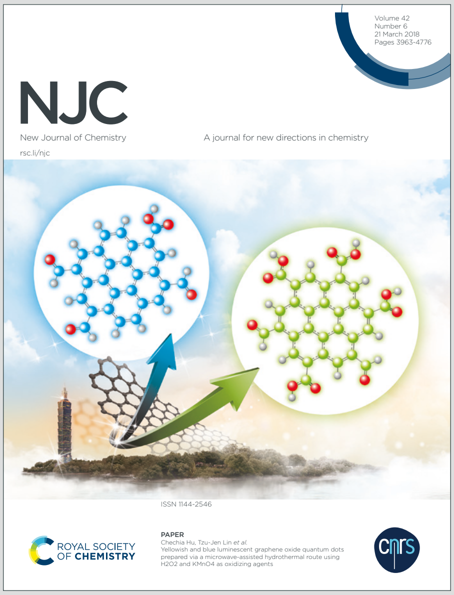
NJC

New Journal of Chemistry

A journal for new directions in chemistry

This article can be cited before page numbers have been issued, to do this please use: S. El Rhabori, M. Alaqarbeh, L. naanaai, Y. Elallouche, A. El Aissouq, M. Bouachrine, H. Zaitan, S. Chtita and F. Khalil, *New J. Chem.*, 2025, DOI: 10.1039/D5NJ02417J.


Volume 42
Number 6
21 March 2018
Pages 3963-4776
NJC
New Journal of Chemistry
A journal for new directions in chemistry
rsc.li/njc
ISSN 1144-2546
ROYAL SOCIETY OF CHEMISTRY
PAPER
Chechia Hu, Tzu-Jen Lin *et al.*
Yellowish and blue luminescent graphene oxide quantum dots prepared via a microwave-assisted hydrothermal route using H2O2 and KMnO4 as oxidizing agents
cnrs

View Article Online
DOI: 10.1039/D5NJ02417J

Integrative computational strategy for anticancer drug discovery: QSAR-ANN modeling, molecular docking, ADMET prediction, molecular dynamics and MM-PBSA simulations, and retrosynthetic analysis

Said El Rhabori[1,*], **Marwa Alaqarbeh**[2], **Lhoucine Naanaai**[1], **Yassine EL Allouche**[1], **Abdellah El Aissouq**[1], **Mohammed Bouachrine**[3], **Hicham Zaitan**[1], **Samir Chtita**[4], **Fouad Khalil**[1]

[1] *Laboratory of Processes, Materials and Environment (LPME), Sidi Mohamed Ben Abdellah University, Faculty of Science and Technology - Fez, Morocco*

[2] *Department of Chemistry, Faculty of Science, Applied Science Private University, Amman, 11931, Jordan*

[3] *MCNS Laboratory, Faculty of Sciences, Moulay Ismail University, Meknes, Morocco*

[4] *Laboratory of Analytical and Molecular Chemistry, Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca, Casablanca, Morocco*

[*]Corresponding author.: *E-mail address: said.elrhabori@usmba.ac.ma (S. El Rhabori)*

Abstract:

Breast cancer constitutes a primary cause of mortality among women. Existing therapeutic targets and treatment modalities are often confronted with drug resistance and the considerable financial expense associated with the development of new therapies, for which the results can be uncertain. Hormone therapy, mainly focused on inhibiting aromatase as a pivotal enzyme in estrogen biosynthesis, remains the preferred approach for treating this type of female cancer while minimizing costs thanks to advanced computer-aided drug design (CADD) methods. In this work, the strategy combines 3D-QSAR, artificial neural networks (ANN), molecular docking, ADMET analysis, molecular dynamics (MD) simulations, and retrosynthesis was applied to design novel anti-breast cancer agents and study their interactions with aromatase to identify potential inhibitors. The predictive models underwent rigorous internal and external validations based on significant statistical parameters, confirming their robustness and friability. As a result, 12 new drug candidates (L1-L12) were designed against breast cancer. Based on the results of virtual screening techniques, only one hit (L5) showed significant potential compared with the reference drug (Exemestane) and previously designed drug candidates (Ligand 5 and C2). Subsequent stability studies and pharmacokinetic evaluations reinforced L5's potential as an effective aromatase inhibitor. Retrosynthesis was used to optimize the synthesis of this candidate, which required in vitro and in vivo validation.

Keywords: Cancer; CADD; ANN; molecular docking; ADMET; molecular dynamic; retrosynthesis

View Article Online
DOI: 10.1039/D5NJ02417J

1. Introduction

Breast cancer remains a primary global health concern, being the second leading cause of cancer-related mortality worldwide, as reported by the World Health Organization (WHO) [1]. Despite significant advancements in therapeutic approaches, including surgery, radiotherapy, and chemotherapy, these modalities have not substantially improved survival rates [2,3]. The urgency of developing new treatment options is further underscored by the increasing prevalence of drug resistance and severe adverse effects associated with current therapies [4]. Breast cancer is predominantly a hormone-dependent disease, with estrogen playing a key role in its initiation and progression. Given that women are primarily affected by this malignancy, targeting estrogen biosynthesis has been established as an effective therapeutic strategy [5]. Aromatase[6–10], a key enzyme in estrogen biosynthesis, catalyzes the conversion of androstenedione into estrogen, making it a critical therapeutic target in the treatment of hormone-dependent breast cancer, particularly in postmenopausal women [11]. Current aromatase inhibitors, such as Exemestane, have demonstrated efficacy in reducing estrogen production; however, they are often associated with challenges such as drug resistance and adverse effects, limiting their long-term clinical success [12]. Moreover, the high costs and lengthy development timelines of novel drug candidates further highlight the need for efficient, cost-effective strategies in anti-cancer drug discovery. Therefore, identifying new molecules with optimized pharmacological profiles, reduced toxicity, and improved affordability remains an urgent priority in breast cancer research [13]. In this context, Benzoxazole derivatives have demonstrated significant potential as pharmacological agents, particularly in anticancer applications. Their unique structural framework, featuring an electron-rich aromatic core and hydrogen bond acceptor sites, allows for strong binding affinity and enhanced specificity toward biological targets such as aromatase [14]. The flexibility of this scaffold supports strategic modifications to optimize potency, selectivity, and pharmacokinetic properties while minimizing off-target effects. Additionally, their cost-effective and straightforward synthesis makes them practical candidates for drug development. These advantages position them as promising molecules for advancing breast cancer treatments through efficient and innovative therapeutic strategies [15–17].

Computational approaches, especially in silico techniques, are crucial to enhance the potential of drug candidates while minimizing development costs. Computer-aided drug design (CADD), which integrates structure-based (SBDD) and ligand-based drug design (LBDD) methodologies, leverages advanced algorithms to accelerate the discovery of new therapeutic agents [18]. Machine learning algorithms, a core component of LBDD, have proven highly

View Article Online
DOI: 10.1039/D5NJ02417J

effective in predicting and optimizing potential drug candidates [19–21]. Additionally, SBDD employs algorithms for molecular docking and molecular dynamics simulations to analyze ligand interactions within enzyme active sites, enabling the identification of key structural features essential for enhancing drug efficacy and specificity [22]. Quantitative structure-activity relationship (3D-QSAR) models, particularly those using comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA), can identify the influence of key molecular descriptors on therapeutic efficacy, guiding the discovery of promising drug candidates [23]. To ensure the reliability of descriptors selected by generated predictive models, advanced artificial intelligence techniques such as artificial neural networks (ANNs) are used for validation and optimization [24]. For this purpose, each model is rigorously validated internally and externally by comparing relevant descriptors with binding interactions via molecular docking simulations [25,26]. In addition, pharmacokinetic profiling through ADMET (absorption, distribution, metabolism, excretion, and toxicity) assessments provide essential information on the similarity and suitability of drug candidates [27]. The molecular dynamics simulations (MD) validate the docking results [28], providing insight into the stability of the protein-ligand complexes studied [29,30]. Finally, retrosynthesis proved decisive in defining viable synthetic pathways for the selected candidates, accelerating the development of new inhibitors against the investigated pathology for in vitro and in vivo evaluation [31,32].

This study aims to accelerate the development of new aromatase inhibitors for treating hormone-dependent breast cancer by adopting a comprehensive computational approach, while optimizing research costs. To achieve this, the therapeutic relevance of 24 benzoxazole derivatives was examined through a combination of advanced in silico methods. Techniques such as 3D-QSAR modeling (CoMFA and CoMSIA), artificial neural networks (ANN), and molecular docking were applied to identify structural features associated with anticancer activity and to explore how these compounds interact with the aromatase active site. Molecular dynamics (MD) simulations were then used to assess the structural stability of the most promising protein–ligand complexes, employing key indicators like RMSD, RMSF, principal component analysis (PCA), and MM-PBSA binding free energy estimates. Pharmacokinetic and toxicity properties were predicted using ADMET profiling, helping to filter the best candidates. Additionally, retrosynthetic analysis was performed to evaluate the feasibility of synthesizing these molecules for future experimental validation in vitro and in vivo.

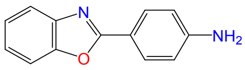
View Article Online
DOI: 10.1039/D5NJ02417J

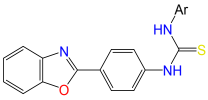
2. Material and methods

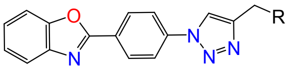
2.1. Data sets

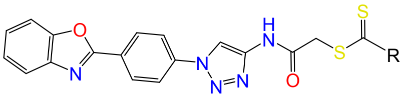
In order to build predictive 3D-QSAR models from a series of 24 benzoxazole derivatives (as shown in Table 1) [14], the dataset of molecular structures with experimental anti-breast cancer activities was divided into two distinct subsets: one for model training and the other for model validation [33,34].

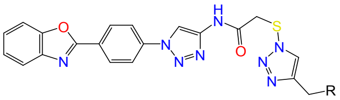
Table 1. Structures of 24 derivatives, IC50, pIC50 and CoMFA/CoMSIA fields (*: test set)

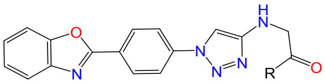

1

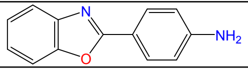
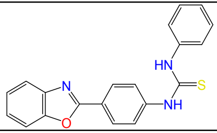
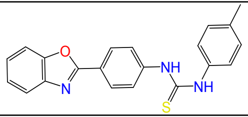
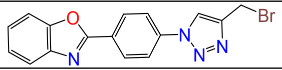
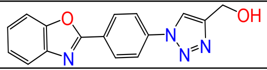
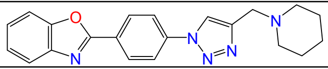
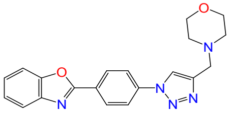
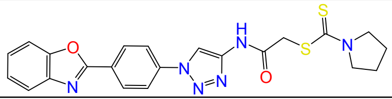
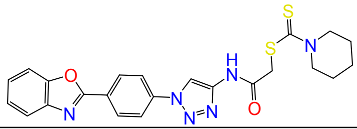

2 /3


4-7

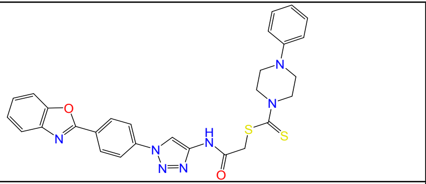
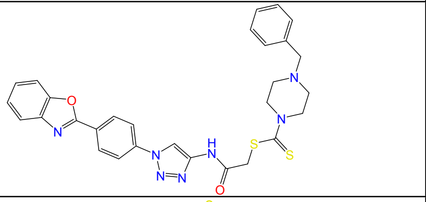
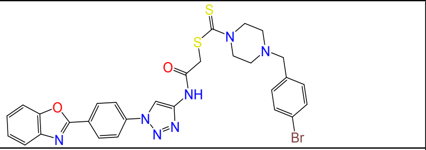
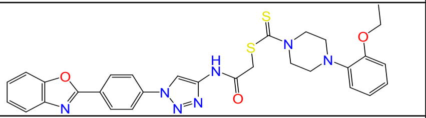
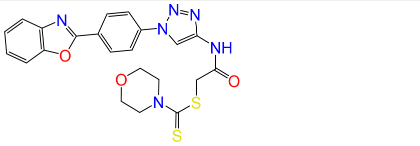
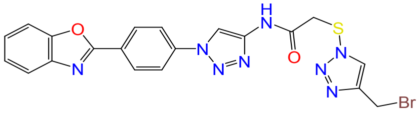
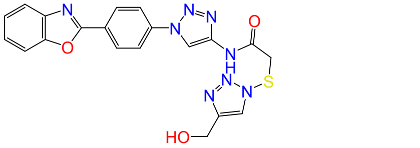
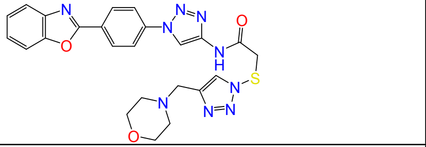
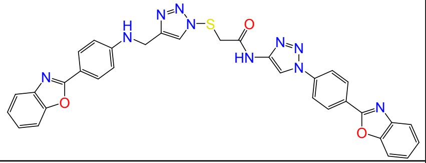
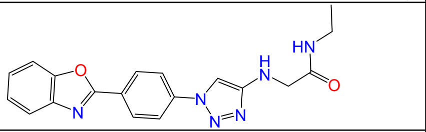
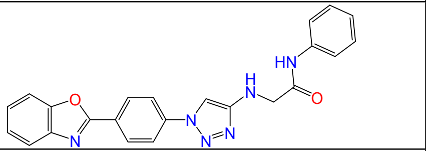
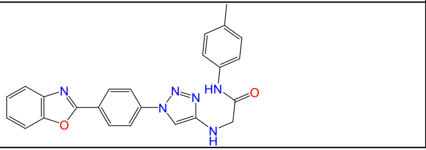

8-14


15-18

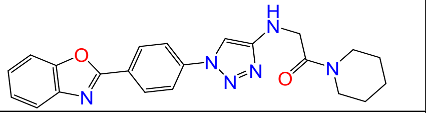
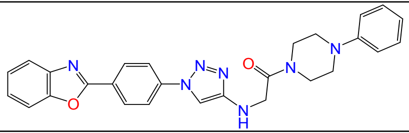
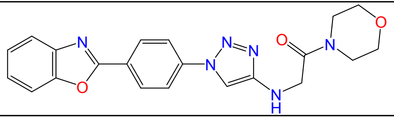

