ACS OMEGA

Open Access

This article is licensed under CC-BY-NC-ND 4.0

http://pubs.acs.org/journal/acsodf

Article

Estrogen Receptor Alpha Binders for Hormone-Dependent Forms of Breast Cancer: e-QSAR and Molecular Docking Supported by X-ray Resolved Structures

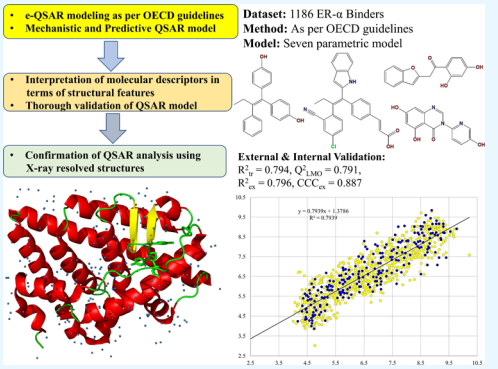
Vijay H. Masand,* Sami A. Al-Hussain, Abdullah Y. Alzahrani, Aamal A. Al-Mutairi, Rania A. Hussien, Abdul Samad, and Magdi E. A. Zaki*

Cite This: *ACS Omega* 2024, 9, 16759−16774

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Cancer, a life-disturbing and lethal disease with a high global impact, causes significant economic, social, and health challenges. Breast cancer refers to the abnormal growth of cells originating from breast tissues. Hormone-dependent forms of breast cancer, such as those influenced by estrogen, prompt the exploration of estrogen receptors as targets for potential therapeutic interventions. In this study, we conducted e-QSAR molecular docking and molecular dynamics analyses on a diverse set of inhibitors targeting estrogen receptor alpha (ER-α). The e-QSAR model is based on a genetic algorithm combined with multilinear regression analysis. The newly developed model possesses a balance between predictive accuracy and mechanistic insights adhering to the OECD guidelines. The e-QSAR model pointed out that sp^2-hybridized carbon and nitrogen atoms are important atoms governing binding profiles. In addition, a specific combination of H-bond donors and acceptors with carbon, nitrogen, and ring sulfur atoms also plays a crucial role. The results are supported by molecular docking, MD simulations, and X-ray-resolved structures. The novel results could be useful for future drug development for ER-α.



1. INTRODUCTION

Cancer is a major global cause of death, driving research to identify receptors for safer and more effective drug development.[1,2] Breast cancer is the most common cancer in women worldwide, with 2.3 million cases and 685,000 deaths in 2020.[3−5] It is often hormone-dependent,[6−10] particularly related to estrogen overproduction. Estrogen, a vital sex hormone, affects various physiological processes[6,11,12] through two main receptors, estrogen receptor alpha (ER-α) and estrogen receptor beta (ER-β).[6−8,10,12−15] ER-α is a key focus for breast cancer drug development due to various advantages.[9,13,16−18]

ER-α is the primary estrogen receptor in the breast, uterus, and ovaries, driving most estrogenic effects in these tissues.[6,8−10,13,17,19] It is also present in bones and fat,[20] regulating bone density and fat metabolism. Estradiol activates ER-α, influencing cell growth, differentiation, and apoptosis,[21−23] with implications for diseases like breast cancer[8,13,14,22] and osteoporosis.[24] ER-β is expressed in various tissues, including the brain, bones, and fat,[25] where it plays roles in regulating bone density,[26] blood pressure,[27] and inflammation.[28] These two receptors have distinct structures and functions (see Figure 1), with ER-α having a strong affinity for estradiol and other ligands such as prostaglandins,[29]

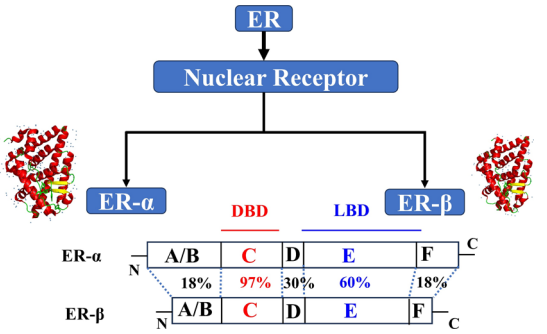
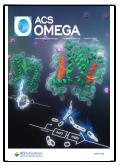


Figure 1. Brief description and comparison of estrogen receptors alpha and beta. The alphabets A−F represent the different domains of the enzyme. The percentage indicates the level of similarity.

thromboxanes,[30] and transcriptional activation,[15] while ER-β has a lower affinity and weaker activation function.[31]

Received: January 28, 2024
Revised: March 16, 2024
Accepted: March 19, 2024
Published: March 29, 2024



© 2024 The Authors. Published by American Chemical Society

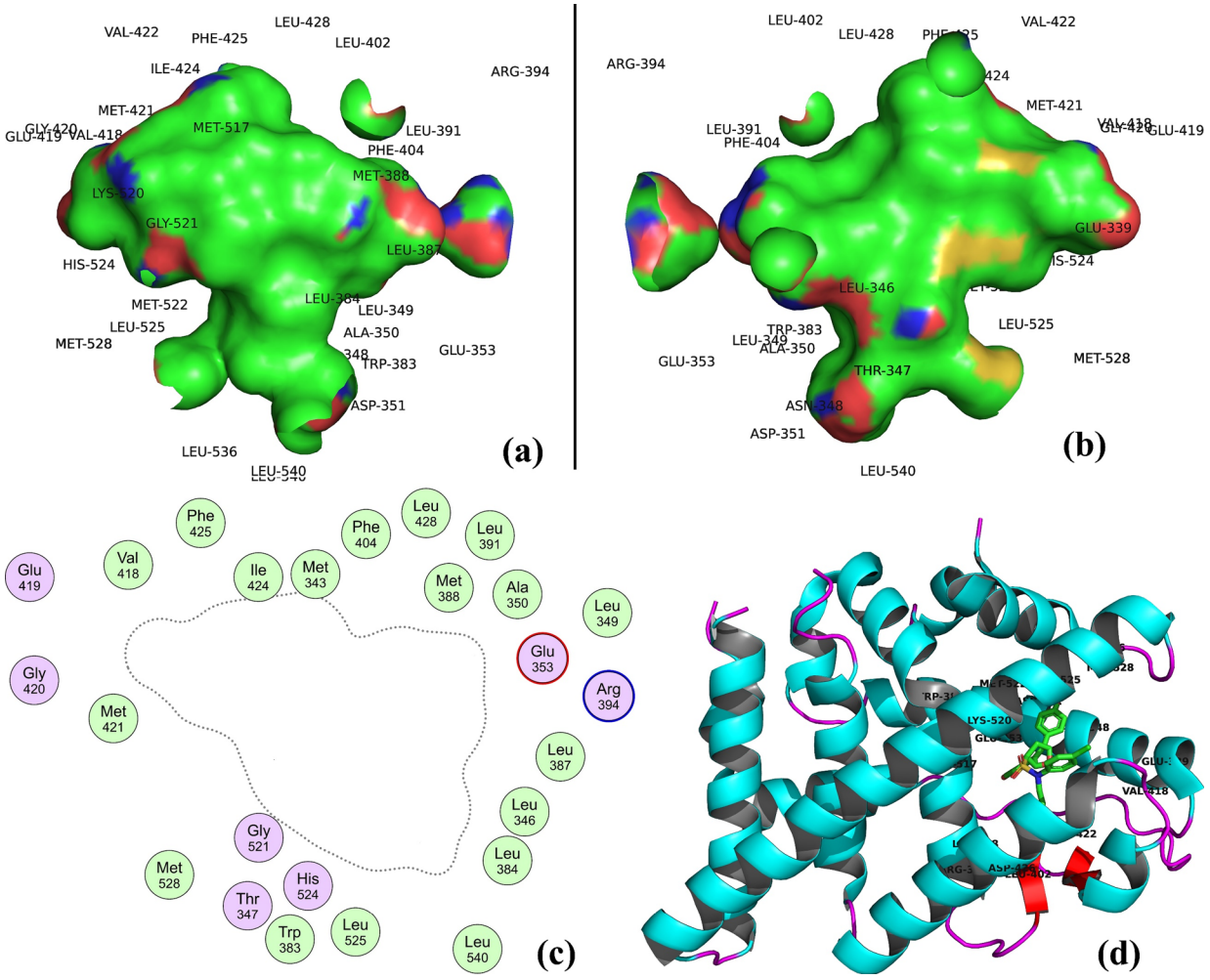


Figure 2. (a, b) 3D representation of the active site of ER-α from different angles with molecular surface area (green: lipophilic, red and blue: H-bond capable regions) along with residues. (c) 2D representation of the active site. The dots represent the proximity contour. (d) Structure of ER-α from pdb 5KCT.

In spite of their differences, the two types of ER have a shared mechanism of action. They bind to estradiol and other ligands, triggering a signaling pathway that activates the target genes. The particular effects of estrogen on a particular tissue or process depend on the presence of different ER isoforms and the specific genes that they regulate.

Dysregulation of the estrogen pathway is linked to hormone-related diseases, including breast cancer,[14,21] ovarian cancer,[32] and endometriosis.[33] In breast cancer, tumor development is tied to ER-α overexpression and ER-β absence.[14,21] Ovarian cancer prognosis worsens with excessive ER-α and ER-β expression.[32] Endometriosis, where uterine tissue grows outside the uterus, shows altered ER-α and ER-β expression within the affected tissue.[33] These expression changes may impact endometriosis progression.

Selective estrogen receptor modulators (SERMs) can selectively bind to either ER-α or ER-β,[34] modulating the estrogen pathway in specific tissues. They are valuable for treating hormone-related diseases without affecting unrelated processes. Tamoxifen, a commonly used SERM, treats breast cancer by inhibiting tumor growth in breast tissue.[35] Raloxifene,[36] another SERM, manages osteoporosis by enhancing bone density and reducing the risk of breast cancer and hot flashes through binding to ER-β.

ER-α, a transmembrane protein with 595 residues, consists of four domains:[37] the transactivation domain (TAD) at the amino-terminal, the hinge region, the ligand-binding domain (LBD), and the DNA-binding domain (DBD) at the carboxy-terminal. The LBD binds estradiol, while the TAD and DBD activate target genes at the transcriptional level. When estradiol binds to the LBD, ER-α is activated, causing a conformational change. This change leads to the TAD and DBD moving to the nucleus, activating target genes through transcriptional activation. The active site of ER-α is illustrated in Figure 2 (PDB: 5KCT[38]), offering a 3D representation from different angles with the molecular surface area and residues for insight into the active site's size, shape, and location.

Four types of transcriptional activities[10,11,18,39] are associated with ER-α: H12 agonist, stable H12 antagonist, mobile H12 antagonist, and highly mobile H12 antagonist are related to specific ligand-dependent coregulator complexes. Previous QSAR studies[11,16,18,40,41] focused on single transcriptional activity, used extensive descriptors, and often suffered from overfitting (see Table 1). To address these limitations, we conducted e-QSAR (easy, efficient, efficacious, and economical QSAR) and molecular docking analysis on a diverse and extensive data set of ER-α binders with dual inhibition mechanisms.[42] This study aims to identify common features

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

Table 1. Some Previous QSAR Studies Reported in the Literature for ER-α

s.n.	model method and end point	data set details	statistical parameters	reference
1	CoMFA, CoMSIA, support vector regression, and linear regression	train: 56	CoMFA: R^2 = 0.816, Q^2 = 0.519 CoMSIA: R^2 = 0.819, Q^2 = 0.511 linear: RMSE = 0.7484 SVR: RMSE = 0.7104	11
2	3D-QSAR model for relative binding affinity	train: 31	R^2 = 0.96, Q^2 = 0.93	16
3	3D pharmacophore for pIC_{50}	train: 26	R^2 = 0.951, Q^2 = 0.826	18
4	classification model VEGA − RBA	train: 148	accuracy: 0.84, specificity: 0.88	40
5	regression for pIC_{50} using 20 molecular descriptors	train: 1231	R^2 = 0.94, Q^2 = 0.73	41

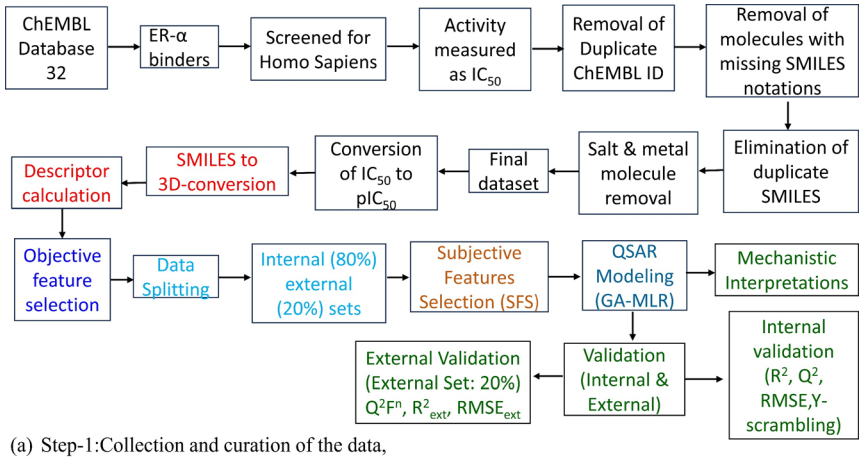
required for next-generation ER-α binders, particularly dual-mechanism estrogen receptor inhibitors (DMERIs).

