Article

Identification of Potential Human Protein Targets for Soybean Isoflavones

*Wilson Maldonado-Rojas, [ORCID] *,a Juan Salinas-Torresa and Jesús Olivero-Verbela*

aEnvironmental and Computational Chemistry Group, School of Pharmaceutical Sciences, Zaragocilla Campus, University of Cartagena, 130014 Cartagena, Colombia

Soy isoflavones have been reported as endocrine disruptors due to their ability to modulate the activity of estrogen receptors (ERs) in mammals; however, its ability to modulate other metabolic pathways is not entirely clear, which makes it necessary to identify new pharmacological targets that interact with these compounds present in soybean. In this work, a virtual screening was executed to identify potential targets of nine soy isoflavones, employing human proteins target from PharmMapper. The best 25 fit scores were selected and prepared for AutoDock Vina docking protocols. The results suggest that equol, daidzein and biochanin A, have the potential to interact with targets such as phenylethanolamine *N*-methyltransferase, sex hormone-binding globulin and vitamin D3 receptor, respectively. The validations of docking protocols showed good pose reproducibility (root-mean-square deviation (RMSD) ranged 0.001-3.854 Å) and a modest correlation between binding affinities and agonist concentration, AC_{50} (correlation coefficient (R) = 0.643, $p < 0.001$). Protein interaction network revealed that predicted targets for soy isoflavones are involved in different pathways, including neurotransmission, metabolism, and cancer remarking the need of a better understanding of the effects of these compounds on human health.

Keywords: soybean, endocrine disruptor, isoflavones, PharmMapper, AutoDock Vina

Introduction

The endocrine disruptors (EDs) are substances or exogenous mixtures that alter the function of the endocrine system, and consequently cause adverse health effects in an organism or its progeny.[1-3] Those include a wide variety of anthropogenic and naturally occurring compounds that work through several mechanisms, from mimicking hormones to interaction with downstream signals.[4] Many EDs are commonly found in the human diet, including fruits, vegetables and beans, among others.[5,6]

Soybean is a common food for adults and infants that is particularly rich in phytoestrogens, in particular isoflavones.[7] These natural products are derivatives of heterocyclic phenols that exhibit similar structure to estrogens. As typical phenolic compounds, they are potent antioxidants.[8] Isoflavones include genistein, daidzein, equol, glycitein, biochanin A, among others.[9,10] Most of these compounds are known as selective estrogen receptor modulators (SERM), and modulator of critical pathways for growth and cell proliferation, an event that affects multiple organs.[11,12] The genistein inhibits the activity of protein tyrosine kinase (PTK).[13,14] *In vitro* assays have

shown that, although phytoestrogens activate transcription of genes dependent on both estrogen receptors, ERα and ERβ, usually have a selective binding affinity for ERβ.[15-18] It has been reported that phytoestrogens may reduce estrogen-dependent prostate cancer,[19] and related to the occurrence of cervical cancer, cardiotoxicity, polycystic ovarian syndrome, infertility, among others.[5,20,21]

In addition to binding to estrogen receptor, isoflavones may be targeting many other biochemical pathways. One way to discover those signaling routes is to use computational methods to find molecular targets. These include: the prediction based on ligand similarity with drugs or evaluated molecules,[22,23] network prediction linked to known proteins,[24] methods based on adverse effects of phenotypes generated by drugs, molecular docking, pharmacophore mapping, and comparison of binding site or fingerprints.[25]

Artificial intelligence methods have emerged as a powerful tool for identifying new targets in drug discovery. However, these methods have limitations when it comes to undeclared proteins as targets, since these methods are based on the principle that similar ligands exert similar biological effects, sometimes presenting erroneous predictions for targets with little information or noisy data available.[26] The identification of new protein targets

*e-mail: wmaldonador@unicartagena.edu.co

for small molecules remains a major challenge in drug discovery, as well as, the study of their adverse effects. In recent years, the tireless search for reliable computational strategies has been growing, since experimental predictions require great expenditure of resources and time.

Virtual screening has made possible the identification suitable biological targets for particular compounds, mainly when conclusive experimental data are not available. There are many web servers that have been used for this purpose, including SwissTargetPrediction,[27] idTarget,[28] INVDOCK,[29] TarFisDock,[30] among others. However, most do not generate the coordinate for protein-ligand binding site, and others base their search on molecular docking, but with a limited number of human target proteins. In this work we combined different criteria of computational chemistry. PharmMapper server[31] was selected as the first filter, which is based on the search for targets based on the receptor pharmacophore mapping, with 23236 proteins, covering 16159 pharmacophore models as druggable binding sites, and then combined with molecular docking and protein interaction network methods.

PharmMapper server is offered as a friendly tool whose purpose is finding potential targets for small molecules. PharmMapper fit score generates a nearly $7{,}000 \times 7{,}000$ score matrix. After this, when a new molecule is submitted, the fit score to each pharmacophore will be calculated first, and then every fit score of a specific pharmacophore will be compared to the fit score matrix to measure its score level among all the scores of the pharmacophore.[32] On the other hand, protein-ligand molecular docking is a computational method, which binds flexible molecules within a rigid or flexible representation of a receptor, allowing to obtain the best energetically-possible conformation and geometric orientation. The program AutoDock Vina[33] has been widely used for this purpose.[34,35]

In this study, combined computational approach PharmMapper server,[31] AutoDock Vina,[33] pharmacophore mapping and protein interaction network were used to perform the virtual screening to identify potential target proteins for soybean isoflavones, with subsequent validation of *in silico* methods. Additionally, it is discussed the interactions and possible role of these compounds on the different metabolic pathways for the identified targets.

Methodology

Ligands optimization

Nine isoflavones present in soybean were evaluated as input ligands to find their plausible targets: genistein,[36] daidzein, equol,[37] glycitein, biochanin A,[38] formononetin,[39] dihydrogenistein (DHG), dihydrodaidzein (DHD) and *O*-desmethylangolensin (O-DMA).[40]

Identification of target proteins for isoflavones with PharmMapper

A three-step approach was applied to identify target proteins for soybean isoflavones. The process includes the optimization of the ligands, virtual screening of new targets based on pharmacophores, and the calculation of the protein-ligand affinity by molecular docking. First, the ligand structures were optimized with the density functional theory (DFT) method at the Becke, 3-parameter, Lee-Yang-Parr (B3LYP)/6-31G (d',p') basis set. The calculations were performed using the Gaussian 09 program.[41] The resulting geometry was converted to Mol2 format with the Open Babel program.[42]

The optimized isoflavone structures were submitted to the PharmMapper server using the option "Human Protein Targets Only". This allows the identification of potential human protein targets for isoflavones using a pharmacophore mapping approach.[43] This server contains 23236 proteins (covering 16159 druggable, 459 of which are human protein targets), and 51431 ligandable pharmacophore models.[44,45] The selection of potential pharmacological targets was carried out using the fit score value, employing the best 25 human proteins targets displayed in the target list (5.4% of human protein targets). The verification of protein binding sites coordinates was carried out visually with Sybyl-X 2.0,[46] in order to ensure that the ligand remained in the cavities of the protein.

