

Rational in silico design and synthetic route elaboration for anti-RCC benzimidazole candidates**Larbi El Mchichi^{a*}, Marwa Alaqrbeh^b and Mohammed Bouachrine^a**^aNatural Substances and Molecular Chemistry Laboratory (SOOM), Faculty of Science, Moulay Ismail University of Meknes, Morocco^bBasic Science Department, Prince Al Hussein bin Abdullah II Academy for Civil Protection, Al-Balqa Applied University, Al-Salt 19117, Jordan**CHRONICLE***Article history:*

Received March 27, 2025

Received in revised form

June 16, 2025

Accepted August 25, 2025

Available online

August 25, 2025

*Keywords:**Benzimidazole**3D-QSAR**CoMSIA**ADMET**Renal cancer carcinoma**Anticancer***ABSTRACT**

Benzimidazole continues to be an intriguing scaffold in recent drug discovery, owing to its broad spectrum of pharmacological effects. In recent years, a variety of its derivatives, which included chalcone imines, hydrazones, and thiosemicarbazones, were actively investigated for their antitumor potential. In the search for new agents capable of treating kidney cancer, an analysis of a small series of 2-substituted benzimidazoles (45) using 3D-QSAR modelling was performed to determine the antiproliferative activities against cancer cell lines A-498. The biological activity was sufficient to establish a meaningful structure–activity relationship, providing a foundation for the design of more potent compounds. The activity-favouring and activity-disfavoring structural regions were clearly revealed using contour maps generated by the models. The CoMSIA/SHD model was one of the best developed, and its high statistical robustness ($q^2 = 0.751$) and predictive power ($R^2_{\text{pred}} = 0.924$) indicated its reliability. We designed five new derivatives of benzimidazole based on the QSAR results, which demonstrated potent inhibitory potential. Molecular docking studies were performed in order to investigate in detail their interaction modes with the aromatic receptor, and stable binding conformations at the active site have been found. The in silico pharmacokinetic studies suggested that these compounds have a favourable ADMET and bioavailability profile, reinforcing their suitability for in vitro testing. Two leads, L15 and L22, with better PKs properties and high-predicted activities, were subjected to a 100-ns MD simulation in complex with the aromatase target to investigate their stability. We also conducted a retrosynthetic analysis for L15 and L22, suggesting potential synthetic routes for experimental validation. Overall, these findings suggest that benzimidazole analogues could be promising candidates for treating RCC and possibly for blocking aromatase.

© 2025 by the authors; licensee Growing Science, Canada.

1. Introduction

Cancer continues to pose a major global health burden, ranked by the World Health Organization as the second leading cause of mortality in the world. Among malignancies, kidney cancer is a significant concern and is the sixteenth cause of cancer-related death. As a proportion of all cancer cases, it represents approximately 2.4%, while it is responsible for 1.7% of the total global cancer burden and 1.7% of deaths due to cancer¹⁻³.

Renal cell carcinoma (RCC) is the most common and lethal subtype of kidney cancer, arising in the renal tubular epithelium. RCC represents almost 90% of all kidney cancer cases and is frequently diagnosed in advanced stages, usually with metastasis⁴. Loss of function in the von Hippel-Lindau (VHL) gene is among the key genetic alterations found in RCC. This mutation leads to the accumulation of hypoxia-inducible factors (HIFs), which further promote tumour development. Patients with RCC still have a grim prognosis, especially in the advanced or high-grade metastatic forms.

Blocking estrogen signalling has been shown to reduce the growth and viability of renal cell carcinoma (RCC) cells⁵. Estrogen exerts its biological effects mainly through binding to estrogen receptors (ERs), which are known to support cell

* Corresponding author

E-mail address: l.elmchichi@edu.umi.ac.ma elmchichi.larbi@gmail.com (L. El Mchichi)

proliferation and survival in several hormone-dependent cancers. In RCC, especially within A498 cell lines, elevated expression of ERs indicates a potential role for estrogen in driving tumour development⁶. Aromatase inhibitors function by inhibiting the enzyme aromatase, which catalyzes the conversion of androgens to estrogens⁷. This enzymatic blockade lowers intracellular estrogen levels, resulting in diminished stimulation of estrogen receptors. As a result, estrogen-driven signalling pathways that promote renal cell carcinoma (RCC) cell proliferation, angiogenesis, and resistance to apoptosis are effectively disrupted.

Numerous studies have shown that disrupting estrogen signalling, either by blocking estrogen receptors or by limiting estrogen production, can markedly suppress the growth of renal cell carcinoma (RCC) cells and may improve the effectiveness of other anticancer treatments⁸⁻¹¹. Therefore, estrogen pathways could serve as potential therapeutic targets for RCC, particularly in tumours with high expressions of ERs. Although the involvement of aromatase in the suppression of RCC is not well-known, an emerging body of data indicates that estrogen can have pleiotropic effects, especially through estrogen receptor β (ER β)-mediated mechanisms, and could contribute importantly to the modulation of tumour biological properties. Aromatase is the enzyme that converts androgens to the estrogens that then act via estrogen receptors to affect multiple cellular functions¹². Activation of estrogen receptor beta (ER β) has been demonstrated to result in decreased tumour cell proliferation and increased apoptosis¹³ and has been linked to reduced metastatic potential in renal cell carcinoma (RCC).

Research work reported in the literature showed that 2-substituted benzimidazole is an important pharmacophore and a significant tool in the discovery of new drugs, which may have led to further investigation¹⁴. The synthesis and biological evaluation of a series of 2-substituted benzimidazoles led to the identification of gastric acid drugs, including rabeprazole and pantoprazole. In recent years, the development of new derivatives has remained an active field of clinical research. Additionally, it is not too difficult to synthesize these compounds, and so they have become an interesting topic to many researchers. Recently, the study of the diverse biological and pharmacological properties of nitrogen containing heterocycles have attracted considerable attention¹⁵⁻¹⁸. Benzimidazoles are important because they have many biological activities and are part of some strong medications. The benzimidazole scaffold has an appreciable structural resemblance to that of natural nucleotides, helping it in its potential to interact with active sites of biopolymers in natural living systems¹⁹.

A lot of patients have used a variety of drugs based on the benzimidazole structure to treat a wide range of diseases. Currently known drugs such as carbendazim, tilomisole, omeprazole, bendamustine, albendazole, and mebendazole are what have profoundly influenced modern treatment regimens (**Fig. 1**)²⁰. Furthermore, benzimidazole derivatives are efficient in interaction with biological macromolecules and also find applications in the synthesis of biologically active analogues that are potential therapeutic agents against different diseases because they are structurally similar to natural nucleotides,²¹ thus playing an important role in the design of novel drugs.

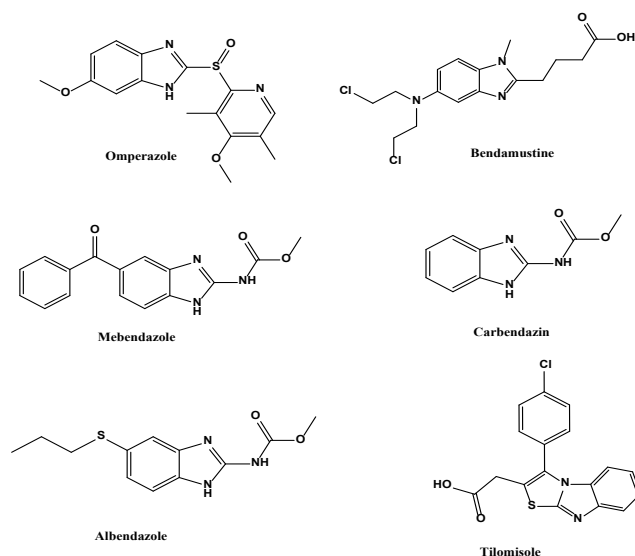


Fig. 1. Examples of marketed anticancer drugs containing benzimidazole rings.

