binding sites in ER_α and ER_β were defined from the bound ligands and were mapped onto 3D energy grids with a resolution of 0.25 Å using the LigFit module of the Cerius^2 [27]. The compounds were docked into the rigid binding site models of the two receptors as flexible molecules by generating conformers of each molecule via randomizing their dihedral angles (10^4 Monte Carlo steps). The generated conformers were fitted to the site models via a flexible fit algorithm by comparing the principal moments of inertia of the sites and the molecule after 50 rigid body minimization steps of 4 possible orientations [32].

The docking score (ligand–receptor interaction energy) was computed as the nonbonding molecular mechanics energy term using the CFF91 force field [33] for each generated conformer after 500 energy minimization steps at the receptor binding site. During the docking score calculation, a grid representation of rigid ER_α and ER_β receptors, a cut-off distance of 15 Å applied to the non-bonded interactions and a dielectric constant of 2 were used. Twenty best-fitting conformers per each molecule were clustered into ten conformational families according to their mutual r.m.s. deviations by means of the Jarvis–Patrick complete linkage clustering method [34], and the best representative of each cluster was further considered. The best conformer in each cluster was selected for virtual screening using the LigScore scoring function, which was selected out of a pool of predefined scoring functions of the Cerius^2 package [27], as it proved to best reproduce the estrogenic activities of the compounds in a preliminary screening run.

The linear QSAR models of estrogenic activities of the type – $\log_{10}\mathrm{EqE}_{10} = f(D_1, D_2, \ldots)$, which combined the LigScore binding affinities to the ER_α ($\mathrm{LigScoreER}_\alpha$) and ER_β ($\mathrm{LigScoreER}_\beta$) receptors with a set of molecular descriptors that characterize the pharmacokinetic properties of the compounds, were prepared by a least-square fitting of the computed descriptors D_i to the activities EqE_{10} measured in the proliferation assays for the MCF-7 and T47D cell lines [22–24]. The statistical significance and the predictive power of the regression equations were assessed by the Fischer *F*-test and by the leave-one-out cross-validation method.

2.4. 3D-QSAR studies based on GRIND descriptors

For the construction of the 3D-QSAR models, the ALMOND program version 3.3.0 was used [35], which was specifically developed for chemometric analysis and calculation of 3D-QSAR models. Not all of the constituents of the initial data set were included in the calculation of the models; the compounds with extremely low activity and the racemic ones were excluded due to their poor contribution to the model in terms of low statistical significance or the ambiguity that they introduced to the model. All of the possible auto- and cross-correlograms were calculated and used for the partial least squares (PLS) analysis. The regression models were validated by means of the leave-one-out cross-correlation procedure. The removal of the outliers continued until the highest predictivity of the models was reached.

The GRIND method transforms the information included in the MIF into alignment-independent descriptors able to describe the different chemical groups in the molecule and their relative spatial position. The calculation of GRIND is a two-step procedure. First, hundreds of thousands of variables, which constitute the original MIF, are filtered to select the most relevant groups of nodes and discard redundant variables. The chosen nodes must fulfill the requirements of having low energy values (corresponding to favorable interactions with a given probe) and being as distant as possible from each other. The second step is the so-called maximum auto- and cross-covariance (MACC) transformation. It is an autocorrelation procedure [36] in which the nodes, selected in the first step, are screened by identifying couples of nodes that are localized at a defined distance. The algorithm uses vectors of length ranging from 1 Å to *n* Å dependent on the size of the molecule under investigation to localize the couple, after which the energy product of the two nodes is calculated. When more than one couple of nodes fulfills the distance requirement, only the vector representing the maximum energy product is conserved. The correlation can be performed between nodes belonging to the same molecular interaction field (MIF) (generated by the same probe), or to different MIFs (generated by two different probes), resulting in auto- or cross-correlation, respectively. Therefore, at the end of this procedure, each molecule of the data set is described by (i) a number of vectors, which link couples of original MIF nodes and (ii) their energy products. The descriptors can be plotted in a correlogram profile, where the distances (the lengths of the vectors) appear on the *x*-axis and the energy product on the *y*-axis. Each correlogram constitutes a sort of fingerprint of the molecule and represents the molecule independently from its position in the space. Ultimately, the correlograms form the input matrix for the multivariate analysis and the construction of the regression models.

[1] Statistical validation techniques (leave-one-out cross-validation) were applied to identify outliers (data points that are not modeled well by a regression equation). In the QSAR module of Cerius^2, an outlier is defined as a point with a residual greater than two times the standard deviation of the residuals generated in the validation procedure [27].

2.5. GRID-PCA analysis of the estrogen receptors

The 3D structures of the receptors were aligned and superposed using the align tool of the MOE software package [37]. By performing this operation, it was possible to define a grid of a defined dimensionality to surround the binding sites of both receptors. The chosen dimensions of the grid were 22 Å × 24 Å × 24 Å, while the grid spacing was 1 Å. By doing so, each grid node represented the same position of the three-dimensional space regardless of the protein under investigation.

The analysis was performed by using the GOLPE program [38], and the procedure was applied to both receptors ER_α and ER_β. The GRID-PCA procedure avails itself of the calculation of GRID-based molecular descriptors and subsequent principal components analysis (PCA). Seven different probes (namely C3, OH, NH, H2O, O, DRY and N1) were used for the calculation of the molecular interaction fields (GRID descriptors). For the PCA, a data pre-treatment was performed by imposing a "cutoff" to positive variables by setting to zero all variables with an absolute value lower than 0.02 and a standard deviation below 0.03. Moreover, the "region cut-out" tool was employed to define the shape of the catalytic site. For this purpose, we used as templates both the estradiol molecule co-crystallized in the active site of ER_α and the genistein molecule co-crystallized in the active site of ER_β.

3. Results and discussion

The estrogenic activities of 55 isoflavonoids and diphenolics (Fig. 1) isolated from Thai medicinal plants *Dalbergia parviflora* R. (Leguminosae) and *Belamcanda chinensis* L. (Iridaceae) were obtained from the works of Umehara and coworkers [22–24]. The activity values were obtained from a proliferation assay using breast cancer cell lines of MCF-7. The results were expressed as equivalent concentrations (EqE_{10}, in μM) of the compound that stimulated cell proliferation equivalent to 10 pM of estradiol.

Two kinds of QSAR analyses of the estrogenic activities were performed: (i) 2D-QSAR analysis by genetic function approximation (GFA) [28] and (ii) 3D-QSAR analysis using GRIND descriptors (GRID independent molecular descriptors) generated by the ALMOND algorithm [35]. Moreover, two different molecular modeling studies of estrogen receptor binding were performed: (i) docking of the phytoestrogenic molecules to the ERs and calculation of the receptor-binding affinities and (ii) GRID-PCA analysis of the two receptors to assess the physicochemical properties of their binding pockets.