The molecular interaction between isoflavones and target proteins selected by PharmMapper was verified by molecular docking protocols with AutoDock Vina. Nine isoflavones optimized structures were docked on first 25 target proteins according to the fit score value using AutoDock Vina program.[33] 3D protein structures were downloaded from Protein Data Bank[47] and prepared with SYBYL-X 2.0 package.[46] The preparation process consisted of removing water molecules and other co-crystallized ligands, repairing and fixing of amides in amino acid side chains, then hydrogens were added to amino acids residues (protonation). Optimization using Powell conjugate gradient method; with dielectric constant value of 1.0, gradient convergence fixed to 0.001 kcal mol^{-1}, maximum number of iterations at 1000, and Kollman united/all-atoms force fields with AMBER charges.[48] The protein-ligand molecular docking was defined by establishing a cube with the dimensions $24 \times 24 \times 24$ Å, fixed at the geometric centre of each evaluated soy isoflavone, and grid point spacing of 1.0, covering the ligand-binding site provided

by PharmMapper. Molecular docking calculations for isoflavones included 20 number modes, an energy range of 1.5, and exhaustiveness equal to 25, and all docking run analyses were performed in triplicate; and their binding affinities (kcal mol^{-1}) reported as mean ± standard deviation.

Identification of interacting protein residues and validation of protocols

The identification of amino acids in target proteins that interact with the isoflavones was performed employing LigandScout 3.0.[49] This tool creates simplified pharmacophores models to detect the number and type of interaction between the ligand and the residue in the binding site of the protein. Images of each protein-ligand complex were displayed using PyMOL.[50]

Protein interaction network (PIN) of potential targets predicted for soy isoflavones

Protein interaction network analysis and pathway enrichment for the potential target of soy isoflavones were performed with STRING 11.0 web server,[51] in order to identify and illustrate the protein-protein interactions and neighbor genes. Names of predicted PharmMapper protein targets for soy isoflavones were searched in Uniprot database and gen codes were extracted. Pathway analyses were processed according to the Kyoto Encyclopaedia of Genes and Genomes (KEGG) and Reactome database.[52]

Validation of docking protocols performed using two computational approaches

The first computational approach, re-docking calculations, was performed with co-crystallized ligand for the hits proteins, in total 100 AutoDock Vina docking runs were performed for each evaluated complex. The RMSD (root-mean-square deviation of atomic positions) values were calculated with Sybyl-X 2.0[46] for the best AutoDock Vina predicted ligand pose, as well as value for experimentally reported pose and calculated RMSD values as histogram graph of co-crystallized ligands.

A second validation approach included a correlation analysis, which was performed in order to establish a relationship between the calculated binding affinity values and the biological data. In this work, only the hormone binding globulin (SHBG) bioassay from PubChem assay ID: 318680[53] was used for validation, because it contained a higher number of compounds evaluated, and unlike the assays recorded for the other targets predicted by PharmMapper, these have structural similarity with soy isoflavones. In addition, this benchmark data set of compounds has been used for validation of popular molecular field-based quantitative structure-activity relationship (QSAR) and molecular docking techniques.[54]

Biological data for a representative set of 60 active ligands of SHBG were extracted from PubChem database.[53] The biological information consisted in the displacement of [3*H*]-5-alpha-dihydrotestosterone from the human SHBG. Theoretical binding affinities for these active compounds on the SHBG (Protein Data Bank (PDB): 1LHO) were obtained using the molecular docking protocols previously described.

Results and Discussion

Virtual screening using PharmMapper and AutoDock Vina

An inverse virtual screening was carried out to predict potential targets of soy isoflavones (Figure 1). The results from PharmMapper, verified using AutoDock Vina are shown in Table 1. According to PharmMapper fit score, most protein targets for isoflavones are nuclear receptors, being the ER highly frequent as a target among the best 25 main hits (Tables S1-S9, Supplementary Information (SI) section). The calculated AutoDock Vina affinity values were considerably poor for those targets above number 25 according to the fit score.
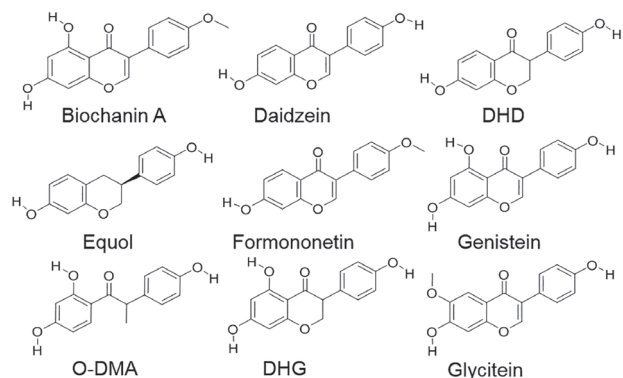

Figure 1. Two-dimensional structures of evaluated soybean isoflavones.

Soy-derived compounds also had good theoretical binding preferences for different enzymes such as transferases, reductases, kinases, phosphatases and dehydrogenases, including also globulin-related proteins. Selected proteins as better targets based on AutoDock Vina calculations were the phenylethanolamine *N*-methyltransferase (PNMT, PDB: 1YZ3),[55] sex hormone-binding globulin, (SHBG, PDB: 1LHO)[56] and vitamin D3 receptor (VDR, PDB: 1DB1)[57] for equol, daidzein and biochanin A, respectively.