The advancement of anticancer drug development is increasingly dependent on computational techniques like quantitative structure–activity relationship (QSAR) modelling, molecular docking, and molecular dynamics (MD) simulations²²⁻²⁵. QSAR models are used to predict the biological activity of chemical compounds by analyzing their structural properties, aiding in the creation of more effective and selective drug candidates²⁶. In the meantime, molecular docking allows for an estimation of the affinities of these compounds to different cancer-related targets, which provides an interesting perspective on how much these compounds might be effective. Molecular-dynamics simulations further complement these predictions by predicting the time-dependent behaviour of drug–target complexes in complex physiological environments, allowing for analysis of long-term stability and key molecular interactions^{27,28}. We have

employed in-silico methods to screen a series of 45 2-substituted benzimidazole derivatives in search of potent anticancer molecules that exhibit antiproliferative activity against the renal cancer cell line A-498. Meanwhile, the retrosynthetic analysis was a key step for the development of the new and practical methods. This approach enables chemists to find non-toxic precursors with the inexpensive and efficient reaction path as well as the synthesis. This new approach of combining computational design with synthetic tractability should facilitate the drug-discovery process and increase the chances of finding novel anticancer agents²⁹⁻³⁴. This study was designed mainly to identify molecules that would inhibit the action of the aromatase protein, for, in doing so, one would halt the growth of cancer cells and their spread to other sites.

2. Experimental

2.1. Methodology of work

This study is structured around four key components: 3D-QSAR analysis, molecular docking, molecular dynamics simulations, and retrosynthetic analysis. An overview of the methodology is illustrated in the flowchart shown in Fig. 2.

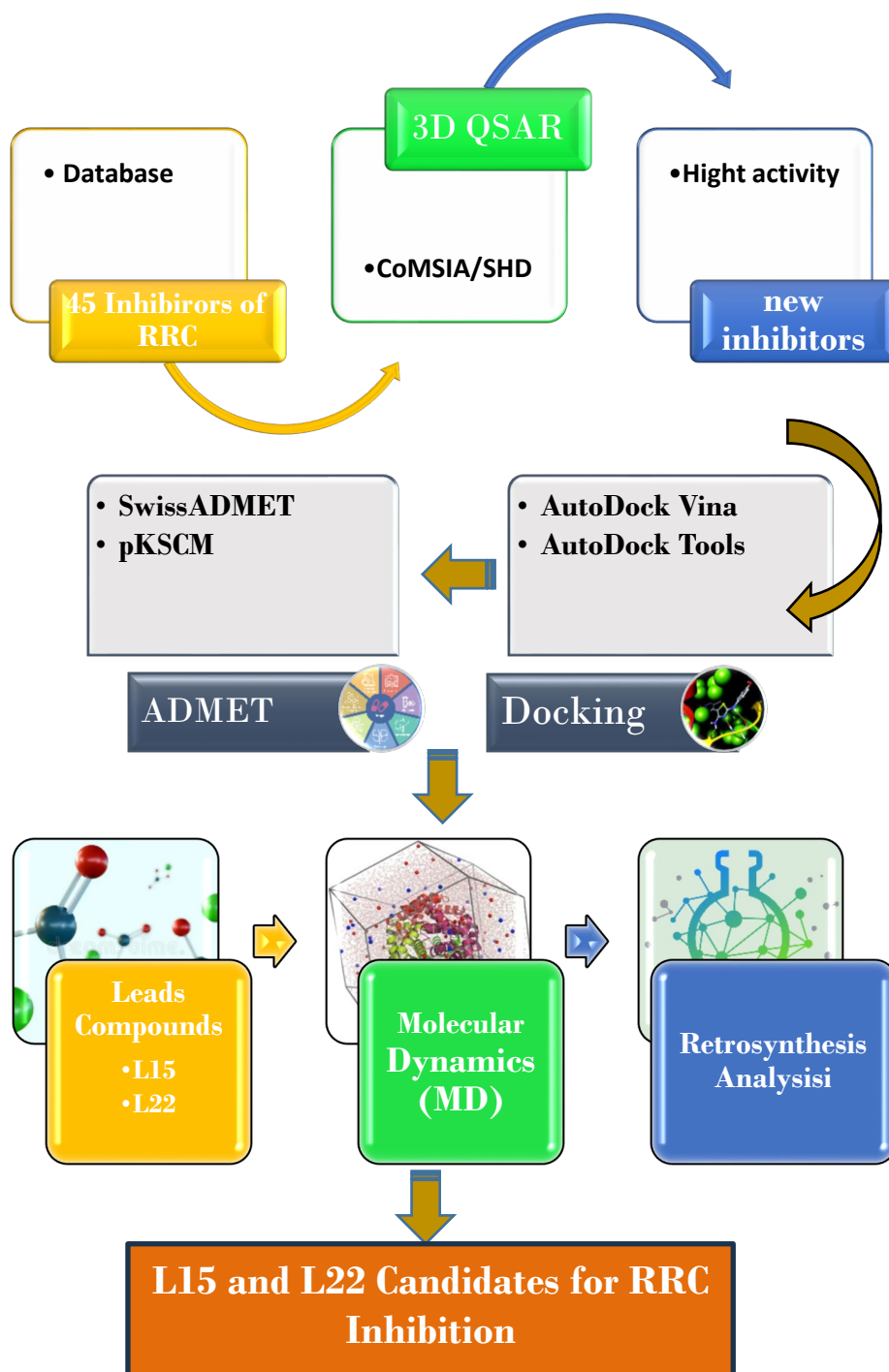


Fig. 2. Process followed for the design of new antiproliferative inhibitors using theoretical approaches.

2.2. Research of database

In this study, we conducted a 3D-QSAR analysis on a set of 45 2-substituted benzimidazole derivatives (**Fig. 3**), obtained from a recent publication³⁵. These molecules exhibited antiproliferative activity against the A-498 renal cancer cell line, with IC₅₀ values in the micromolar range, which were subsequently converted into their corresponding pIC₅₀ values (**Table 1**). Data was divided into two subsets for model development and validation with a split of 80% for training and 20% for testing³⁶. The choice of this compound series is based on the known therapeutic potential and synthetic amenability of the benzimidazole scaffolds, which makes them interesting candidates for the development of potent and selective antitumoral agents³⁷.

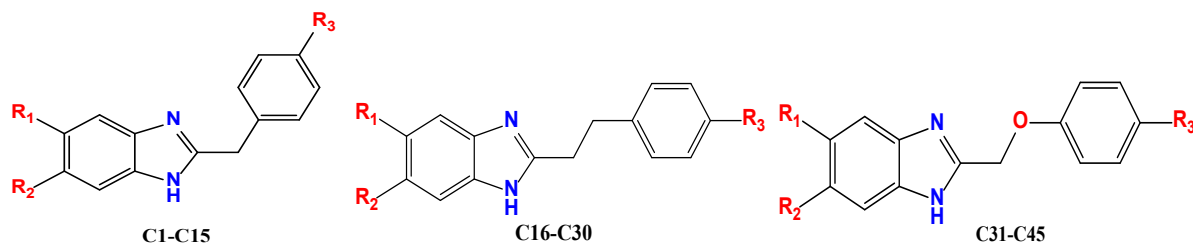


Fig. 3. Chemical structure of the studied compounds.

Table 1. Benzimidazole derivatives with antiproliferative activity against renal cancer cell line A-498.