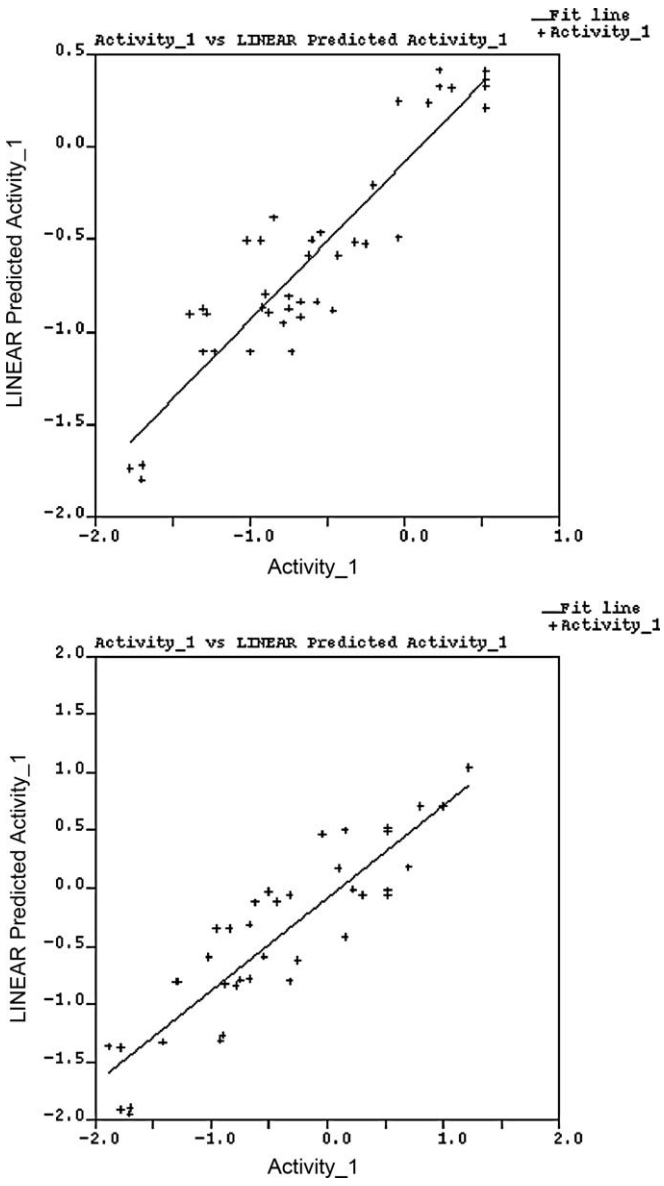

Fig. 3. Plot of correlation of estrogenic activity measured in MCF-7 cells versus the activity predicted from QSAR regression equations (Table 1): (A) $\log_{10} EqE_{10}[MCF7] = -1.1052 - 2.2404Q_{13} + 0.8079Q_{32} + 0.0659D_{OO}$. (B) $\log_{10} EqE_{10}[MCF7] = 2.8177 - 1.3931Q_{13} - 0.5272HB_{acc}$.

3.1. 2D-QSAR analysis

The GFA method [28] of the Cerius^2 package [27] was employed to generate and optimize linear multiparameter correlation equations involving a large number of atomic and molecular descriptors. Fig. 2 shows the structural formula of a hypothetical supermolecule with an estrogen-like core, which was obtained by manual superposition of the 2D chemical structures of the studied compounds. Atoms in all considered compounds were labeled according to this scheme. Net atomic charges in all of the considered positions as well as additional atomic and molecular properties and descriptors related to molecular size, shape, polarity and hydrogen bonding abilities (for details, see Section 2) were used as descriptors in the 2D-QSAR analysis.

The GFA algorithm generated an optimized set of QSAR regression equations, which were highly correlated to the experimental activities. The twenty-five best equations were further analyzed by iterative removal of outliers and leave-one-out cross-validation. During each iterative cycle, a small number of outliers automatically identified by the QSAR module of Cerius^2 as points with a residual greater than two times the standard deviation of the residuals generated in the preceding leave-one-out validation procedure were removed [27]. The process continued until no more outliers were present in the data set. Due to the heterogeneity of the chemical structures of the considered estrogen-like isoflavonoids and diphenolics (Fig. 1) and the simple compound superposition scheme (Fig. 2), a number of outliers were removed. Finally the six best QSAR models for the MCF-7 biological activity were recorded (Table 1, Fig. 3). It follows from the models shown in the Table 1 that the estrogenic potencies of the studied compounds depend mainly upon the presence/absence of the hydroxyl groups attached to carbons C_{13} and C_{15} (Fig. 2), which is reflected by the magnitudes of the atomic net charges at these positions or atoms attached to them as well as upon the distance between the hydroxyl groups at positions C_6 and C_{14} (distance D_{OO}). In fact, the presence of hydroxyl groups at the longitudinal extremities of the molecular scaffold is impor-

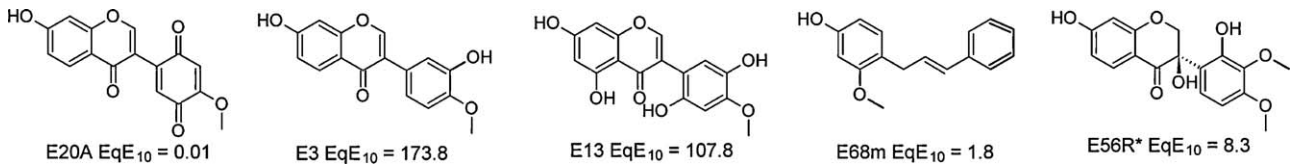

Fig. 4. Estrogenic compounds of the validation set (not included into the QSAR models derivation). Compound code and experimental estrogenic activity EqE_{10} measured in the MCF-7 cells proliferation assay are also shown.

tant for binding to ER_α and ER_β receptors, as expressed by the inclusion of the D_{OO} descriptor in the best QSAR model. On the other hand, hydroxyl or methoxy substitutions in the ortho and meta positions of the benzene ring of the isoflavone scaffold, e.g., in C_{13} (net atomic charge of C_{13} is positive) and C_{15} (net atomic charge of O_{32} is negative), seem to reduce the estrogenic activity of the compounds. The selected isoflavonoids were ordered by decreasing activity (increasing EqE_{10}) in MCF-7 cells as follows: $E_7 > E_8 > E_{16} > E_{15} > E_{14} > E_{12} > E_{10} > E_{17}$; these isoflavonoids, in fact, show an approximate increase in the number of *o*- and *m*-substitutions, although other factors must also play a role (as E_{10} and E_{17} do not fully comply with this trend). The role of individual substitutions upon the activity can be more precisely elucidated by modeling the ligand–estrogenic receptor interactions (see Section 2.3). Similar conclusions were presented by Fang et al. [39], who previously studied a large set of estrogenic molecules. Other molecular descriptors that were shown to be related to the estrogenic activity were molecular weight (M_w) and the number of hydrogen bond (proton) acceptors (HB_{acc}), as shown in Table 1. Both M_w and HB_{acc} are related to the number of hydroxyl and methoxy substituents attached to the molecular scaffold.