2. METHODOLOGY

We sequentially performed the following steps to build a widely applicable and thriving QSAR model for ER-α binding activity:[43,44] (a) collection and curation of the data, (b) generation of 3D structures and calculation of molecular descriptors, (c) conducted objective feature selection (OFS), (d) split the data set into training (80%) and external validation (20%) sets, (e) conduction of subjective feature selection (SFS), (f) building a regression model, and (g) validation of the developed model (see Figure 3). Thus, in this work, we adhered to the OECD-recommended guidelines for deriving a QSAR model for ER-α binding activity.

2.1. Data Collection and Curation. The data set of ER-α binders, which was used to construct, train, and validate the QSAR model, was obtained from ChEMBL,[45] a publicly accessible database (https://ebi.ac.uk/chembl/ accessed on 31/05/2023). The downloaded data set consists of a diverse range of structurally distinct 1828 molecules that have undergone experimental testing to determine their activity toward ER-α. These molecules include estrogen receptor modulators (SERMs) and selective ER-α downregulators (SERDs), which act as mixed agonists/antagonists and full antagonists of ER-α, respectively. As part of the data curation process, molecules with uncertain IC_{50} values, duplicates, salts, metal-based inhibitors, and not tested for *Homo sapiens*, were excluded.[46] Eventually, the data set comprises a total of 1186 molecules, exhibiting a significant variety of structural scaffolds, all of which were experimentally tested for potency in terms of IC_{50} (nM). Additional information and the complete list of molecules can be found in the Supporting Information (see Figure SF1). The experimental IC_{50} values encompass a wide range, spanning from 0.06 to 97,982 nM. Subsequently, the IC_{50} values were transformed into their negative logarithmic form ($pIC_{50} = -\log_{10}IC_{50}$) to facilitate comparison. Table 1 and Figure 4 include a selection of the most and least active molecules, serving as illustrative examples only.

2.2. Calculation of Molecular Descriptors and Objective Feature Selection (OFS). The SMILES notations were then converted into 3D-optimized (MMFF94 force field, cutoff 0.01) structures using Openbabel 3.1[47] prior to the calculation of molecular descriptors. The success of a QSAR analysis greatly relies on accurately calculating diverse and a large number of molecular descriptors to enhance mechanistic interpretation while also reducing the risk of overfitting by eliminating noisy and redundant descriptors. To achieve these objectives, *PyDescriptor*[48] was employed to calculate over 30,000 molecular descriptors, encompassing both 1D- to 3D-descriptors. Subsequently, an objective feature selection (OFS) was conducted using QSARINS-2.2.4[49] to remove molecular descriptors that were nearly constant (90% molecules), constant, or highly intercorrelated ($|R| > 0.90$). The resulting set consists of 1706 molecular descriptors, which still offer a wide range of descriptors, ensuring comprehensive coverage of the descriptor space.



(a) Step-1:Collection and curation of the data,
(b) Step-2:Generation of 3D-structures and calculation of molecular descriptors,
(c) Step-3: Conducted objective feature selection (OFS),
(d) Step-4: Split the dataset into training (80%) and external validation (20%) sets,
(e) Step-5: Conduction of subjective feature selection (SFS),
(f) Step-6: Building a regression model, and
(g) Step-7: Validation and mechanistic interpretations of the developed model

Figure 3. Methodology used in the present work for robust and thriving QSAR analysis

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

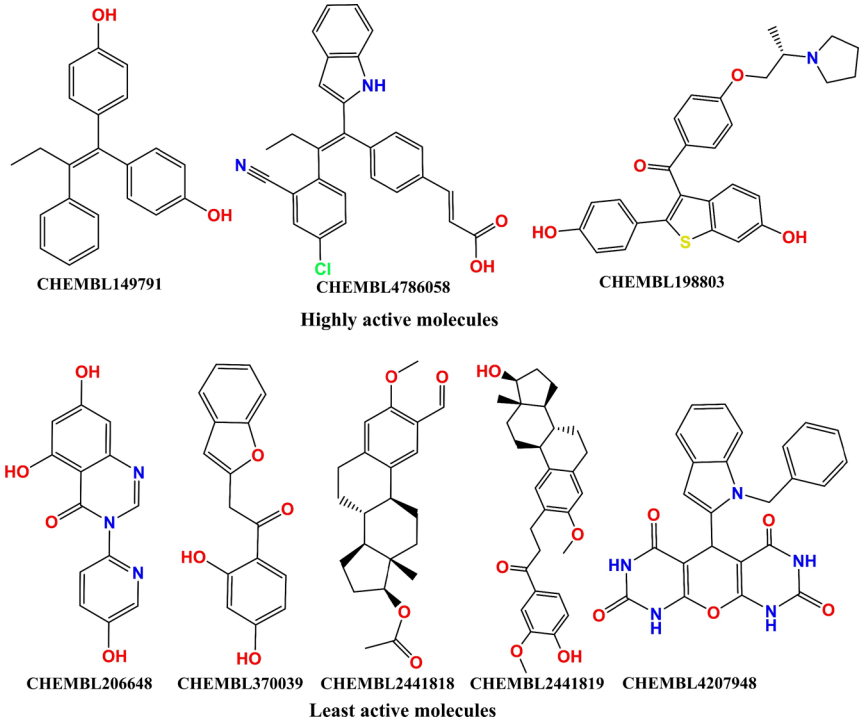


Figure 4. Selected high and low active ER-α binders from the present data set as representative examples only.

2.3. Splitting the Data Set into Training and External Sets and Subjective Feature Selection (SFS). Prior to conducting an exhaustive SFS, it is crucial to divide the data set into appropriate training and prediction sets (also known as external or test sets) to prevent any information leakage.[50] To ensure an unbiased approach, the data set was randomly split into training (80% = 949 molecules) and prediction/external (20% = 237 molecules) sets. The only use of the training set was to select an optimum number of molecular descriptors, while the prediction/external set was entirely set aside for the external validation of the model (Predictive QSAR). For SFS, the QSARINS-2.2.4 software employed the genetic algorithm unified with multilinear regression (GA-MLR) method, using Q^2_{LOO} as the fitness parameter, to identify relevant descriptors. It is crucial to strike a balance in developing a successful QSAR model that avoids overfitting while maintaining acceptable interpretability, which necessitates an appropriate number of molecular descriptors.

In this study, a graph (refer to Figure 5) was plotted to illustrate the relationship between the number of molecular descriptors included in the model and the corresponding Q^2_{LOO} values, aiming to determine the so-called breaking point.[43] The number of molecular descriptors corresponding to the breaking point was deemed to be optimal for model construction. Analysis of Figure 5 indicates that the breaking point corresponds to seven variables. Consequently, any QSAR models with more than seven descriptors were deemed unsuitable and rejected.

2.4. Building Regression Model and Its Validation. A well-validated QSAR model, which has undergone thorough assessment using various methods including cross-validation, external validation, Y-randomization, and applicability domain analysis (Williams plot), proves valuable for future applications such as virtual screening, molecular optimization, and decision-making.[51,52] The validation process incorporates several statistical parameters and their recommended threshold values

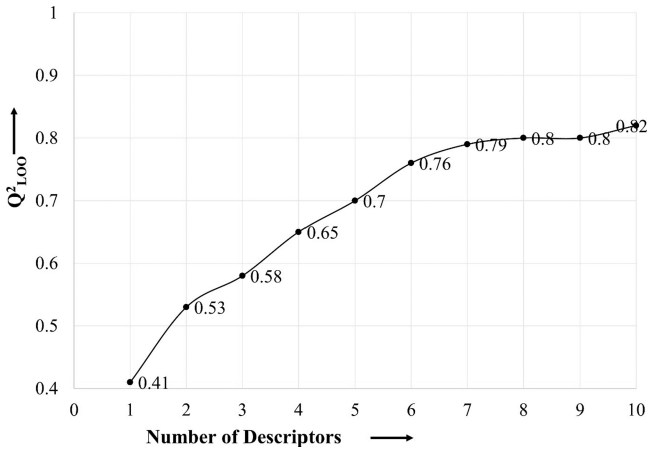


Figure 5. Identification of optimal numbers of descriptors (number of descriptors vs Q^2_{LOO}).

as standard criteria. These parameters are routinely employed to evaluate and validate the model. The following parameters and their corresponding thresholds are typically considered:[49,52,53]

$R^2_{tr} \geq 0.6$, $Q^2_{LOO} \geq 0.5$, $Q^2_{LMO} \geq 0.6$, $R^2 > Q^2$, $R^2_{ex} \geq 0.6$, $RMSE_{tr} < RMSE_{cv}$, $\Delta K \geq 0.05$, CCC ≥ 0.80, $Q^2 - F^n \geq 0.60$, $r^2_m \geq 0.5$, $(1 - r^2/r_o^2) < 0.1$, $0.9 \leq k \leq 1.1$ or $(1 - r^2/r'_o{}^2) < 0.1$, $0.9 \leq k' \leq 1.1$, $| r_o^2 - r'_o{}^2| < 0.3$, a low value of $RMSE_{ex}$, and MAE_{ex} along with high value of R^2_{ex}.

The formulas for calculating these statistical parameters can be found in the Supporting Information. Additionally, a Williams plot was generated to assess the applicability domain of the QSAR model.

2.5. Molecular Docking and 2D-Interaction Diagrams. For molecular docking, DockingApp, a freeware that uses Autodock-Vina as the docking engine, was used with default settings (exhaustiveness = 24). The X-ray resolved structure of

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

ER-α (PDB 7UJO) was downloaded from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (www.rcsb.org), due to a good resolution of 1.45 Å and a sequence length of 263. Then, protein preparation was carried out using AutoDock4, followed by the determination of the active site. The native ligand and literature were used to identify the active site of ER-α (see Figure 2). The grid center coordinates of ER-α were 16.626, − 26.928, and −72.026 along the *X*, *Y*, and *Z* axis, respectively, and grid sizes were 208, 20, and 25 along *X*, *Y*, and *Z* axis, respectively, with grid spacing 0.375. After that virtual screening was accomplished by using DockingApp. The docking poses with the highest docking score and protein in the pdbqt format were imported in PyMOL 2.5 for visual inspection of binding poses, followed by the export of the protein−ligand complex in PDB. The 2D-interaction between different ligands and receptors has been generated using Biovia Discovery Studio 2020.