Compounds	R ₁	R ₂	R ₃	pIC ₅₀ obs	CoMSIA/SHD	
					pIC ₅₀ (Pred)	Residual
C1	H	H	H	7.096	7.250	-0.154
C2	H	H	CH ₃	6.698	6.985	-0.287
C3*	H	H	OH	6.495	6.488	0.007
C4	H	H	Cl	6.495	6.795	-0.300
C5	H	H	F	6.698	6.699	-0.001
C6	Cl	H	H	6.495	7.154	-0.659
C7*	Cl	H	Me	7.096	7.007	0.089
C8	Cl	H	OH	6.795	6.864	-0.069
C9	Cl	H	Cl	7.096	7.754	-0.658
C10	Cl	H	F	6.698	6.678	0.020
C11*	Cl	Cl	H	6.698	6.954	-0.256
C12	Cl	Cl	Me	6.495	6.495	0.000
C13	Cl	Cl	OH	6.795	6.801	-0.006
C14	Cl	Cl	Cl	6.795	7.957	-1.162
C15	Cl	Cl	F	6.795	5.995	0.800
C16*	H	H	H	6.397	6.185	0.212
C17	H	H	CH ₃	7.096	7.854	-0.758
C18	H	H	OH	7.698	8.254	-0.556
C19	H	H	Cl	5.591	6.591	-1.000
C20	H	H	F	7.096	7.396	-0.300
C21*	Cl	H	H	6.397	6.457	-0.060
C22	Cl	H	Me	6.096	6.745	-0.649
C23	Cl	H	OH	6.397	6.438	-0.041
C24	Cl	H	Cl	7.698	7.698	0.000
C25	Cl	H	F	6.795	5.867	0.928
C26	Cl	Cl	H	7.096	7.077	0.019
C27*	Cl	Cl	Me	6.795	6.673	0.122
C28	Cl	Cl	OH	7.397	8.198	-0.801
C29	Cl	Cl	Cl	7.698	8.096	-0.398
C30	Cl	Cl	F	6.795	6.814	-0.019
C31	H	H	H	6.096	7.096	-1.000
C32	H	H	CH ₃	7.698	7.896	-0.198
C33	H	H	OH	7.397	7.318	0.079
C34	H	H	Cl	7.397	8.108	-0.711
C35*	H	H	F	6.397	6.124	0.273
C36	Cl	H	H	6.397	6.687	-0.290
C37	Cl	H	Me	7.096	7.097	-0.001
C38	Cl	H	OH	7.096	7.196	-0.100
C39	Cl	H	Cl	6.397	6.397	0.000
C40*	Cl	H	F	6.397	6.411	-0.014
C41	Cl	Cl	H	7.397	8.297	-0.900
C42	Cl	Cl	Me	7.096	7.091	0.005
C43*	Cl	Cl	OH	6.795	6.795	0.000
C44	Cl	Cl	Cl	7.397	7.784	-0.387
C45	Cl	Cl	F	5.592	6.386	-0.794

Test*

2.3. Minimization and alignment procedure

The molecular alignment process, carried out using SYBYL-X 2.0 software, plays a crucial and sensitive role in generating reliable 3D-QSAR models³⁸⁻⁴⁰. In this approach, all compounds in the dataset were energy-minimized to achieve their most stable conformations, employing the standard Tripos force field with a convergence threshold set at 0.01 kcal/mol⁴¹. For alignment purposes, **compound 18** was selected as the reference structure, as illustrated in **Fig. 4**.

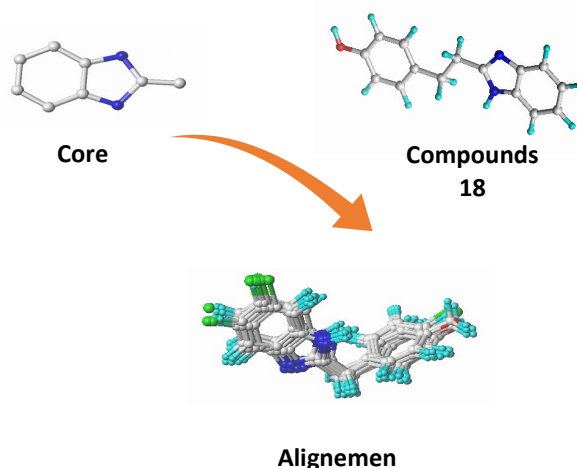


Fig. 4. Alignment of 45 studied molecules using molecule 18 as a template: (a). Molecular core:(b).

2.4. Procedure of CoMFA and CoMSIA descriptors

The development of the 3D-QSAR models was carried out using the CoMFA and CoMSIA methodologies, which are widely utilized for analyzing molecular fields and similarities⁴². These techniques were applied to estimate the antiproliferative activity (pGI_{50}^{pred}) of the investigated dipeptide salicylamides, based on various molecular interaction fields: steric, electrostatic, hydrophobic, and hydrogen bond donor and acceptor contributions. In the CoMFA approach, all compounds were aligned within a 3D grid framework with a resolution of 2.0 Å. An sp^3 hybridized carbon atom carrying a +1 charge was employed as the probe to calculate both electrostatic and steric field energies. Interaction energy cut-offs were set at 30 kcal/mol, and a column filtering threshold of 2.0 kcal/mol was used to refine the dataset⁴³. For the CoMSIA approach, descriptor calculations were carried out using a grid with a spacing of 2.0 Å. Interaction energy thresholds were limited to approximately 30 kcal/mol, and an attenuation factor of 0.3 was applied to modulate the distance-dependent decay of field values. Additionally, column filtering was employed to exclude molecular field energy variations below 2.0 kcal/mol, ensuring the relevance of retained descriptors⁴⁴.

2.5. Partial least squares (PLS) analysis

The construction of all 3D-QSAR models was carried out using the partial least squares (PLS) regression technique, which is commonly employed to investigate correlations between two variable sets⁴⁵. Here, these sets comprise the CoMSIA molecular descriptors and the pIC_{50} values representing antiproliferative activity. To determine the optimal number of components (NOC) and the cross-validated correlation coefficient (Q^2), the leave-one-out cross-validation approach was applied⁴⁶. Following this, a non-cross-validated method was used to calculate key statistical parameters, including the coefficient of determination (R^2) and other performance indicators reflecting model accuracy and predictive reliability⁴⁷.

2.6. Validation of the models

Accurate and dependable predictions from 3D-QSAR models rely on fulfilling several essential criteria. These include demonstrating robust statistical performance, validating the model's predictive accuracy with an external test set, and clearly outlining the domain of applicability within which the model's predictions are considered reliable⁴⁸. To evaluate the predicted anticancer activities using the CoMSIA model developed from the training set, the test compounds were first aligned using the same procedure applied to the training molecules. Following alignment, the coefficient of determination (R^2_{test}) was calculated to assess the model's external predictive capability⁴⁹.

$$R^2_{test} = 1 - \frac{\sum_{i=1}^{test} (Y_i^{Obs(test)} - Y_i^{Pred(test)})^2}{\sum_{i=1}^{training} (Y_i^{Obs(test)} - \bar{Y}_i^{training})^2} \quad (1)$$

Recent studies^{50,51} have highlighted that relying solely on R² test and Q² values is insufficient for fully evaluating the predictive reliability of QSAR models. As a result, additional validation parameters collectively known as the Golbraikh and Tropsha criteria are often utilized to provide a more comprehensive assessment of model robustness and external predictivity⁵².

$$\frac{(r^2-r_0^2)}{r^2} < 0.1 \text{ Or } \frac{(r^2-r_0'^2)}{r_2} < 0.1 \quad (2)$$

$$0.85 < k < 1.15 \text{ Or } 0.85 < k' < 1.15 \quad (3)$$

The statistical parameters used in the regression analysis, namely the correlation coefficient (r^2), the regression slopes (k and k'), along with the coefficients (r_0^2 and $r_0'^2$), were calculated using the following formulas:

$$r^2 = \frac{\sum(Y_{\text{test}} - \bar{Y}_{\text{test}})(Y_{\text{Pred}(\text{test})} - \bar{Y}_{\text{Pred}(\text{test})})^2}{\sqrt{\sum(Y_{\text{test}} - \bar{Y}_{\text{test}}) \sum(Y_{\text{Pred}(\text{test})} - \bar{Y}_{\text{Pred}(\text{test})})^2}} \quad (4)$$

$$r_0^2 = 1 - \frac{\sum(Y_{\text{Pred}(\text{test})} - kxY_{\text{Pred}(\text{test})})^2}{\sum(Y_{\text{Pred}(\text{test})} - kx\bar{Y}_{\text{Pred}(\text{test})})^2} \quad (5)$$

$$r_0'^2 = 1 - \frac{\sum(\bar{Y}_{\text{Pred}(\text{test})} - kx\bar{Y}_{\text{Pred}(\text{test})})^2}{\sum(Y_{\text{Pred}(\text{test})} - kxY_{\text{Pred}(\text{test})})^2} \quad (6)$$

$$K = \frac{\sum(Y_{\text{Pred}(\text{test})} - kxY_{\text{Pred}(\text{test})})^2}{\sum(Y_{\text{Pred}(\text{test})} - kx\bar{Y}_{\text{Pred}(\text{test})})^2} \quad (7)$$

$$K' = \frac{\sum(Y_{\text{Pred}(\text{test})} \times Y_{\text{test}})^2}{\sum(Y_{\text{test}})^2} \quad (8)$$