To validate the most promising 2D QSAR model (Eq. (1), Table 1) with an external set of test compounds, we examined the correlation between observed and computed activities for a small series of estrogen-like compounds (Fig. 4) not included in the model generation for which homogeneous experimental MCF-7 activity data were available [22–24]. The correlation of the measured EqE_{10} values and activities predicted from the model was satisfactory ($R^2 = 0.93$).

3.2. Molecular modeling of estrogen receptors binding

Molecular modeling of phytoestrogen binding to the estrogen receptors was carried out by docking the 55 isoflavonoids and diphenolics molecules to the 3D models of binding sites of human ER_α and ER_β receptors and estimating the receptor binding affinities by means of scoring functions (for details, see Section 2). The computed binding affinities, which were assumed to correspond to the extent of the receptors activation, were correlated with observed cell growth stimulation effects.

All the studied phytoestrogens were docked to the rigid binding sites using the LigFit tool of the Cerius^2 program [27]. Fig. 5 shows the native ligand estradiol at the binding site of the ER_α and displays the existing hydrogen bonds between the hydroxyl groups positioned at the extremities of the ligand scaffold and residues of the binding site. The studied phytoestrogens exhibit relatively high structural diversity and acquired a variety of binding conformations in the rigid estradiol (ER_α) and genistein (ER_β) binding sites, which were controlled mainly by the hydrogen bonding interactions taking place at the longitudinal extremities of the scaffolds and hydrophobic interactions of the core portion of the molecular scaffolds with the ligand binding pockets of the ERs. The typical pose of the docked phytoestrogens corresponded well with the binding mode of the native ligand, with the 7-OH group of the benzopyrone ring forming hydrogen bonds with Arg 394 and Glu 353 and phenolic hydroxyls making direct contact with His 524, while the body of the compound participated in hydrophobic contacts with receptor residues lining the closed binding pocket, such as Leu 346, Leu 349, Ala 350, Phe 404 and Leu 525 (Fig. 5).

Out of nine scoring functions available in the virtual screening module of the Cerius^2 program, the LigScore function [40] led to the best correlation between the receptor affinities, representing simultaneous binding to both ER_α and ER_β receptors and the

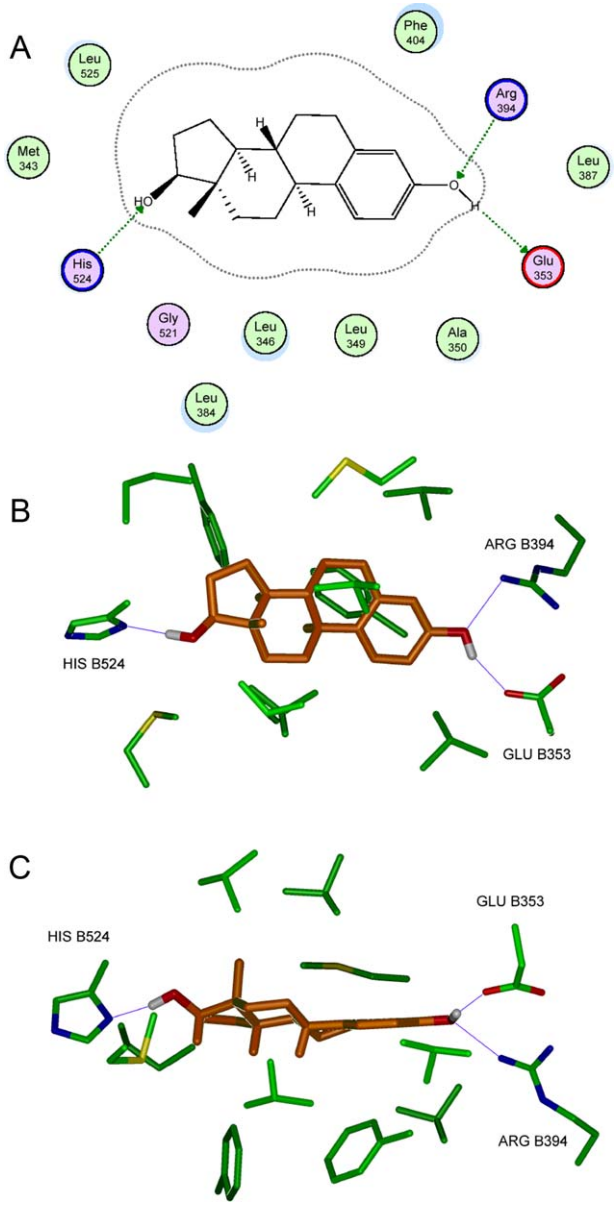

Fig. 5. Estradiol at the binding site of estrogen receptor ER_α. (A) Essential hydrogen bonds between OH groups located at the extremities of the ligand scaffold and residues of the receptor lining the binding site are shown. (B) 3D-structure of the ligand at the ER_α active site shown in stick representation. Hydrogen bonding interactions are shown as yellow lines, ligand hydroxyl oxygens are shown in red and hydrogens in white. Front view. (C) Side view.

Table 2
Computed binding affinities of selected isoflavonoids with increasing number of hydroxyl and methoxy substitutions on the benzene ring of the isoflavonoid scaffold to the ER_α and ER_β receptors expressed as $LigScoreER_\alpha$ and $LigScoreER_\beta$ scores.