2.6. Molecular Dynamic and Simulation Study. ER-α docking complexes were studied by doing molecular dynamics (MD) simulations using the Desmond 2020.1 program developed by Schrodinger, LLC. The focus was on investigating the interaction between ER-α and the most potent ligand CHEMBL149791. To ensure the reproducibility of the findings, samples were obtained using the same conditions for each measurement and data-collecting iteration.[54] The OPLS-2005 force field was used with an explicit solvent model including SPC water molecules across the whole system.[54] The solvated system was minimized by the use of the OPLS-2005 force field, followed by a relaxation process.[55] The Desmond 2020.1 software was used to establish the initial parameters inside the explicit SPC water model orthorhombic box, which had dimensions of 7.0 × 7.0 × 7.0, with the aim of accomplishing this objective. Sodium ions (Na+) were introduced into the system to establish physiological equilibrium, whereas 0.15 M sodium chloride (NaCl) solutions were used to mimic the conditions seen in humans. The system underwent an initial equilibration phase lasting 50 ns, during which an NVT ensemble was used to stabilize the protein-olmesartan combination. After the previous phase, a 12 ns equilibration and minimization process were conducted utilizing an NPT ensemble. The NPT ensemble was established using the Nose-Hoover chain coupling technique, ensuring the temperature was kept constant at 37 °C. The simulations were conducted at a constant pressure of 1 bar, with a relaxation period of 1.0 ps. The duration used was 2 fs.[56] The Martyna-Tuckerman-Klein chain coupling technique was used to regulate pressure, with a barostat technology with a relaxation period of 2 ps.[57] We used the particle mesh Ewald technique to calculate the long-range electrostatic interactions. The radius for Coulomb interactions was maintained at 9 Å.[58] The cohesive forces were determined for each trajectory using a RESPA integrator with a time step of 2 fs. Multiple metrics were used to assess the stability of the molecular dynamics (MD) simulations. The mentioned parameters are the root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), and the amount of hydrogen bonds (H-bonds).[59]

3. RESULTS

In the present work, comprehensive e-QSAR analysis is performed to identify significant pharmacophoric features associated with binding with ER-α. The robust QSAR model is built using easily interpretable molecular descriptors to correlate them with structural features required for effective ER-α binding. The seven-parametric GA-MLR model has good external predictive ability with the presence of easily understandable molecular descriptors along with interpretation in terms of structural features. A good number of model-fitting, y-scrambling, cross-validation, and external validation parameters associated with the newly developed QSAR model were calculated. The model-A validation parameters are presented in Table 2.

Table 2. Model-A Validation Parameters

internal validation parameter	value	cross-validation parameter	value	external validation parameter	value
R^2_{tr}	0.797	R^2_{cv} (Q^2_{LOO})	0.793	$RMSE_{ex}$	0.705
$R^2_{adj.}$	0.795	$RMSE_{cv}$	0.697	MAE_{ex}	0.559
LOF	0.492	MAE_{cv}	0.56	$PRESS_{ext}$	117.122
Kxx	0.237	$PRESS_{cv}$	461.061	R^2_{ex}	0.786
Δ*K*	0.067	CCC_{cv}	0.885	Q^2-F^1	0.785
$RMSE_{tr}$	0.691	Q^2_{LMO}	0.792	Q^2-F^2	0.783
MAE_{tr}	0.555	R^2_{Yscr}	0.007	Q^2-F^3	0.788
RSS_{tr}	453.27	Q^2_{Yscr}	−0.01	CCC_{ex}	0.882
CCC_{tr}	0.887			R^2-ExPy	0.786
s	0.694			$r'o^2$	0.741
F	526.951			*K*′	1

Model-A: pIC_{50} = 3.393 (±0.214) + 0.134 (±0.01) × nLipo + −1.158 (±0.097) × fsp3Cdon1B + 0.124 (±0.01) × da_ringC_5B −0.552 (±0.069) × sp2O_sp3O_9B −0.069 (±0.01) × sp2C_sp2N_5B + 0.635 (±0.12) × fsp2Csp3N4B −3.007 (±0.559) × ringS_don_6Ac

The formulas and brief description of above validation parameters are available in the Supporting Information. The details and description of molecular descriptors are presented in Table S1 in the Supporting Information.

4. DISCUSSION

As per the OECD guidelines,[60] which emphasize to develop QSAR models for regulatory use should possess (1) a defined end point, (2) an unambiguous algorithm, (3) a defined domain of applicability, (4) appropriate measures of goodness-of-fit, robustness, and predictivity, (5) a mechanistic interpretation, if possible. Thus, expanding knowledge on the mechanistic aspects involves pharmacophore or structure-oriented linking of molecular descriptors along with properly validating the correlation between significant molecular features (molecular descriptors) and desired bioactivity. In this analysis, we compared the pIC_{50} values of different molecules to investigate the impact of a specific molecular descriptor. However, it is important to consider that other molecular descriptors, structural features, or unknown factors may also exert analogous or opposite effects, significantly influencing the overall pIC_{50} value of a molecule for ER-α. In essence, a single molecular descriptor alone cannot fully explain the experimental pIC_{50} value for a diverse range of molecules for ER-α. Therefore, the successful use of the developed QSAR model relies on simultaneously incorporating multiple constituent molecular descriptors.

Of the seven descriptors in the QSAR model, there are four descriptors that are based on different types of carbon atoms. This indicates that carbon, in turn, the lipophilicity of the molecule is a significant factor governing the ER-α binding profile. The descriptor nLipo also vindicates this observation.

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

Therefore, increasing the overall lipophilicity of a molecule helps to increase the level of binding of a molecule with ER-α.

4.1. nLipo. The positive correlation of molecular descriptor nLipo with pIC_{50} points out that increasing lipophilic atoms could lead to better binding to ER-α. The positive coefficient of nLipo in the newly developed QSAR model further reinforces this observation. This also raises curiosity for clogP as a molecular descriptor. Replacing nLipo with clogP in the QSAR model led to a model with significantly lower statistical performance (R^2_{tr} = 0.72). The descriptor nLipo has a better association with binding affinity than clog *P*. A comparison of molecule CHEMBL4228649 with CHEMBL4224880 reinforces this observation (Figure 6).

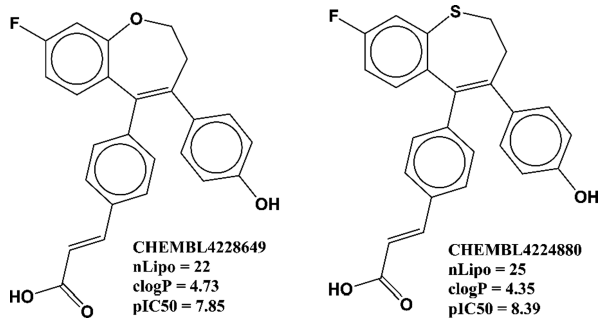


Figure 6. Comparison of nLipo and clogP on ER-α binding affinity.

Therefore, nLipo appears as a better choice. Accordingly, the present work successfully revealed a new observation that instead of molecular lipophilicity (clog *P*) a better parameter to enhance ER-a binding affinity is the number of lipophilic atoms.

The importance of nLipo is further confirmed and supported by the fact that the active site of ER-α is significantly lipophilic (see Figure 2). The lipophilic nature of the active site has been reported by a good number of researchers.[9,12,61,62] In the present data set, highly active molecules possessing pIC_{50} from 10.22 to 8 have nLipo in the range of 11−47. This indicates that increasing the number of lipophilic atoms is a good strategy to have higher ER-α binding. The direct influence of nLipo is further confirmed when the following pairs of molecules are compared: CHEMBL4169456 with CHEMBL4164944, CHEMBL4225599 with CHEMBL4228388 (see Figure 7), CHEMBL4170073 with CHEMBL4160682, CHEMBL362020 with CHEMBL178665 and CHEMBL180707, CHEMBL363412 with CHEMBL196836, CHEMBL196131 with CHEMBL370282, etc.

The importance of nLipo is strengthened and further confirmed on comparing the X-ray resolved structure of CHEMBL3775378 (PDB: 5fqp[63] and IC_{50} = 19 nM) with CHEMBL3775824 (PDB: 5fqr[63] and IC_{50} = 26 nM). The former has an additional methyl group, which leads to a higher number of lipophilic atoms and is also responsible for lipophilic interactions with Leu428 and Phe404 (see Figure 8), which are absent in the case of CHEMBL3775824. This difference could be accountable for the difference in their IC_{50} profile.

4.2. da_ringC_5B. The descriptor highlights the importance of ring carbon atoms and their relation with H-bond donor and acceptor atoms in constituting the pharmacophoric pattern. The positive coefficient for da_ringC_5B in the developed model signifies that a higher number of ring carbon atoms close to H-bond capable groups is a good strategy to increase pIC_{50} for a molecule. This in turn again emphasizes the role of the lipophilic region, as carbons are considered to be lipophilic. However, da_ringC_5B further clarifies that the balance of lipophilic (ring carbons) and H-bond donor and acceptor (polar groups) requires relatively a high proportion of ring carbon atoms with respect to H-bond donor and acceptor groups for better binding. From this discussion, the impression is that the number of ring carbon atoms (ringC) alone is significant. However, replacing da_ringC_5B with ringC as a descriptor in the QSAR model significantly reduced the statistical performance from R^2_{tr} = 0.79 to 0.68. Therefore, a combination of ring carbons with H-bond donor and acceptor groups is necessary. The following pairs of molecules further support the influence of da_ringC_5B on the activity profile: CHEMBL101807 with CHEMBL100617 (see Figure 9),

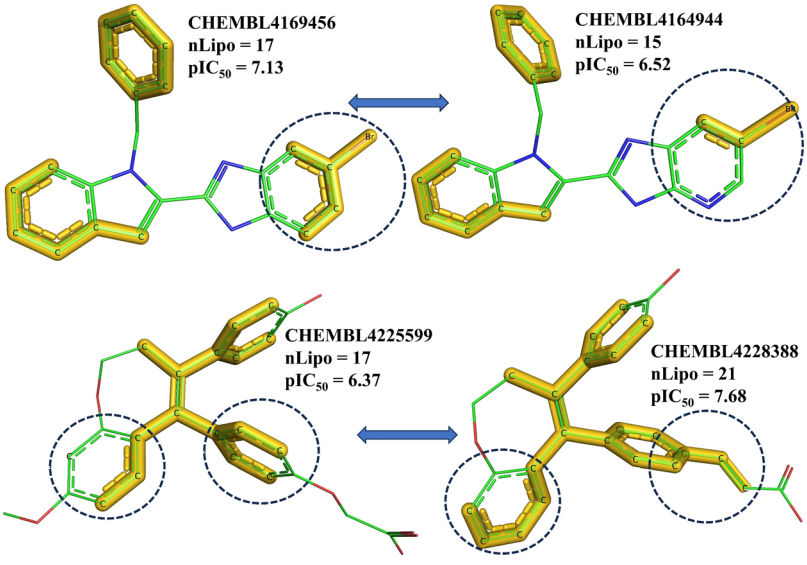


Figure 7. Comparison of ER-α binding of selected molecules with respect to nLipo as representative examples (lipophilic atoms are highlighted by gold-colored sticks).