According to the recommendations provided by Roy and colleagues^{52,53}, confirming the external predictive power of a QSAR model necessitates the use of additional statistical indicators, which are derived using the equations below:

$$r_m^2 = r^2(1 + \sqrt{r^2 - r_0^2}) \quad (9)$$

$$r_m'^2 = r^2(1 + \sqrt{r^2 - r_0'^2}) \quad (10)$$

2.7. Applicability domain (AD)

Assessing the applicability domain (AD) of a QSAR model is essential to determine whether the model's predictions are valid for a given set of compounds. The AD defines a theoretical boundary within the chemical space of the training data, based on both the descriptors and the modeled biological activity⁵⁴. This evaluation ensures that predictions are only made for compounds whose properties fall within the defined limits of the model. If a compound lies outside this boundary, its predicted activity may be unreliable. According to⁵⁵ Roy et al., several approaches can be used to define and assess the applicability domain, but the most frequently used method is the determination of the leverage values effect h_i for each compound.

$$h_i = xT_i (XTX)^{-1} x_i \quad (11)$$

The Williams plot is commonly employed to define the applicability domain (AD) within a specified square area⁵⁶, using the leverage threshold h^* as a reference point^{40,57}. If a compound exhibits a leverage value (h_i) exceeding the warning threshold h^* , it is considered to exert an undue influence on the model and is thus classified as falling outside the applicability domain.

$$h^* = 3 \times (K+1) n \quad (12)$$

where **n** represents the number of compounds in the training set, and **k** denotes the number of descriptors selected for the model.

2.8. ADMET profile and drug-likeness analysis

Understanding the pharmacokinetic and toxicity-related properties of potential drug candidates commonly referred to as ADME/T (Absorption, Distribution, Metabolism, Excretion, and Toxicity) is a key aspect of successful drug development^{58,59}. Challenges related to these properties are believed to contribute to approximately 40% of drug candidate failures during the development process. As a result, computational tools that predict ADME/T characteristics have become increasingly valuable for guiding early-stage drug discovery. A variety of freely accessible online platforms now support such predictions. Notably, tools like SwissADME and pkCSM enable researchers to evaluate a compound's pharmacokinetic behavior and assess its overall drug-likeness based on its molecular structure^{60,61}. Access to comprehensive and up-to-date databases is essential for building accurate computational models that predict the drug-like potential of biomolecules. Online tools designed to evaluate ADME/T properties offer valuable guidance in distinguishing compounds with favorable pharmacokinetic profiles from those unlikely to succeed as drug candidates.

In this study, we employed two widely used web-based platforms **SwissADME** (<http://www.swissadme.ch/>) and **pkCSM** (<http://structure.bioc.cam.ac.uk/pkcsm>) to predict and analyze the pharmacokinetic characteristics of selected molecules⁶²

In the context of bioavailability, drug-likeness describes the probability that a given compound possesses the necessary characteristics to become a viable drug candidate. This concept is typically assessed by analyzing the structural and physicochemical properties of molecules to determine their therapeutic potential. In this study, we utilized **SwissADME**, a computational tool that applies five distinct rule-based filters to evaluate key molecular features⁶³. These filters help determine whether a compound aligns with commonly accepted criteria for drug-like behavior.

2.9. Molecular Docking

To explore the notable selectivity of the compounds in inhibiting cell proliferation in renal cell carcinoma (RCC), specifically the A498 cell line, we investigated their potential molecular interactions. This analysis was performed using AutoDock Vina⁶⁴ and AutoDock Tools 1.5.6⁶⁵⁻⁶⁸, with particular attention given to the aromatase receptor (PDB ID: 3S7S), a protein closely associated with RCC pathology. It is important to highlight that the aromatase receptor plays a crucial role in suppressing the proliferation of renal cell carcinoma by inhibiting estrogen synthesis. Aromatase inhibitors, which block estrogen production, have been shown to significantly reduce RCC cell growth. Targeting aromatase has emerged as a key strategy in anticancer therapy, positioning these inhibitors as critical agents in the fight against cancer cell proliferation.

2.10. Molecular dynamic simulation

2.10.1. Molecular dynamic simulation of protein-ligand complex

The protein–ligand complex underwent an all-atom molecular dynamics (MD) simulation. We prepared the system using the CHARMM-GUI server⁶⁹ and the Solution Builder protocol. The protein was protonated at physiological pH (7.4). Ligand parameters were generated using the CGenFF tool, while the protein was parametrized with the CHARMM36m force field. The system was solvated by the TIP3P water model in a periodic wall cubic box with 10 Å space between the protein and the wall. System neutrality and physiological ionic strength were achieved by the insertion of counter-ions at 0.15 M. The electrostatics and van der Waals interactions were simulated using the Verlet cutoff scheme. Bond lengths were constrained with the LINCS algorithm for the system. The long-range electrostatic forces were accurately computed using the particle mesh Ewald (PME) method. Before the production run, energy minimization was carried out using the steepest descent (SD) algorithm to eliminate unfavourable atomic contacts and reduce the system's potential energy to below 1000 kJ·mol⁻¹·nm⁻¹. This step was succeeded by a two-stage equilibration, the first to fix the temperature using a thermostat [NVT ensemble] and the second to stabilize the pressure using a barostat [NPT ensemble]. This process fixed the system's temperature (310.15 K) and pressure (1 bar) to allow for energy interchange, resulting in the achievement of thermodynamic equilibrium. The molecular dynamics (MD) simulation was then carried out for 100 ns with GROMACS software (version 2024.5)⁷⁰. Simulation parameters for the L15 and L22 protein–ligand complexes are summarized in **Table 2**.

Table 2. Molecular Dynamics Simulation Parameters for Protein-Ligand Complexes L15 and L22.

Protein-Ligand Complex		Complex L15-3s7s		Complex L22-3s7s	
MD Simulation Type		Conventional All-atom MD simulation			
Parametrization Forcefield		Charmm36m (Charmm-gui Solution Builder)			
System Solvation		TIP3P water molecule			
		10 Å extended from the protein			
System Neutralization		Sodium ions		66	
		Chloride ions		70	
Nstep Energy Minimization		not exceeding 1000 Kj/mol/nm			
System Equilibration		NVT and NPT ensemble			
MD Simulation time, nstep		100 ns (5000000 nstep)			
MD Simulation Temperature		310.15 K			
Gromacs Software Version		2024.5			

2.10.2. Trajectory Analysis and Binding free energy calculation

Following a 100 ns molecular dynamics (MD) simulation, trajectory analysis was conducted utilizing the `gmx_energy`, `gmx_rms`, `gmx_rmsf`, `gmx_gyrate` and `gmx_sasa` modules to compute total energy, temperature, root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), and radius of gyration (Rg) calculation for the protein. The calculation of MMPBSA of the protein-molecule complex was calculated at last 20 ns of md simulation using `gmx_MMPBSA` Ver 1.6.4 developed by Valdés-Tresanco et al.⁷¹.

2.11. Retrosynthesis

In the realm of synthetic organic chemistry, formulating an efficient method for molecule construction is essential for producing compounds across different scales. Such planning facilitates the creation of viable synthetic routes, enabling chemists to optimize both efficiency and cost when targeting specific molecules. In this study, we employed the ASKCOS platform, an open-source and freely accessible tool developed to assist with retrosynthetic planning of the leads compounds L15 and L22, available at <https://askcos.mit.edu>. ASKCOS features a one-step retrosynthesis module that identifies direct precursors of the target molecule. It also includes an interactive pathway planning tool, allowing users to manually combine one-step retrosynthesis results to construct a complete synthetic route by suggesting the required reaction conditions. Additionally, the Tree-Builder module can automatically generate multi-step retrosynthetic pathways⁷².

3. Results and discussions

3.1. CoMSIA statistical results

Several CoMSIA models were built to consider distinct combinations of the five molecular fields (steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D), and hydrogen bond acceptor (A)). Only those models that satisfied the desired tools validation criteria were kept, and their statistics are shown in **Table 3**. All chosen models showed significant results, and the CoMSIA model with SHD fields remained the best (**Fig. 5**). This model had a very high R^2 (0.984), satisfactory cross-validation Q^2 (0.751), and superior predictive ability for the test set ($R^2_{\text{pred}} = 0.924$), which implied that the CoMSIA/SHD model was robust and reliable.

Table 3. Statistical metrics from PLS analysis of optimized CoMSIA models.