Selected Isoflavonoids	$LigScoreER_\alpha$[a]	$LigScoreER_\beta$
E_7	3.9	5.4
E_8	−7.8	3.9
E_{16}	−1.2	5.1
E_{15}	−4.4	4.4
E_{14}	−17.8	3.9
E_{12}	−19.7	4.9
E_{10}	−11.0	4.0
E_{13}	−2.4	4.4
E_{17}	−16.8	4.3

[a] LigScoreERi is the ligand–receptor binding affinity of the estrogen receptor ER_i computed by LigScore scoring function of Cerius [2,27]. A higher value of the $LigScoreER_i$ implies a higher binding affinity of the given ligand to the ER_i receptor.

estrogenic activities measured in MCF-7 and T47D cells. When the descriptors $LigScoreER_\alpha$ and $LigScoreER_\beta$ were considered independently, they did not yield statistically significant QSAR models. With some exemptions, the substitution pattern of the benzene ring of selected isoflavonoids can be approximately related to the fitness to binding sites of the ER_α and ER_β receptors (Table 2).

However, when binding to both receptors expressed via $LigScoreER_\alpha$ and $LigScoreER_\beta$, scores were considered; then, statistically significant correlations were obtained (Table 3). Augmenting the descriptor selection using a pharmacokinetics-related parameter *A* log *P*98, a water/octanol partitioning coefficient that characterizes the ability of the compound to permeate a cellular membrane towards the site of action, improved the final regression equations even further, as shown in Table 3 and depicted in Fig. 6. The different magnitudes of the coefficients multiplying the $LigScoreER_\alpha$ and $LigScoreER_\beta$ descriptors and, in the case of the T47D activity, also the different signs of these two coefficients may be interpreted in terms of a dominant role or concentration of the ER_α and ER_β receptors in the two types of cells (MCF-7 and T47D) and competitive binding to the ER_α over ER_β receptor in the T47D cell line, which leads to the estrogenic effect of the compounds.

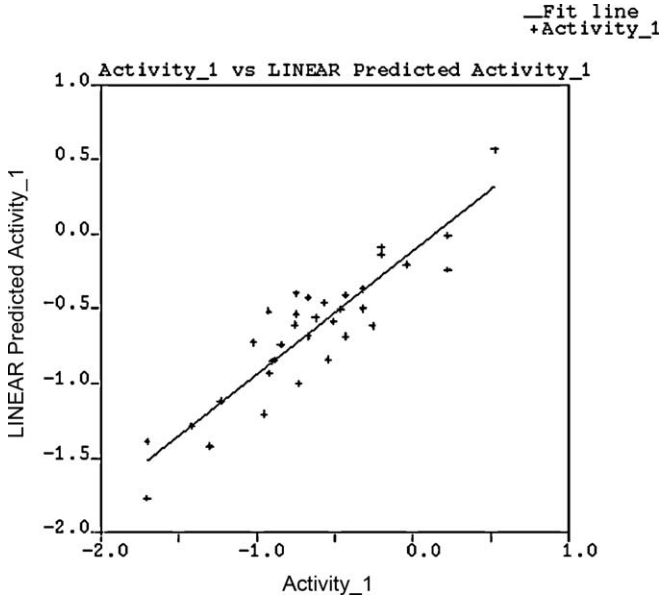

Fig. 6. Plot of the observed estrogenic activity measured in MCF-7 cells versus activity predicted from regression equation linking the activity to the ER_α and ER_β binding activities (Table 3): $\log_{10} EqE_{10}[MCF7] = -3.6636 + 0.0095 LigScoreER_\alpha + 0.4462 LigScoreER_\beta + 0.5575 A \log P98$.

In the literature, human ER_α and ER_β have been shown to share less than 20% sequence homology between their N-terminal transactivation domains and 55% homology between their ligand-binding domains but well-conserved DNA-binding domains (95%) [43]. The ligand-binding domains of the two ER subtypes have very similar three-dimensional structures, although their amino acids lining the ligand-binding cavities differ in two positions [44]. One of these is the replacement of the leucine at position 336 in ER_α by a bulkier methionine in ER_β, causing the ligand-binding cavity of ER_β to be significantly smaller (~20%) than that of ER_α, which may have implications on the selective affinity and activity of ligands [44]. We have used semi-flexible docking algorithm [32], which allows for conformational flexibility of the ligands at the active site of both ER_α and ER_β, in addition, for ligands which contain a chiral center and are known to be present in both enantiomeric forms in the plant extracts, both *R*- and *S*-forms of the compounds were considered during the docking. From the computed ligand–receptor interaction energy, we were able to predict which stereoisomer displays higher binding affinity towards a particular type of the estrogenic receptor and compute the LigScore scoring function [27] for the preferred enantiomer (Fig. 1). We can namely assume that the observed estrogenic activity can be attributed mainly to the receptor binding of this favored stereoisomer.

To validate the QSAR model of estrogenic activity of MCF-7 cells involving computed binding affinities to human ER_α and ER_β receptors (Eq. (1), Table 3) by means of external test compounds (Fig. 4), we examined the correlation between observed and predicted activities. Correlation with an R^2 value of 0.89 further confirmed the validity of this model.

3.3. 3D-QSAR analysis

The initial 3D-QSAR analysis was performed with the CoMFA (comparative molecular field analysis) approach [41], which samples steric and electrostatic fields surrounding a set of ligands to determine a quantitative structure-activity relationship. The estradiol co-crystallized in the active site of the ER_α [29] was chosen as the template to perform a flexible alignment procedure for the superimposition of all the constituents of the data set, a crucial preliminary step of the CoMFA analysis. However, this approach did not lead to satisfactory results, revealing a lack of the models' predictive and interpretive capabilities, possibly due to problems in the alignment procedure, which could be inapt for the present case. Therefore, we adopted an alternative, alignment-independent approach to calculate the 3D-QSAR models, namely the Grid Independent Descriptors (GRIND) [35] method. This approach relies on the chemical description of the ligands by means of the GRID program [42], which is a computational procedure that calculates the interaction between the ligand and a small chemical probe. Energies are calculated in all the nodes of a three-dimensional grid, which spans the structure of the substrates considered. The output is called a molecular interaction field (MIF), which can be used as input for a multivariate analysis to generate QSAR models.

The GRIND method was used to calculate the MIF descriptors for both ER_α and ER_β ligand conformers, which were derived by docking the ligands to the ER_α and ER_β receptors using the LigFit tool of the Cerius^2 program. The following probes were used in the MIF derivation: DRY (hydrophobic probe), O (hydrogen bond acceptor), N1 (hydrogen bond donor) and TIP (shape probe). A multitude of the resulting variables required a multivariate statistical strategy for data analysis. The standard NIPALS algorithm in its GRIND implementation was used to perform partial least square (PLS) regression. Four QSAR models were calculated (Table 4), and five principal components were extracted for each of the models. The predictivity of the models was evaluated by cross-validation

Table 3
QSAR models for MCF-7 and T47D activity containing receptor binding affinities and pharmacokinetic descriptors.