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

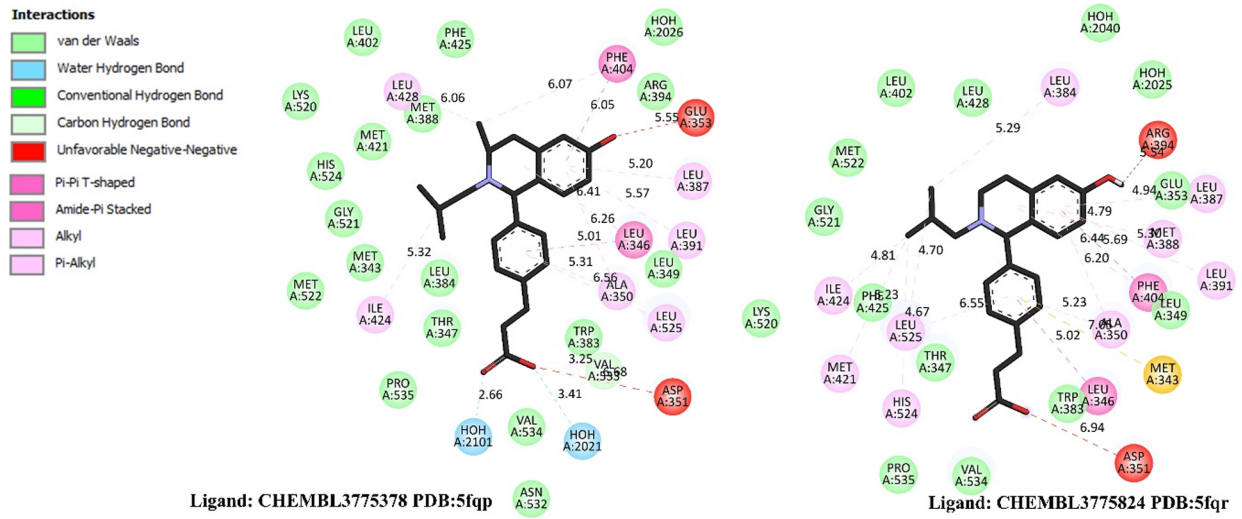


Figure 8. 2D- interaction diagram for CHEMBL3775378 (PDB: 5fqp) and CHEMBL3775824 (PDB: 5fqr).

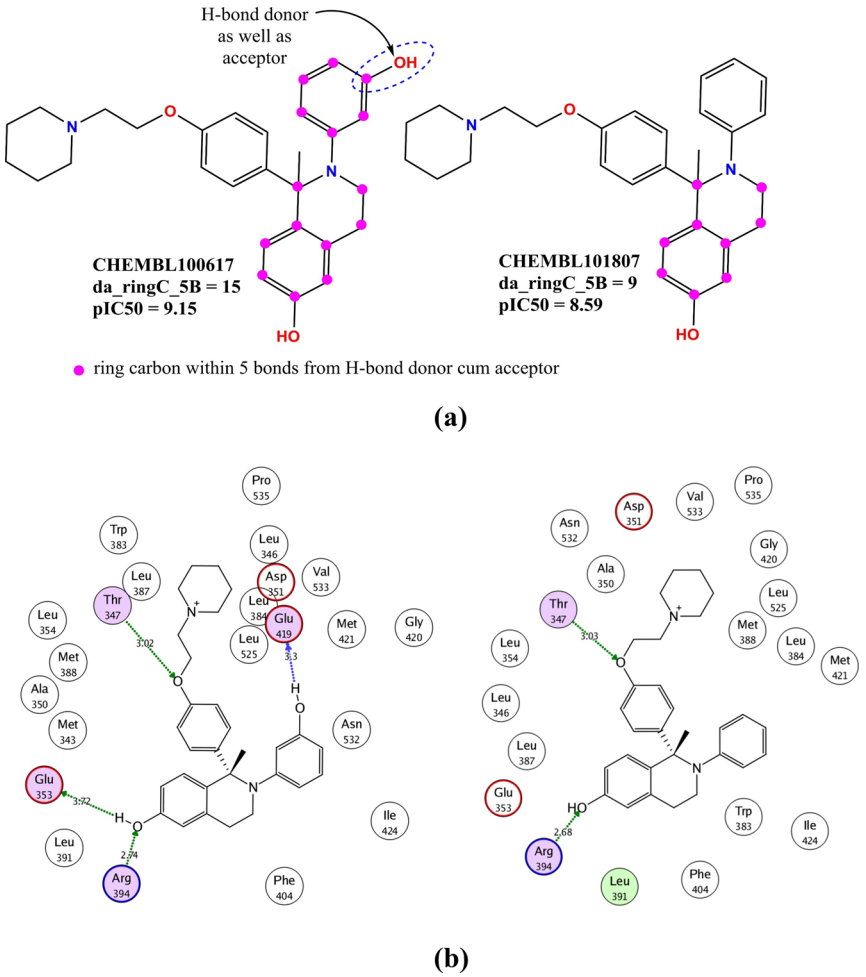


Figure 9. (a) Representation of da_ringC_5B. (b) 2D-interaction diagram after molecular docking for selected molecules CHEMBL101807 and CHEMBL100617.

CHEMBL100595 with CHEMBL100617, CHEMBL363412 with CHEMBL196131, CHEMBL4228649 with CHEMBL4228617, CHEMBL179929 with CHEMBL183162, a few to mention.

In addition, the molecular docking analysis also confirmed that the −OH attached to the meta position of the benzene ring in CHEMBL100617 is responsible for H-bonding with Glue419 with a distance of 3.3 Å. Such an important interaction is absent in case of CHEMBL101807 (see Figure 9). Thus, QSAR and molecular docking provided equivalent results.

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

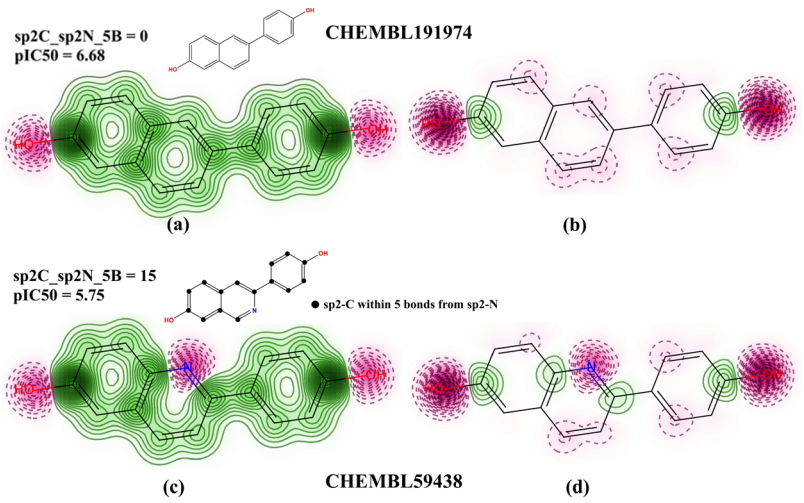


Figure 10. Some selected examples to understand the influence of sp2C_sp2N_5B on the activity profile. (a, c) Similarity maps with respect to lipophilic contribution. (b, d) Similarity maps with respect to partial charges (generated using RDKit).

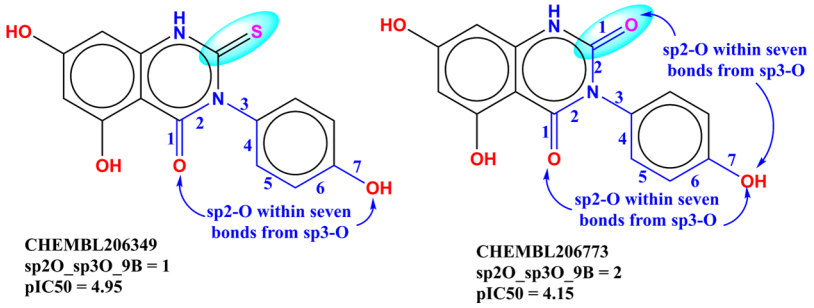


Figure 11. Depiction of sp2O_sp3O_9B using CHEMBL206349 with CHEMBL206773.

4.3. sp2C_sp2N_5B and fsp2Csp3N4B. fsp2Csp3N4B counts the occurrence of sp^3-hybridized nitrogen atoms exactly at 4 bonds from sp^2-hybridized carbon atoms. If the same sp^3-hybridized nitrogen atom is also present within ≤3 bonds from any other sp^2-hybridized carbon atoms, then it was excluded. The positive coefficient for fsp2Csp3N4B in the developed model indicates that a higher value of fsp2Csp3N4B could lead to better binding with the target receptor. This is further supported by its positive correlation with pIC_{50}.

The inverse is true for sp2C_sp2N_5B, which indicates that the occurrence of sp^2-N within five bonds from sp^2-C does not favor binding with ER-α. For example, CHEMBL360498 with CHEMBL362425, CHEMBL33625 with CHEMBL286115, CHEMBL191974 with CHEMBL59438, CHEMBL5074047 with CHEMBL5075144 (see Figure S4 in the Supporting Information), etc. Some selected examples to understand the influence of sp2C_sp2N_5B on the activity profile are depicted in Figure 10.

The two molecular descriptors viz. fsp2Csp3N4B and sp2C_sp2N_5B highlight the importance of sp^2-carbon, when in combination with sp^3-nitrogen atoms (fsp2Csp3N4B) or sp^2-nitrogen atoms (sp2C_sp2N_5B). Generally, sp^2-carbon is possible only when carbonyl, vinyl, imine groups, or Schiff bases are present in a molecule. That is, the presence of these groups could be used for significantly influencing the binding of a molecule with ER-α. The two descriptors fsp2Csp3N4B and sp2C_sp2N_5B have not only opposite sign coefficients in the QSAR model but also reverse correlation with pIC_{50}, which gives interesting information about the pharmacophoric requirements to be an ER-α binder. The molecule must possess an sp^2-carbon separated from the sp^3-nitrogen atom, mostly either present as 1° or 2° amine to act as a donor as well as an acceptor, exactly by four bonds. The requirement of an sp^2-carbon with sp^3-nitrogen also points out that a balance of flatness with flexibility (tetrahedral geometry) is required to be an effective ER-α binder. This is further vindicated by another observation that a combination of sp^2-carbon with sp^2-nitrogen is not favorable for ER-α binding as a higher value of sp2C_sp2N_5B leads to a lower pIC_{50}. Nantasenamat and co-workers[41] have also reported that amine, 2° carbon, conjugated double bonds, and carboxylic acid group are important structural features associated with ER-α binding ability. All of these features ultimately point to sp^2-carbon or 1° or 2° amine. The present work successfully identified not only similar features but also extended them by identifying the distance between them, their inter-relations, and the constraints to be followed while selecting the nitrogen-based group to have a better binding with ER-α.

4.4. sp2O_sp3O_9B. The descriptor sp2O_sp3O_9B has a negative correlation with pIC_{50} and also a negative coefficient in model 1, which indicates that the presence of sp^2-O within nine bonds from sp^3-O in a molecule leads to a lower pIC_{50}. Therefore, such types of oxygen atoms should be kept at a large distance, preferably 10 bonds or more in the molecule. A plausible reason for this could be the crucial hydrogen bond formation or polar interactions with the respective amino acids for adequate ligand-binding affinity. This further points out to an interesting aspect of ER-α binders that the polar regions

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

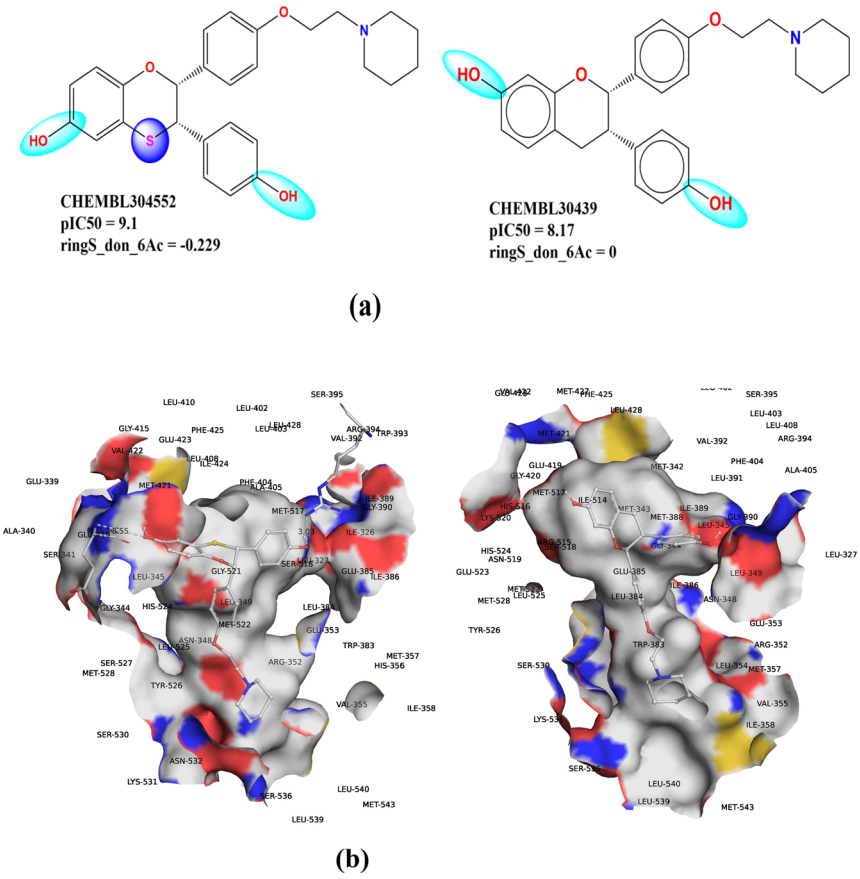


Figure 12. (a) Pictorial representation for ringS_don_6Ac using selected examples. (b) 3D-docking poses for CHEMBL304552 and CHEMBL30439.

must be present at the periphery of the molecule and separated by a longer distance. This provides a plausible reason for the difference in the binding potential of positional isomers. For example, a comparison of CHEMBL206349 with CHEMBL206773 (depicted in Figure 11), and CHEMBL240227 with CHEMBL396118, further strengthened this observation.