QSAR Models	R^2	Q^2	F	SEE	NOC	R^2_{pred}	Fraction				
							S	E	H	D	A
CoMSIA/S	0.937	0.523	98.125	0.245	6	0.701	1	-	-	-	-
CoMSIA/E	0.889	0.517	88.746	0.199	5	0.679	-	1	-	-	-
CoMSIA/H	0.965	0.531	89.823	0.163	6	0.699	-	-	1	-	-
CoMSIA/D	0.887	0.509	75.890	0.266	6	0.788	-	-	-	1	-
CoMSIA/A	0.940	0.566	92.673	0.116	5	0.899	-	-	-	-	1
CoMSIA/SH	0.901	0.611	124.245	0.123	5	0.854	0.487	-	0.513	-	-
CoMSIA/SE	0.915	0.563	101.265	0.214	5	0.632	0.430	0.570	-	-	0.430
CoMSIA/SD	0.905	0.603	106.575	0.110	5	0.760	0.488	-	-	0.512	-
CoMSIA/SA	0.907	0.543	89.645	0.187	6	0.630	0.573	-	-	-	0.427
CoMSIA/EA	0.899	0.534	96.853	0.157	6	0.612	-	0.530	-	-	0.470
CoMSIA/HA	0.911	0.518	87.5245	0.254	5	0.763	-	-	0.562	-	0.438
CoMSIA/HD	0.931	0.600	110.231	0.234	5	0.637	-	0.544	-	0.456	-
CoMSIA/HE	0.955	0.579	99.756	0.187	5	0.875	-	0.489	0.611	-	-
CoMSIA/DA	0.908	0.533	93.785	0.163	5	0.701	-	-	-	0.645	0.355
CoMSIA/DE	0.944	0.569	112.756	0.172	5	0.733	-	0.392	-	0.608	-
CoMSIA/SHD	0.984	0.751	145.85	0.096	5	0.924	0.300	-	0.280	0.420	-
CoMSIA/SHA	0.921	0.502	135.756	0.258	5	0.811	0.342	-	0.360	-	-
CoMSIA/SHE	0.912	0.511	99.785	0.121	5	0.685	0.321	0.256	0.423	-	-
CoMSIA/SAD	0.925	0.574	77.555	0.241	5	0.772	0.412	-	-	0.356	0.232
CoMSIA/EAD	0.909	0.509	80.643	0.273	6	0.621	-	0.265	-	0.395	0.340
CoMSIA/EDH	0.925	0.577	90.472	0.263	6	0.786	-	0.000	0.000	0.000	-
CoMSIA/ESA	0.877	0.519	76.333	0.288	6	0.644	0.320	0.351	-	-	0.329
CoMSIA/EHA	0.868	0.457	65.875	0.247	6	0.610	-	0.195	0.410	-	0.395
CoMSIA/HAD	0.935	0.588	78.587	0.231	5	0.688	-	0.278	-	0.368	0.354
CoMSIA/HDA	0.970	0.546	121.654	0.142	5	0.855	-	-	0.310	0.402	0.288
CoMSIA/SEHD	0.939	0.683	140.257	0.103	5	0.914	0.254	0.115	0.286	0.345	-
CoMSIA/SEHA	0.926	0.512	145.278	0.099	5	0.901	0.287	0.201	0.305	-	0.207
CoMSIA/SHAD	0.941	0.656	122.532	0.117	5	0.918	0.325	0.178	-	0.286	0.211
CoMSIA/EHDA	0.909	0.532	102.98	0.213	5	-	0.249	0.181	0.208	0.251	0.111
CoMSIA/SEDA	0.931	0.511	109.77	0.145	5	0.123	0.263	-	0.187	0.238	0.189
CoMSIA/SEHDA	0.944	0.688	121.214	0.122	5	0.918	0.270	0.137	0.257	0.210	0.136

The best model was found to include three terms: steric, hydrophobic, and hydrogen bond donor: 30%, 28%, and 42%. This model had a high Fisher test value ($F = 145.85$), a low standard error of estimate (SEE) around 0.092, and an optimal

number of components ($n = 5$). CoMSIA/SHD, CoMSIA/SHAD, CoMSIA/SEHD, and CoMSIA/SEHDA models with steric, hydrophobic, and donor fields had also shown good performance. This work shows a direct correlation between the three domains and antiproliferative activity.

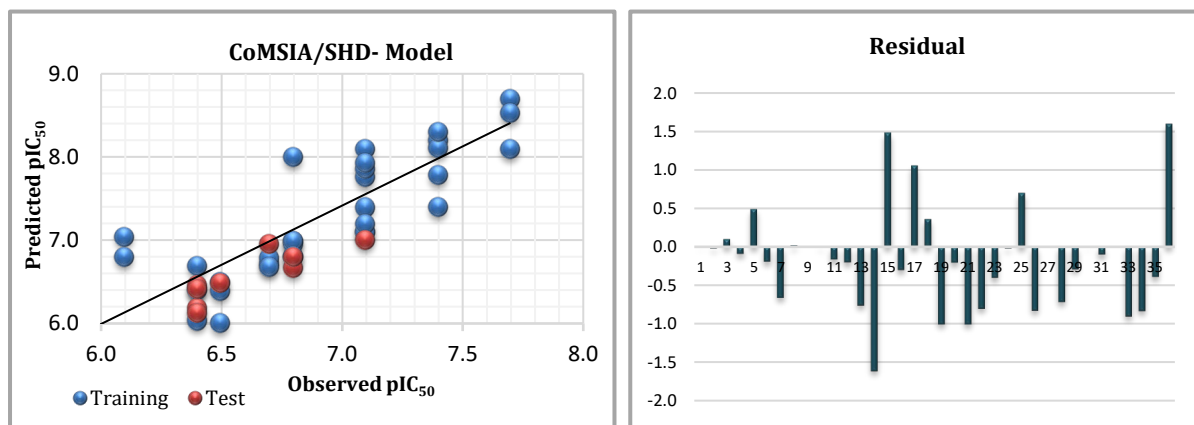


Fig. 5. Correlation and residual plots for CoMSIA/SHD predictions.

3.2. Validation of the models

Four models were selected according to the statistics parameters of R^2 and Q^2 obtained during the process of model development and internal validation phases and with satisfactory performance ($Q^2 > 0.6$) (Table 3). The quality of predictions of the most reliable developed model was assessed further during external validation according to Golbraikh and Tropsha. For this purpose, nine compounds were designated as a test set to assess the model's predictive capability. As presented in Table 4, the external validation results confirm that the CoMSIA model complies with all of Golbraikh and Tropsha's requirements. The results confirm the strong predictive performance of the model, supporting the selection of the CoMSIA/SHD model as the optimal candidate for further analysis.

Table 4. Golbraikh, Tropsha and Roy criteria of the selected CoMSIA models.

Parameter	SEHDA	SEHD	SHD	SHDA	Threshold
R^2	0.944	0.939	0.984	0.941	$R^2 > 0.6$
Q^2	0.688	0.683	0.751	0.656	$Q^2 > 0.5$
r^2_{pred}	0.918	0.914	0.924	0.918	$r^2 > 0.6$
r_0^2	0.878	0.889	0.910	0.900	$r_0^2 > 0.5$
$r_0'^2$	0.899	0.897	0.914	0.903	$r_0'^2 > 0.5$
$\frac{(r^2 - r_0^2)}{r^2} < 0.1$	0.043	0.027	0.015	0.020	< 0.1
$\frac{(r^2 - r_0'^2)}{r^2} < 0.1$	0.020	0.019	0.010	0.016	< 0.1
τ_2					
$0.85 < k$	0.911	0.931	0.952	0.922	$0.85 < k < 1.15$
$0.85 < k'$	1.101	1.058	1.033	1.041	$0.85 < k' < 1.15$
r_m^2	0.734	0.769	0.814	0.795	$r_m^2 > 0.5$
$r_m'^2$	0.791	0.795	0.831	0.805	$r_m'^2 > 0.5$
Δr_m^2	0.057	0.026	0.017	0.010	$\Delta r_m^2 < 0.2$
Δr_0^2	0.021	0.008	0.004	0.003	$\Delta r_0^2 < 0.03$

3.3. Applicability domain (AD)

Additionally, the applicability domain (AD) of the test compounds was assessed using the leverage method, and the outcomes were visualized using a Williams plot to examine predictive reliability. Hence, the applicability domain of the QSAR model developed using CoMSIA method was evaluated by examining the relationship between standardized residuals and leverage values.