Table 1. Results of virtual screening for main soybean isoflavones with PharmMapper server,[31] and docking calculations with AutoDock Vina program[33]

Compound	Protein		PharmMapper	AutoDock Vina[a] ± SD /
	Name	PDB code	Fit score	(kcal mol⁻¹)
Biochanin A	vitamin D3 receptor	1DB1	3.72	−10.1 ± 0.0
	aldose reductase	2DUX	3.91	−9.7 ± 0.1
	estradiol 17-beta dehydrogenase 1	1I5R	4.44	−9.1 ± 0.0
Daidzein	sex hormone binding globulin	1LHO	3.68	−10.4 ± 0.0
	estrogen receptor	1L2I	3.93	−9.4 ± 0.0
	estrogen-related receptor gamma	1S9P	3.89	−9.2 ± 0.1
Dihydrodaidzein	sex hormone-binding globulin	1LHO	3.93	−10.2 ± 0.0
	estrogen related receptor gamma	1S9P	3.81	−9.4 ± 0.0
	estrogen receptor beta	2Z4B	3.81	−9.4 ± 0.1
Dihydrogenistein	dihydroorotate dehydrogenase mitocondrial	1D3H	4.23	−9.6 ± 0.2
	sex hormone-binding globulin	1D2S	3.74	−9.1 ± 0.0
	estrogen receptor beta	1L2J	3.79	−9.1 ± 0.0
Equol	phenylethanolamine *N*-methyltransferase	1YZ3	3.51	−10.8 ± 0.0
	eukaryotic translation initiation factor 4E	1IPB	3.63	−9.3 ± 0.1
	transthyretin	1RLB	3.74	−9.3 ± 0.0
Formononetin	estradiol 17-beta-dehydrogenase 1	1I5R	4.41	−8.9 ± 0.0
	tyrosine-protein phosphatase non-receptor type 1	1Q1M	3.71	−8.9 ± 0.0
	transthyretin	1RLB	3.50	−8.9 ± 0.1
Genistein	vitamin D3 receptor	1DB1	3.76	−10.0 ± 0.0
	aldose reductase	2DUX	3.89	−9.5 ± 0.0
	cell division proteinkinase 2	1DI8	3.72	−9.4 ± 0.0
Glycitein	ribosyldihydronicotinamide dehydrogenase	1SG0	3.78	−10.0 ± 0.0
	serine/threonine-protein kinase Chk1	2AYP	3.83	−9.5 ± 0.0
	transthyretin	1RLB	3.69	−9.1 ± 0.0
O-Desmethylangolensin	vitamin D3 receptor	1DB1	3.77	−8.9 ± 0.1
	estrogen-related receptor gamma	1S9P	3.80	−8.8 ± 0.0
	estrogen-related receptor gamma	2P7G	3.78	−8.8 ± 0.0

[a]AutoDock Vina binding affinities are expressed as mean ± standard deviation (SD) for three docking. PDB: Protein Data Bank.

Interactions between protein residues and soy isoflavones on PharmMapper predicted hits

The greatest AutoDock Vina affinity value for PNMT was obtained for the isoflavone equol (−10.8 kcal mol⁻¹). The binding site on PNMT-equol complex is characterized by interactions between the ligand and residues Val-23, Ala-24, Tyr-27, Phe-30, Phe-102, Leu-103, His-160 and Val-187 (Figure 2), with mostly hydrogen bonds and aromatic interactions.

In the case of SHBG, daidzein produced the best result with a binding affinity value of −10.4 kcal mol⁻¹. The interacting residues identified were Ser-42, Thr-40, Asp-65 (hydrogen bond), Phe-56, Phe-67, Leu-80, Ile-141, Val-105, Met-139 and Leu-171 (hydrophobic interactions) (Figure 3).

Biochanin A was the best ligand for VDR, with a binding affinity value of −10.1 kcal mol⁻¹ (Table 1). The predicted residues with LigandScout program for the complex VDR-biochanin A established the presence of a hydrogen bond with Ser-237 and aromatic interactions with Trp-286 (Figure 4).

Isoflavones DHD and genistein also showed promising affinity values. The complex SHBG-DHD displayed an affinity binding value of −10.2 kcal mol⁻¹, and the binding site comprised Phe-56, Val-105, Ser-42, Thr-40, Leu-171, Phe-67, Leu-80, Met-139 and Asn-82 (Figure S1, SI section). The *in silico* binding of genistein on VDR resulted on an affinity of −10.0 kcal mol⁻¹, and interactions with Ser-237 and Trp-286 (Figure S2, SI section).

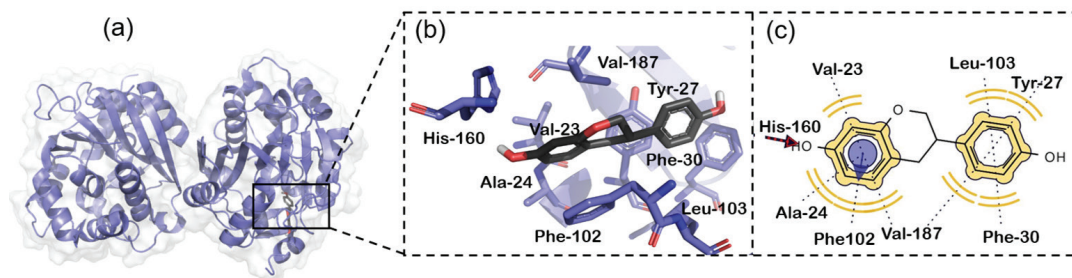Soy isoflavones are phenolic/non-steroidal bioactive compounds classified as phytoestrogens. The exposure to

Figure 2. (a) PNMT-equol complex. Interacting residues with equol are shown in (b) 3D (Pymol) and (c) 2D (LigandScout).
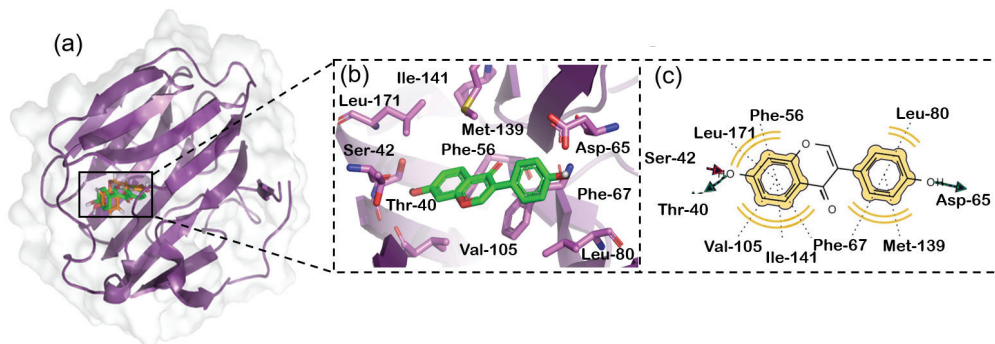