Mbachu et al.[61] have reported that ER-α binding increases due to the presence of two optimally positioned hydroxyl groups (or bioisosteres in some cases), about 11 Å apart, separated by a lipophilic core. A similar feature has been identified in the present work. This encouraged us to study a similar descriptor sp3O_sp3O_9B (total number of sp^3-hybridized oxygen atoms within 9 bonds from sp^3-hybridized oxygen atoms). This descriptor sp3O_sp3O_9B has a positive correlation of 0.340 with pIC_{50}. However, replacing sp2O_sp3O_9B with sp3O_sp3O_9B led to a model with a low statistical performance of R^2_{tr} = 0.74. Therefore, sp2O_sp3O_9B appears as a better choice from a statistical view. A plausible reason could be that one of the oxygens from the two oxygens, which are separated by a large distance, is a H-bond acceptor through its interaction with His524.[16] Thus, the present work specifies that the presence of two −OH group (sp^3-oxygens) is not mandatory, instead two oxygens with a combination of sp^3- and sp^2- hybridization are preferred.

4.5. ringS_don_6Ac. The descriptor ringS_don_6Ac has a negative coefficient in the QSAR model; henceforth, its value must be kept as low as possible. In the present data set, its value ranges from 0.469 to 0.303, which indicates that in some molecules the value of ringS_don_6Ac is positive. There are 42 such molecules, which are low active with pIC_{50} in the range of 4.33−5.66. On the other side, there are 159 molecules with a negative value of ringS_don_6Ac. Of these 159 molecules, 150 molecules have $pIC_{50} \geq 6.27$. This indicates that the donor atom present within 6 Å from the ring Sulfur atom must be negatively charged. This restriction cum requirement of the existence of only negatively charged H-bond donor indicates that either the donor should be −OH, or a 1°/2° amine distant from the oxygen atom. From a matched molecular pair point of view, there are very few examples to illustrate the direct influence of ringS_don_6Ac on activity. For example, CHEMBL304552 with CHEMBL30439 (see Figure 12a).

Molecular docking analysis of CHEMBL304552 and CHEMBL30439, depicted in Figure 12b, indicates that the molecule CHEMBL304552 binds with the receptor through H-bonding with Glu419 and Arg394 with a distance of 3.55 and 3.03 Å respectively. However, compound CHEMBL30439 lacks such interactions.

4.6. fsp3Cdon1B. Of the 298 highly active molecules possessing pIC_{50} from 10.22 to 8, only 23 possess fsp3Cdon1B. The highest value of fsp3Cdon1B = 3 is only for two molecules viz. CHEMBL1453264 and CHEMBL382632, which have a low pIC_{50} value of 4.93 and 5.20, respectively. This indicates that for a higher binding with ER-α, such a combination must be avoided.

The descriptor fsp3Cdon1B is associated with carbon and H-bond donor features, but it further emphasizes that the carbon must be sp^3-hybridized to influence the binding with ER-α. Generally, N and O atoms, which are more electro-

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

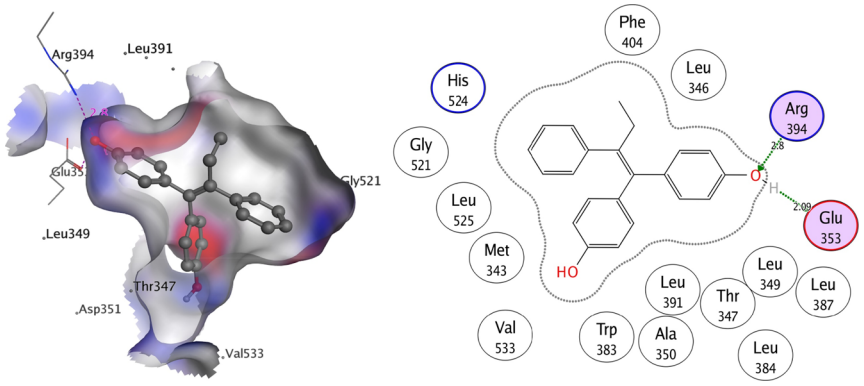


Figure 13. Docking pose for the most active molecule CHEMBL149791 in the active site of ER-α.

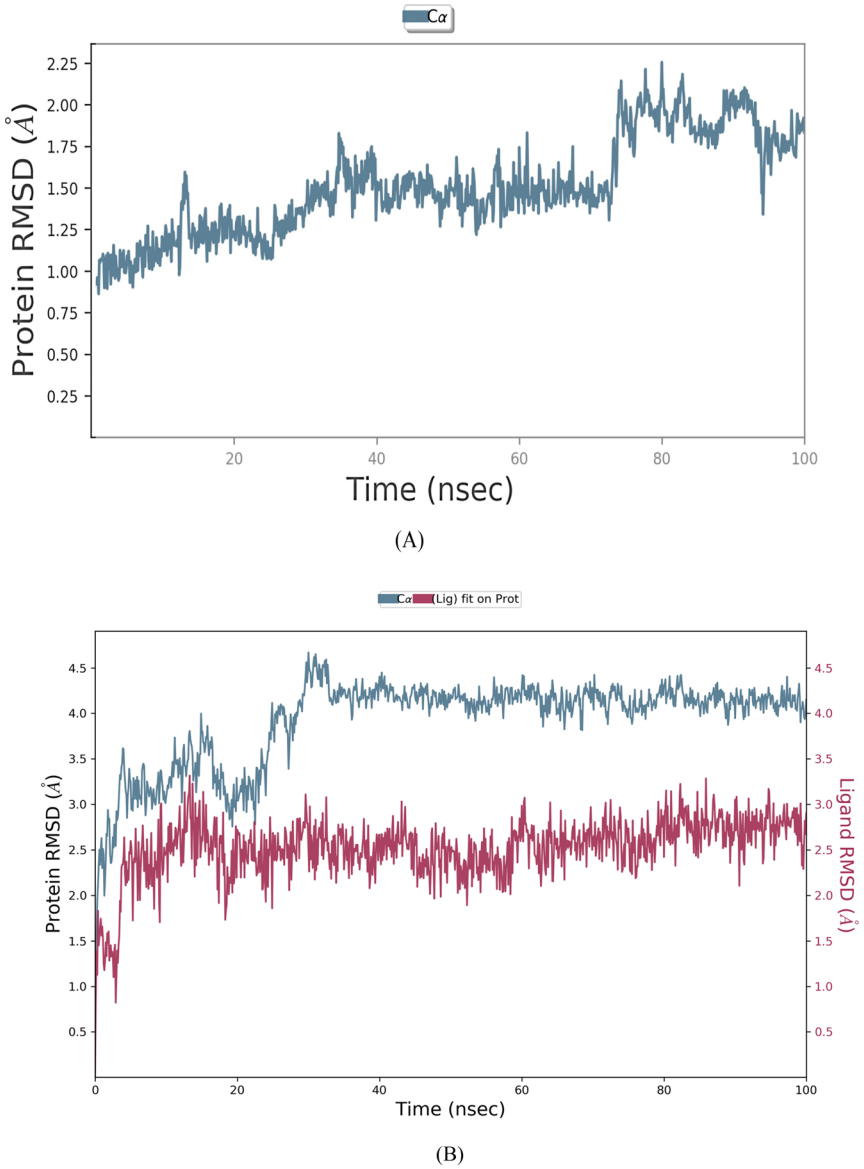


Figure 14. RMSD plots for (A) ER-α apoprotein and (B) ER-α with the most active ligand CHEMBL149791.

negative than carbon, act as donors in the form of −NH, $-NH_2$, and −OH. As sp^3-C is less electronegative than sp^2- and sp-C, the descriptor fsp3Cdon1B indicates that a polar moiety negatively influences binding with ER-α. Further, a close inspection of molecules with fsp3Cdon1B reveals that in the majority of molecules, such a combination is present near to central part of a molecule. All of these observations and the fact that the active site of ER-α is significantly lipophilic (Figure 2) specify that the number of polar carbon-donor bonds must be kept as low as possible.

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

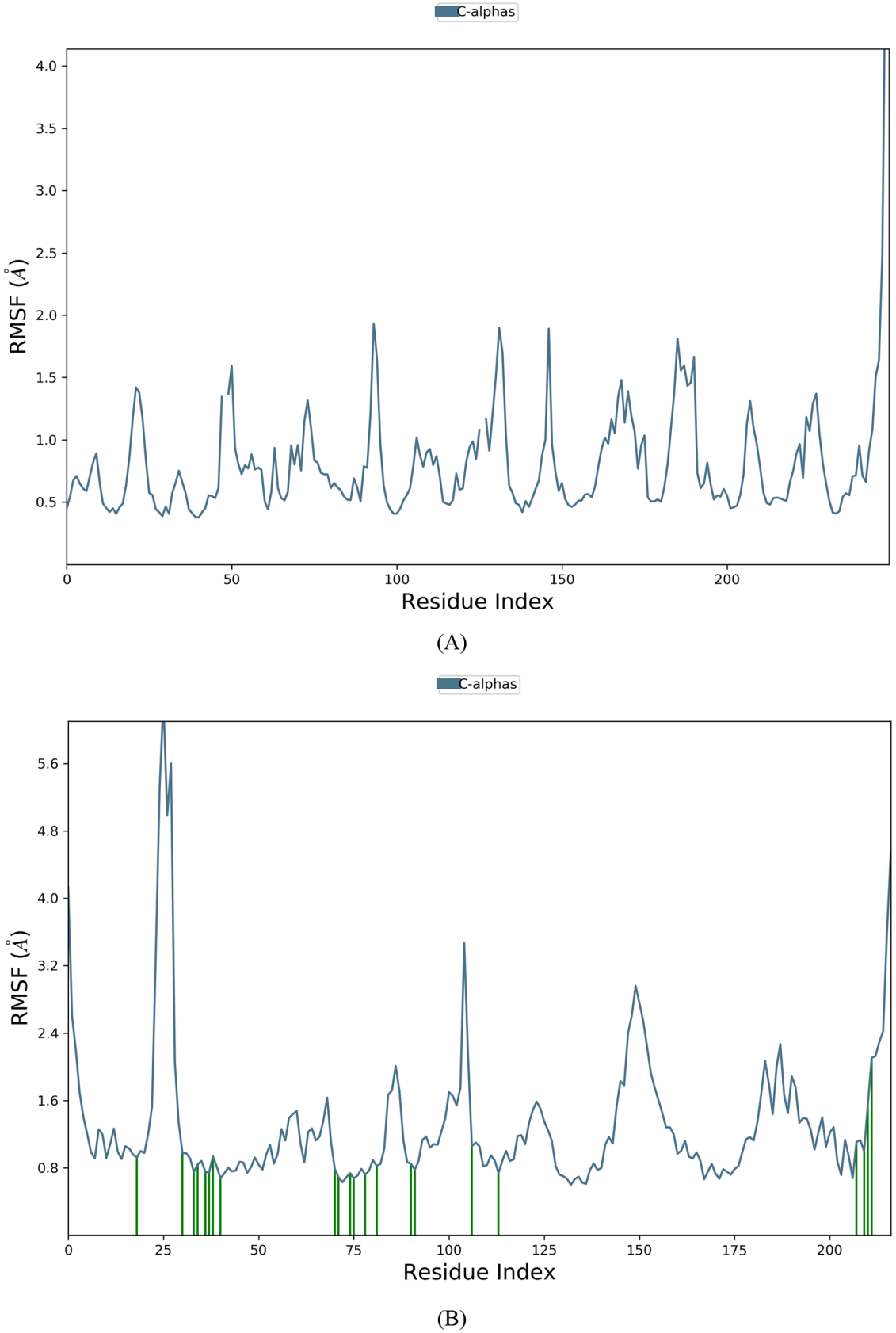


Figure 15. RMSF plots for (A) ER-α apoprotein and (B) ER-α-most active ligand.