19-24

N°	Structure	IC_{50} (µM)	pIC_{50}	CoMFA	S	E	H	D	A
1		132.190	3.879	84.000	3.122	0.304	3.092	0.927	0.932
2		42.016	4.377	124.000	4.158	0.464	3.822	0.933	0.928
3		219.894	3.658	138.000	4.341	0.471	4.033	0.917	0.930
4		7.647	5.117	114.000	3.584	0.640	3.472	0.000	1.306
5*		1.000	6.000	112.000	3.624	0.638	3.190	0.399	1.870
6		23.860	4.622	148.000	4.449	0.671	3.449	0.000	1.333
7*		0.271	6.567	142.000	4.336	0.710	3.551	0.000	1.984
8		26.642	4.574	168.000	4.576	0.939	3.668	0.671	2.229
9		39.321	4.405	172.000	4.799	0.951	3.786	0.663	2.221

View Article Online
DOI: 10.1039/D5NJ02417J

10		3.782	5.422	202.000	5.105	0.996	4.247	0.652	2.184
11		1.179	5.928	210.000	5.230	1.000	4.334	0.657	2.212
12		5.465	5.262	212.000	5.419	0.993	4.898	0.666	1.9[illegible]
13		132.619	3.877	218.000	5.352	1.015	4.120	0.648	2.206
14*		1.206	5.919	172.000	4.745	0.945	3.893	0.659	2.661
15		12.395	4.907	160.000	4.444	0.965	3.734	0.671	2.396
16		5.631	5.249	158.000	4.460	1.064	3.474	0.839	2.626
17		8.273	5.082	184.000	5.032	1.124	3.863	0.672	2.877
18		5.415	5.266	234.000	5.348	1.112	4.608	0.935	2.572
19		33.054	4.481	140.000	4.218	0.730	3.446	0.942	1.679
20*		3.293	5.482	160.000	4.411	0.823	3.665	1.054	2.29[illegible]
21*		1.071	5.970	160.000	4.563	0.736	3.818	0.935	1.676

View Article Online
DOI: 10.1039/D5NJ02417J

22		3.665	5.436	164.000	4.708	0.832	3.587	0.671	2.287
23		483.306	3.316	174.000	5.308	0.883	4.403	0.659	2.259
24		5.495	5.260	156.000	4.597	0.857	3.831	0.671	2.722

pIC_{50}=-log IC_{50}

2.2. *Methodology*

2.2.1. *CADD using LBDD and SBDD approaches*

Building CoMFA and CoMSIA models using partial least squares (PLS) regression necessitates accurate molecular alignment [35,36]. For this, the molecular structures were designed with SYBYL-X.2.1, optimized using the Tripos force field and Gasteiger-Hückel charges, and stabilized through the Powell gradient method [37]. CoMFA utilized an sp3 carbon probe to generate steric and electrostatic fields, while CoMSIA employed a probe atom to derive additional descriptors, including hydrophobicity, hydrogen bond donor and acceptor properties, steric, and electrostatic fields, with specific initial attenuation and column filtering settings [38]. Various PLS regression models were developed to correlate these fields with biological activity[39]. Artificial neural networks (ANN) were deployed to assess descriptor significance and validate those identified by the 3D-QSAR model [40].

Validation of the models was conducted through multiple techniques, including cross-validation and data partitioning [41]. Model performance was evaluated using statistical metrics such as the coefficient of determination (R^2), mean squared error (MSE), Fisher's value, p-value, and cross-validated coefficient of determination (Q^2) [42]. Predictive accuracy was determined by the external coefficient of determination (R^2_{pred})[43]. Robustness was verified via Y-randomization tests, comparing the results of randomized models with non-randomized counterparts, and additional validation criteria were applied [44–46]. The applicability domain (AD) was established using the leverage approach and William's plot, identifying the chemical space for reliable predictions [47].

Molecular docking simulations were used to study ligand interactions with the aromatase active site (PDB: 3S7S) [48]. Receptor preparation and docking analyses were conducted using Discovery Studio and AutoDock [49]. To validate the docking procedure, the co-

View Article Online
DOI: 10.1039/D5NJ02417J

crystallized ligand was re-docked, and the root mean square deviation (RMSD) was calculated, aiming for a value of less than two angstroms [50].

Pharmacokinetic properties of the drug candidates were assessed using pkCSM [51] and SwissADME [52] were used to examine drug similarity characteristics and address aspects of absorption, distribution, metabolism, excretion, and toxicity (ADMET). To this end, in silico predictions were merged concerning intestinal absorption, blood-brain barrier permeability, central nervous system penetration, biotransformation, clearance, and AMES testing for potential drug candidates [53].

The stability of newly designed candidates has been compared with Ligand 5 [54] and C2 [55], which are prominent compounds from previous studies, as well as the reference drug in breast cancer hormone therapy (Exemestane) [56]. To achieve this, dynamic evaluations were conducted on optimal docking poses to assess and compare the stability of protein-ligand interactions [57]. Input files for MD calculations were generated via the CHARMM-GUI solution generator, using CHARMM force field parameters for proteins. The Param-Chem server established Ligand topology using the general CHARMM force field. The CHARMM-GUI solution generator comprises a sequence of five steps. In the first step, the tool reads the coordinates of the protein-ligand complex. The second step involves solving the protein-ligand complex and determining the shape and size of the system. Na^+ and Cl^- ions are introduced in this step to neutralize the system. The third step defines periodic boundary conditions (PBC) to mimic an extended system using a unit cell replicated in all directions. The simulation exclusively considers the atoms inside the PBC box, eliminating erroneous contacts by brief minimization. The fourth and fifth steps include system balancing and production. Balancing takes place in the NVT and NPT assemblies to achieve the desired temperature and pressure. Input files for balancing and production are acquired, and adjustments are made, such as the definition of MD execution steps, trajectory save frequency, and energy calculation. All MD calculations, balancing, and production cycles were performed using GROMACS 2020.2. Initially, all complexes were immersed in a cubic box of TIP3P water. Na^+ and Cl^- ions were randomly substituted for the water molecules to neutralize the system's net atomic charge. PBC was applied, considering the system's shape and size. Unbound interactions were managed with a cut-off distance of 12 Å, while the neighbor search list was buffered using Verlet's cut-off scheme. Long-range electrostatic interactions were handled using Ewald's particle mesh method (PME). The protein-ligand complex adhered to the CHARMM36 force field. Before production simulation, the system underwent energy minimization via the steepest descent algorithm (5000

View Article Online
DOI: 10.1039/D5NJ02417J

steps). Next, balancing was performed using the NVT and NPT packages, simulating for 125 ps at 300.15 K with positional constraints of 400 kJ/mol.nm^2 and 40 kJ/mol.nm^2 on the backbone and side chains. Finally, the complex was subjected to a 100 ns production simulation in an NPT assembly at 300.15 K and 1 bar. The Hoover nose thermostat maintained temperature, and the Parrinello-Rahman barostat maintained pressure. H-bonds were constrained using the LINCS algorithm based on CHARMM-GUI data. A V-scale thermostat at 300 K with a coupling constant of 1 ps was used, and trajectories were saved every 2 ps during 100 ns simulations in the NPT assembly [58,59]. GROMACS tools were employed to scrutinize molecular dynamics (MD) simulations, enabling a comprehensive understanding of the dynamic behavior of atoms within the protein-ligand complex [60]. The gmx_rms subprogram, known for its precision, facilitated the computation of the root mean square deviation (RMSD) between the positions of protein and ligand atoms. Additionally, the gmx_rmsf function was applied to assess the root mean square fluctuations (RMSF) centered on the C-alpha atoms of the protein [61]. Furthermore, the gyration radius of each protein atom was accurately determined using the gmx_gyrate tool. The calculation of hydrogen bond occurrences within the protein-ligand interaction was carried out with the gmx_hbond tool. Additionally, throughout the simulation, the center of mass distance between the protein and ligand was systematically assessed using gmx_distance [62]. To gain further insights into the system's behavior, the VMD molecular graphics application was utilized for trajectory analysis and to investigate the frequency of protein-ligand interactions, offering valuable information about the system's dynamic behavior [30].

The selected systems for further investigation were subjected to Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) calculations using the g_mmpbsa tool within the GROMACS software suite, as outlined in Eq. 1 [63] :

$$\Delta G_{binding} = \Delta G_{complex} - (\Delta G_{protein} + \Delta G_{ligand}) \qquad \text{(Eq. 1)}$$

Here, ΔG complex signifies the overall free energy of the protein-ligand complex, while ΔG protein and ΔG ligand concurrently represent the total free energy of the separated protein and ligand in the solvent, respectively. The g_mmpbsa tool enables the determination of each residue's energy contribution to the binding energy, offering a breakdown of the binding energy. Specifically, ΔEMM, ΔGpolar, and ΔGnon-polar are individually computed for each residue. These values are then aggregated to ascertain the cumulative contribution of each residue to the overall binding energy. Notably, the g_mmpbsa tool is proficient in handling files generated by

View Article Online
DOI: 10.1039/D5NJ02417J

specified GROMACS versions; thus, GROMACS 5.1.4 was employed to recreate the binary run input file (.tpr). The molecular structure file (.gro), topology file (.top), and MD-parameter file (.mdp), all derived from the MD process, serve as inputs to regenerate the binary run input file [64].

2.2.2. Retrosynthesis of new drug-candidates

Retrosynthesis method was employed to design synthetic pathways for the drug candidates. This approach, based on the structural breakdown of target compounds, allows the identification of multiple synthetic routes [65,66]. For this, we used advanced computational tools, including the IBM RXN for Chemistry database, reflecting the progress in computer-assisted synthesis planning [67,68].

3. Results and discussion

3.1. Alignment of molecular structures

Molecular 7 was chosen as the template compound, illustrated in Fig. 1, to align the chemical compounds studied and create contour maps for CoMFA and CoMSIA analyses of reliable 3D-QSAR models.

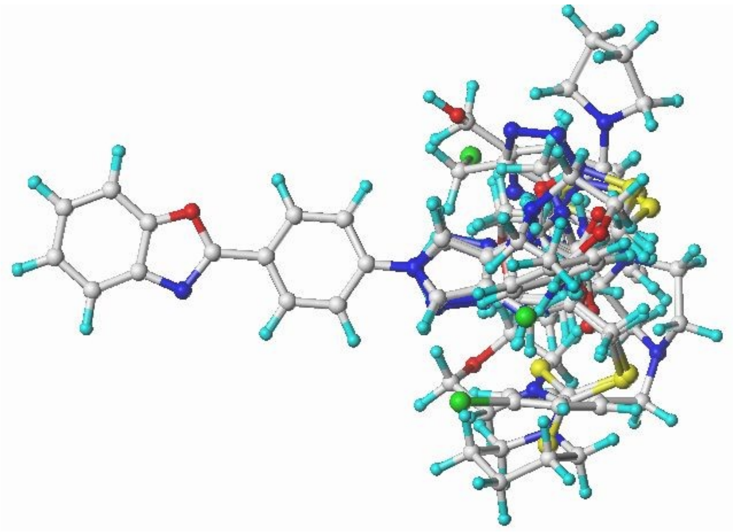


Fig. 1. Alignment of compounds

3.2. Generation of 3D-QSAR models

In this research, ten different field combinations were utilized to create the CoMFA and CoMSIA models. The detailed statistical performance of these models is presented in Table 2. The selection of the best model was based on the highest coefficients of determination (R^2) and

View Article Online
DOI: 10.1039/D5NJ02417J

cross-validation values (Q^2), coupled with the lowest standard error of estimation (SEE), number of principal components (N), and F-test significance level.

Table 2. Models' statistical parameters and fields of models (S: Steric - E: Electrostatic- H: Hydrophobic- D: Donnor of HBond - A: Acceptor of HBond)

Fields of models (CoMFA/CoMSIA)	Q^2	N	SEE	R^2	F
S-E (CoMFA)	0.391	1	0.396	0.696	38.855
S-E-H	0.483	2	0.275	0.862	49.911
S-E-D	0.476	2	0.300	0.836	40.733
S-E-A	0.417	1	0.416	0.646	33.644
E-H-A	0.546	2	0.084	0.989	244.998
S-E-H-D	0.523	2	0.268	0.869	53.080
S-E-H-A	0.465	2	0.284	0.853	46.410
S-E-D-A	0.441	2	0.308	0.827	38.246
E-H-D-A	0.492	2	0.285	0.851	45.837
S-E-H-D-A	0.489	2	0.279	0.857	48.135

Analysis of the statistical parameters presented in Table 2 indicates that the model with the highest Q^2 (0.546) and R^2 (0.989), the lowest standard error of estimate (SEE) (0.084), and the most notable F-value (244.998) with two principal components demonstrated optimal performance. This model incorporated electrostatic, hydrophobic, and hydrogen bond acceptor (HBA) fields and achieved the highest coefficient of determination for external prediction (R^2_{pred}= 0.915). In contrast, the CoMFA model and other CoMSIA models were excluded due to their insufficient internal validation parameters (Q^2 less than 0.5) and high SEE value (0.396).

To validate the descriptors selected using the CoMSIA/EHA model, the artificial neural network (ANN) technique with a 3-3-1 architecture was employed, and the parameter ρ was calculated using equation (Eq. 2) as follows [69]:

$$\rho = \frac{N}{H(I+O+1)+O} \qquad \text{(Eq. 2)}$$

Here, N, H, I, and O represent the number of molecules in the training set, the hidden layers, the input layers, and the output layers.

View Article Online
DOI: 10.1039/D5NJ02417J

A ρ value of 1.25 indicates that the number of hidden layers (3) must be proportional to the number of descriptors (input layers) to predict pIC50, which is represented by a single output layer (Fig. 2)[70–72].

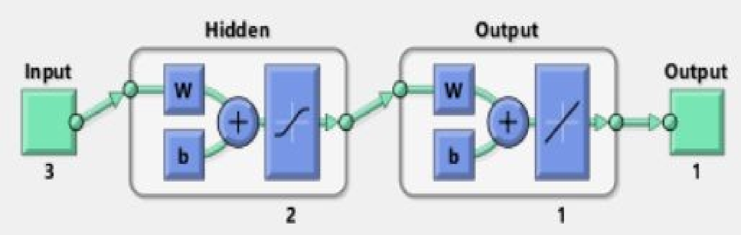


Fig. 2. ANN architecture

The substantial coefficient of determination (R² = 0.754), low mean square error (MSE = 0.127) and high validation-test coefficient (R^2_{test} = 0.887) collectively confirm the effectiveness of the ANN model in predicting the biological activities studied. As a result, the CoMSIA/EHA model may exhibit a remarkable degree of stability and predictability. To assess the predictive power of optimal models (3D-QSAR and ANN), it is imperative to compare predicted pIC50 values with observed values, as shown in Table 3.