Figure 3. (a) SHBG-daidzein complex. Interacting residues with daidzein are shown in (b) 3D (Pymol) and (c) 2D (LigandScout).
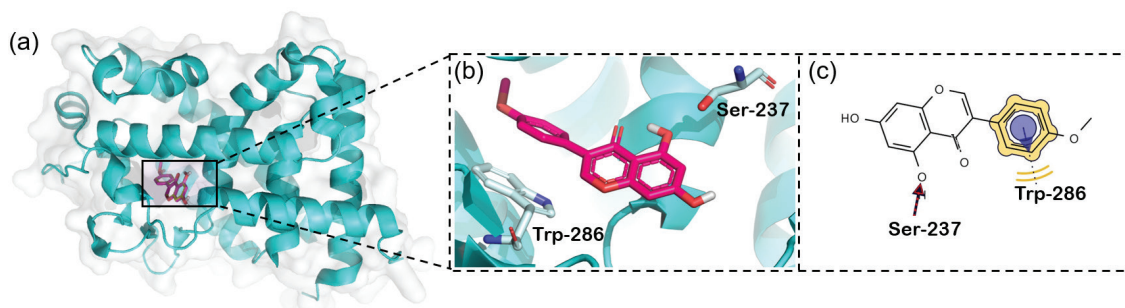

Figure 4. (a) VDR-biochanin A complex. Interacting residues with biochanin A are shown in (b) 3D (Pymol) and (c) 2D (LigandScout).

these chemical agents occurs mainly through the intake of soy and its derivatives, these are being consumed at varying rates according to the region and dietary preferences. Moreover, these phytoestrogens are also provided in food products for cattle.[43] Isoflavones present in soybeans have been widely described as phytoestrogens,[58-60] as a result of their ability to modulate hormonal function via ER,[19,61] making them well-known endocrine disruptors. However, additional pathways by which these chemical agents can exert various health effects are still unknown.

The results showed that soy-derived products could have health effects through their interaction with potential molecular targets. The best binding affinity for tested molecules was found for equol (−10.8 kcal mol⁻¹) on PNMT, an enzyme important on the catecholamine biosynthesis. This protein participates on the conversion of norepinephrine to epinephrine.[62] The latter is related to the sympathetic nervous system and adrenal medulla,

involved in many physiological functions, including blood pressure, vasoconstriction, cardiac stimulation and regulation of blood glucose level.[63] The protein-ligand interaction takes place near a binding site occupied by the cofactor S-adenosyl-L-homocysteine. In fact, equol also interacts with Phe-102 and Val-187 (Figure 2), residues that also bind to the purine ring in the cofactor.[55]

Daidzein, another recognized soy phytoestrogen, presented an affinity value of −10.4 kcal mol⁻¹ when binding on the sex hormone binding globulin. SHBG is a glycoprotein, biosynthesized mainly in the liver, which participates in the regulation of free estradiol and testosterone in plasma.[64,65] Moreover, it has been reported[66] that isoflavones or phytoestrogens increase the serum levels of SHBG. In addition, *in vitro* studies[14,18] have showed that isoflavones displace testosterone and 17β-estradiol from the binding sites of SHBG, potentially altering the bioavailability of free steroids and the androgen/estrogen

balance. Likely, as shown by the docking results, the potential interaction of daidzein with SHBG may result on hormone displacement from SHBG. The docking SHBG-daidzein complex displayed the same binding site reported for the native ligand, 5α-androstan-3β,17β-diol (1LHO).[56] This site includes Thr-40, Ser-42, Gly-58, Asp-65, Phe-67, Asn-82, Leu-131 and Met-107 (Figure 3). The DHD, a derivative of daidzein, presented an affinity value of –10.2 kcal mol^{-1}, occupying the same pocket in the protein, with small differences in protein-ligand interaction residues.

Molecular docking for biochanin A and genistein on VDR, a ligand-dependent transcription regulator, presented an affinity value of –10.1 and –10.0 kcal mol^{-1}, respectively. This receptor is involved in the transcription of a wide variety of genes like miRNAs, MIR181a, CYP19A1, among others. In addition, it controls the transcription of messenger ribonucleic acid (RNA), the regulation of the amount of translated proteins, the differentiation of hematopoietic cells and the production of aromatase, a key enzyme on the biosynthesis of estrogens.[67] The interaction VDR-biochanin A (Figure 4) involves residues Ser-237 and Trp-286, amino acids that have been reported[68,69] to have a critical role in the positioning of their natural ligand 1α,25-dihydroxyvitamin D3 (1,25 (OH)$_2$D$_3$) on VDR. Trp-286 is also important as it allows an adjustment of the small conjugated chain (C=C–C=C) that connects the rings A and C of 1,25 (OH)$_2$D$_3$, facilitating the processing of the ligand binding through a network of hydrogen bonds.[57,68,70]

Protein-protein interactions (PPI) for predicted soy isoflavones target proteins (Table 1) were evaluated again by using STRING 11.0[51] and these seed proteins were integrated into the total PIN with nodes 25, number of edges

was 70, average node degree was 5.6, average clustering coefficient of 0.676, expected number of edges = 24, and PPI enrichment p-value = 3.83 × 10^{-14}. Pathways of hub neighbors were obtained from GO (Gene Ontology) proteins involved in the same pathway except NRH: quinone oxidoreductase 2 (NQO2), dihydroorotate dehydrogenase (DHODH), transthyretin (TTR), estrogen-related receptor gamma (ESRRG) (Figure 5). However, TTR and ESRRG participate in hormonal regulation processes.[71]

The KEGG analysis revealed the participation of these proteins in different pathways including cyclin-dependent protein serine/threonine kinase activity, hormone binding, kinase binding, nuclear receptor activity, among others (Table S10, SI section). On the other hand, Reactome pathways analysis showed the participation of these proteins in critical routes including transcription of genes involved in G1 cell cycle arrest, G1/S deoxyribonucleic acid (DNA) damage checkpoints, generic transcription pathway, and diseases of signal transduction (Table S11, SI section). These findings suggest that isoflavones could modulate different signaling pathways, explaining some of the effects observed for the evaluated compounds.[72] However, some KEGG signaling pathways revealed the close relationship of these proteins in cancer-related mechanisms such as CDK2, CCEN1 and RB1, opening a door for the isoflavones present in soy to be used as a platform for the design of anticancer agents.[73]