4.7. Molecular Docking Analysis. Molecular docking is a branch of computer-aided drug design to identify the prominent structural features that enhance the binding of a molecule with a receptor. In the present work, all 1186 molecules were docked inside the active site of the receptor ER-α. For the sake of convenience, herein, we describe the

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

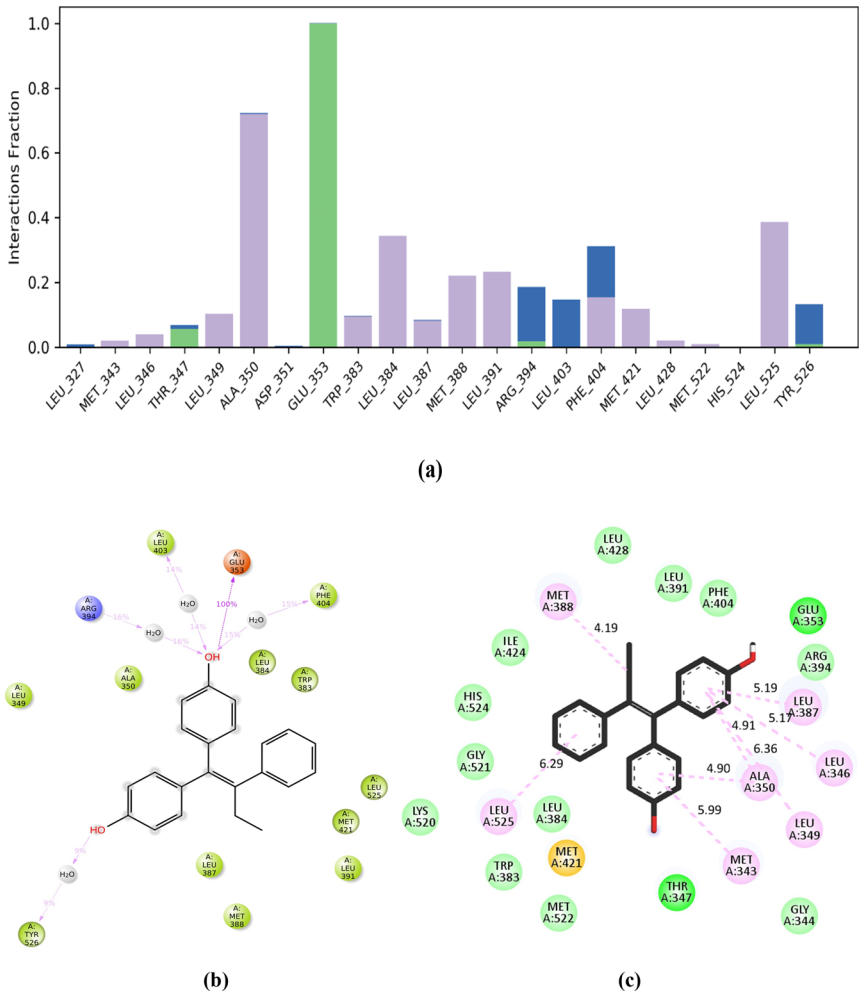


Figure 16. (a) Protein−ligand contact histogram (H-bond (green), hydrophobic (purple), and water bridge (blue)). 2D interactions of the most active ligand with protein 7ujo recorded in a 100 ns simulation interval with distances in Å. (b) Only H-bonds (purple) and water bridges (blue) in percentage. (c) Hydrophobic interactions (pink) with distances.

docking pose of the most active molecule CHEMBL149791 having pIC_{50} = 10.22. The docking pose, depicted in Figure 13, specifies that the molecule has lipophilic, polar, and H-bonding interactions with the receptor. One of the phenolic −OH groups is responsible for H-bonding with Glu353 and Arg394 with a distance of 2.09 and 2.80 Å respectively. Interestingly, the molecule has adopted an almost propeller-like shape with the three benzene rings oriented toward different pockets of the active site, the central vinyl group as a hub/boss of the propeller, and the ethyl acting as a shaft. The central core, ethyl chain, as well as the benzene rings led to lipophilic interactions with nearby residues Gly521, Leu525, Met343, Val533, Trp383, Ala350, Leu391, Thr347, Leu349, Leu387, Leu384, Phe404 and Leu346 (see Figure 14).

4.8. Molecular Dynamic and Simulation Analysis. Molecular dynamics (MD) simulation is widely recognized as a reliable technique for assessing the structural and functional stability of proteins and protein−ligand complexes. Two molecular dynamics and simulation (MD) studies were conducted to assess the stability and spatial proximity of protein complexes attached to a molecule. During the inquiry, a series of two simulations was conducted. The first simulation was formulated as a negative control, including just the proteins. The second simulation served as a control test using the pdb: 7ujo protein and a ligand. After examining the root-mean-square deviation (RMSD) data, it was observed that every 100 ns simulation consistently displayed the same conformation, as seen in Figure 14. In order to get a comprehensive understanding of the stability and convergence of the systems under investigation, we conducted an analysis of the root-mean-square deviation (RMSD) values pertaining to the atoms in the protein's backbone. The C-α backbone of ER-α had a variation of 2.06 Å after contact in its apo form (Figure 14A), while the ligand with the highest activity (CHEMBL149791) demonstrated a departure of 3.90 Å (Figure 14B). The RMSD value for the ligand with the highest activity bound to 7ujo is higher than the RMSD value of the apoprotein (apoprotein; RMSD: 2.06 Å).

Remarkably, the apoprotein ER-α displayed substantial oscillations from 70 to 90 ns and then stabilized for the rest of the experiment, with minor variations persisting at 100 ns. Moreover, when the highest active ligand is present, there were no observed variations, suggesting that the protein is more stable when it is attached to the ligand than when it is free. The results suggest that the ligand with the highest activity, when bound to 7ujo, has much greater stability in the complex compared with the apoprotein. The ability of proteins to maintain their structural stability, both before and during simulation, may be shown by the allowable range of root-mean-square deviation (RMSD) plots. The enhanced affinity

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

of the ligand indicates that the complexes including ER-α (7ujo) coupled with the most potent ligand exhibit considerable stability. The RMSD charts demonstrate the stability of the proteins in both the apo and ligand-bound states during the whole 100 ns simulation period. These findings indicate that the ligand with the highest level of activity forms a stable complex when bound to the ER-α.

In order to assess the remaining variance in a 100 ns molecular dynamics (MD) simulation, we have analyzed the root-mean-square fluctuation (RMSF) plot for two specific elements: the apoprotein and the 7ujo-bound most active ligand (see Figure 14). The RMSF analysis of the apoprotein revealed substantial modifications in several regions. The specified sections include the loop region including amino acids number 178−216, with particular attention to amino acids GLN502, TYR526, SER527, MET528, LYS529, SER530, and LYS531. In addition, the area including amino acids 15−40 showed higher fluctuations which uniquely comprises amino acids GLU330, TYR331, ASP332, PRO333, THR334, ARG335, PRO336, SER341, and MET342. Particularly, the residues THR334, PRO333, and PRO336 fluctuate more. Moreover, it was observed that the protein in apo form is more stable and fluctuates more when bound to a ligand. Conversely, the ligand that is now attached to pdb: 7ujo is very active and shows notable differences at residues such as PRO333, GLU419, and SER464. In addition, it was observed that the root-mean-square fluctuations (RMSF) plots for the highly active ligand attached to the protein structure 7ujo exhibited a significant rise in fluctuations in certain amino acid residues. The increase in variations remained constant throughout the whole 100 ns experiment. Conversely, the remaining residues had comparatively diminished levels of variance, as shown in Figure 15. Significantly, the residues SER530, LYS531, SER463, and THR334 exhibited more pronounced changes during the course of the experiment. Throughout the whole simulation, the study of ER-α' s root-mean-square fluctuation (RMSF) revealed a limited range of variation, namely between 1 and 30 residues. The discovery suggests that the arrangements of amino acids in ER-α stayed constant over the whole duration of the simulation (Figure 15).

The analysis of the root-mean-square fluctuation (RMSF) plots clearly shows that the ER-α apoprotein and the ER-α bound to the most powerful ligand have similar fluctuation patterns. Furthermore, the ER-α protein associated with the most potent ligand exhibits a similar pattern of variation throughout the course of the 100 ns simulation.

The original structure of the most active ligand in the simulation included around three hydrogen bonds. Furthermore, it exhibited more than 10 hydrophobic interactions during a simulated duration of 100 ns. The bonding was seen for the whole 100 ns simulation period, as seen in Figure 16. Throughout the 100 ns of simulation, the predominant observations were hydrogen bonding, hydrophobic, and water bridge interactions. In addition, we observed sustained hydrogen bonding interactions throughout the whole 100 ns simulation. Throughout the experiment, the ligand that exhibited the greatest degree of activity and engaged in a hydrogen bond with Leu346 in a typical fashion occurred with a frequency of 10%. In addition, it formed a hydrogen bond with Asp351 (90%) and Leu391 (5%) respectively. Water molecules inside the ER-α protein enabled the creation of water bridge connections, with Leu391 (20%), Arg394 (15%), Leu403 (10%), and Leu525 (10%) being engaged throughout the simulation period.

In addition, the formation of stable complexes was seen to include water bridges, hydrogen bonding, and hydrophobic contacts (Figure 16). The pi-cation interaction significantly enhanced the stability of the drug−protein combination inside the binding pocket of 7ujo. The ligand with the highest activity has a strong attraction to ER-α, leading to the creation of a substantial quantity of conventional hydrogen bonds.

5. CONCLUSIONS

The present work successfully led to the development of an e-QSAR model satisfying OECD guidelines. The developed model has a high predictive ability reflected from high values of different parameters viz. R^2_{ex}, Q^2-F^n, CCC_{ex}, R^2-ExPy, etc. The QSAR model emphasized some known pharmacophoric features such as lipophilicity and the preference for separation of two H-bond capable groups by a large distance to have effective binding with ER-α. The present work successfully extends that the ring and sp^2-carbon atoms, which are lipophilic in nature, are associated with the binding ability of a molecule with ER-α. Likewise, the work also points out that oxygen atoms, in turn, H-bond capable groups, separated by a large distance of more than nine bonds, enhance ER-α binding ability. The presence of a H-bond donor in the vicinity of ring sulfur is a novel and unreported structural feature reported in the present work. Similarly, the work also clarifies the position of sp^3-N (i.e., 1 and 2° amine) with respect to sp^2-carbon atoms required to increase binding with ER-α. The molecular docking and molecular dynamics simulations also support the above results. The results could be useful in the future optimization of ER-α binders.

■ ASSOCIATED CONTENT

Data Availability Statement

The data of molecules used in the present work along with formulas to calculate various validation parameters are available as Supporting Information as a PDF file.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c00906.