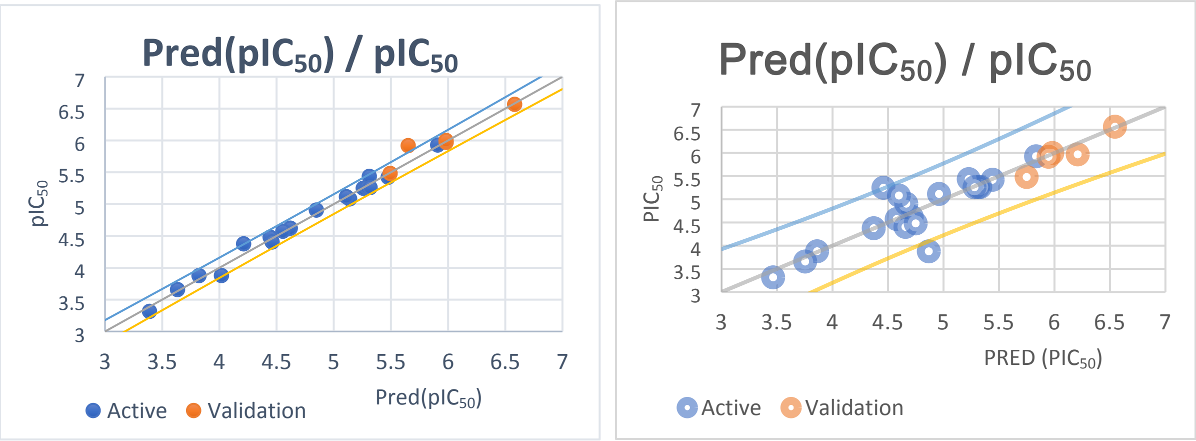
Table 3. Compounds' predicted pIC_{50} ($pIC_{50\ pr.}$) for training and test set (*)

N°	pIC_{50}	$pIC_{50\ pr.}$ (CoMSIA)	$pIC_{50\ pr.}$ (ANN)	-	N°	pIC_{50}	$pIC_{50\ pr.}$ (CoMSIA)	$pIC_{50\ pr.}$ (ANN)	-	N°	pIC_{50}	$pIC_{50\ pr.}$ (CoMSIA)	$pIC_{50\ pr.}$ (ANN)
1	3.879	4.018	3.861	-	**9**	4.405	4.461	4.657	-	**17**	5.082	5.139	4.598
2	4.377	4.212	4.371	-	**10**	5.422	5.476	5.443	-	**18**	5.266	5.288	5.331
3	3.658	3.634	3.753	-	**11**	5.929	5.909	5.834	-	**19**	4.481	4.444	4.749
4	5.117	5.111	4.959	-	**12**	5.262	5.308	5.311	-	**20***	5.482	5.493	5.743
5*	6.000	5.982	5.981	-	**13**	3.877	3.822	4.867	-	**21***	5.970	5.983	6.209
6	4.622	4.620	4.714	-	**14***	5.919	5.651	5.947	-	**22**	5.436	5.311	5.227
7*	6.567	6.582	6.545	-	**15**	4.907	4.848	4.664	-	**23**	3.316	3.389	3.464
8	4.574	4.555	4.578	-	**16**	5.249	5.258	4.458	-	**24**	5.260	5.318	5.279

To assess the accuracy of the models studied, it is conventional to construct a graphical representation of predicted values (pIC50pred) versus observed biological activities (pIC50). A line graph represents a favorable result in the positive quadrant of the pIC_{50} and pIC_{50pred} axes.

View Article Online
DOI: 10.1039/D5NJ02417J

Furthermore, the almost equal slopes of each model confirm the strength of the correlation, indicating a high correlation. This correlation is visually demonstrated in Fig. 3 (A, B) for the training and test sets.



(A) **(B)**

Fig. 3. Graphical comparison of observed vs. predicted pIC50 values for CoMSIA/EHA model (A) and ANN model (B)

As shown in Fig. 3 (A, B) and reported in Table 3, there is a notable correlation between observed and predicted pIC50 values. This highlights the model's robust ability to predict pIC$_{50}$ values for new compounds suitable for breast cancer treatment.

Before implementing the optimal model for predicting new breast cancer drug candidates, it is imperative to validate its predictive ability according to the statistical criteria established by Golbraikh, Tropsha, and Roy (Table 4).

Table 4. CoMSIA/EHA statistical criteria

Indicator of statistics	Score	Threshold	Validation Score
R^2_{pred}	0.915	> 0.600	Validated
R_0^2	0.915	> 0.600	Validated
$R_0'^2$	0.915	> 0.600	Validated
$\left\|R_0^2 - R_0'^2\right\|$	0.000	< 0.300	Validated
$\frac{R^2 - R_0^2}{R^2}$	0.075	< 0.100	Validated
$\frac{R^2 - R_0'^2}{R^2}$	0.075	< 0.100	Validated
K	1.001	$0.850 \leq K \leq 1.150$	Validated
K’	0.992	$0.850 \leq K' \leq 1.150$	Validated
$R_m^2 = R^2(1-\sqrt{(R^2-R_0'^2})$	0.718	> 0.600	Validated
$R_m'^2 = R^2(1-\sqrt{(R^2-R_0'^2})$	0.718	> 0.600	Validated

View Article Online
DOI: 10.1039/D5NJ02417J

R_0^2 : Determination coefficient for the zero-intercept line in the plot comparing predicted versus observed activities.

$R_0'^2$: Determination coefficient for the zero-intercept line in the plot comparing observed versus predicted activities.

K: Zero-intercept slope for the relationship between predicted and observed activities in the test set.

K': Zero-intercept slope for the relationship between observed and predicted activities in the test set.

In accordance with the statistical criteria described in Table 4, the CoMSIA/EHA model is considered validated, to the exclusion of any characterization as a random model. To ensure this, an assessment of the stability of the proposed CoMSIA/EHA model was carried out using the Y randomization test, as shown in Table 5.

Table 5. Statistics data of Y-randomization test

Y-randomization Iterations (It.)	Random statistical parameters		
	$\boldsymbol{Q^2_{rand}}$	$\boldsymbol{R^2_{rand}}$	$\boldsymbol{cR^2_p}$
It.1	0.174	0.523	0.675
It. 2	-0.063	0.401	0.758
It. 3	0.077	0.422	0.745
It. 4	-0.306	0.326	0.656
It. 5	-0.645	0.424	0.743

The results shown in Table 5 validate the reliability of the model. Specifically, the values of Q^2_{rand}, R^2_{rand} and cR^2_p demonstrate that the model's predictions are not due to random correlation. Therefore, the model is considered reliable for predicting the efficacy of potential new drug-candidates for breast cancer within a specified applicability domain.

3.3. Applicability domain of CoMSIA/EHA model

The applicability domain (AD) of the CoMSIA/EHA model was assessed using William's plot (Fig. 4), which analyzes leverage and normalized residuals for each compound in both the training and test sets. The analysis indicates that, with the exception of one outlier, compound 3, which shows an unusual, normalized residual, the leverage values for all other compounds are below the alert threshold (h* = 0.632). This confirms that all compounds fall within the model's applicability domain. Consequently, the reliability of the predicted activity values for the compounds under study is affirmed, allowing for further analysis.

View Article Online
DOI: 10.1039/D5NJ02417J

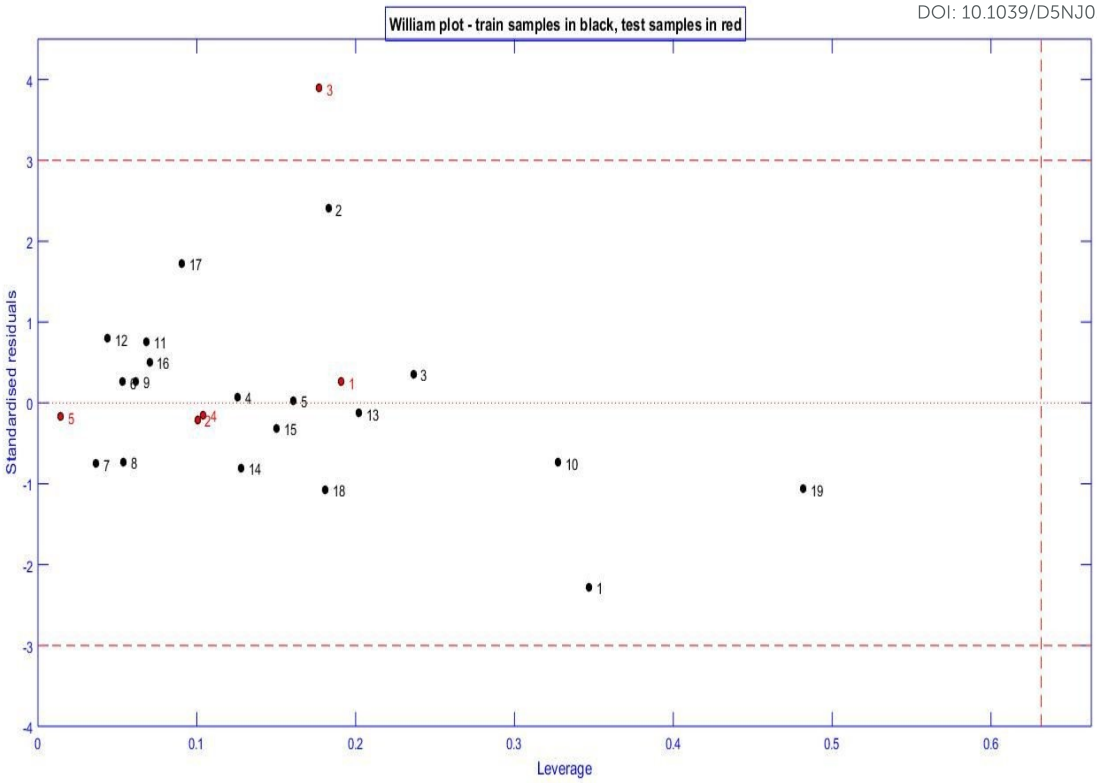


Fig. 4. Applicability domain of the model

3.4. Contour maps visualization of CoMSIA/EHA model

The most accurate 3D-QSAR model was illustrated through CoMSIA/EHA contour maps, using the most active compound (7) as the reference. Fig. 5 (a-c) displays the electrostatic (a), hydrophobic (b), and hydrogen bond acceptor (c) fields. The electrostatic field is depicted with blue and red contours; blue contours highlight regions with positive electrostatic interactions, while red contours indicate areas with negative electrostatic interactions. The hydrophobic field is illustrated with yellow and white contours, where yellow contours represent regions that favor hydrophobic interactions, and white contours denote areas with less steric influence. The hydrogen bond acceptor field is shown with magenta and cyan contours, with magenta representing regions conducive to hydrogen bond acceptance and cyan indicating regions that are less favorable for such interactions. These maps are crucial for understanding the molecular interactions within the system..

View Article Online
DOI: 10.1039/D5NJ02417J

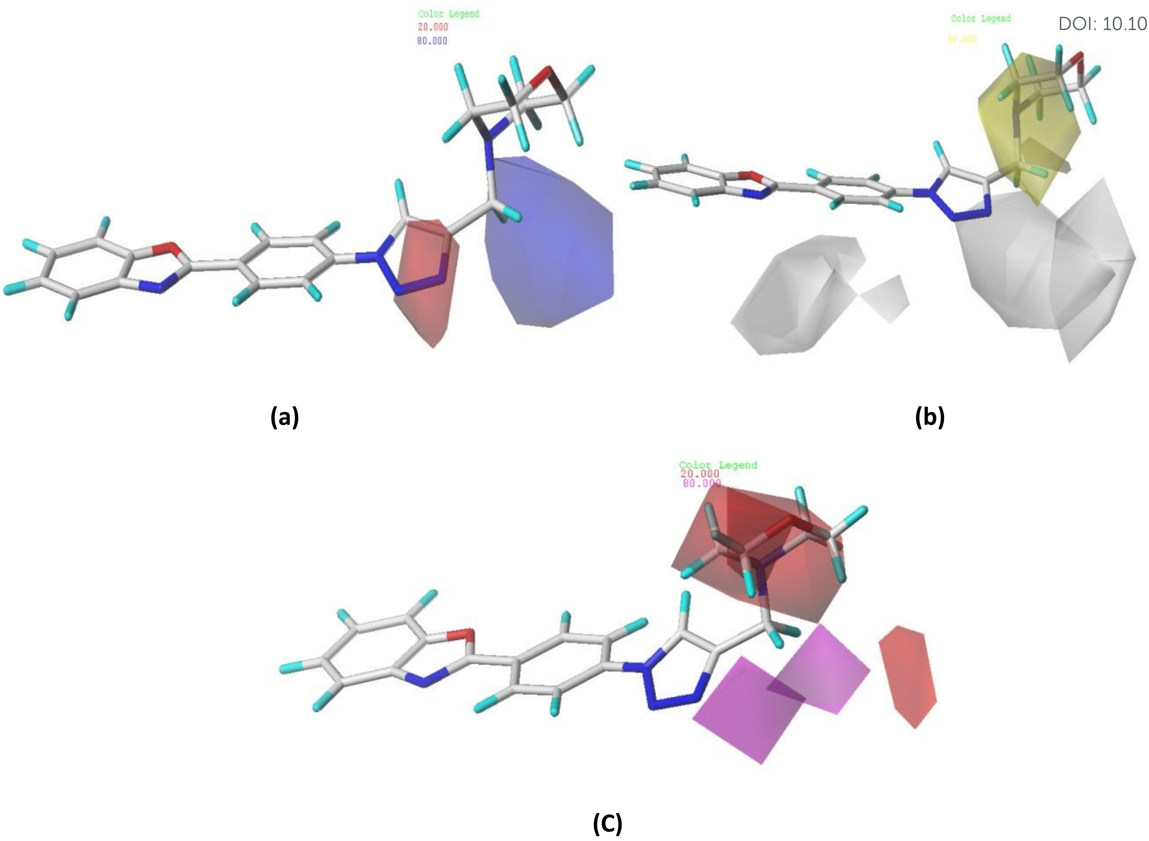


Fig. 5. CoMSIA/EHA Fields of the Most Active Molecule (14): (a) Hydrophobic Interactions (Blue: Favored, Red: Disfavored), (b) Electrostatics (Yellow: Favored, White: Disfavored), and (c) H-Bond Acceptors (Cyan: Favored, Magenta: Disfavored)

Examining the electrostatic field contour maps from CoMSIA shows a clear alignment with those from CoMFA. The blue contour surrounding the altered R group suggests that electron-withdrawing groups could boost biological activity, indicating that strongly electronegative atoms or groups may enhance the desired effect. In contrast, the red contour around the 1,2,3-triazole ring implies that substituting hydrogen with electron-donating groups might improve activity, suggesting that electropositive groups in this area could increase the compound's effectiveness against breast cancer. Yellow contours around the modified R group suggest that introducing hydrophobic groups in this area could decrease activity. White contours around regions distant from bulky groups indicate a preference for smaller substituents to potentially enhance anti-breast cancer activity. Cyan contours at the 1,2,3-triazole ring near the R group suggest that adding a hydrogen bond acceptor in these positions could improve breast cancer inhibition. Similarly, magenta contours suggest that placing a hydrogen bond acceptor in the most critical positions, away from the R group, could further boost activity. These observations are clarified by the fact that the most potent compound (7), whose molecular size is smaller than that of the least active molecule (23), has a more electrostatic, hydrophobic and hydrogen-

View Article Online
DOI: 10.1039/D5NJ02417J

bond-acceptor group in the R position that is less steric. In contrast, the less effective compound (23) has a less electrostatic group at the larger, more voluminous R-position, which may crowd the 1,2,3-triazole ring (the most hydrogen-bond-acceptor).