Descriptor	Cells	Compound[a]	R^2	R^2_{xv}	F-test	Regression equation
LigScoreER$_\alpha$[b]	MCF7	34	0.83	0.79	47.27	$\log_{10}$ EqE$_{10}$[MCF7] = −3.6636 + 0.0095LigScoreER$_\alpha$ + 0.4462LigScoreER$_\beta$ + 0.5575*A* log *P*98
LigScoreER$_\beta$						
A log *P*98[c]						
Compounds included		BC_{16}, BC_2, E_{33}, E_{35}, E_{36}, E_{37}, E_{38}, E_{41}, E_{42}, E_{43}, E_{45}, E_{10}, E_{12}, E_{14}, E_{15}, E_{16}, E_{17}, E_4, E_5, E_6, E_{40}, E_{56}, E_{51}, E_{54}, E_{48}, E_{49}, E_{53}, E_{23}, BC_{13}, $BC_{12,4A,7A}$, Daidzein				
LigScoreER$_\alpha$	T47D	23	0.80	0.72	24.63	$\log_{10}$ EqE$_{10}$[T47D] = −6.9752 + 0.6286LigScoreER$_\alpha$ − 0.0348LigScoreER$_\beta$ + 1.8832*A* log *P*98
LigScoreER$_\beta$						
A log *P*98						
Compounds included		BC_{13}, BC_5, E_{28}, E_{33}, E_{35}, E_{37}, E_{38}, E_{41}, E_{42}, E_{43}, E_{45}, E_{10}, E_{12}, E_{16}, E_{17}, E_{18}, E_{40}, E_{54}, E_{47}, E_{53}, E_{11A}, Daidzein				

[a] Number of compounds in the correlation after removal of outliers, squared correlation coefficient of the regression (R^2), predictive leave-one-out cross-validated squared correlation coefficient (R^2_{xv}), statistical significance of the correlation expressed as Fischer-test (F-test).
[b] LigScoreER$_i$ is the ligand–receptor binding affinity of the estrogen receptor ER$_i$ computed by LigScore scoring function of Cerius [2,27].
[c] *A* log *P*98 – water/octanol partitioning coefficient descriptor of the QSAR module of the Cerius^2 is a pharmacokinetic descriptor, which characterizes the ability of the compound to permeate cellular membrane.

via the leave-one-out method, which is ultimately expressed in terms of a predictive correlation coefficient (q^2). In all cases, the descriptive and predictive capacity of the models was augmented by elimination of the outliers and the selection of the most informative variables. For ER$_\alpha$ conformers, the best q^2 obtained by the regression of the MIF variables with their MCF-7 activity was 0.70, whereas using their T47D activity as the dependent variable, the PLS regression yielded a model with a q^2 of 0.61. Analogously, two QSAR models were calculated for ER$_\beta$ conformers, in which case the regression of the MIF descriptors versus their MCF-7 and T47D activity values generated models with q^2 values of 0.62 and 0.55, respectively.

The necessity of two separate analyses for the components of the data set, i.e., ER$_\alpha$ and ER$_\beta$ conformers, derives from the outcome of the docking procedure. In fact, the algorithm calculates different conformations of the same ligand based on the specific macromolecular environment at the active site of the two receptor proteins.

Therefore, the starting point for the QSAR analyses was two distinct sets of conformers that were inevitably considered in two distinct analyses. As a matter of fact, the resulting models indicate a better relationship between ER$_\alpha$ conformers and the abovementioned activities, expressed as higher q^2 values, possibly suggesting a dominant role in the observed estrogenic activity or higher expression level of the ER$_\alpha$ receptor in both cancer cell lines.

Further analysis of the auto- and cross-correlograms calculated for the models with the activity of the ligands (both ER$_\alpha$ and ER$_\beta$ conformers) in the MCF-7 cell proliferation assay indicated that, in this case, the most informative probe was the O probe, which simulates the hydrogen bond acceptor areas. The results indicate that the most active compounds are those that lack hydrogen bond donors (like hydroxyl groups) in the central part of the molecular scaffold, though they contain groups capable of hydrogen bonding on the extremities of the structure (C_6 and C_{14}), as shown in Fig. 7. These results were expected from the initial analysis obtained by the GFA approach, as well as by molecular docking, which predicts the bound conformation of the ligand and assigns the highest scores to highly active conformers that are capable of establishing two hydrogen bonds at the extremity of their molecular structure with the residues of the active sites in both ER$_\alpha$ and ER$_\beta$ receptors.

On the other hand, when the activity of the ligands in the T47D cell proliferation assay was considered, the most informative probe was the TIP probe, which explores the shape of the

Table 4
3D-QSAR models calculated for ER$_\alpha$ and ER$_\beta$ conformers for both cell growth assay activities. The models are characterized by high predictive capacity expressed in terms of q^2.

Activity	Data set	Compound[a]	r^2	q^2 (LOO)
MCF-7	ER$_\alpha$ conformers	24	0.99	0.70
	ER$_\beta$ conformers	27	0.99	0.62
T47D	ER$_\alpha$ conformers	18	0.99	0.61
	ER$_\beta$ conformers	19	0.99	0.55

[a] Number of compounds in the model after removal of outliers, squared correlation coefficient of the regression (R^2), predictive leave-one-out (LOO) cross-validated squared correlation coefficient (q^2 or R^2_{xv}).