Validation of molecular docking protocols

Two approaches were carried out to validate docking protocols. The first one included a superposition using the ligand in its X-ray complex and that predicted employing AutoDock Vina. For several complexes
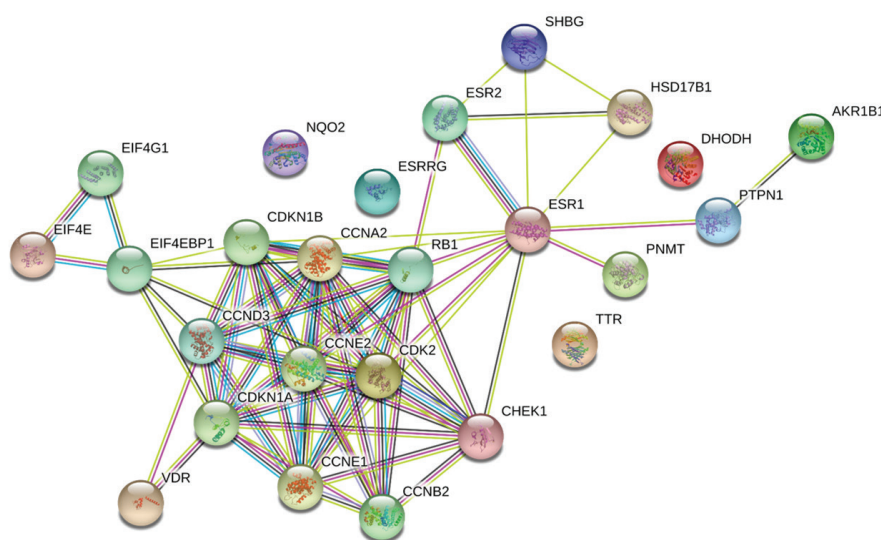

Figure 5. Protein interaction network analysis with STRING website 11.0,[51] showing results obtained upon entering a set of 27 theoretical targets proteins for soy isoflavones.

available in the PDB containing isoflavones as ligands or substrates, 100 re-docking runs were carried out using the co-crystallized ligands previously extracted from the protein complex (Table S12, SI section). The capability of the molecular docking protocols to reproduce the crystallographic pose of the ligands were examined by RMSD calculation (Å).[74] Calculated mean RMSD values ranged from 0.001 to 3.864 Å (Table 2). The maximum RMSD value (3.864 Å) was observed for the complex estradiol 17-beta-dehydrogenase 1/EM-1745 ($C_{37}H_{51}N_5O_7$), probably as a result of its bulky nature with a large number of rotatable bonds.[33]

Mean RMSD distributions are presented in Figure 6 for PDB crystallographic structures, showing the reproducibility of poses by the AutoDock Vina, where trends are also presented with RMSD values for pose

replication analyses, as well as some values far from the mean value. The histograms were drawn using PAST,[75] adjusted to eight class marks, with kernel density estimation (KDE), used to estimate the probability density function of a nonparametric random variable.[75]

Mean affinity values obtained for co-crystallized ligands in the analyzed proteins (VDR, PNMT and SHBG) showed similar magnitude to those presented for soy isoflavones, these affinity values ranged between –6.2 and 11.7 kcal mol⁻¹. However, for the PNMT protein, the affinity with the co-crystallized ligand was much lower (–6.2 kcal mol⁻¹) than with the soy isoflavone equol (–10.1 kcal mol⁻¹). It is important to clarify that in most cases the binding sites of the co-crystallized ligands were different from those predicted by PharmMaper server for the soy isoflavones.

Table 2. RMSD values for superposition of AutoDock Vina predicted pose and co-crystallized ligand pose for protein target of soy isoflavones (n = 100)

Protein	PDB code	Co-crystallized ligand	RMSD / Å	AV affinity / (kcal mol⁻¹)
Vitamin D3 receptor	1DB1	1,25 dihydroxy vitamin D3	1.359	–11.7 ± 0.7
Dihydroorotate dehydrogenase mitochondrial	1D3H	antiproliferative agent A771726	1.031	–7.6 ± 0.9
Sex hormone binding globulin	1LHO	5-alpha-androstane-3-beta,17beta-diol	0.001	–11.3 ± 0.0
Phenylethanolamine *N*-methyltransferase	1YZ3	7,8-dichloro-1,2,3,4-tetrahydroisoquinoline	0.001	–6.2 ± 0.1
Ribosyldihydronicotinamide dehydrogenase	1SG0	resveratrol	0.754	–9.2 ± 0.3
Estradiol 17-beta-dehydrogenase 1	1I5R	EM-1745	3.864	–11.6 ± 1.0

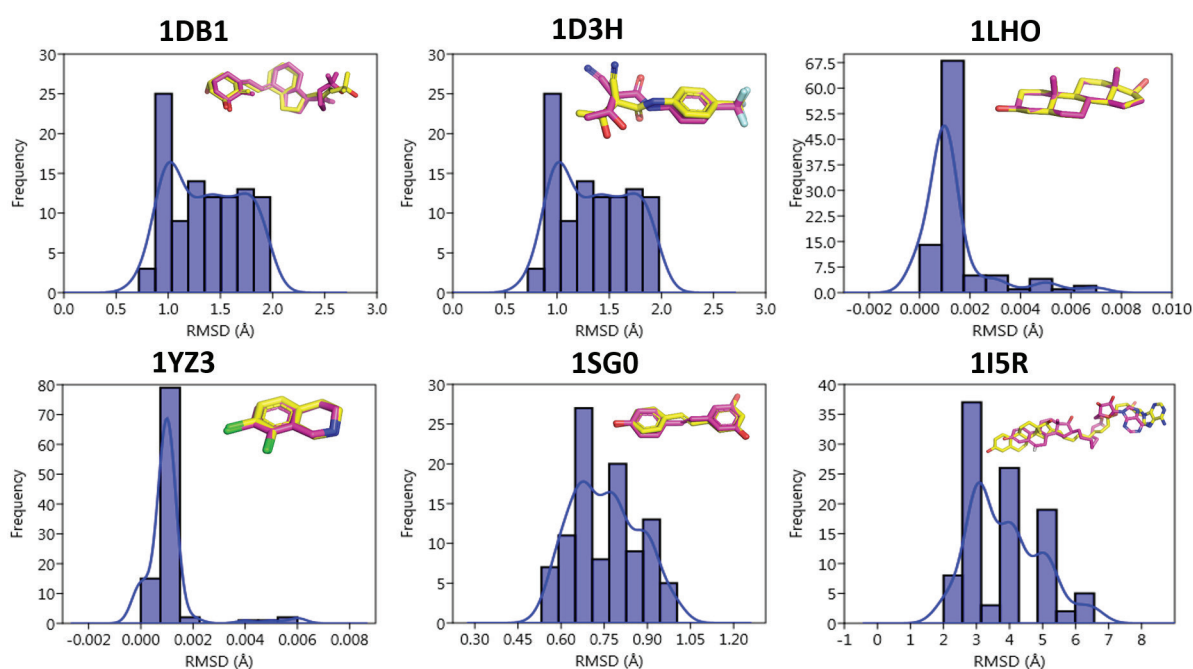PDB: Protein Data Bank; RMSD: root-mean-square deviation; AV: AutoDock Vina.