According to the results of Fig. 5 and based on the CoMSIA field fraction significance values presented in Table 6, enhancing acceptor hydrogen bonding and hydrophobic and electrostatic interactions at a smaller group level could potentially increase efficacy against breast cancer. This can be achieved by substituting the most active molecule's modified (R) group.

Table 6. Comparative fraction analysis of selected fields

COMSIA/ EHA Fields	Fraction
ELECTROSTATIC (**E**)	0.431
Hydrophobic (**H**)	0.221
ACCEPTOR (**A**)	0.348

3.5. Studies via molecular docking method

Molecular docking was performed to validate the CoMSIA/EHA analysis results and to thoroughly examine the binding interactions between various compounds and the aromatase enzyme. We first focused on the interactions between the reference ligand (exemestane) and the receptor (3S7S), which were visualized using Discovery Studio software, as illustrated in Fig. 6. Study identified the active site of the target protein (PDB ID: 3S7S) and highlighted key amino acids, including ARG115, MET374, ILE133, PHE134, PHE221, LEU477, VAL370, VAL373, ALA306, and TRP224. These amino acids are crucial for understanding the binding mechanism and assessing the docking interactions, as detailed in prior literature.

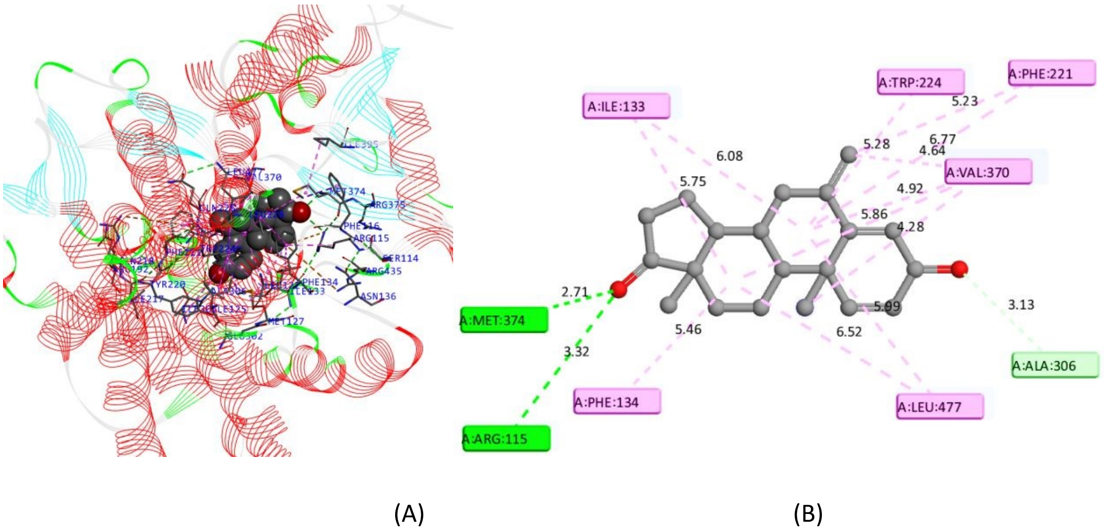


Fig. 6. Interactions 3D (A) and 2D (B) of the 3S7S-Exemestane complex

View Article Online
DOI: 10.1039/D5NJ02417J

Fig. 6 shows that exemestane, the standard aromatase inhibitor, exhibits significant hydrogen-bonding and hydrophobic interactions with key amino acids, which is vital for its effectiveness in breast cancer treatment. Redocking with a new co-ligand was conducted at the same active site (PDB ID: 3S7S) to verify the accuracy of the molecular docking approach used in this study. Grid maps were created with dimensions of size_x = 48, size_y = 52, and size_z = 44, employing a default grid spacing of 0.375 Å. The central grid box coordinates were set to 86.031 Å, 54.004 Å, and 46.404 Å, based on the ligand's initial position. Subsequently, a 3D visualization of the binding interactions between the superimposed ligands and the protein was generated using Discovery Studio software, as depicted in Fig. 7.

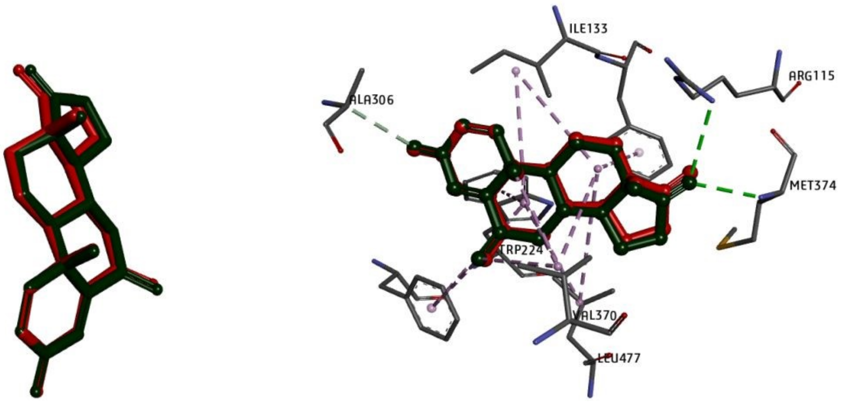


Fig. 7. Docking validation via superposition re-docked (green) and original (red) ligands

To evaluate the accuracy of the molecular docking procedure, we compared the lowest energy conformation obtained from the docking simulations with that of the original ligand. Root-mean-square deviation (RMSD) between the superimposed structures of these two ligands was calculated to be 0.251Å, indicating that the docking method is reliable for subsequent analyses. Consequently, we selected the most active molecule (7) and the least active molecule (23) for docking studies to elucidate their key interactions and various binding modes with the aromatase enzyme, which could contribute to breast cancer inhibition, as illustrated in Fig.8 .

View Article Online
DOI: 10.1039/D5NJ02417J

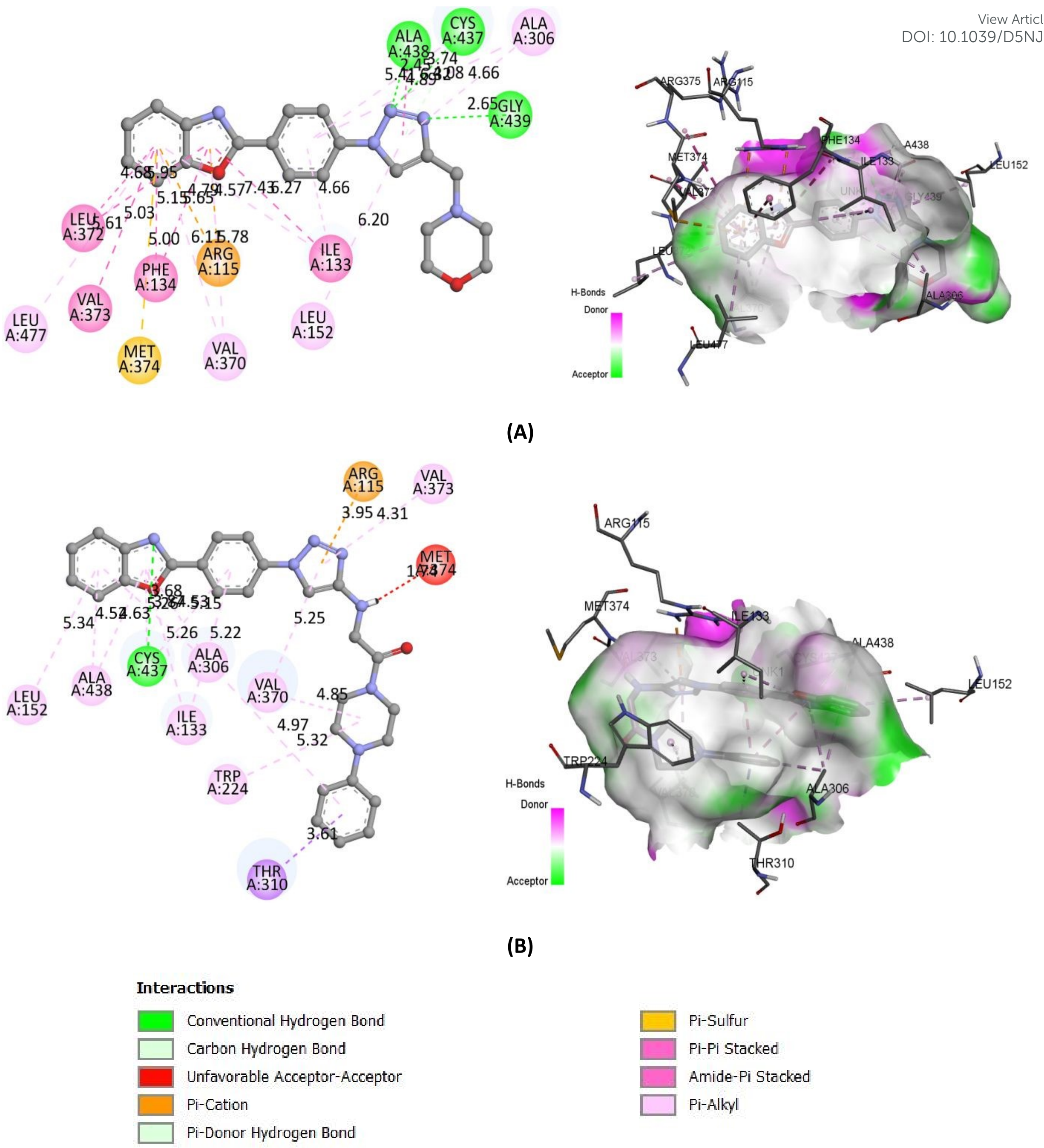


Fig. 8. 2D /3D interactions of the most active molecule 7 (A) and least active molecule 23 (B) with the enzyme's binding site

Fig. 8 depicts the interactions of selected compounds with the aromatase enzyme, showcasing their hydrogen bonding, π-interactions, electrostatic, and hydrophobic interactions with critical amino acids within the enzyme's active site. The most active molecule (7) established three conventional hydrogen bond acceptors, one π-donor hydrogen bond interaction, and three electrostatic and multiple hydrophobic interactions with essential amino acids in the active site, as shown in Fig. 8(A). In contrast, the less active molecule (23) interacted with the same active site through one hydrogen bond

View Article Online
DOI: 10.1039/D5NJ02417J

acceptor interaction, one electrostatic interaction, and various hydrophobic interactions, including π-alkyl, π-π, π-sigma, and alkyl interactions, as illustrated in Fig. 8(B). Additionally, the analysis revealed that the more active molecule (7) had a lower binding energy (-8.900 kcal/mol) compared to the less active molecule (23) with a binding energy of -8.400 kcal/mol, correlating with its higher experimental pIC50 value. Hence, improved affinities for hydrogen bonds (conventional hydrogen bond acceptor and unconventional hydrogen bond donor (C and pi)) and electrostatic interactions, as well as enhanced hydrophobic interactions, may enhance the breast cancer-fighting activity of newly designed agents. These results reinforce and complement CoMSIA's current findings (the best CoMSIA/EHA model) in terms of reliability for the discovery of the most potent drug candidates.

3.6. Identification of new anti-beast cancer candidates

3.6.1. Novel molecules as drug-candidates and selection of hits

To optimize bioactivity and streamline drug development, 12 compounds (L1-L12) were designed based on predictive models and computational data. The zinc database provided a rich source of chemically viable and non-toxic molecular fragments, minimizing the risk of adverse effects [73]. These fragments were carefully aligned with the predictions of the generated models to enhance activity through targeted substitutions. Indeed, using well-characterized fragments from the database guarantees the feasibility of synthesis and simplifies retrosynthetic analysis. This approach reconciles improved pharmacological potential, reduced toxicity, and cost-effective synthesis, making these compounds suitable candidates for further experimental validation. Subsequently, the newly designed ligands were optimized and aligned within a database to predict their activities and facilitate comparison with the reference drug, Exemestane, as well as previously designed drug candidates, Ligand 5 and C2 (Fig. 9a, Fig. 9b and Table 7) [54–56].