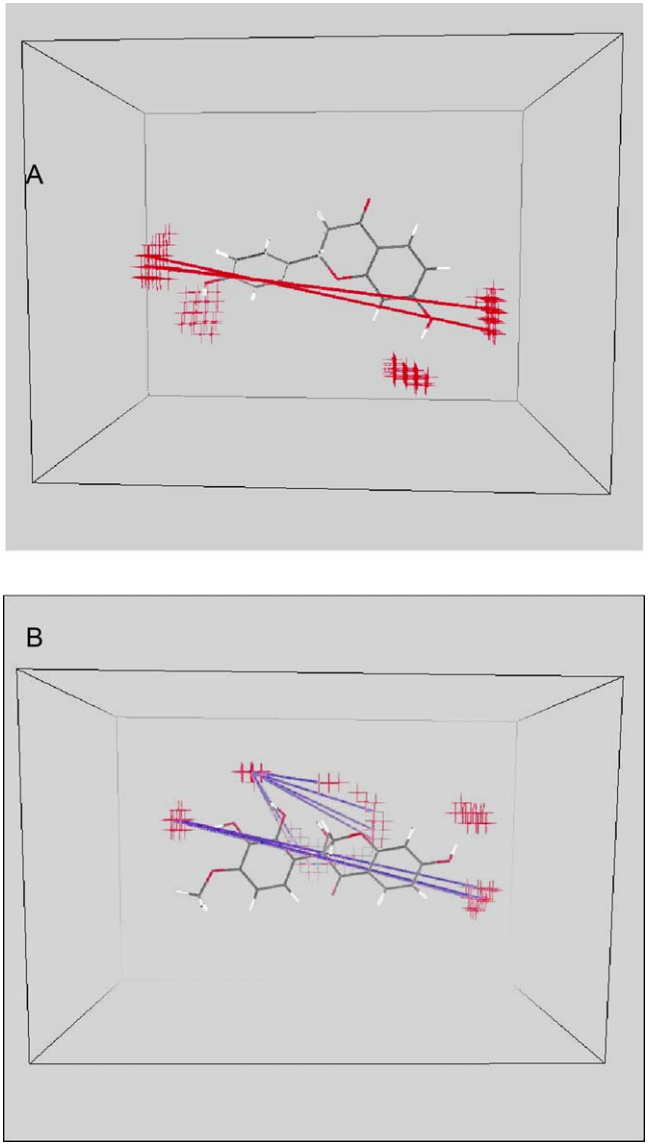

Fig. 7. GRIND molecular descriptors derived from probe O (carbonyl oxygen, an H bond acceptor) in the model describing MCF-7 activity. In the upper part of the figure, one of the highly active molecules is represented: the vectors connecting the hydrogen bond donors at the extremities of the compound are the most informative descriptors and are correlated to strong affinity for the receptor. In the lower part, an inactive molecule is represented: a much more complex pattern of hydrogen bonds donor areas is evident.

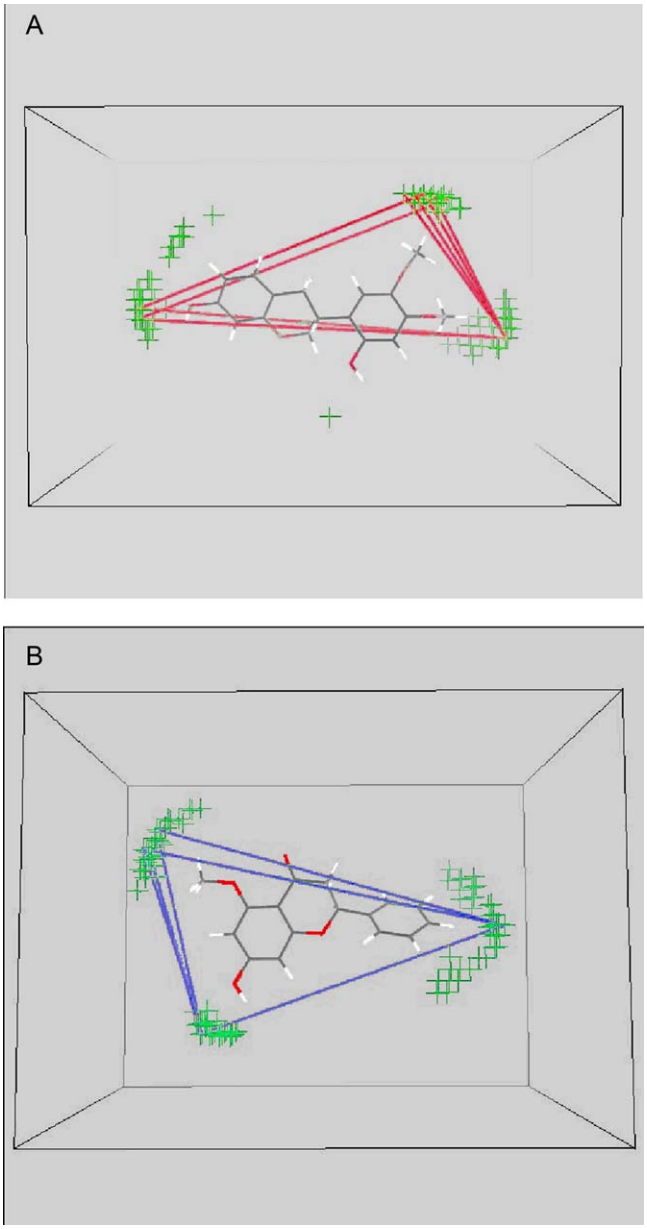

Fig. 8. GRIND molecular descriptors derived from the TIP probe (shape descriptor) in the model for the T47D activity. (A) In the upper part of the figure, one of the highly active molecules is represented. The preferred shape of the active compounds is coplanar. (B) The shape of the less active compounds shows a rotation of the phenyl functionality of approximately 30 degrees with respect to the benzopyrone component of the scaffold.

studied molecules. Therefore, we can draw the conclusion that substrates need to have particular structural requirements in order to exhibit high estrogen-like activity in the proliferation of the T47D breast cancer cell line. The preferred shape of the active compounds is coplanar, while the shape of the less active compounds shows a rotation of the phenyl functionality of approximately 30 degrees with respect to the benzopyrone core of the scaffold (Fig. 8).

3.4. GRID-PCA

The GRID-PCA computational approach combines a molecular mechanics technique for the description of protein–ligand interactions with multivariate statistical analysis. The latter makes it possible to distinguish the zones of structurally related proteins that exhibit different interactions, in terms of nature and strength, with a defined ligand. Therefore, the GRID-PCA approach was used to study the active sites of both ER_α and ER_β receptors to better understand the physicochemical environment in which the ligands are placed upon docking.

After the calculation of the GRID descriptors using a larger number of probes, a principal components analysis (PCA) of the resulting variables was performed. The PCA approach was specifically developed to reduce the dimensionality of statistical models and to maximize the information present in the variables. In our study, the active site areas of the two receptors were described by the interaction energies between the probes and their molecular structure (MIF), which in turn represents the variables of the system. The method was able to point out the major differences, in terms of physicochemical properties, between the binding sites of the two receptors. As can be seen in Fig. 9, the receptors differ mostly in terms of hydrogen bond donors and acceptor areas (interactions