Description of molecular descriptors present in the QSAR model and their range; Williams plot for model-A to define applicability domain; plot of experimental vs predicted pIC_{50} for model-A; plot of residuals vs predicted pIC_{50} for model-A; additional selected examples to understand the influence on sp2C_sp2N_5B on activity profile; statistical parameters used for validation of QSAR models; and CHEMBL ID, 2D-representation, experimental pIC_{50}, and values of molecular descriptors for all molecules used in the present work (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Vijay H. Masand − *Department of Chemistry, Vidya Bharati Mahavidyalaya, Amravati 444 602 Maharashtra, India;* orcid.org/0000-0001-9300-4147; Email: vijaymasand@gmail.com

Magdi E. A. Zaki − *Department of Chemistry, College of Science, Imam Mohammad Ibn Saud Islamic University,*

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

Riyadh 11623, Saudi Arabia; Email: mezaki@imamu.edu.sa

Authors

Sami A. Al-Hussain − *Department of Chemistry, College of Science, Imam Mohammad Ibn Saud Islamic University, Riyadh 11623, Saudi Arabia*

Abdullah Y. Alzahrani − *Department of Chemistry, Faculty of Science and Arts, King Khalid University, Mohail 61421, Saudi Arabia*

Aamal A. Al-Mutairi − *Department of Chemistry, College of Science, Imam Mohammad Ibn Saud Islamic University, Riyadh 11623, Saudi Arabia*

Rania A. Hussien − *Department of Chemistry, Faculty of Science, Al-Baha University, Al-Baha 65799, Kingdom of Saudi Arabia*; orcid.org/0000-0002-5268-7161

Abdul Samad − *Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Tishk International University, Erbil 44001, Iraq*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c00906

Author Contributions

V.H.M., M.E.A.Z., and S.A.A-H. conceived and supervised this study. A.A.A.-M. collected and curated the data. V.H.M., M.E.A.Z., and A.Y.A. processed the data and analyzed the results. S.A.A-H., A.Y.A., and V.H.M. created images, writing and drafted manuscript. All authors reviewed and approved the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia (Grant number IMSIU-RG23013) for its support of this research work.

■ REFERENCES

(1) Siegel, R. L.; Miller, K. D.; Wagle, N. S.; Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin* **2023**, *73* (1), 17−48.

(2) Marangoni, F.; Zhakyp, A.; Corsini, M.; Geels, S. N.; Carrizosa, E.; Thelen, M.; Mani, V.; Prussmann, J. N.; Warner, R. D.; Ozga, A. J.; et al. Expansion of tumor-associated Treg cells upon disruption of a CTLA-4-dependent feedback loop. *Cell* **2021**, *184* (15), 3998−4015.

(3) Giaquinto, A. N.; Sung, H.; Miller, K. D.; Kramer, J. L.; Newman, L. A.; Minihan, A.; Jemal, A.; Siegel, R. L. Breast Cancer Statistics, 2022. *CA Cancer J. Clin* **2022**, *72* (6), 524−541.

(4) Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin* **2022**, *72* (1), 7−33.

(5) WHO. *Breast cancer*; WHO, 2023. https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed 2023).

(6) Bolego, C.; Vegeto, E.; Pinna, C.; Maggi, A.; Cignarella, A. Selective agonists of estrogen receptor isoforms: new perspectives for cardiovascular disease. *Arterioscler Thromb Vasc Biol.* **2006**, *26* (10), 2192−2199.

(7) Xu, X.; Yang, W.; Li, Y.; Wang, Y. Discovery of estrogen receptor modulators: a review of virtual screening and SAR efforts. *Expert Opin Drug Discov* **2010**, *5* (1), 21−31.

(8) Vrtacnik, P.; Ostanek, B.; Mencej-Bedrac, S.; Marc, J. The many faces of estrogen signaling. *Biochem Med. (Zagreb)* **2014**, *24* (3), 329−342.

(9) Mottamal, M.; Kang, B.; Peng, X.; Wang, G. From Pure Antagonists to Pure Degraders of the Estrogen Receptor: Evolving Strategies for the Same Target. *ACS Omega* **2021**, *6* (14), 9334−9343.

(10) Fernando, T. M.; Moore, H. M.; Wongchenko, M. J.; Metcalfe, C. Next-Generation Estrogen Receptor−Targeted Therapeutics. *Annual Review of Cancer Biology* **2023**, *7* (1), 313−330.

(11) Chang, Y. H.; Chen, J. Y.; Hor, C. Y.; Chuang, Y. C.; Yang, C. B.; Yang, C. N. Computational study of estrogen receptor-alpha antagonist with three-dimensional quantitative structure-activity relationship, support vector regression, and linear regression methods. *Int. J. Med. Chem.* **2013**, *2013*, No. 743139.

(12) Dai, S. Y.; Burris, T. P.; Dodge, J. A.; Montrose-Rafizadeh, C.; Wang, Y.; Pascal, B. D.; Chalmers, M. J.; Griffin, P. R. Unique ligand binding patterns between estrogen receptor alpha and beta revealed by hydrogen-deuterium exchange. *Biochemistry* **2009**, *48* (40), 9668−9676.

(13) Sharma, D.; Kumar, S.; Narasimhan, B. Estrogen alpha receptor antagonists for the treatment of breast cancer: a review. *Chem. Cent J.* **2018**, *12* (1), 107.

(14) Clusan, L.; Ferriere, F.; Flouriot, G.; Pakdel, F. A Basic Review on Estrogen Receptor Signaling Pathways in Breast Cancer. *Int. J. Mol. Sci.* **2023**, *24* (7), 6834.

(15) Paterni, I.; Granchi, C.; Katzenellenbogen, J. A.; Minutolo, F. Estrogen receptors alpha (ERalpha) and beta (ERbeta): subtype-selective ligands and clinical potential. *Steroids* **2014**, *90*, 13−29.

(16) Lee, S.; Barron, M. G. Structure-Based Understanding of Binding Affinity and Mode of Estrogen Receptor alpha Agonists and Antagonists. *PLoS One* **2017**, *12* (1), No. e0169607.

(17) Xue, Q.; Liu, X.; Liu, X. C.; Pan, W. X.; Fu, J. J.; Zhang, A. Q. The Effect of Structural Diversity on Ligand Specificity and Resulting Signaling Differences of Estrogen Receptor alpha. *Chem. Res. Toxicol.* **2019**, *32* (6), 1002−1013.

(18) Kurtanovic, N.; Tomasevic, N.; Matic, S.; Proia, E.; Sabatino, M.; Antonini, L.; Mladenovic, M.; Ragno, R. Human Estrogen Receptor Alpha Antagonists, Part 3:3-D Pharmacophore and 3-D QSAR Guided Brefeldin A Hit-to-Lead Optimization toward New Breast Cancer Suppressants. *Molecules* **2022**, *27* (9), 2823.

(19) Candelaria, N. R.; Liu, K.; Lin, C. Y. Estrogen receptor alpha: molecular mechanisms and emerging insights. *J. Cell Biochem* **2013**, *114* (10), 2203−2208.

(20) Foryst-Ludwig, A.; Kintscher, U. Metabolic impact of estrogen signalling through ERalpha and ERbeta. *J. Steroid Biochem Mol. Biol.* **2010**, *122* (1−3), 74−81.

(21) Chimento, A.; De Luca, A.; Avena, P.; De Amicis, F.; Casaburi, I.; Sirianni, R.; Pezzi, V. Estrogen Receptors-Mediated Apoptosis in Hormone-Dependent Cancers. *Int. J. Mol. Sci.* **2022**, *23* (3), 1242.

(22) Lewoniewska, S.; Oscilowska, I.; Forlino, A.; Palka, J. Understanding the Role of Estrogen Receptor Status in PRODH/POX-Dependent Apoptosis/Survival in Breast Cancer Cells. *Biology (Basel)* **2021**, *10* (12), 1314.

(23) Tran, A.; Scholtes, C.; Songane, M.; Champagne, C.; Galarneau, L.; Levasseur, M. P.; Fodil, N.; Dufour, C. R.; Giguere, V.; Saleh, M. Estrogen-related receptor alpha (ERRalpha) is a key regulator of intestinal homeostasis and protects against colitis. *Sci. Rep* **2021**, *11* (1), 15073.

(24) Cheng, C. H.; Chen, L. R.; Chen, K. H. Osteoporosis Due to Hormone Imbalance: An Overview of the Effects of Estrogen Deficiency and Glucocorticoid Overuse on Bone Turnover. *Int. J. Mol. Sci.* **2022**, *23* (3), 1376.

(25) (a) Zhao, C.; Dahlman-Wright, K.; Gustafsson, J. A. Estrogen receptor beta: an overview and update. *Nucl. Recept Signal* **2008**, *6*, No. e003. (b) Mal, R.; Magner, A.; David, J.; Datta, J.; Vallabhaneni, M.; Kassem, M.; Manouchehri, J.; Willingham, N.; Stover, D.; Vandeusen, J.; et al. Estrogen Receptor Beta (ERbeta): A Ligand Activated Tumor Suppressor. *Front Oncol* **2020**, *10*, No. 587386. (c) Zhou, Y.; Liu, X. The role of estrogen receptor beta in breast cancer. *Biomark Res.* **2020**, *8*, 39.

(26) Khalid, A. B.; Krum, S. A. Estrogen receptors alpha and beta in bone. *Bone* **2016**, *87*, 130−135.

(27) Aryan, L.; Younessi, D.; Zargari, M.; Banerjee, S.; Agopian, J.; Rahman, S.; Borna, R.; Ruffenach, G.; Umar, S.; Eghbali, M. The Role