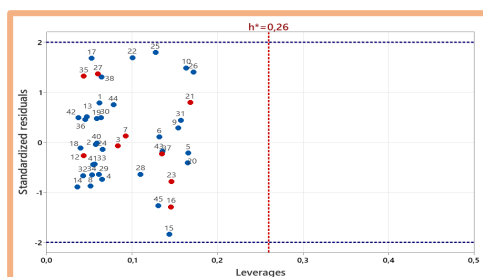


Fig. 6. Williams plot for CoMFA/SHD model.

The Williams plot was employed to visualize the model's applicability domain. The threshold for the leverage value was determined to be $h^* = 0.26$. Both the distribution of standardized residuals and leverage values are presented in **Fig. 6**. Since all compounds exhibit leverage values below the critical threshold ($h_i < h^*$), the CoMSIA/SHD demonstrates robust statistical performance and strong predictive reliability.

3.4. Graphical result of the CoMSIA/SHD model

To obtain more information on the essential molecular requirements described by the CoMSIA/SHD model, we generated steric, hydrophobic, and hydrogen bond donor contour maps, using the most active compound of the training set (Molecule 18) as reference. The contour plots obtained are shown in **Fig. 7**.

The steric contour map (**Fig. 7a**) is shown in two colours, green and yellow, covering 80% and 20% of the mapped region, respectively. The green contours represent areas where bulky substituents are favoured, and the yellow contours show where the presence of such groups is unfavourable. A large green area is seen around the R3 position and its nearby substituent position, suggesting that adding a large substituent group there could boost biological activity.

The hydrophobic contour map of the CoMSIA model (**Fig. 7b**) is shown with yellow and grey contours, indicating hydrophobic and hydrophilic substituents, respectively. The calculation shows that a significant yellow contour exists around the aliphatic linker connecting the benzimidazole nucleus to the aromatic ring, along with two moderate yellow regions at the R2 and R3 substitution patterns. These yellow regions indicate that it is unfavourable (i.e., the hydrophobic group stabilizes the regions) to add the hydrophobic group to these areas. The grey area near the Meta position of the R3 substituent indicates that making hydrophobic changes at this spot could help improve biological activity.

The hydrogen bond donor contour map of the CoMSIA (**Fig. 7c**) shows cyan and magenta contours, representing favourable interaction regions for the hydrogen bond donor and acceptor groups, respectively. A large cyan contour is located on map (C) at the R3 substitution position, with a smaller one at the Meta of R3. These results raise the possibility that by incorporating hydrogen bond donors in these regions, the biological activity of the compound could be improved.

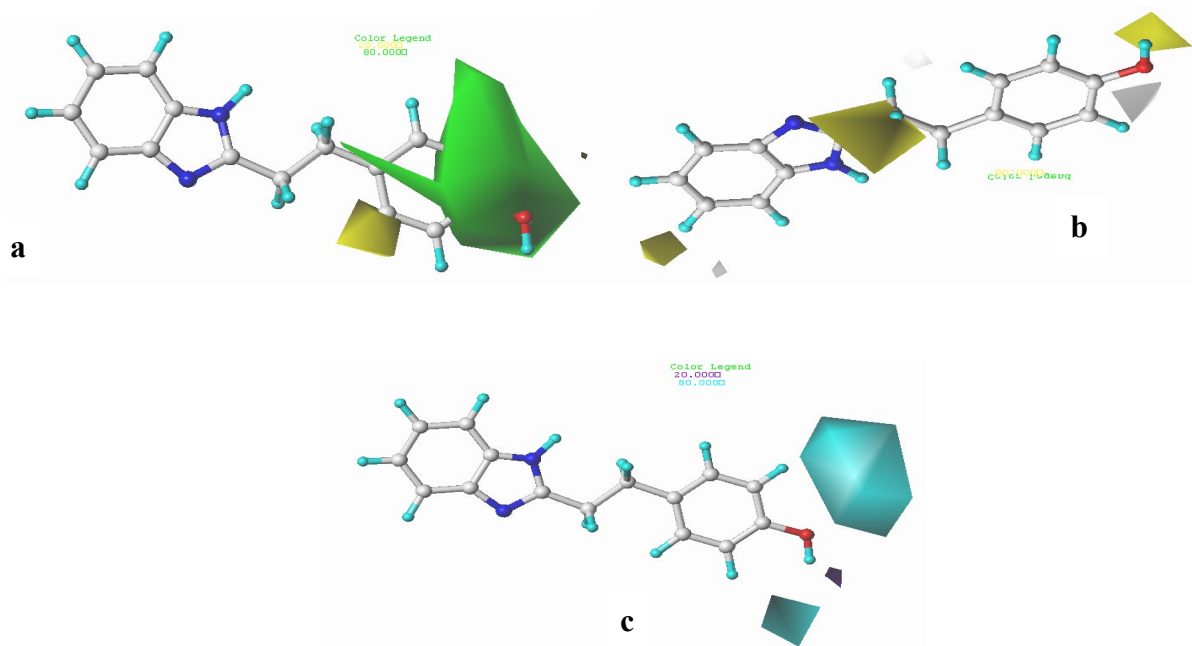


Fig. 7. Steric: (a), hydrophobic: (b) and H-Bond Donor (c) contour maps from CoMSIA/SHD analysis.

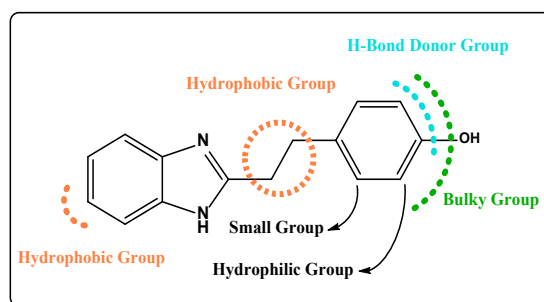


Fig. 8. Main structure-activity relationships derived from this study.

Moreover, further information on the structure–activity relationship can be extracted by the analysis of the key feature of the CoMSIA/SHD contour maps (Fig. 8). With this in mind, we were motivated to create and suggest five new types of 2-substituted benzimidazoles (L7, L15, L19, L22, and L39), which showed strong ability to stop the growth of the renal cancer cell line A-498.

3.5. Design for new Benzimidazole as anticancer agents

The aim of this work is to find out new and potent anticancer leads compounds. The method relies on 3D QSAR modelling and comparative molecular field analyses of benzimidazole analogs (see Fig. 8). The hydrogen bond donor field was highlighted as the most important factor affecting the anticancer activity of the compounds, covering 42% of the observed differences as indicated by the QSAR model. Based on this, numerous donor groups, such as -OH, were designedly installed at R3 and the meta-position to R3. However, in medicinal chemistry, alcohol is usually seen as a hydrogen bond donor, meaning the whole -OH group when it interacts with biological targets like proteins or DNA. The rational design strategy led to the discovery of five new compounds that are predicted to have pIC_{50} (pred) values even greater than the most active compound in the initial set (C18 (pIC_{50} (obs) = 7.698)), and this knowledge is shown in Fig. 9.