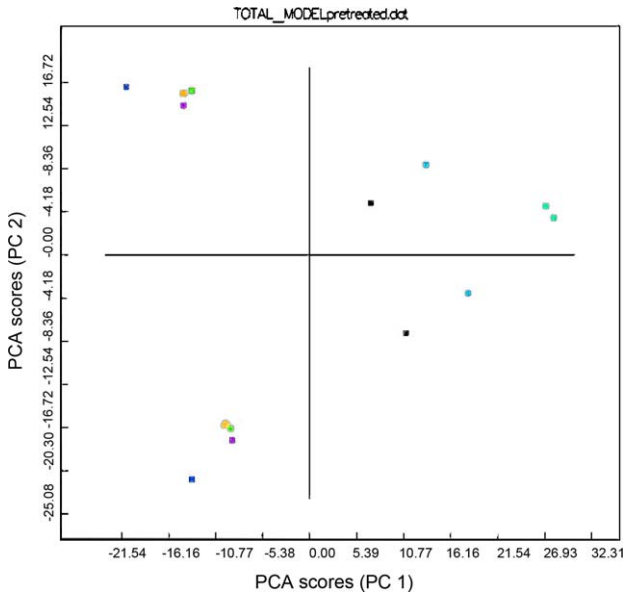

Fig. 9. PCA scores plot of the GRID-PCA-based comparison of the receptors. The squares in the plot represent the interaction of a specific probe with one of the receptors. The squares clustered together in the plot indicate similar interactions, whereas the distant ones indicate significant differences in terms of the receptors' physicochemical features. Probe color code: white – C3, magenta – OH, yellow – NH, blue – H_2O, orange – O, red – DRY, green – N1.

with N1, OH, NH, H_2O probes), though they share the same steric interactions pattern, which is not very informative.

The structural comparison of the two receptors by GRID-PCA confirmed the results derived from the previous study of modeling of the estrogen receptors binding. In fact, the molecular size and shape as well as the presence/absence of hydroxyl groups at positions 6 and 13, 14, 15 of the molecular scaffold play a dominant role in the activity of the ligands. This might represent the key features that play a dominant role in estrogenic potency of the studied compounds as well as a guide for a rational design of novel and possibly more effective SERMs.

4. Conclusions

Our computational studies of the natural estrogen-like isoflavonoids and diphenolics isolated from Thai medicinal plants aimed at gaining a mechanistic insight into molecular structure–estrogenic potency relationships with a perspective to rationally derivatize the most potent natural compounds and identify a potential lead candidate for drug development. The 2D and 3D QSAR models and GRID-PCA analysis of the ER_α and ER_β receptors revealed the importance of hydroxyl and methoxy groups substitutions at the C_6 and C_{14} positions (Fig. 2) and the unfavorable role of substitutions in the ortho and meta positions of the benzene ring of the isoflavonoid scaffold. Similar conclusions can be made about other classes of studied phytoestrogens, which also assume similar active conformations at ER receptor binding sites characterized by maximum overlap and alignment to the estradiol steroid scaffold – the native ligand of the estrogenic receptors.

The LigScore scoring function, which was used to describe the receptor affinities of simultaneous binding to the ER_α and ER_β led to the best correlation between the observed activities and computed $LigScoreER_\alpha$ and $LigScoreER_\beta$ descriptors. We can thus conclude that simultaneous and possibly competitive interaction of the compounds with the ER_α and ER_β receptors, in which the size and presence of hydroxyl groups at the abovementioned positions of the molecular scaffold play a dominant role, may determine the estrogenic potency of the considered phytochemicals.

Acknowledgements

We would like to thank the International Centre for Science and High Technology, UNIDO, Trieste, Italy and Thailand's Commission on Higher Education, Bangkok, for their joint financial support. W.D. would also like to thank the National Center for Genetic Engineering and Biotechnology (BIOTEC, grant no. 2/2545) for their partial support on the previous phytochemical work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.01.001.

References

[1] T. Cornwell, W. Cohick, I. Raskin, Dietary phytoestrogens and health, Phytochemistry 65 (2004) 995–1016.
[2] W. Wuttke, H. Jarry, S. Westphalen, V. Christoffel, D. Seidlová-Wuttke, Phytoestrogens for hormone replacement therapy? J. Steroid Biochem. Mol. Biol. 83 (2003) 133–147.
[3] S.O. Mueller, Overview of in vitro tools to assess the estrogenic and antiestrogenic activity of phytoestrogens, J. Chromatogr. B 77 (2002) 155–165.
[4] B.L. Riggs, L.C. Hartmann, Selective estrogen-receptor modulators-mechanisms of action and application to clinical practice, N. Engl. J. Med. 348 (2003) 618–629.
[5] K. Pettersson, J.A. Gustafsson, Role of estrogen receptor beta in estrogen action, Annu. Rev. Physiol. 63 (2001) 165–192.
[6] J.A. Gustafsson, What pharmacologists can learn from recent advances in estrogen signalling, Trends Pharmacol. Sci. 24 (2003) 479–485.
[7] S. Mosselman, J. Polman, R. Dijkema, ERβ: identification and characterization of a novel human estrogen receptor, FEBS Lett. 392 (1996) 49–53.
[8] G.G. Kuiper, B. Carlsson, K. Grandien, E. Enmark, J. Haggblad, S. Nilsson, J.A. Gustafsson, Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta, Endocrinology 138 (1997) 863–870.
[9] T. Barkhem, B. Carlsson, Y. Nilsson, E. Enmark, J.A. Gustafsson, S. Nilsson, Differential response of estrogen receptor α and estrogen receptor β to partial estrogen agonists/antagonists, Mol. Pharmacol. 54 (1998) 105–112.
[10] K. Pettersson, K. Grandien, G.G. Kuiper, J.A. Gustafsson, Mouse estrogen receptor β forms estrogen response element-binding heterodimers with estrogen receptor α, Mol. Endocrinol. 11 (1997) 1486–1496.
[11] S.M. Cowley, S. Hoare, S. Mosselman, M.G. Parker, Estrogen receptors α and β form heterodimers on DNA, J. Biol. Chem. 272 (1997) 19858–19862.
[12] S. Ogawa, S. Inoue, T. Watanabe, H. Hiroi, A. Orimo, T. Hosoi, Y. Ouchi, M. Muramatsu, Biochem. Biophys. Res. Commun. 272 (1998) 122–126.
[13] H. Hiroi, S. Inoue, T. Watanabe, W. Goto, A. Orimo, M. Momoeda, O. Tsutsumi, Y. Taketani, M. Muramatsu, Differential immunolocalization of estrogen receptor alpha and beta in rat ovary and uterus, J. Mol. Endocrinol. 22 (1999) 37–44.
[14] M. Sar, F. Welsch, Differential expression of estrogen receptor-p and estrogen receptor-a in the rat ovary, Endocrinology 140 (1999) 963–971.
[15] M. Sar, F. Welsch, Estrogen receptor alpha and beta in rat prostate and epididymis, Andrologia 32 (2000) 295–301.
[16] A.H. Taylor, F. Al-Azzawi, J. Mol. Immunolocalisation of oestrogen receptor beta in human tissues, Endocrinology 24 (2000) 145–155.
[17] V.C. Jordan, The strategic use of antiestrogens to control the development and growth of breast cancer, Cancer 70 (1992) 977–982.
[18] R.R. Love, R.B. Mazess, H.S. Barden, S. Epstein, P.A. Newcomb, V.C. Jordan, P.P. Carbone, D.L. DeMets, Effects of tamoxifen on bone mineral density in postmenopausal women with breast cancer, N. Engl. J. Med. 326 (1992) 852–856.
[19] R.P. Kedar, T.H. Bourne, T.J. Powles, W.P. Collins, S.E. Ashley, D.O. Cosgrove, S. Campbell, Effects of tamoxifen on uterus and ovaries of postmenopausal women in a randomised breast cancer prevention trial, Lancet 343 (1994) 1318–1321.
[20] L.J. Black, M. Sato, E.R. Rowley, D.E. Magee, A. Bekele, D.C. Williams, G.J. Cullinan, R. Bendele, F.R. Kaufman, W.R. Bensch, Raloxifene (LY139481 HCl) prevents bone loss and reduces serum cholesterol without causing uterine hypertrophy in ovariectomized rats, J. Clin. Invest. 93 (1994) 63–69.
[21] L.A. Paige, D.J. Christensen, H. Gron, J.D. Norris, E.B. Gottlin, K.M. Padilia, C.Y. Chang, L.M. Ballas, P.T. Hamilton, D.P. McDonnell, D.M. Fowlkes, Estrogen receptor (ER) modulators each induce distinct conformational changes in ER alpha and ER beta, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 3999–4004.
[22] K. Umehara, K. Nemoto, K. Kimijima, A. Matsushita, E. Terada, O. Monthakantirat, W. De-Eknamkul, T. Miyase, T. Warashina, M. Degawa, H. Noguchi, Estrogenic constituents of the heartwood of *Dalbergia parviflora*, Phytochemistry 69 (2008) 546–552.
[23] K. Umehara, K. Nemoto, A. Matsushita, E. Terada, O. Monthakantirat, W. De-Eknamkul, T. Miyase, T. Warashina, M. Degawa, H. Noguchi, Flavonoids from the heartwood of Thai medicinal plant *Dalbergia parviflora* and their effects on estrogenic responsive human breast cancer cells, J. Nat. Prod. 72 (2009) 2163–2168.