Figure 6. RMSD values distribution for 100 re-docking runs for crystallographic structures of vitamin D3 receptor (PDB: 1DB1), dihydroorotate dehydrogenase mitocondrial (PDB: 1D3H), sex hormone binding globulin (PDB: 1LHO), phenylethanolamine *N*-methyltransferase (PDB: 1YZ3), ribosyldihydronicotinamide dehydrogenase (PDB: 1SG0), and estradiol 17-beta-dehydrogenase 1 (PDB: 1IR5). For illustrative purposes AutoDock Vina predicted pose (purple) and co-crystallized ligand pose (yellow) are presented on each PDB structure.

A second validation approach consisted of a correlation between experimental and predicted biological data for SHBG agonists based on their binding affinities. Experimental agonist concentration 50 (AC_{50}) were recovered from PubChem BioAssay Database for 60 compounds. 3D structures were docked on receptor sex hormone-binding globulin structure, and PDB: 1LHO with AutoDock Vina program. The PubChem chemical structure identifier (CID), AC_{50}, AutoDock Vina affinity and logarithm of biological activity (log AC_{50}) for compound set evaluated are shown in the Supplementary Information (Table S13).

Linear correlation analysis showed the acceptable prediction capacity of protein-ligand interaction and biological activity of SHBG agonists with molecular docking protocols used, as shown in Figure 7. Linear Pearson correlation obtained for AutoDock Vina was 0.643 ($p < 0.001$), with a positive relation between binding affinity values (kcal mol^{-1}) and log AC_{50}. Details of statistical parameters for correlation analysis are described in Supplementary Information (Table S14).
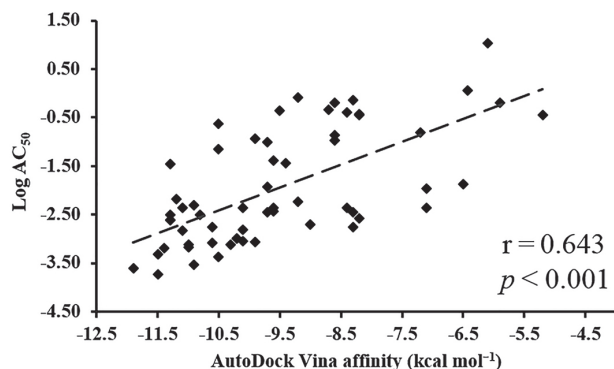

Figure 7. Linear correlation analysis between obtained AutoDock Vina affinities and their half maximal agonist molar concentration for sixty compounds experimentally evaluated (log AC_{50} assay ID: 318680). For illustrative purposes a regression dots line was added.

The internal validation of molecular docking protocols generated RMSD values, mostly lower than 2 Å (Figure 6), as it has been reported for other studies using AutoDock Vina;[76,77] supporting the reliability of this protein-ligand docking protocols. Moreover, external validation showed that binding affinity values from AutoDock Vina displayed a good correlation with biological data (Figure 7), suggesting the simulated protein-ligand dockings are likely to occur at the molecular level. The correlation coefficients obtained during the validation are similar to those generated in similar theoretical studies.[48,78,79]

The identification of potential targets for natural products using PharmMapper and AutoDock Vina, combining pharmacophore mapping, molecular docking and protein interaction network provides a theoretical strategy to detect molecular targets for small natural compounds. The results showed that soy isoflavones can interact with proteins present in different routes of hormonal metabolism, which agrees with the reported evidence that classifies these compounds as phytoestrogens. The advantages of these methodologies combine in a complementary way, including the identification of potential targets and the exploration of potential molecular mechanisms associated with bioactive natural compounds.[22,60,80]

Conclusions

Inverse virtual screening and molecular docking protocols were employed to discover theoretical human protein targets for isoflavones found in soybean such as equol, daidzein and biochanin A which have the potential to interact with proteins such as PNMT, SHBG, and VDR. These interactions were mostly of type aromatic, hydrophobic and hydrogen bonds; which is favored by the structure of isoflavones. Predicted macromolecules are present in critical pathways including neurotransmission, cell differentiation metabolism, cancer and homeostasis of estradiol/testosterone, among others. AutoDock Vina binding affinities for soy isoflavones were validated using *in silico* and experimental data. This information should promote additional pharmacological testing for isoflavones found in many food products, especially those recommended for children.

Supplementary Information

Supplementary information (fit score values and binding affinities for the best 25 protein targets predicted by PharmMapper, KEGG and Reactome pathways for soy isoflavones targets using STRING 11.0, data set for validation of docking protocols with experimental data, interactions for SHBG-DHD and VDR-genistein complexes) is available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

References

1. Schneider, M.; Pons, J. L.; Labesse, G.; Bourguet, W.; *Endocrinology* **2019**, *160*, 2709.

2. Hampl, R.; Kubátová, J.; Stárka, L.; *J. Steroid Biochem. Mol. Biol.* **2014**, *155*, 217.

3. Jung, H.; Hong, Y.; Lee, D.; Pang, K.; Kim, Y.; *Environ. Toxicol. Pharmacol.* **2013**, *35*, 278.

4. Tabb, M. M.; Blumberg, B.; *Mol. Endocrinol.* **2006**, *20*, 475.

5. Bowes, D. A.; Halden, R. U.; *Curr. Pathobiol. Rep.* **2019**, *7*, 41.

6. Harvey, P. A.; Leinwand, L. A.; *Mol. Cell. Endocrinol.* **2015**, *399*, 330.

7. He, F. J.; Chen, J. Q.; *Food Sci. Hum. Wellness* **2013**, *2*, 146.

8. Mazur, W.; Adlercreutz, H.; *Pure Appl. Chem.* **1998**, *70*, 1759.

9. Rietjens, I. M.; Louisse, J.; Beekmann, K.; *Br. J. Pharmacol.* **2017**, *174*, 1263.

10. Mariko, U.; *J. Clin. Biochem. Nutr.* **2013**, *52*, 202.

11. Lipovac, M.; Pfitscher, A.; Hobiger, S.; Laschitz, T.; Imhof, M.; Chedraui, P.; Jungbauer, A.; *Fitoterapia* **2015**, *105*, 93.