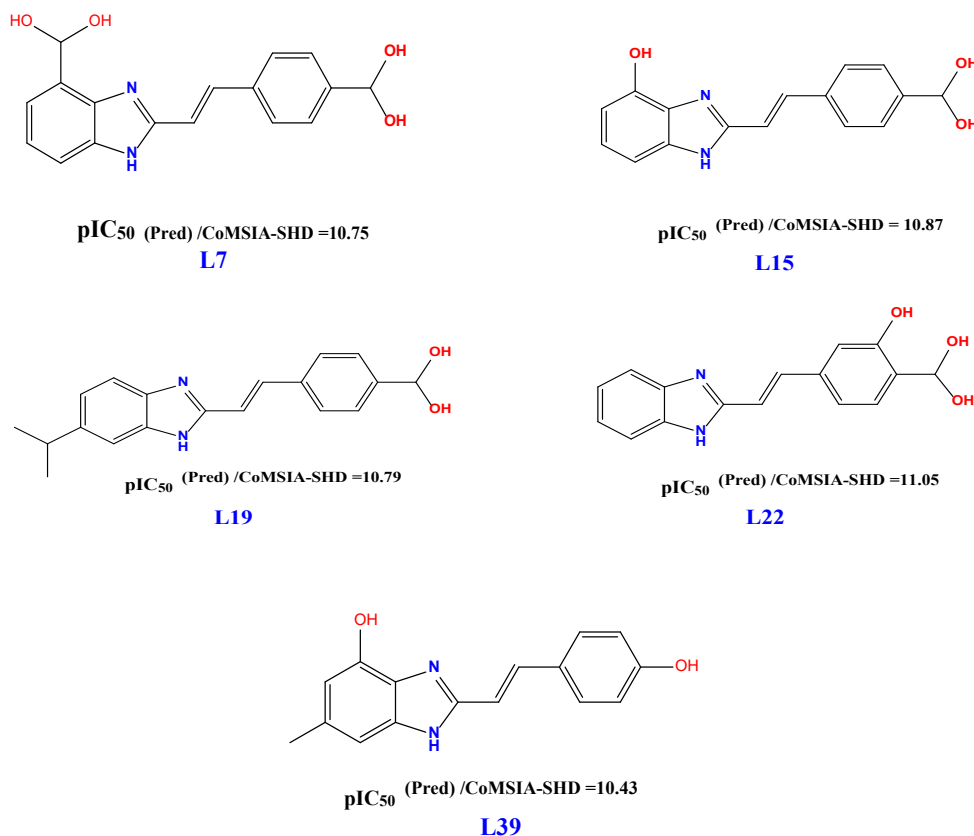


Fig. 9. Structures of the newly compounds and their activity based on CoMSIA/SHD model.

3.6. Molecular Docking Analysis

In this study, we also did a molecular docking analysis to estimate how well five new compounds (L7, L15, L19, L22, and L39) can block renal cell carcinoma (RCC). Regimens were chosen on the basis of their predicted IC_{50} values, which suggested good inhibitory potential. Model prediction and docking AutoDock Vina performed the docking process, using the aromatase receptor structure (PDB ID: 3s7s) as the reference. To compare the binding of the candidate compounds with the co-crystallized reference ligand, the binding pose was used to evaluate their binding affinities and interaction profiles. To do this, the detailed structure of the aromatase receptor (PDB ID: 3S7S, 2.2 Å) was obtained from the Protein Data Bank (<http://www.rcsb.org>) after careful selection to ensure it was suitable and accurate. The receptor is chosen on the basis of its purported role in promoting the growth, proliferation, and development of cancer cells. Reports also indicate that aromatase inhibitors, which inhibit estrogen synthesis, inhibit the proliferation of RCC. Consequently, investigation of the binding affinity of course to aromatase of our compounds (L7, L15, L19, L22, and L39) could shed light on the anticancer activity of these compounds.

Initially, the active site of the protein was identified by looking at how it interacts with the enzyme's natural ligand, as shown in Fig. 10. This analysis found that the amino acids ARG115, ALA306, ILE133, MET374, PHE134, VAL370, TRP224, VAL373, PHE221, and LEU477 are important parts of the likely active site of aromatase (PDB ID: 3s7s). These findings illuminate the molecular interactions at play and establish the foundation for comprehending the potential inhibition

of the enzyme by novel drug candidates. Consequently, these particular amino acids are vital to the molecular docking processes employed in this study. To check how accurate the molecular docking method was in this study, the co-ligand was docked again at the same active site (PDB ID: 3S7S). Grid maps were generated with a span of 80 Å along each axis (X, Y, Z), with a grid spacing of 0.375 Å. The centre of the grid box was set by finding where the original ligand (Exemestane) was located in the protein at the coordinates 85.834, 51.078, and 45.723 Å. This step was essential to confirm the docking method's reliability in probing interactions within the defined active site.

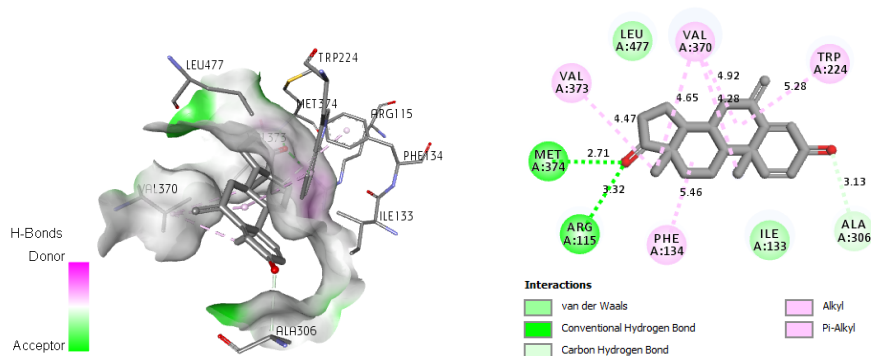


Fig. 10. (a) 2D and (b) 3D) interactions Complex 3S7S-Exemestane.

Following the molecular re-docking, nearly perfect overlay of the native and the re-docked ligands within the active site cavity of the aromatase enzyme are shown in **Fig. 11**. The RMSD of 1.428 Å is however below the 2 Å limit, which indicates that the AutoDock Vina software is reliable and accurate in fitting the ligand's original position. Based on this validation, we conducted molecular docking investigations in the aromatase active site using the selected compounds (L7, L15, L19, L22, and L39). **Figure 12** shows that the original ligand (Exemestane) attaches to a spot that is outside the main binding area, interacting with important active site residues: ARG115, PHE134, ILE133, MET374, ALA306, VAL370, TRP224, VAL373, PHE221, and LEU477. The re-docking study also showed that exemestane remains in touch with those amino acids, similar to how it was positioned in the crystal structure of the exemestane-3S7S complex (**Fig. 10**).

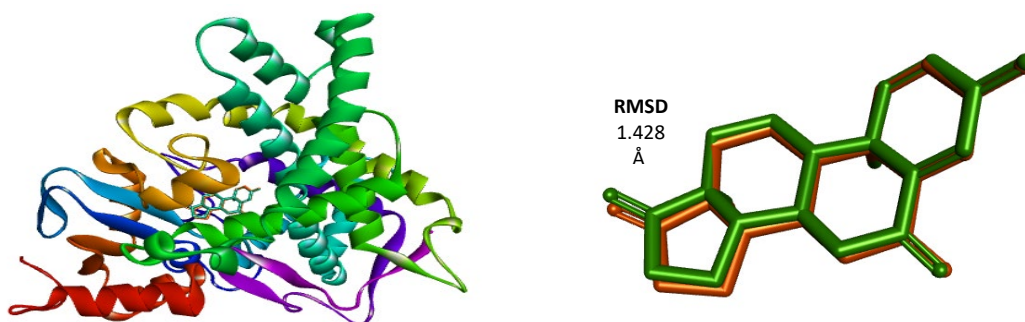


Fig. 11. Re-docking pose with RMSD value it 1.428 Å (Green = Original, Brown = Docked).

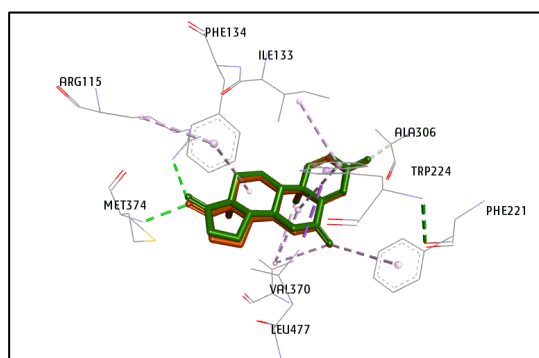


Fig. 12. 3D interactions of Exemestane-3S7S complex after Re-docking.

For the first time, we identify the residues of the active site (ARG115, ALA306, ILE133, MET374, PHE134, VAL370, TRP224, VAL373, PHE221, and LEU477) directly involved in the binding of exemestane, and we have a map of those molecular interactions that are essential for the inhibition of the enzymatic activity. These residues will be used as landmarks in exploring the enzymology of renal cell carcinoma. Therefore, if any interaction occurs between these key sites and the selected ligands (L7, L15, L19, L22, and L39), it will imply that it might inhibit the aromatase.