View Article Online
DOI: 10.1039/D5NJ02417J

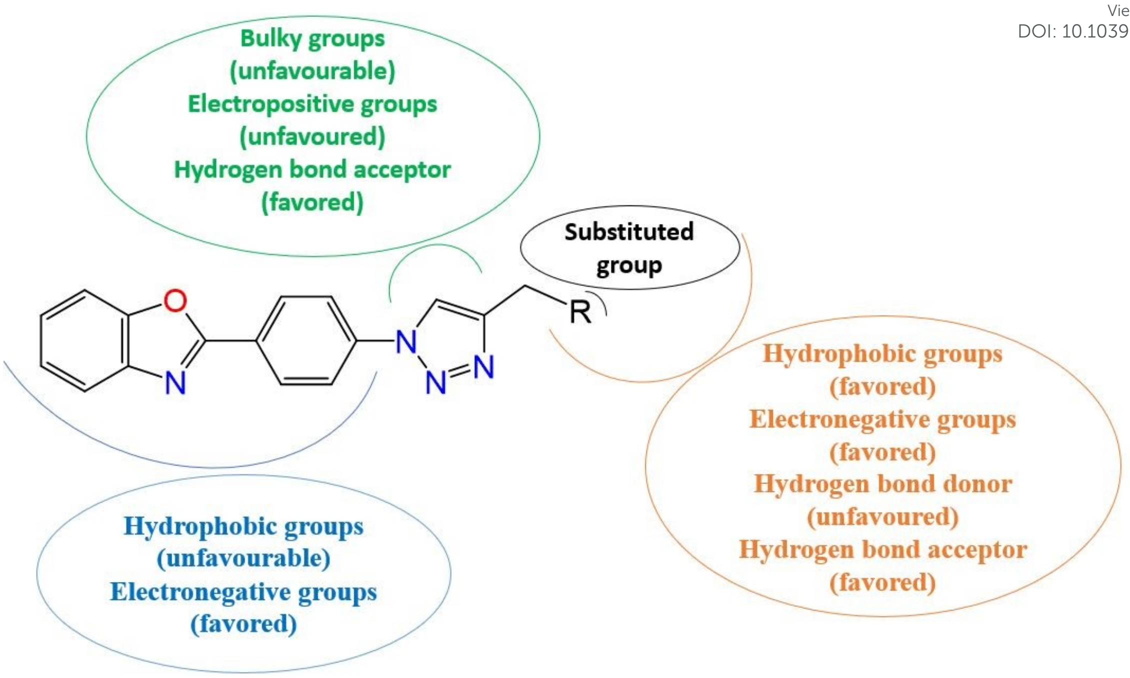
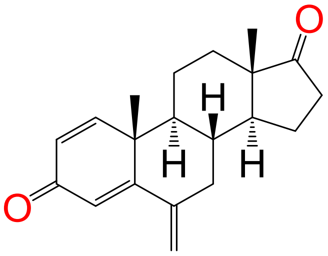
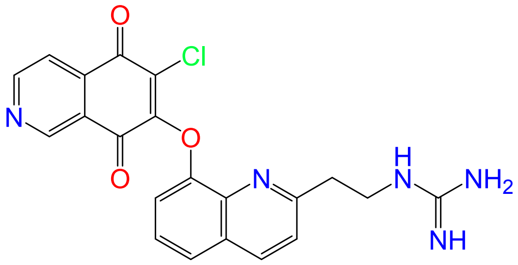

Bulky groups (unfavourable)
Electropositive groups (unfavoured)
Hydrogen bond acceptor (favored)
Substituted group
R
O
N
N
N=N
Hydrophobic groups (favored)
Electronegative groups (favored)
Hydrogen bond donor (unfavoured)
Hydrogen bond acceptor (favored)
Hydrophobic groups (unfavourable)
Electronegative groups (favored)

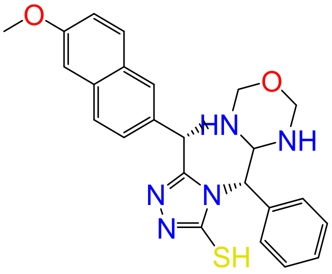
Fig. 9a. Structural requirements based on CoMSIA/EHA contour maps and molecular docking interactions



Exemestane



Ligand 5



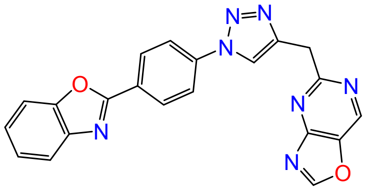
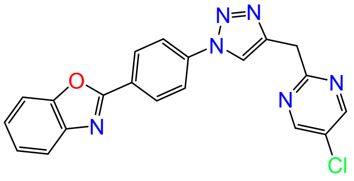
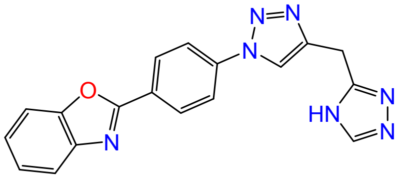
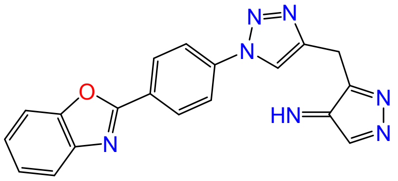
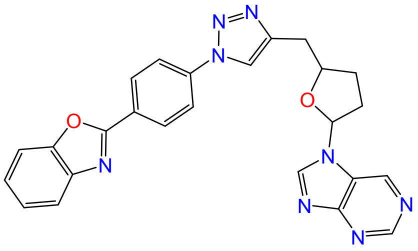
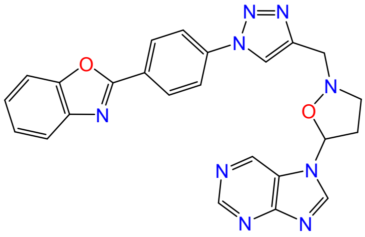
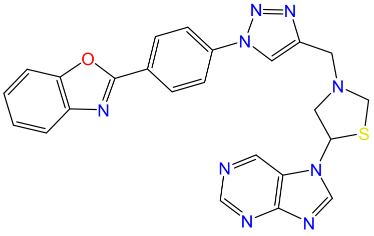
C2

Fig. 9b. Structurs of reference drug, Ligand 5 and C2 used for comparison

The results indicated that most newly designed drug candidates demonstrated significant biological activity (Table 7). It is crucial to note that derivatives falling outside the acceptable applicability domain (where hi exceeds h*) are excluded from consideration as viable drug candidates. Additionally, drug similarity was assessed using Lipinski's rule of five, and synthetic accessibility (SA) was evaluated with the SwissADME online tool. These evaluations are critical in the pharmaceutical industry when selecting promising drug candidates. Table 7 summarizes the biological activities, applicability domain values (hi), and the desirable drug-like properties predicted for the designed ligands, including Lipinski's rule validation and synthetic accessibility.

View Article Online
DOI: 10.1039/D5NJ02417J

Table 7. Straucture and pIC_{50} of drug-candidates, hi, Lipinsky'validation and SA

N°	Structure	pIC_{50pred}	h_i	AD (h*= 0.632)	Lipinski	SA
L1		6.010	0.461	Inside (h_1 less than h*)	Validated	3.450
L2		6.530	0.252	Inside (h_2 less than h*)	Validated	3.240
L3		6.180	0.528	Inside (h_3 more than h*)	Validated	3.210
L4		6.758	0.626	Inside (h_4 more than h*)	Validated	3.500
L5		7.200	0.560	Inside (h_5 less than h*)	Validated	4.010
L6		6. 508	0.081	Inside (h_6 less than h*)	Validated	4.370
L7		6.841	0.174	Inside (h_7 less than h*)	Validated	4.350

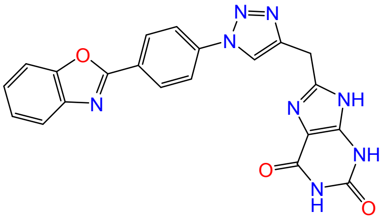
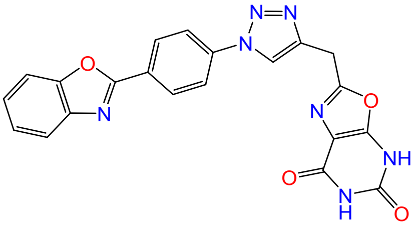
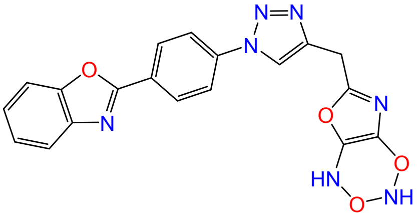
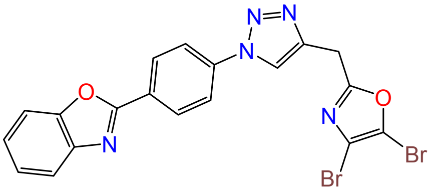
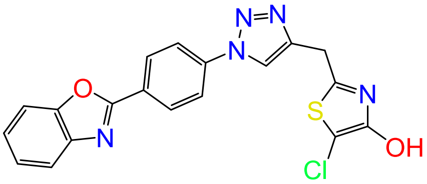
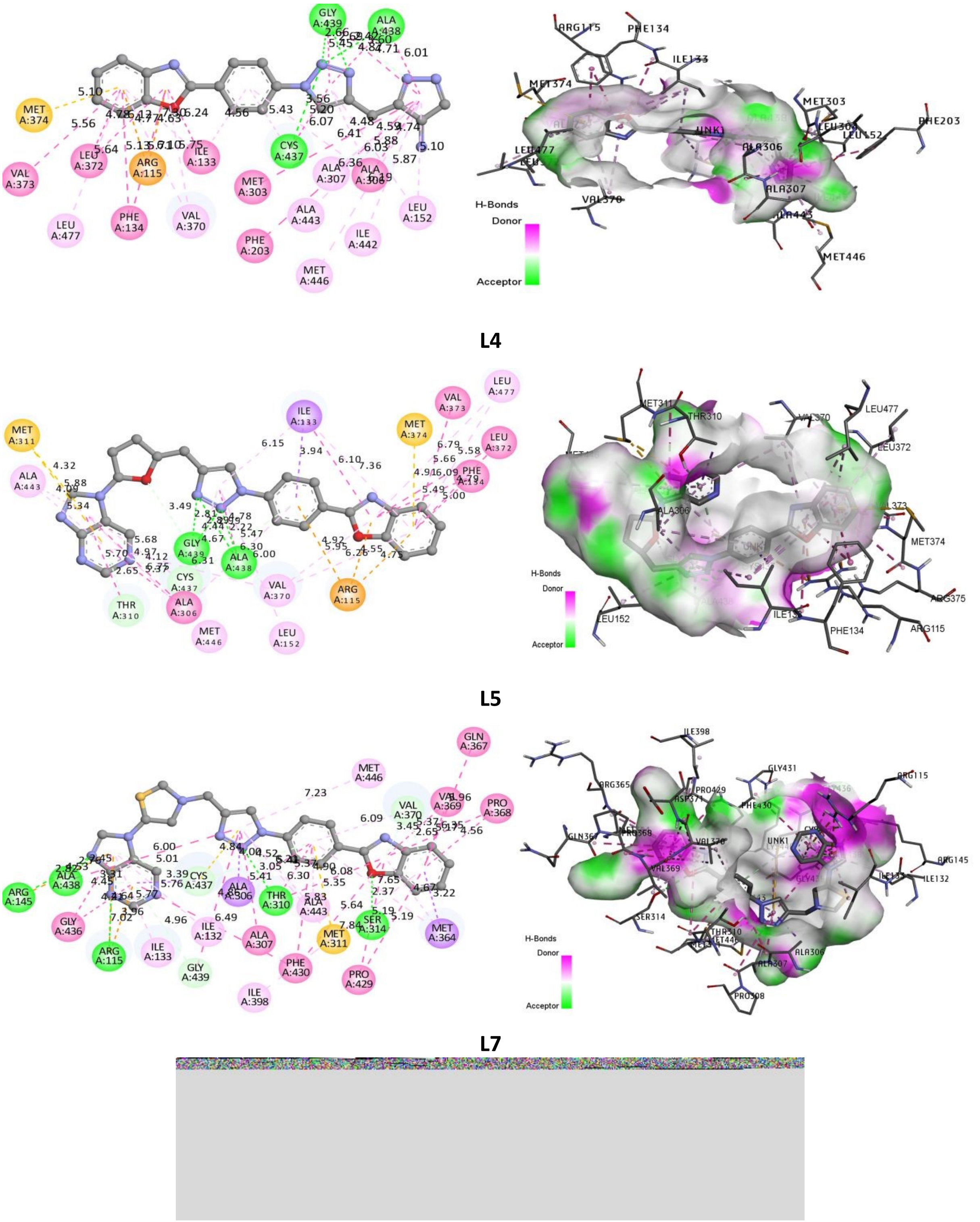
L8		6.506	0.372	Inside (h_8 less than h*)	Validated	3.440
L9		6.380	0.544	Inside (h_9 less than h*)	Validated	3.650
L10		6.089	0.064	Inside (h_{10} less than h*)	Validated	3.850
L11		6.270	0.222	Inside (h_{11} less than h*)	Validated	3.570
L12		6.250	0.203	Inside (h_{12} less than h*)	Validated	3.500

Table 7 indicates that all 12 proposed ligands (L1-L12) conform to Lipinski's Rule of Five. Furthermore, the synthetic accessibility values for these compounds range from 3.380 to 3.960, which is within the acceptable range of 1 to 10. This suggests that the compounds are feasible for synthesis and have potential as drug candidates. All the designed molecules (L1-L12) also meet the criterion for leverage values (hi), which must be less than the threshold value of h* (h* = 0.632), confirming their suitability within the domain of applicability (DA). Comparison of pIC50 values reveals that ligands L4, L5, and L7 exhibit higher anti-breast cancer activity compared to molecule 7. These compounds show promise as potential new drug candidates against breast cancer and require further investigation into their binding interactions with the aromatase active site using molecular docking, as well as stability assessments through molecular dynamics simulations and in silico ADMET evaluations.

View Article Online
DOI: 10.1039/D5NJ02417J

3.6.2. Molecular docking studies of hits compounds

To enhance the reliability of docking interpretations, a validated 3D grid, previously utilized for docking validation, was employed to dock all the most effective ligands to the aromatase active site. Fig. 10 and Table 8 explain the interactions with the aromatase target and binding energies of L4, L5 and L7, as well as reference molecule 7.