of Estrogen Receptors in Cardiovascular Disease. *Int. J. Mol. Sci.* **2020**, *21* (12), 4314.
(28) Indukuri, R.; Hases, L.; Archer, A.; Williams, C. Estrogen Receptor Beta Influences the Inflammatory p65 Cistrome in Colon Cancer Cells. *Front Endocrinol (Lausanne)* **2021**, *12*, No. 650625.
(29) Monsivais, D.; Dyson, M. T.; Yin, P.; Coon, J. S.; Navarro, A.; Feng, G.; Malpani, S. S.; Ono, M.; Ercan, C. M.; Wei, J. J.; et al. ERbeta- and prostaglandin E2-regulated pathways integrate cell proliferation via Ras-like and estrogen-regulated growth inhibitor in endometriosis. *Mol. Endocrinol.* **2014**, *28* (8), 1304−1315.
(30) Jobe, S. O.; Ramadoss, J.; Wargin, A.; Magness, R. R. The Prostacyclin-Thromboxane Balance: Role of ER-alpha Versus ER-beta in Estradiol-17beta and Its CYP450- and COMT-Derived Metabolites-Stimulated Prostacyclin and Thromboxane Production in Uterine Artery Endothelial Cells. *Biol. Reprod.* **2011**, *85* (Suppl_1), 471.
(31) Dahlman-Wright, K.; Koehler, K.; Gustafsson, J.-Å. Estrogen Receptor-β Structure and Function. *In Encyclopedia of Hormones* **2003**, 599−608.
(32) Halon, A.; Materna, V.; Drag-Zalesinska, M.; Nowak-Markwitz, E.; Gansukh, T.; Donizy, P.; Spaczynski, M.; Zabel, M.; Dietel, M.; Lage, H.; et al. Estrogen receptor alpha expression in ovarian cancer predicts longer overall survival. *Pathol Oncol Res.* **2011**, *17* (3), 511−518.
(33) Chantalat, E.; Valera, M. C.; Vaysse, C.; Noirrit, E.; Rusidze, M.; Weyl, A.; Vergriete, K.; Buscail, E.; Lluel, P.; Fontaine, C.; et al. Estrogen Receptors and Endometriosis. *Int. J. Mol. Sci.* **2020**, *21* (8), 2815.
(34) An, K. C. Selective Estrogen Receptor Modulators. *Asian Spine J.* **2016**, *10* (4), 787−791.
(35) (a) Sanchez-Spitman, A. B.; Swen, J. J.; Dezentje, V. O.; Moes, D.; Gelderblom, H.; Guchelaar, H. J. Clinical pharmacokinetics and pharmacogenetics of tamoxifen and endoxifen. *Expert Rev. Clin Pharmacol* **2019**, *12* (6), 523−536. (b) Girault, I.; Bieche, I.; Lidereau, R. Role of estrogen receptor alpha transcriptional coregulators in tamoxifen resistance in breast cancer. *Maturitas* **2006**, *54* (4), 342−351.
(36) Bryant, H. U.; Glasebrook, A. L.; Yang, N. N.; Sato, M. An estrogen receptor basis for raloxifene action in bone. *J. Steroid Biochem Mol. Biol.* **1999**, *69* (1−6), 37−44.
(37) (a) Huang, W.; Peng, Y.; Kiselar, J.; Zhao, X.; Albaqami, A.; Mendez, D.; Chen, Y.; Chakravarthy, S.; Gupta, S.; Ralston, C.; et al. Multidomain architecture of estrogen receptor reveals interfacial cross-talk between its DNA-binding and ligand-binding domains. *Nat. Commun.* **2018**, *9* (1), 3520. (b) Hewitt, S. C.; Korach, K. S. Estrogen receptors: structure, mechanisms and function. *Rev. Endocr Metab Disord* **2002**, *3* (3), 193−200. (c) Pham, T. H.; Lecomte, S.; Efstathiou, T.; Ferriere, F.; Pakdel, F. An Update on the Effects of Glyceollins on Human Health: Possible Anticancer Effects and Underlying Mechanisms. *Nutrients* **2019**, *11* (1), 79.
(38) Srinivasan, S.; Nwachukwu, J. C.; Bruno, N. E.; Dharmarajan, V.; Goswami, D.; Kastrati, I.; Novick, S.; Nowak, J.; Cavett, V.; Zhou, H. B.; et al. Full antagonism of the estrogen receptor without a prototypical ligand side chain. *Nat. Chem. Biol.* **2017**, *13* (1), 111−118.
(39) Metivier, R.; Stark, A.; Flouriot, G.; Hubner, M. R.; Brand, H.; Penot, G.; Manu, D.; Denger, S.; Reid, G.; Kos, M.; et al. A dynamic structural model for estrogen receptor-alpha activation by ligands, emphasizing the role of interactions between distant A and E domains. *Mol. Cell* **2002**, *10* (5), 1019−1032.
(40) (a) Cotterill, J. V.; Palazzolo, L.; Ridgway, C.; Price, N.; Rorije, E.; Moretto, A.; Peijnenburg, A.; Eberini, I. Predicting estrogen receptor binding of chemicals using a suite of in silico methods - Complementary approaches of (Q)SAR, molecular docking and molecular dynamics. *Toxicol. Appl. Pharmacol.* **2019**, *378*, No. 114630. (b) Ribay, K.; Kim, M. T.; Wang, W.; Pinolini, D.; Zhu, H. Predictive Modeling of Estrogen Receptor Binding Agents Using Advanced Cheminformatics Tools and Massive Public Data. *Front. Environ. Sci.* **2016**, *4*, 12. (c) Worachartcheewan, A.; Mandi, P.; Prachayasittikul, V.; Toropova, A. P.; Toropov, A. A.; Nantasenamat, C. Large-scale QSAR study of aromatase inhibitors using SMILES-based descriptors. *Chemometrics and Intelligent Laboratory Systems* **2014**, *138*, 120−126. (d) Zekri, A.; Harkati, D.; Kenouche, S.; Saleh, B. A. QSAR modeling, docking, ADME and reactivity of indazole derivatives as antagonizes of estrogen receptor alpha (ER-α) positive in breast cancer. *J. Mol. Struct.* **2020**, *1217*, No. 128442.
(41) Suvannang, N.; Preeyanon, L.; Malik, A. A.; Schaduangrat, N.; Shoombuatong, W.; Worachartcheewan, A.; Tantimongcolwat, T.; Nantasenamat, C. Probing the origin of estrogen receptor alpha inhibition via large-scale QSAR study. *RSC Adv.* **2018**, *8* (21), 11344−11356.
(42) Min, J.; Nwachukwu, J. C.; Min, C. K.; Njeri, J. W.; Srinivasan, S.; Rangarajan, E. S.; Nettles, C. C.; Sanabria Guillen, V.; Ziegler, Y.; Yan, S.; et al. Dual-mechanism estrogen receptor inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (35), No. e2101657118.
(43) (a) Masand, V. H.; Al-Hussain, S. A.; Rathore, M. M.; Thakur, S. D.; Akasapu, S.; Samad, A.; Al-Mutairi, A. A.; Zaki, M. E. A. Pharmacophore Synergism in Diverse Scaffold Clinches in Aurora Kinase B. *Int. J. Mol. Sci.* **2022**, *23* (23), 14527. (b) Bukhari, S. N. A.; Elsherif, M. A.; Junaid, K.; Ejaz, H.; Alam, P.; Samad, A.; Jawarkar, R. D.; Masand, V. H. Perceiving the Concealed and Unreported Pharmacophoric Features of the 5-Hydroxytryptamine Receptor Using Balanced QSAR Analysis. *Pharmaceuticals (Basel)* **2022**, *15* (7), 834.
(44) Zaki, M. E. A.; Al-Hussain, S. A.; Al-Mutairi, A. A.; Masand, V. H.; Samad, A.; Jawarkar, R. D. Mechanistic Analysis of Chemically Diverse Bromodomain-4 Inhibitors Using Balanced QSAR Analysis and Supported by X-ray Resolved Crystal Structures. *Pharmaceuticals (Basel)* **2022**, *15* (6), 745.
(45) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Felix, E.; Magarinos, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47* (D1), D930−D940.
(46) (a) Zaki, M. E. A.; Al-Hussain, S. A.; Bukhari, S. N. A.; Masand, V. H.; Rathore, M. M.; Thakur, S. D.; Patil, V. M. Exploring the Prominent and Concealed Inhibitory Features for Cytoplasmic Isoforms of Hsp90 Using QSAR Analysis. *Pharmaceuticals (Basel)* **2022**, *15* (3), 303. (b) Fourches, D.; Muratov, E.; Tropsha, A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J. Chem. Inf. Model.* **2010**, *50* (7), 1189−1204.
(47) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform* **2011**, *3* (1), 33.
(48) Masand, V. H.; Rastija, V. PyDescriptor: A new PyMOL plugin for calculating thousands of easily understandable molecular descriptors. *Chemometrics and Intelligent Laboratory Systems* **2017**, *169*, 12−18.
(49) Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem.* **2013**, *34* (24), 2121−2132.
(50) Masand, V. H.; Mahajan, D. T.; Nazeruddin, G. M.; Hadda, T. B.; Rastija, V.; Alfeefy, A. M. Effect of information leakage and method of splitting (rational and random) on external predictive ability and behavior of different statistical parameters of QSAR model. *Medicinal Chemistry Research* **2015**, *24* (3), 1241−1264.
(51) (a) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; et al. QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57* (12), 4977−5010. (b) Gramatica, P. Principles of QSAR Modeling. *International Journal of Quantitative Structure-Property Relationships* **2020**, *5* (3), 61−97. (c) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; et al. QSAR without borders. *Chem. Soc. Rev.* **2020**, *49* (11), 3525−3564.
(52) (a) Gramatica, P. Principles of QSAR models validation internal and external, QSAR & Combinatorial Science Volume 26,

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774

Issue 5. *QSAR & Combinatorial Science* **2007**, *26* (5), 694−701. (b) Gramatica, P. On the development and validation of QSAR models. *Methods in molecular biology* **2013**, *930*, 499−526.
(53) (a) Chirico, N.; Gramatica, P. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *J. Chem. Inf. Model.* **2011**, *51* (9), 2320−2335. (b) Chirico, N.; Gramatica, P. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *J. Chem. Inf. Model.* **2012**, *52* (8), 2044−2058. (c) Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the definition of the Q2 parameter for QSAR validation. *J. Chem. Inf. Model.* **2009**, *49* (7), 1669−1678. (d) Gutlein, M.; Helma, C.; Karwath, A.; Kramer, S. A Large-Scale Empirical Evaluation of Cross-Validation and External Test Set Validation in (Q)SAR. *Mol. Inform* **2013**, *32* (5−6), 516−528. (e) Tropsha, A.; Gramatica, P.; Gombar, V. K. The Importance of Being Earnest Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science* **2003**, *22* (1), 69−77.
(54) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6* (5), 1509−1519.
(55) Ouassaf, M.; Belaidi, S.; Chtita, S.; Lanez, T.; Abul Qais, F.; Md Amiruddin, H. Combined molecular docking and dynamics simulations studies of natural compounds as potent inhibitors against SARS-CoV-2 main protease. *J. Biomol. Struct. Dyn.* **2022**, *40* (21), 11264−11273.
(56) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(57) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé−Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97* (4), 2635−2643.
(58) Toukmaji, A. Y.; Board, J. A. Ewald summation techniques in perspective: a survey. *Comput. Phys. Commun.* **1996**, *95* (2−3), 73−92.
(59) (a) Ghosh, A.; Mukerjee, N.; Sharma, B.; Pant, A.; Kishore Mohanta, Y.; Jawarkar, R. D.; Bakal, R. L.; Terefe, E. M.; Batiha, G. E.-S.; Mostafa-Hedeab, G.; et al. Target Specific Inhibition of Protein Tyrosine Kinase in Conjunction With Cancer and SARS-COV-2 by Olive Nutraceuticals. *Frontiers in Pharmacology* **2022**, *12*, No. 812565. (b) Daoui, O.; Nour, H.; Abchir, O.; Elkhattabi, S.; Bakhouch, M.; Chtita, S. A computer-aided drug design approach to explore novel type II inhibitors of c-Met receptor tyrosine kinase for cancer therapy: QSAR, molecular docking, ADMET and molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **2023**, *41* (16), 7768−7785. (c) Ouassaf, M.; Daoui, O.; Alam, S.; Elkhattabi, S.; Belaidi, S.; Chtita, S. Pharmacophore-based virtual screening, molecular docking, and molecular dynamics studies for the discovery of novel FLT3 inhibitors. *J. Biomol. Struct. Dyn.* **2023**, *41* (16), 7712−7724. (d) Al-Jumaili, M. H. A.; Siddique, F.; Abul Qais, F.; Hashem, H. E.; Chtita, S.; Rani, A.; Uzair, M.; Almzaien, K. A. Analysis and prediction pathways of natural products and their cytotoxicity against HeLa cell line protein using docking, molecular dynamics and ADMET. *J. Biomol. Struct. Dyn.* **2023**, *41* (3), 765−777.
(60) Liu, H.; Papa, E.; Gramatica, P. QSAR prediction of estrogen activity for a large set of diverse chemicals under the guidance of OECD principles. *Chemical research in toxicology* **2006**, *19* (11), 1540−1548.
(61) Mbachu, O. C.; Howell, C.; Simmler, C.; Malca Garcia, G. R.; Skowron, K. J.; Dong, H.; Ellis, S. G.; Hitzman, R. T.; Hajirahimkhan, A.; Chen, S. N.; et al. SAR Study on Estrogen Receptor alpha/beta Activity of (Iso)flavonoids: Importance of Prenylation, C-Ring (Un)Saturation, and Hydroxyl Substituents. *J. Agric. Food Chem.* **2020**, *68* (39), 10651−10663.
(62) Liao, Z. Q.; Dong, C.; Carlson, K. E.; Srinivasan, S.; Nwachukwu, J. C.; Chesnut, R. W.; Sharma, A.; Nettles, K. W.; Katzenellenbogen, J. A.; Zhou, H. B. Triaryl-substituted Schiff bases are high-affinity subtype-selective ligands for the estrogen receptor. *J. Med. Chem.* **2014**, *57* (8), 3532−3545.
(63) Scott, J. S.; Bailey, A.; Davies, R. D.; Degorce, S. L.; MacFaul, P. A.; Gingell, H.; Moss, T.; Norman, R. A.; Pink, J. H.; Rabow, A. A.; et al. Tetrahydroisoquinoline Phenols: Selective Estrogen Receptor Downregulator Antagonists with Oral Bioavailability in Rat. *ACS Med. Chem. Lett.* **2016**, *7* (1), 94−99.

https://doi.org/10.1021/acsomega.4c00906
ACS Omega 2024, 9, 16759−16774