12. Zhou, T.; Meng, C.; He, P.; *Curr. Drug Metab.* **2019**, *20*, 46.

13. López, T.; López, S.; Arias, C. F.; *Virus Res.* **2015**, *210*, 141.

14. Patisaul, H. B.; Jefferson, W.; *Front. Neuroendocrinol.* **2010**, *31*, 400.

15. Moreira, A. C.; Silva, A. M.; Santos, M. S.; Sardão, V. A.; *J. Steroid Biochem. Mol. Biol.* **2014**, *143*, 61.

16. Pons, D. G.; Vilanova-Llompart, J.; Gaya-Bover, A.; Alorda-Clara, M.; Oliver, J.; Roca, P.; Sastre-Serra, J.; *Int. J. Food Sci. Nutr.* **2019**, *70*, 941.

17. Jiang, Y.; Gong, P.; Madak-Erdogan, Z.; Martin, T.; Jeyakumar, M.; Carlson, K.; Khan, I.; Smillie, T. J.; Chittiboyina, A. G.; Rotte, S. C. K.; Helferich, W. G.; Katzenellenbogen, J. A.; Katzenellenbogen, B. S.; *FASEB J.* **2013**, *27*, 4406.

18. Cederroth, C. R.; Zimmermann, C.; Nef, S.; *Mol. Cell. Endocrinol.* **2012**, *355*, 192.

19. Thelen, P.; Wuttke, W.; Seidlová-Wuttke, D.; *J. Steroid Biochem. Mol. Biol.* **2014**, *139*, 290.

20. Chen, H. H.; Chen, S. P.; Zheng, Q. L.; Nie, S. P.; Li, W. J.; Hu, X. J.; Xie, M. Y.; *J. Cancer* **2018**, *9*, 288.

21. Patisaul, H. B.; Mabrey, N.; Adewale, H. B.; Sullivan, A. W.; *Reprod. Toxicol.* **2014**, *49*, 209.

22. Soares, F. A.; Ceschi, M. A.; Franceschini, D. B.; do Canto, V. P.; Netz, P. A.; Campo, L. F.; *J. Braz. Chem. Soc.* **2019**, *30*, 2125.

23. Niijima, S.; Yabuuchi, H.; Okuno, Y.; *J. Chem. Inf. Model.* **2011**, *51*, 15.

24. Hopkins, A. L.; *Nat. Chem. Biol.* **2008**, *4*, 682.

25. Rognan, D.; *Mol. Inf.* **2010**, *29*, 176.

26. Rodrigues, T.; Bernardes, G. J.; *Curr. Opin. Chem. Biol.* **2020**, *56*, 16.

27. Daina, A.; Michielin, O.; Zoete, V.; *Nucleic Acids Res.* **2019**, *47*, 357.

28. Wang, J. C.; Chu, P. Y.; Chen, C. M.; Lin, J. H.; *Nucleic Acids Res.* **2012**, *40*, 393.

29. Chen, Y. Z.; Zhi, D. G.; *Proteins* **2001**, *43*, 217.

30. Li, H.; Gao, Z.; Kang, L.; Zhang, H.; Yang, K.; Yu, K.; Luo, X.; Zhu, W.; Chen, K.; Shen, J.; Wang, X.; Jiang, H.; *Nucleic Acids Res.* **2006**, *34*, 219.

31. *PharmMapper*; http://www.lilab-ecust.cn/pharmmapper/, accessed in November 2020.

32. Liu, X.; Ouyang, S.; Yu, B.; Liu, Y.; Huang, K.; Gong, J.; Zheng, S.; Li, Z.; Li, H.; Jiang, H.; *Nucleic Acids Res.* **2010**, *38*, 5.

33. Trott, O.; Olson, A. J.; *J. Comput. Chem.* **2010**, *31*, 455.

34. Maldonado-Rojas, W.; Olivero-Verbel, J.; Ortega-Zuñiga, C.; *J. Braz. Chem. Soc.* **2011**, *22*, 2250.

35. Maldonado-Rojas, W.; Noguera-Oviedo, K.; Olivero-Verbel, J.; *Rev. Acad. Colomb. Cienc. Exactas, Fís. Nat.* **2012**, *140*, 411.

36. Perez, M.; Otero, J. A.; Barrera, B.; Prieto, J. G.; Merino, G.; Alvarez, A. I.; *Vet. J.* **2013**, *196*, 203.

37. Sekikawa, A.; Ihara, M.; Lopez, O.; Kakuta, C.; Lopresti, B.; Higashiyama, A.; Aizenstein, H.; Chang, Y. F.; Mathis, C.; Miyamoto, Y.; Kuller, L.; *Curr. Cardiol. Rev.* **2019**, *15*, 114.

38. Cui, M.; Wu, D.; Bao, K.; Wen, Z.; Hao, Y.; Luo, L.; *Anal. Bioanal. Chem.* **2019**, *411*, 3091.

39. Hohenstein, J. D.; Studham, M. E.; Klein, A.; Kovinich, N.; Barry, K.; Lee, Y. J.; MacIntosh, G. C.; *Front. Plant Sci.* **2019**, *10*, 310.

40. Franke, A. A.; Lai, J. F.; Halm, B. M.; *Arch. Biochem. Biophys.* **2014**, *559*, 5.

41. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; *Gaussian 09, Revision A.02*; Gaussian Inc., Wallingford, CT, 2009.

42. Guha, R.; Howard, M. T.; Hutchison, G. R.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J.; Willighagen, E. L.; *J. Chem. Inf. Model.* **2006**, *46*, 991.