L4

L5

L7

Fig. 10. 2D/3D interactions between aromatase and the newly developed ligands (L4, L5 and L7)

View Article Online
DOI: 10.1039/D5NJ02417J

Table 8. Kinds of interactions between aromatase and designed ligands, molecule 7 and exemestane

Ligands	Ligand - 3S7S	Energy of Binding (Kcal /mol)	HB- interactions		Hydrophobic interactions	Electrostatic Interactions
			HB Acceptors	HB Donors (C and Pi)	Alkyl and Pi (Hydrophobic/ Steric)	Pi-ions/ Pi-Sulfur
L 4	L 4 - 3S7S	-9.200	2GLY 439 ; ALA 438 ; CYS 437	GLY 439	VAL 373 ;2 LEU 372 ; LEU 477 ; 2PHE 134 ; 2VAL 370 ; 3ILE 133 ; 2 CYS 437 ; MET 303 ; PHE 203 ; ALA 443 ; ALA 307 ; MET 446 ; ILE 442 ;2LEU 152 ; 3 ALA 306 ; 2 ALA 438 ; 2GLY 439	MET 374 ; 2 ARG 115
L 5	L 5 - 3S7S	-10.700	2 ALA 438 ; 3 GLY 439	3 CYS 437; THR 310	ALA 443 ; THR 310 ; 4 ALA 306 ; CYS 437 ; ALA 438 ; 3 VAL 370 ; LEU 152 ; 2 PHE 134 ; 2 LEU 372 ; 2 LEU 477 ; VAL 373 ; 4 ILE 133	2 MET 311 ; 3 ARG 115 ; MET 374

L 7	L 7 - 3S7S	-10.300	ARG 115;
			ALA 438;
			ARG 115;
			THR 310;
			ER 314
				2 ALA438;
				2 CYS 437;
				GLY 439;
				2 VAL 370
					ALA 438; 2 GLY 436; ILE 133; ILE 132; GLY 435; ALA 307; 3 PHE 430; MET 311; ILE 398; 2 PRO 429; 3MET 364; 2 ALA 443; 2 THR 310; 2 ALA 306; CYS 437; 2 PRO 368; 2 VAL 369; GLN 367; VAL 370; MET 449	2 ARG 115 ;
						ARG 145 ;
						CYS 437 ;
						MET 311

Molecule 7	Molecule 7 - 3S7S	-8.900	CYS 437 ; ALA 438 ; GLY 439	CYS 437	CYS 437 ; 2ALA 306 ; 3 ILE 133 ; LEU 152 ; 2 VAL 370 ; 2 PHE 134 ; VAL 373 ; LEU 477 ; 2 LEU 372	2 ARG 115 ; MET 374
Exemestane	Exemestane - 3S7S	-8.800	MET 374 ; ARG 115 ; ALA 306	-	PHE 134 ; 2 LEU 477 ; 4 VAL 370 ; 2 PHE 221 ; TRP 224 ; 2 ILE 133	-

In this study, the binding energy of the reference molecule was found to be lower (-8.900 kcal/mol) compared to that of the proposed candidates, which had binding energies of -9.200 kcal/mol, -10.700 kcal/mol, and -10.300 kcal/mol for L4, L5, and L7, respectively. This difference implies that the higher pIC50 values observed for L4 (6.758), L5 (7.200), and L7 (6.841) relative to the reference molecule (7) with a pIC_{50} of 6.567 may be due to the enhanced stability of the ligand-receptor complexes with lower binding energies. The increased stability is likely a result of the types and number of interactions between these ligands and the active site of the receptor. Specifically, the stability and biological efficacy of the ligand-receptor complex are strongly correlated with the number of hydrogen bonds (acceptors, C-donors, and pi-donors), as well as electrostatic and hydrophobic interactions with key amino acids that contribute to breast cancer inhibition. The designed ligands (L4, L5, and L7) demonstrate a higher number of these interactions compared to the reference molecule. These results are consistent with findings from previous 3D-QSAR and docking studies.

3.6.3. Study of pharmacokinetic properties

To investigate the pharmacokinetic characteristics of compounds L4, L5 and L7, as well as reference molecule 7 and exemestane, the pkCSM online tool was used [42,49]. Table 9 summarizes the ADMET values for these compounds. Human intestinal absorption (HIA) was categorized into three ranges: low (0-20%), moderate (20-70%), and high (70-100%). The high absorption rates observed for the designed ligands suggest they have strong potential for effective absorption in the human intestine. Regarding the volume of distribution (VDss), a value exceeding 0.45 indicates significant distribution capability. Thus, all the designed compounds are expected to have considerable distribution potential throughout the body. For central nervous system (CNS) permeability and blood-brain barrier (BBB) penetration, substances with a LogBB value below -1 are unlikely to distribute well in the brain, whereas those with a LogBB value above 0.3 are more likely to cross the BBB. Similarly, compounds with a LogPS value above -2 can generally penetrate the CNS, while those with a LogPS value below -3 might face challenges in doing so. Therefore, only some of the proposed compounds are anticipated to effectively cross these physiological barriers.

Among the enzyme families involved in drug metabolism, CYP3A4 is an important inhibitor. In contrast to reference molecule 7, the ligands designed (L4, L5 and L7) proved to be either inhibitors or substrates of CYP3A4. Drug clearance, which assesses the efficiency with which substances are eliminated from the body, is not a problem for these compounds according to Table 9. Toxicity assessment, a critical step in early drug development, was conducted using the AMES test. The results in Table 9 indicate that the designed ligands are non-toxic, whereas the most active compound (molecule 7) exhibited some level of toxicity. Overall, according to the ADMET analysis, the most promising ligands (L4, L5 and L7) show favorable pharmacokinetic properties, with L5 standing out as the most active candidate, requiring further study of its stability in relation to the therapeutic target.

Table 9. Selected compounds' ADMET properties, with molecule 7 and exemestane

Ligands	**ADMET** Proprieties												
	Absorption	**D**istribution			**M**etabolism							**E**xcretion	**T**oxicity
	Intestinal absorption	**VDss**	**BBB**	**CNS**	**CYP**							**Total clearance**	**AMES toxicity**
					Substrate		Inhibitor						
					2D6	**3A4**	**1A2**	**2C19**	**2C9**	**2D6**	**3A4**		
	Numeric (% Absorbed)	Numeric (Log L/kg)	Numeric (Log BB)	Numeric (Log PS)	Categorical (Yes/No)							Numeric (log mL min^{-1} kg^{-1})	Categorical (Yes/No)
L4	100	0.335	-1.054	-2.444	No	Yes	Yes	Yes	Yes	No	Yes	0.866	No
L5	100	-0.043	-1.979	-3.634	No	Yes	No	No	Yes	No	Yes	0.733	No
L7	100	0.009	-2.081	-3.380	No	Yes	No	Yes	Yes	No	Yes	0.520	No
Molecule 7	98.684	0.850	-0.715	-2.448	No	Yes	Yes	No	No	No	No	0.835	Yes
Exemestane	100	0.472	0.142	-2.267	No	Yes	No	Yes	No	No	No	1.015	No

View Article Online
DOI: 10.1039/D5NJ02417J

3.7. Binding stability assessment for ligand- protein complexes

Binding stability of the 3S7S protein complexes with Ligand5, Exemestane, C2, and L5 was assessed using molecular dynamics (MD) simulations over a period of 100 nanoseconds at ambient temperature. The analysis following these simulations indicated that all ligands remained consistently bound within the ligand-binding groove of the protein pocket. Stability evaluations for each complex included measurements of the radius of gyration, root mean square fluctuation (RMSF), root mean square deviation (RMSD), average center of mass (COM) distance between the protein and ligand, hydrogen bonding, and binding free energy (MM/PBSA).

Fig. 11A illustrates the RMSD values for the protein-ligand complex, the protein backbone, and the ligand structures. The RMSD curves for both the complex and the backbone indicate stability, with low values observed after 10 nanoseconds. The RMSF of the protein complex, calculated using the GROMACS algorithm and focusing on 'C-alpha' atoms, typically remains below 2.0 Å, except at residues corresponding to loops or turns (Fig. 11B). Analysis of the radius of gyration (Fig. 11C) shows minimal variation in Rg values (less than 1 Å) throughout the simulation, reflecting the compactness and stability of the protein-ligand system, with Rg values ranging from 22.4 Å to 23.2 Å. Fig. 12A depicts the total number of hydrogen bonds formed between the ligand and protein during the 100-nanosecond simulation. Exemestane and C2 exhibit predominantly weak hydrogen bond interactions with noticeable intervals of separation. Ligands Ligand5 and L5 maintain average hydrogen bond counts of 3.85 and 2.439, respectively.

The average center-of-mass distance separating the ligand from the protein over the 100 ns simulation is shown in Fig. 12B. The COM distance fluctuates minimally within 2-3 Å, suggesting that the ligands remain bound to their binding sites, as confirmed by visual inspection of the trajectories.

To further assess the binding interactions between the 3S7S protein and Ligand5, contact frequency (CF) analysis was performed using VMD software. The contact frequency was calculated using the contactFrEquation.tcl module, defining a contact as an amino acid within a 4 Å cutoff distance. Fig. 13 presents the contact frequency analysis results for the 3S7S protein with Ligand5, as well as comparative data for Ligand5, Exemestane (reference drug), and C2. The analysis highlights protein residues with significant hydrogen bonding and high CF%, notably Phe116, Met364, Pro429, and Met447.

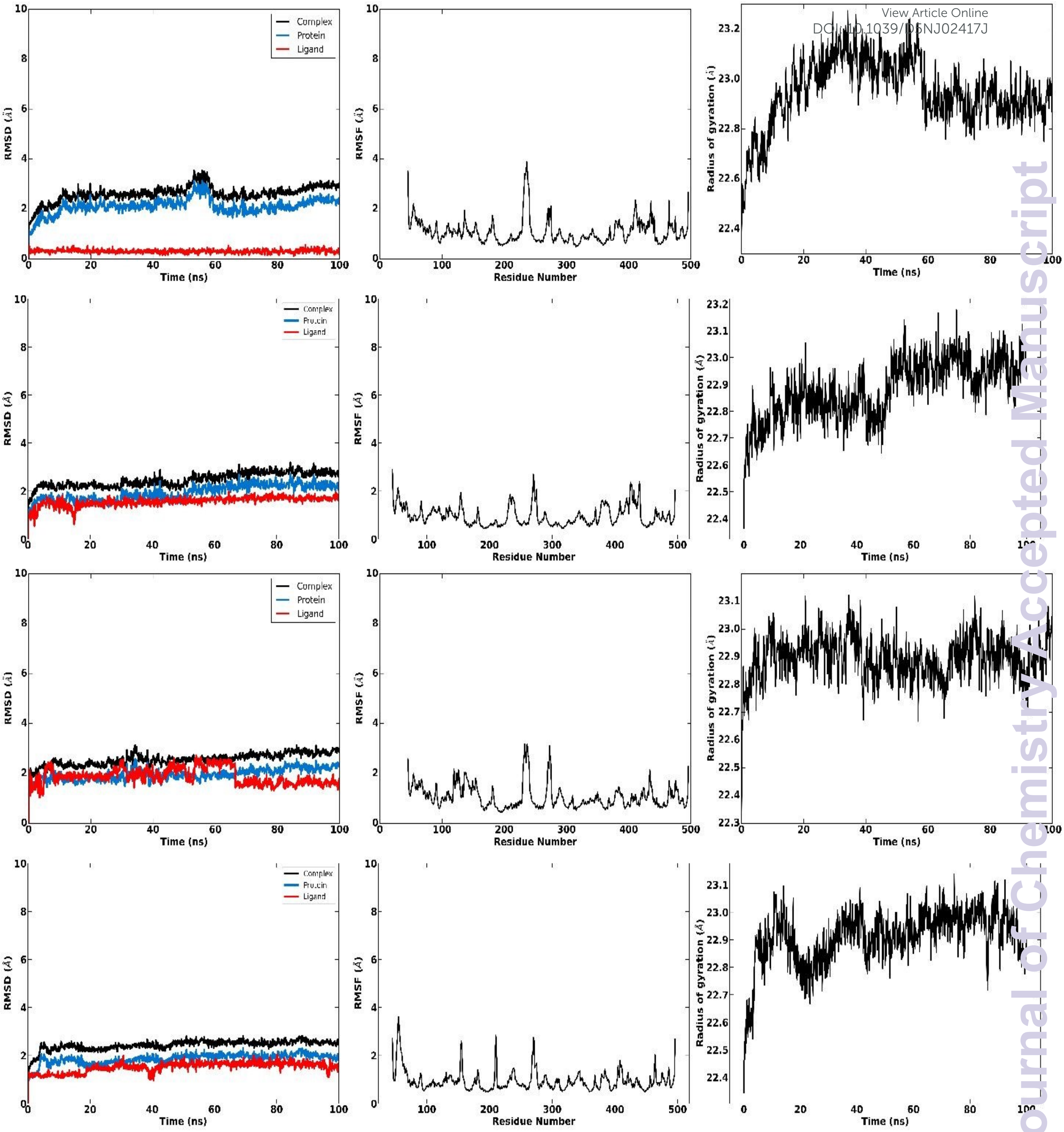
View Article Online
DOI: 10.1039/D5NJ02417J

For complex re-evaluation, the MM/PBSA method was chosen for its efficiency in calculating binding free energy using force-field approaches. It offers greater computational efficiency compared with other free energy calculation methods, such as free energy perturbation (FEP) and thermodynamic integration (TI). The g_mmpbsa tool from the GROMACS suite was used to perform these calculations, the results of which are detailed in Table 10.