[24] O. Monthakantirat, W. De-Eknamkul, K. Umehara, Y. Yoshinaga, T. Miyase, T. Warashina, H. Noguchi, Phenolic constituents of the rhizomes of the Thai medicinal plant *Belamcanda chinensis* with proliferative activity for two breast cancer cell lines, J. Nat. Prod. 68 (2005) 361–364.
[25] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.
[26] B.H. Besler, J.K.M. Mertz, P.A. Kollman, Atomic charges derived from semiempirical methods, J. Comput. Chem. 11 (1990) 431.
[27] Cerius^2 Life Sciences, version 4.5, Accelrys, San Diego, CA, 2000.
[28] D. Rogers, A.J. Hopfinger, Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure–property relationships, J. Chem. Inf. Comput. Sci. 34 (1994) 854–866.
[29] V. Frecer, QSAR analysis of antimicrobial and haemolytic effects of cyclic cationic antimicrobial peptides derived from protegrin-1, Bioorg. Med. Chem. 14 (2006) 6065–6074.
[30] D.M. Tanenbaum, Y. Wang, S.P. Williams, P.B. Sigler, Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 5998–6003.
[31] E.S. Manas, Z.B. Xu, R.J. Unwalla, W.S. Somers, Understanding the selectivity of genistein for human estrogen receptor-beta using X-ray crystallography and computational methods, Structure 12 (2004) 2197–2207.
[32] K.P. Peters, J. Fauck, C. Frommel, The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria, J. Mol. Biol. 256 (1996) 201–213.
[33] J.R. Maple, M.J. Hwang, T.P. Stockfish, U. Dinur, M. Waldman, C.S. Ewing, A.T. Hagler, Derivation of class II force fields. I. methodology and quantum force field for the alkyl of functional group and alkane molecules, J. Comput. Chem. 15 (1994) 162–182.
[34] P. Willett, in: P.M. Dean (Ed.), Molecular Similarity in Drug Design, Chapman and Hall, Glasgow, 1994, pp. 110–137.
[35] M. Pastor, G. Cruciani, I. McLay, Grid-independent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors, J. Med. Chem. 43 (2000) 3233–3243.
[36] D. Riganelli, R. Valigi, G. Costantino, M. Baroni, S. Wold, Autocorrelation as a tool for a congruent description of molecules in 3D QSAR studies, Pharm. Pharmacol. Lett. 3 (1993) 5–8.
[37] Molecular operating environment version 2006.08, distributed by Chemical Computing Group, Montreal, Canada.
[38] G. Cruciani, S. Clementi, M. Pastor, GOLPE-guided region selection in 3D-QSAR, in: H. Kubinyi, G. Folkers, Y.C. Martin (Eds.), Drug Design. Recent Advances, KLUWER/ESCOM, Dordrecht, 1998, pp. 71–86.
[39] H. Fang, W. Tong, M.L. Shi, R. Blair, R. Perkins, W. Branham, B.S. Hass, Q. Xie, S.L. Dial, C.L. Moland, D.M. Sheehan, Structure–activity relationships for a large diverse set of natural synthetic, and environmental estrogens, Chem. Res. Toxicol. 14 (2001) 280–294.
[40] H.J. Böhm, The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure, J. Comput. Aided Mol. Des. 8 (1994) 243–256.
[41] R.D. Cramer III, D.E. Patterson, J.D. Brounce, Comparative molecular field analysis (COMFA). 1. Effect of shape on binding of steroid to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[42] P.J. Goodford, A computational procedure for determining energetically favorable binding site on biologically important macromolecules, J. Med. Chem. 28 (1985) 849–857.
[43] C. Zhao, K. Dahlman-Wright, J.A. Gustafsson, Estrogen receptor β: an overview and update, Nucl. Recept. Signal 6 (2008), doi:10.1621/nrs.06003.
[44] A.C. Pike, A.M. Brzozowski, R.E. Hubbard, T. Bonn, A.G. Thorsell, O. Engstrom, J. Ljunggren, J.A. Gustafsson, M. Carlquist, Structure of the ligand-binding domain of oestrogen receptor β in the presence of a partial agonist and a full antagonist, EMBO J. 18 (1999) 4608–4618.