43. Bhuvaneshwari, S.; Sankaranarayanan, K.; *SAR QSAR Environ. Res.* **2019**, *30*, 81.

44. Wang, X.; Shen, Y.; Wang, S.; Li, S.; Zhang, W.; Liu, X.; Lai, L.; Pei, J.; Li, H.; *Nucleic Acids Res.* **2017**, *45*, W356.

45. Georgieva, M.; Zlatkov, B.; Zlatkov, A.; *World J. Pharm. Pharm. Sci.* **2014**, *3*, 94.

46. *Sybyl-X, Molecular Modeling Software Packages*, v. 2.0; Tripos Associates, Inc., St. Louis, MO, USA, 2012.

47. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E.; *Nucleic Acids Res.* **2000**, *28*, 235.

48. Maldonado-Rojas, W.; Olivero-Verbel, J.; Marrero-Ponce, Y.; *J. Mol. Graphics Modell.* **2015**, *60*, 43.

49. Wolber, T.; Langer, G.; *J. Chem. Inf. Model.* **2005**, *45*, 160.

50. DeLano, W. L.; *The PyMOL Molecular Graphics System*; Schrödinger, LLC, San Carlos, CA, USA, 2003.

51. *STRING*; https://string-db.org/, accessed in November 2020.

52. Safaei, A.; Tavirani, M. R.; Oskouei, A. A.; Azodi, M. Z.; Mohebbi, S. R.; Nikzamir, A. R.; *Gastroenterol. Hepatol. Bed Bench* **2016**, *9*, 114.

53. https://pubchem.ncbi.nlm.nih.gov/bioassay/318680, accessed in November 2020.

54. Cherkasov, A.; Ban, F.; Santos-Filho, O.; Thorsteinson, N.; Fallahi, M.; Hammond, G. L.; *J. Med. Chem.* **2008**, *51*, 2047.

55. Wu, Q.; Gee, C. L.; Lin, F.; Tyndall, J. D.; Martin, J. L.; Grunewald, G. L.; McLeish, M. J.; *J. Med. Chem.* **2005**, *48*, 7243.

56. Grishkovskaya, I.; Avvakumov, G. V.; Hammond, G. L.; Catalano, M. G.; Muller, Y. A.; *J. Biol. Chem.* **2002**, *277*, 32086.

57. Rochel, N.; Wurtz, J. M.; Mitschler, A.; Klaholz, B.; Moras, D.; Structurale, L. D. B.; Fries, L.; Illkirch, F.; *Mol. Cell* **2000**, *5*, 173.

58. Qin, H.; Lin, Z.; Vásquez, E.; Luan, X.; Guo, F.; Xu, L.; *Nutr. Res.* **2019**, *71*, 30.

59. Sarasquete, C.; Úbeda-Manzanaro, M.; Ortiz-Delgado, J. B.; *Histol. Histopathol.* **2019**, *34*, 875.

60. Moreira, E. D. T.; Pinto, L.; Gomes, A. A.; Goicoechea, H. C.; Araújo, M. C. U.; *J. Braz. Chem. Soc.* **2015**, *26*, 1573.

61. Cornwell, T.; Cohick, W.; Raskin, I.; *Phytochemistry* **2004**, *65*, 995.

62. Hou, Q. Q.; Wang, J. H.; Gao, J.; Liu, Y. J.; Liu, C. B.; *Biochim. Biophys. Acta, Proteins Proteomics* **2012**, *1824*, 533.

63. Álvarez-Diduk, R.; Galano, A.; *J. Phys. Chem. B* **2015**, *119*, 3479.

64. Rosner, W.; *Steroids* **2015**, *99*, 113.

65. Liao, C. H.; Li, H. Y.; Yu, H. J.; Chiang, H. S.; Lin, M. S.; Hua, C. H.; Ma, W. Y.; *Clin. Chim. Acta* **2012**, *413*, 803.

66. Adlercreutz, H.; Mousavi, Y.; Clark, J.; Höckerstedt, K.; Hämäläinen, E.; Wähälä, K.; Mäkelä, T.; Hase, T.; *J. Steroid Biochem. Mol. Biol.* **1992**, *41*, 8012.

67. Carlberg, C.; Campbell, M. J.; *Steroids* **2013**, *78*, 127.

68. Sicinska, W.; Rotkiewicz, P.; *J. Steroid Biochem. Mol. Biol.* **2009**, *113*, 253.

69. Eduardo-Canosa, S.; Fraga, R.; Sigüeiro, R.; Marco, M.; Rochel, N.; Moras, D.; Mouriño, A.; *J. Steroid Biochem. Mol. Biol.* **2010**, *121*, 7.

70. Sokolowska, K.; Carballa, D.; Seoane, S.; Pérez-Fernández, R.; Mouriño, A.; Sicinski, R. R.; *J. Org. Chem.* **2014**, *80*, 165.

71. Kurubanjerdjit, N.; *Gene Rep.* **2020**, 100656.

72. Campos, M. G.; *Handbook of Dietary Phytochemicals*; Xiao, J.; Sarker, S.; Asakawa, Y., eds.; Springer: Singapore, 2020, p. 1-38.

73. Ito, C.; Matsui, T.; Miyabe, K.; Hasan, C. M.; Rashid, M. A.; Tokuda, H.; Itoigawa, M.; *Phytochemistry* **2020**, *175*, 112376.

74. Baber, J. C.; Thompson, D. C.; Cross, J. B.; Humblet, C.; *J. Chem. Inf. Model.* **2009**, *49*, 1889.

75. Hammer, Ø.; Harper, D. A.; Ryan, P. D.; *Palaeontol. Electron.* **2001**, *4*, 9.

76. Olivero-Acosta, M.; Maldonado-Rojas, W.; Olivero-Verbel, J.; *Molecules* **2017**, *22*, 308.

77. Maldonado-Rojas, W.; Olivero-Verbel, J.; *J. Mol. Graphics Modell.* **2011**, *30*, 157.

78. Utsintong, M.; Talley, T. T.; Taylor, P. W.; Olson, A. J.; Vajragupta, O.; *J. Biomol. Screening* **2009**, *14*, 1109.

79. Mandal, S.; Moudgil, M.; Mandal, S. K.; *Eur. J. Pharmacol.* **2009**, *625*, 90.

80. Tang, Y.; Zhu, W.; Chen, K.; Jiang, H.; *Drug Discovery Today: Technol.* **2006**, *3*, 307.