Table 10. Energies' values of compared compounds

Complexes	**Energies' values [kJ/mol]**				
3S7S-Ligands	ΔG	**Van der Waal**	**Electrostatic**	**Polar solvation**	**SASA**
3S7S-Exemestane	-132.023 +/- 22.179	-133.406 +/- 11.864	-78.462 +/- 23.651	96.561 +/- 13.790	-16.716 +/- 0.957
3S7S-Ligand5	-154.206 +/- 53.601	-171.208+/- 12.197	-113.278 +/- 100.473	153.019 +/- 20.303	-22.738 +/- 0.845
3S7S-C2	-177.903 +/- 18.150	-209.937 +/- 17.614	-102.209 +/- 19.365	157.542 +/- 28.670	-23.299 +/- 1.471
3S7S -L5	-201.756 +/- 44.840	-228.020 +/- 24.529	-194.342 +/- 59.192	245.125 +/- 27.196	-24.519 +/- 1.034

View Article Online
DOI: 10.1039/D5NJ02417J



(A) (B) (C)

Fig. 11. RMSD (A), RMSF(B), and Radius of gyration (C) during 100ns MD simulation. Rows 1 (Exemestane), 2 (Ligand 5), 3 (C2) and 4 (L5)

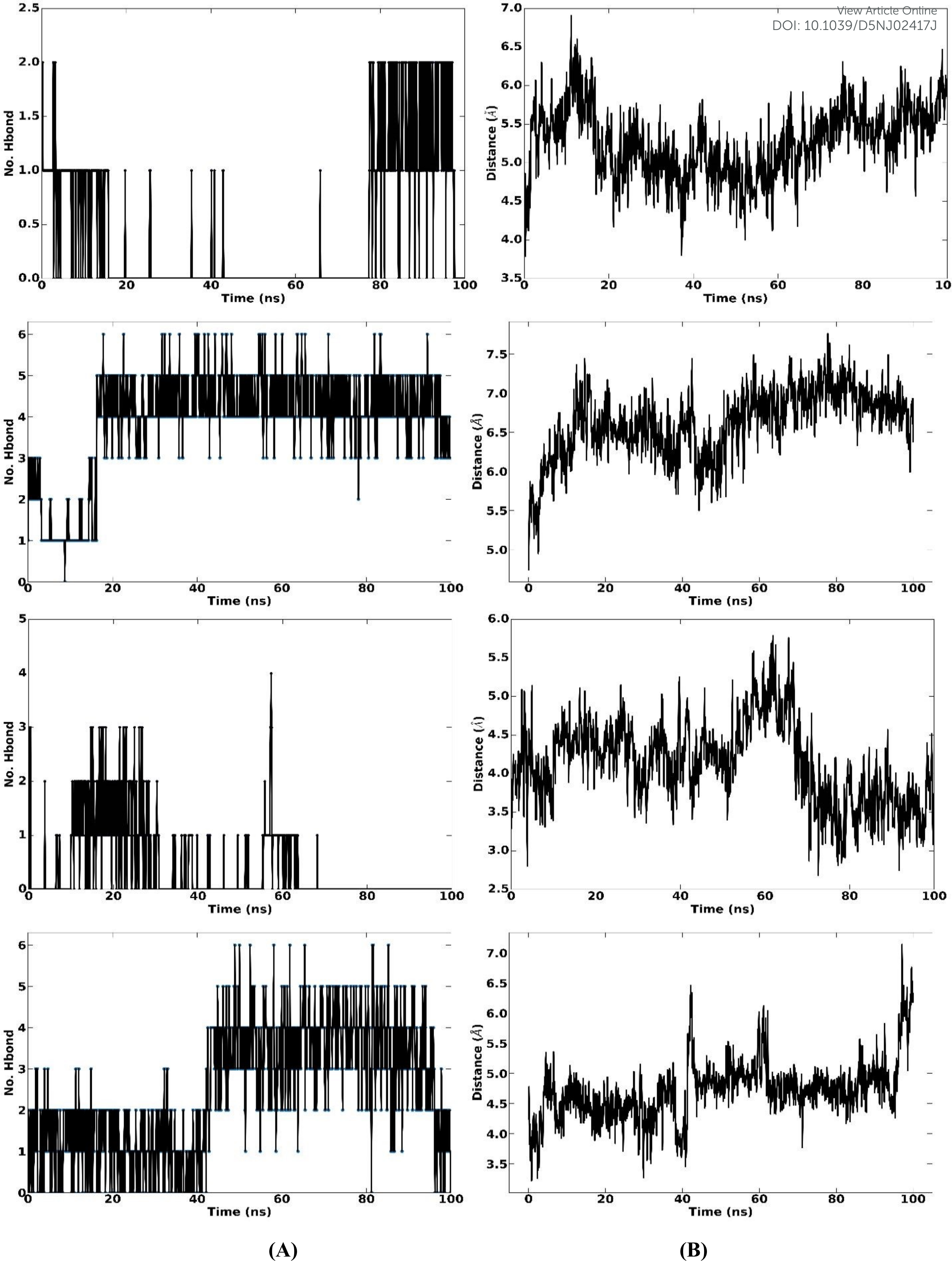


Fig. 12. Hbonds (A) and Average distance between Ligand and the Protein (B). Rows 1 (Exemestane), 2 (Ligand 5), 3 (C2) and 4 (L5)

View Article Online
DOI: 10.1039/D5NJ02417J

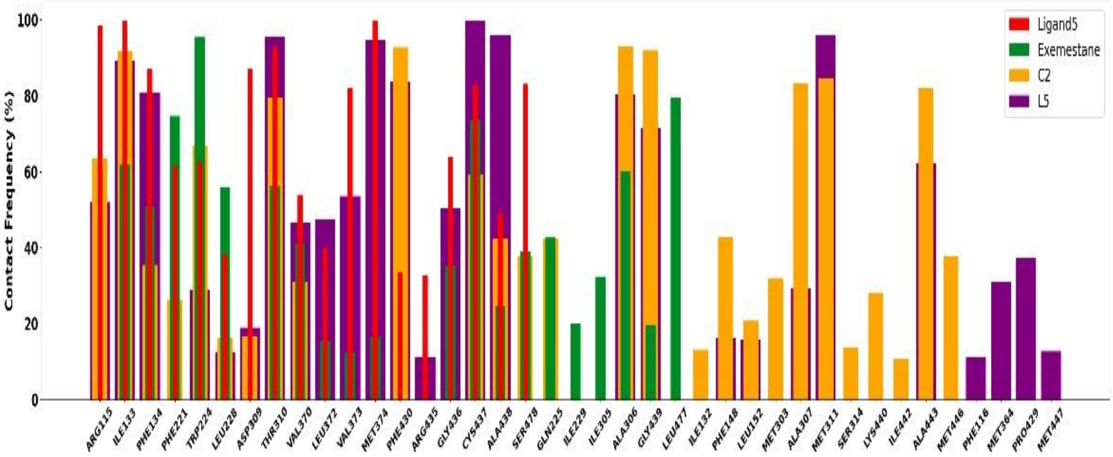


Fig. 13. Contact Frequency Analysis

According to molecular dynamics (MD) simulation results and our previous findings with Ligand5 and C2, Ligand L5 emerges as the most promising candidate for inhibiting aromatase. Its superior stability and more consistent interactions with the aromatase active site, confirmed by MD simulations, highlight its potential as a highly effective inhibitor. The stable hydrogen bonds formed between L5 and key residues in the active site throughout the simulations further strengthen its binding affinity, making it stand out compared to Exemestane and the other drug candidates.

Identifying drug candidates like L5, which improve binding stability and enhance selectivity and drug-likeness, is crucial for advancing breast cancer treatment. In silico methods, such as those used to refine binding affinity and pharmacokinetic properties, significantly contribute to minimizing development costs and time. This demonstrates the importance of computational approaches in streamlining the drug discovery process.

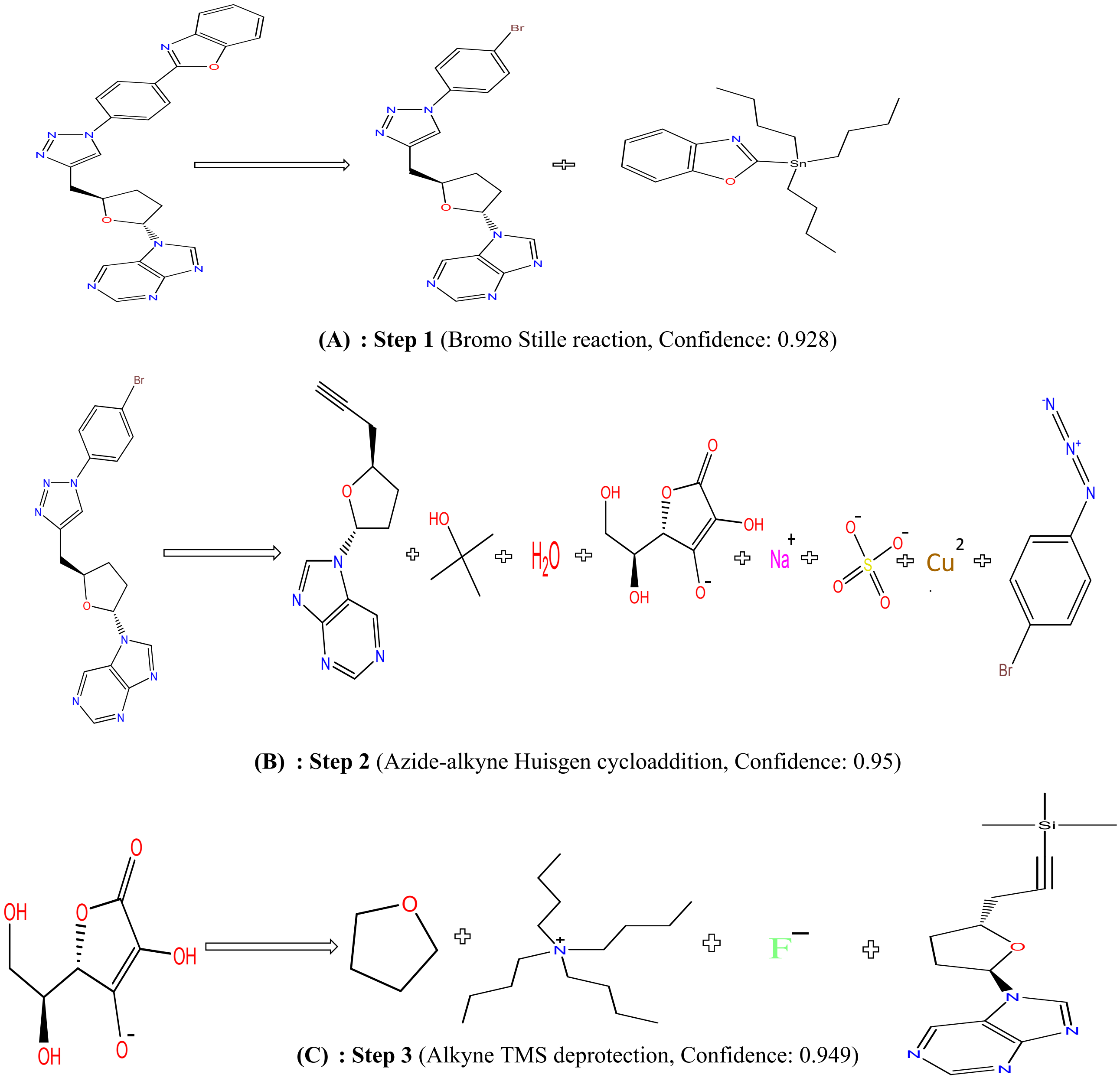
Furthermore, incorporating techniques such as retrosynthesis in drug development can facilitate the synthesis of promising candidates like L5, making the transition to in vitro and in vivo testing more efficient. Combining in silico strategies with synthetic chemistry can accelerate the identification and optimization of novel, effective drug candidates, with L5 serving as an exemplary model of how these methods can enhance drug development.

3.8. Retrosynthesis of selected drug-candidate

As indicated by the significant value of the synthetic acceptability parameter confirming the possibility of synthesis, and to facilitate this task, we employed the retrosynthesis of L5 using the IBM RXN for Chemistry platform. From the available routes, we selected the one with the highest score. As described in Fig. 14, the methodology consists of three main steps (A, B, and

View Article Online
DOI: 10.1039/D5NJ02417J

C), adapted from analogous synthetic strategies described in the literature. In the initial step, the Bromo-Stille coupling reaction is carried out between the compound 7-((2R,5R)-5-((1-(4-bromophenyl)-1H-1,2,3-triazol-4-yl)methyl)tetrahydrofuran-2-yl)-7H-purine and 2-(tributylstannyl)benzo[d]oxazole [74]. The second step involves Huisgen's azide-alkyne cycloaddition, in which 7-((2R,5R)-5-(prop-2-yn-1-yl)tetrahydrofuran-2-yl)-7H-purine reacts with 3-(4-bromophenyl)triaz-1-en-1-ide to form the triazole derivative [75]. Finally, the third step involves deprotecting the protected alkyne group in 7-((2R,5R)-5-(3-(trimethylsilyl)prop-2-yn-1-yl)tetrahydrofuran-2-yl)-7H-purine using trimethylsilanol (TMS) as the deprotecting agent [76].



(A) : Step 1 (Bromo Stille reaction, Confidence: 0.928)

(B) : Step 2 (Azide-alkyne Huisgen cycloaddition, Confidence: 0.95)

(C) : Step 3 (Alkyne TMS deprotection, Confidence: 0.949)

Fig. 14. Steps for retrosynthesis of drug-candidate using IBM RXN platform

View Article Online
DOI: 10.1039/D5NJ02417J

Based on the procedure illustrated in Fig. 14 and the specified experimental conditions, L5 can be efficiently synthesized, enabling in vitro and in vivo evaluation for the treatment of breast cancer. Among the compounds evaluated, L5 showed greater stability and better binding affinity than C2 and Ligand5, confirming its potential as a lead aromatase inhibitor. Retrosynthetic analysis not only rationalizes the synthesis pathway, but also enables targeted structural modifications to be made in order to optimize pharmacological efficacy.

Ultimately, the integration of advanced in silico strategies - combining machine learning algorithms and molecular modeling - has enabled the reliable identification of high-potential candidates, significantly reducing the need for costly and time-consuming experimental procedures while accelerating the overall drug discovery and development process.

4. Conclusion

This study represents a significant advance in breast cancer research, identifying benzoxazole derivatives as promising therapeutic candidates through a comprehensive computational approach. The integration of 3D-QSAR modeling with artificial neural networks (ANNs) has resulted in a highly predictive and interpretable model that provides valuable insights into the molecular descriptors governing anticancer activity. Molecular docking analyses highlighted key ligand-enzyme interactions, strengthened the reliability of the CoMSIA/EHA model and confirmed L5's selectivity as a potent aromatase inhibitor. In addition, ADMET predictions facilitated the identification of compounds with favorable pharmacokinetic profiles, ensuring enhanced bioavailability and safety. Molecular dynamics (MD) simulations validated the structural stability and binding affinity of this candidate (L5), comparing it with Ligand5, C2 and the reference drug, exemestane. Compared with these references, the compound demonstrated superior stability and more consistent interactions in the aromatase active site, underlining its potential for further research.

Beyond computational validation, retrosynthetic analysis has provided a strategic framework for optimizing synthetic accessibility, reducing costs, and facilitating progression to experimental validation. Collectively, this integrated approach accelerates the identification of potent, drug-like candidates, providing a cost- and time-efficient platform for drug development. Future efforts should focus on experimental validation through in vitro and in vivo assays to confirm biological efficacy and safety. In addition, structure-guided optimization, supported by artificial intelligence techniques, could enable the design of more potent analogues.

View Article Online
DOI: 10.1039/D5NJ02417J

Conflicts of interest

All authors declare that they have no conflict of interest in this work.

Acknowledgements

The authors want to thank the Moroccan Association of Theoretical Chemistry (MATC) for its relevant help regarding to the Software.

Authors' contributions

Said El Rhabori: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Marwa Alaqarbeh: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Lhoucine Naanaai: Writing – review, Visualization, Validation, Conceptualization.

Yassine El Allouche: Visualization, Validation, Conceptualization.

Abdellah El Aissouq: Visualization, Validation, Investigation, Conceptualization.

Mohammed Bouachrine: Visualization, Validation, Investigation, Conceptualization.

Samir Chtita: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Hicham Zaitan: Visualization, Validation, Conceptualization.

Fouad Khalil: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

References

1 H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, *CA: A Cancer Journal for Clinicians*, 2021, **71**, 209–249.

2 C. Laplagne, M. Domagala, A. Le Naour, C. Quemerais, D. Hamel, J. J. Fournié, B. Couderc, C. Bousquet, A. Ferrand and M. Poupot, *International Journal of Molecular Sciences*, , DOI:10.3390/IJMS20194719.

3 U. Anand, A. Dey, A. K. S. Chandel, R. Sanyal, A. Mishra, D. K. Pandey, V. De Falco, A. Upadhyay, R. Kandimalla, A. Chaudhary, J. K. Dhanjal, S. Dewanjee, J. Vallamkondu and J. M. Pérez de la Lastra, *Genes & Diseases*, 2023, **10**, 1367–1401.

View Article Online
DOI: 10.1039/D5NJ02417J

4 M. Chehelgerdi, M. Chehelgerdi, O. Q. B. Allela, R. D. C. Pecho, N. Jayasankar, D. P. Rao, T. Thamaraikani, M. Vasanthan, P. Viktor, N. Lakshmaiya, M. J. Saadh, A. Amajd, M. A. Abo-Zaid, R. Y. Castillo-Acobo, A. H. Ismail, A. H. Amin and R. Akhavan-Sigari, *Molecular Cancer*, 2023, **22**, 1–103.

5 J. Caciolla, A. Bisi, F. Belluti, A. Rampa and S. Gobbi, *Molecules 2020, Vol. 25, Page 5351*, 2020, **25**, 5351.

6 K. Gandhi, U. Shah, S. Patel, M. Patel, S. Patel, A. Patel, S. Patel, N. Solanki, A. Shah and D. Baria, *Indian Journal of Chemistry (IJC)*, 2022, **61**, 192–200.

7 U. Shah, S. Patel, M. Patel, K. Gandhi and A. Patel, *Indian Journal of Chemistry -Section B (IJC-B)*, 2020, **59**, 283–293.

8 U. Shah, S. Patel, M. Patel and J. Upadhayay, *Letters in Drug Design & Discovery*, 2017, **14**, 1267–1276.

9 U. Shah, S. Patel, M. Patel, N. Jain, N. Pandey, A. Chauhan, A. Patel and S. Patel, *Anti-Cancer Agents in Medicinal Chemistry*, 2021, **22**, 1370–1385.

10 U. Shah, A. Patel, S. Patel, M. Patel, A. Patel, S. Patel, S. Patel, R. Maheshwari, A. G. Mtewa and K. Gandhi, *Anti-Cancer Agents in Medicinal Chemistry*, 2021, **22**, 2063–2079.

11 T. L. Rižner and A. Romano, *Frontiers in Pharmacology*, 2023, **14**, 1155558.

12 G. A. Walker, M. Xenophontos, L. C. Chen and K. L. Cheung, *Patient preference and adherence*, 2013, **7**, 245.

13 P. Kumar, B. Mangla, S. Javed, W. Ahsan, P. Musyuni, D. Sivadasan, S. S. Alqahtani and G. Aggarwal, *Frontiers in pharmacology*, , DOI:10.3389/FPHAR.2023.1149554.

14 A. M. M. E. Omar, O. M. AboulWafa, M. S. El-Shoukrofy and M. E. Amr, *Bioorganic Chemistry*, 2020, **96**, 103593.

15 C. K. Ryu, R. Y. Lee, N. Y. Kim, Y. H. Kim and A. L. Song, *Bioorg Med Chem Lett*, 2009, **19**, 5924–5926.

16 S. M. Sondhi, N. Singh, A. Kumar, O. Lozach and L. Meijer, *Bioorg Med Chem*, 2006, **14**, 3758–3765.

17 S. Kakkar, S. Tahlan, S. M. Lim, K. Ramasamy, V. Mani, S. A. A. Shah and B. Narasimhan, *Chemistry Central Journal*, 2018, **12**, 1–16.

18 W. Yu and A. D. Mackerell, *Methods in Molecular Biology*, 2017, **1520**, 85–106.

19 T. A. Soares, A. Nunes-Alves, A. Mazzolari, F. Ruggiu, G. W. Wei and K. Merz, *Journal of Chemical Information and Modeling*, 2022, **62**, 5317–5320.

20 J. Mao, J. Akhtar, X. Zhang, L. Sun, S. Guan, X. Li, G. Chen, J. Liu, H. N. Jeon, M. S. Kim, K. T. No and G. Wang, *iScience*, 2021, **24**, 103052.

21 D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia and R. K. Tekade, *Drug Discovery Today*, 2021, **26**, 80–93.

22 A. Vidal-Limon, J. E. Aguilar-Toalá and A. M. Liceaga, *Journal of Agricultural and Food*

View Article Online
DOI: 10.1039/D5NJ02417J

Chemistry, 2022, **70**, 934–943.

23 O. Méndez-Lucio, J. Pérez-Villanueva, A. Romo-Mancillas and R. Castillo, *MedChemComm*, 2011, **2**, 1058–1065.

24 T. Kim, B. H. You, S. Han, H. C. Shin, K. C. Chung and H. Park, *International Journal of Molecular Sciences*, 2021, **22**, 10995.

25 P. H. M. Torres, A. C. R. Sodero, P. Jofily and F. P. Silva-Jr, *International Journal of Molecular Sciences 2019, Vol. 20, Page 4574*, 2019, **20**, 4574.

26 O. Trott and A. J. Olson, *Journal of Computational Chemistry*, 2010, **31**, 455–461.

27 C. S. Kumar, M. L. Narasu and C. R. Singh, *Phytomedicine Plus*, 2023, **3**, 100422.

28 M. R. ; Tomás-Alvarado, E. ; Espinoza-Baigorria, A. ; León-Figueroa, D. A. ; Sah, R. ; Rodriguez-Morales, A. J. ; Barboza, M. R. Challapa-Mamani, E. Tomás-Alvarado, A. Espinoza-Baigorria, D. A. León-Figueroa, R. Sah, A. J. Rodriguez-Morales and J. J. Barboza, *Tropical Medicine and Infectious Disease 2023, Vol. 8, Page 457*, 2023, **8**, 457.

29 Y. Moukhliss, Y. Koubi, M. Alaqarbeh, N. Alsakhen, S. Hamzeh, H. Maghat, A. Sbai, M. Bouachrine and T. Lakhlifi, *New Journal of Chemistry*, 2022, **46**, 10154–10161.

30 M. A. El Alaouy, M. Alaqarbeh, M. Ouabane, H. Zaki, M. ElBouhi, H. Badaoui, Y. Moukhliss, A. Sbai, H. Maghat, T. Lakhlifi and M. Bouachrine, *Journal of Biomolecular Structure and Dynamics*, , DOI:10.1080/07391102.2023.2252116.

31 W. Zhong, Z. Yang and C. Y. C. Chen, *Nature Communications 2023 14:1*, 2023, **14**, 1–14.

32 R. O. M. A. de Souza, L. S. M. Miranda and U. T. Bornscheuer, *Chemistry – A European Journal*, 2017, **23**, 12040–12063.

33 A. Mumuni and F. Mumuni, *Array*, 2022, **16**, 100258.

34 Z. Zhong, L. Zheng, G. Kang, S. Li and Y. Yang, *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, 2020, 13001–13008.

35 S. Wold, M. Sjöström and L. Eriksson, *Chemometrics and Intelligent Laboratory Systems*, 2001, **58**, 109–130.

36 A. Zifarelli, M. Giglio, G. Menduni, A. Sampaolo, P. Patimisco, V. M. N. Passaro, H. Wu, L. Dong and V. Spagnolo, *Analytical Chemistry*, 2020, **92**, 11035–11043.

37 H. Hadni and M. Elhallaoui, *Heliyon*, 2020, **6**, e03580.

38 P. Labrie, S. P. Maddaford, S. Fortin, S. Rakhit, L. P. Kotra and R. C. Gaudreault, *Journal of Medicinal Chemistry*, 2006, **49**, 7646–7660.

39 M. S. Callén, I. Martínez, G. Grasa, J. M. López and R. Murillo, *Biomass Conversion and Biorefinery*, 2022, **1**, 1–21.

40 A. Tayarani, A. Baratian, M.-B. N. Sistani, M. R. Saberi and Z. Tehranizadeh, *Iranian Journal of Basic Medical Sciences*, 2013, **16**, 1196.

41 U. Michelucci and F. Venturini, *Machine Learning and Knowledge Extraction 2021, Vol. 3, Pages 357-373*, 2021, **3**, 357–373.

View Article Online
DOI: 10.1039/D5NJ02417J

42 S. El Rhabori, A. El Aissouq, S. Chtita and F. Khalil, *Structural Chemistry*, 2023, **34**, 585–603.

43 H. Hadni and M. Elhallaouia, *Heliyon*, 2022, **8**, e11537.

44 K. Roy and I. Mitra, *Combinatorial Chemistry & High Throughput Screening*, 2011, **14**, 450–474.

45 K. Roy, P. Ambure and S. Kar, *ACS Omega*, 2018, **3**, 11392–11406.

46 P. P. Roy, S. Paul, I. Mitra and K. Roy, *Molecules 2009, Vol. 14, Pages 1660-1701*, 2009, **14**, 1660–1701.

47 S. Kar, K. Roy and J. Leszczynski, *Methods in Molecular Biology*, 2018, **1800**, 141–169.

48 D. Ghosh, J. Lo, D. Morton, D. Valette, J. Xi, J. Griswold, S. Hubbell, C. Egbuta, W. Jiang, J. An and H. M. L. Davies, *Journal of medicinal chemistry*, 2012, **55**, 8464–8476.

49 S. El Rhabori, A. El Aissouq, S. Chtita and F. Khalil, *Journal of the Indian Chemical Society*, 2022, **99**, 100675.

50 D. Méndez-Álvarez, M. F. Torres-Rojas, E. E. Lara-Ramirez, L. A. Marchat and G. Rivera, *Molecules*, 2023, **28**, 4389.

51 D. E. V. Pires, T. L. Blundell and D. B. Ascher, *Journal of Medicinal Chemistry*, 2015, **58**, 4066–4072.

52 A. Daina, O. Michielin and V. Zoete, *Scientific Reports 2017 7:1*, 2017, **7**, 1–13.

53 S. El Rhabori, A. El Aissouq, S. Chtita and F. Khalil, *Anti-Cancer Drugs*, 2022, **33**, 789–802.

54 S. El Rhabori, M. Alaqarbeh, A. El Aissouq, M. Bouachrine, S. Chtita and F. Khalil, *Chemical Physics Impact*, 2024, **8**, 100455.

55 S. El Rhabori, Y. El Allouche, L. Naanaai, A. El Aissouq, F. Khalil, M. Alaqarbeh, M. Bouachrine and S. Chtita, *Journal of Molecular Structure*, 2025, **1320**, 139500.

56 M. Untch and C. Jackisch, *Therapeutics and Clinical Risk Management*, 2008, **4**, 1295.

57 A. El Aissouq, A. Lachhab, S. El Rhabori, M. Bouachrine, A. Ouammou and F. Khalil, *New Journal of Chemistry*, 2022, **46**, 20786–20800.

58 H. Guterres and W. Im, *Journal of Chemical Information and Modeling*, 2023, **63**, 4772–4779.

59 X. Zhang, L. Li and Q. Zheng, *Journal of Chemical Information and Modeling*, 2023, **63**, 4762–4771.

60 H. Hajji, M. Alaqarbeh, T. Lakhlifi, M. A. Ajana, N. Alsakhen and M. Bouachrine, *Computers in Biology and Medicine*, 2022, **150**, 106209.

61 A. Belhassan, H. Zaki, S. Chtita, M. Alaqarbeh, N. Alsakhen, M. Benlyas, T. Lakhlifi and M. Bouachrine, *Computers in Biology and Medicine*, 2021, **136**, 104758.

62 M. Pieroni, F. Madeddu, J. Di Martino, M. Arcieri, V. Parisi, P. Bottoni and T. Castrignanò, *International Journal of Molecular Sciences 2023, Vol. 24, Page 11671*, 2023, **24**, 11671.

63 R. Kumari, R. Kumar and A. Lynn, *Journal of Chemical Information and Modeling*, 2014, **54**, 1951–1962.

64 M. OUABANE, K. TABTI, H. HAJJI, M. ELBOUHI, A. KHALDAN, K. ELKAMEL, A. SBAI,

View Article Online
DOI: 10.1039/D5NJ02417J

M. Aziz AJANA, C. SEKKATE, M. BOUACHRINE and T. LAKHLIFI, *Arabian Journal of Chemistry*, 2023, **16**, 105207.

65 Y. Jiang, Y. Yu, M. Kong, Y. Mei, L. Yuan, Z. Huang, K. Kuang, Z. Wang, H. Yao, J. Zou, C. W. Coley and Y. Wei, *Engineering*, 2023, **25**, 32–50.

66 A. Toniato, J. P. Unsleber, A. C. Vaucher, T. Weymuth, D. Probst, T. Laino and M. Reiher, *Digital Discovery*, 2023, **2**, 663.

67 W. Gao, P. Raghavan and C. W. Coley, *Nature Communications*, 2022, **13**, 1075.

68 J. M. Smith, S. J. Harwood and P. S. Baran, *Accounts of chemical research*, 2018, **51**, 1807–1817.

69 S. E. R. et al, *RHAZES: Green and Applied Chemistry*, 2022, **16**, 63–71.

70 D. Van Tilborg, A. Alenicheva and F. Grisoni, *Journal of Chemical Information and Modeling*, 2022, **62**, 5938–5951.

71 K. V. Chuang, L. M. Gunsalus and M. J. Keiser, *Journal of Medicinal Chemistry*, 2020, **63**, 8705–8722.

72 B. Dou, Z. Zhu, E. Merkurjev, L. Ke, L. Chen, J. Jiang, Y. Zhu, J. Liu, B. Zhang and G. W. Wei, *Chemical reviews*, 2023, **123**, 8736.

73 J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, *Journal of Chemical Information and Modeling*, 2012, **52**, 1757.

74 A. Babu, M. N. Joy, K. Sunil, A. M. Sajith, S. Santra, G. V. Zyryanov, O. A. Konovalova, I. I. Butorin and K. Muniraju, *RSC Advances*, 2022, **12**, 22476–22491.

75 I. Cano, E. Álvarez, M. C. Nicasio and P. J. Pérez, *Journal of the American Chemical Society*, 2011, **133**, 191–193.

76 P. Bhattacharya and A. Basak, *Tetrahedron Letters*, 2013, **54**, 5137–5139.

View Article Online
DOI: 10.1039/D5NJ02417J

Data Availability Statement

The data and software used are included in this manuscript.