Suppression of aromatase activity may inhibit the growth of renal cancer cells. In addition, the binding energy for the ligands will be calculated to find out which compound (ligand) makes the best complex in the binding pocket of the aromatase. As plotted in **Fig. 13**, molecular-docking results indicate that ligands L7, L15, L19, L22, and L39 successfully docked into the aromatase receptor pocket and can interact with key residues such as ARG115, ALA306, ILE133, ILE132, ARG145, ALA307, TRP141, LEU152, MET311, MET303, and CYS437. A summary of the interactions of the ligands (L7, L15, L19, L22, and L39) and the reference ligand (Exemestane) with the active sites of the aromatase receptor with their corresponding binding energies is presented in **Table 5**.

In the 3S7S-L22 complex, the ligand L22 ($pIC_{50} = 11.05$) is observed to interact with four key residues known to play a critical role in the inhibition of renal cell carcinoma. These key binding sites include ARG115 (2.39 Å), ALA306 (3.85 Å), ILE133 (4.94 Å), and again ALA306 (4.00 Å). The interactions involved are as follows: a hydrogen bond is formed with ARG115, two Pi-Sigma interactions occur with ALA306, and a Pi-alkyl interaction is established with ILE133. Furthermore, ligand LA01 establishes three additional hydrogen bonds with the receptor's active site, specifically involving residues ARG145 (2.60 Å), TRP141 (2.60 Å), and ILE132 (2.60 Å). However, these particular residues were not previously identified as critical for the inhibition of aromatase enzyme activity (refer to **Fig. 12**). It is also worth noting that ligand L22 exhibits a binding energy of -10.58 kcal/mol with the aromatase receptor. This strong affinity may be attributed to the presence of three donor groups (-OH) substituted at the R3 position, including one at the adjacent Ortho position. Previous studies have highlighted that introducing donor groups at this specific location significantly influences the compound's inhibitory effectiveness.

In the 3S7S-L15 complex, ligand L15 ($pIC_{50} = 10.87$) forms interactions with five key amino acid residues that are considered crucial for inhibiting aromatase enzyme activity. These interactions occur with ILE133 at distances of 2.31 Å and 5.07 Å, as well as with ALA306 at distances of 2.94 Å, 3.82 Å, and 4.02 Å. The interaction between ligand LA02 and these residues involves several types of molecular contacts: hydrogen bonds are formed with both ILE133 (2.31 Å) and ALA306 (2.94 Å), Pi-alkyl interactions occur with the same two residues, and a Pi-sigma interaction is specifically observed with ALA306 (3.82 Å). Ligand L15 also establishes two hydrogen bond interactions with the residues TRP141 (1.98 Å) and ARG435 (2.29 Å). However, these amino acids were not previously identified as key sites involved in the inhibition of aromatase enzyme activity, as indicated in **Fig. 12**. Additionally, the binding energy of L15 with the aromatase receptor is calculated to be -10.4 kcal/mol, reflecting a strong binding affinity.

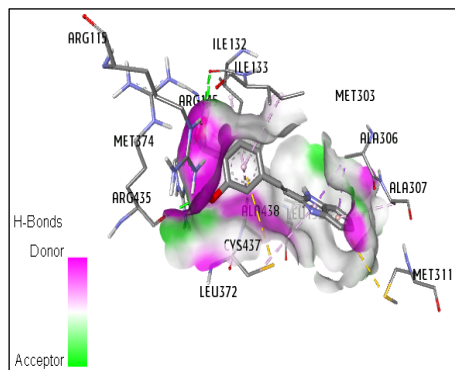
In the 3S7S-L19 complex, ligand L19 ($pIC_{50} = 10.79$) interacts with three key reference sites involved in the inhibition of renal cell carcinoma. These interactions include hydrogen bonds with ARG115 (2.29 Å) and ILE133 (2.10 Å), as well as a Pi-alkyl interaction with ILE133 (2.29 Å). Additionally, LA03 binds to the aromatase receptor through four hydrogen bonds with residues ARG435 (1.88 Å and 2.27 Å), TRP141 (2.29 Å), and ILE132 (2.40 Å), although these amino acids were not identified as crucial for aromatase inhibition. The optimal docking pose of L19 has a binding score of -10.3 kcal/mol.

In the 3S7S-L7 complex, three significant interactions are observed between ligand L7 ($pIC_{50} = 10.55$) and the aromatase receptor, which are similar to those seen with Exemestane as the reference ligand. These interactions include hydrogen bonds with ILE133 (2.51 Å) and ALA306 (2.52 Å). In addition, L7 forms hydrogen bonds with residues TRP141 (1.84 Å) and ARG145 (2.34 Å), although these residues have not been previously identified as essential for the inhibition of aromatase enzyme activity. Furthermore, ligand L7 exhibits a binding energy of -10.3 kcal/mol, which is lower than that of Exemestane -10.2 kcal/mol.

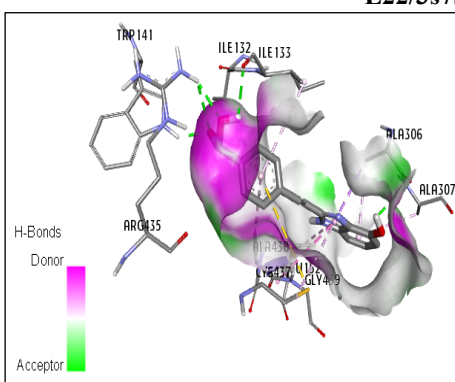
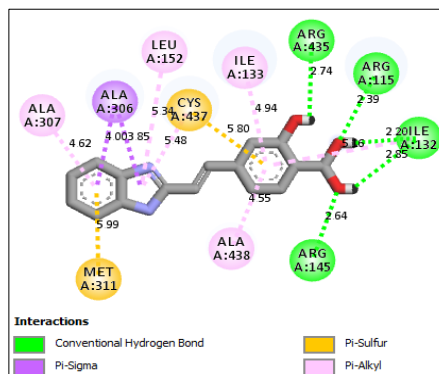
In the 3S7S-L39 complex, ligand L39 ($pIC_{50} = 9.69$) interacts with two key reference sites involved in the inhibition of aromatase enzyme activity. These interactions include hydrogen bonding and Pi-alkyl interactions with ILE133 at distances of 2.30 Å and 2.86 Å, respectively. Additionally, L39 binds to the aromatase receptor through other hydrophobic interactions with residues ARG435 (1.88 Å), TRP141 (2.29 Å), ARG435 (2.27 Å), and ILE132 (2.40 Å), although these residues have not been previously identified as important for inhibiting aromatase enzyme activity. The affinity for the optimal interaction pose is -9.7 kcal/mol.

A comparison of the key interaction sites involved in aromatase inhibition reveals that ligands L22 and L15 engage with the highest number of critical residues (five in total) while L19 and L7 each interact with four, and L39 with only two. Regarding binding affinity, all proposed ligands (L22, L15, L19, L7, and L39) exhibit lower binding energy values than Exemestane (-10.2 kcal/mol), with the exception of L39, which shows a binding energy of -9.7 kcal/mol. These results suggest that the designed ligands, particularly L22 and L15, demonstrate greater stability and stronger binding within the aromatase receptor pocket compared to Exemestane, indicating their potential as more effective inhibitors. Based on the molecular docking analysis, ligands L22 and L15 demonstrate the strongest potential as aromatase inhibitors among the

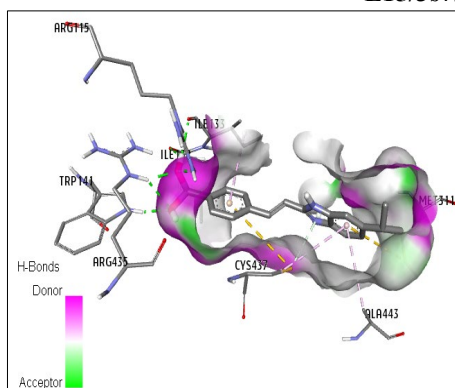
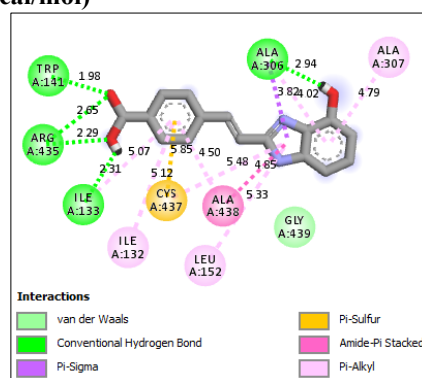
five tested compounds (L22, L15, L19, L7, and L39). This conclusion is supported by their interaction with the highest number of key residues and their notably favorable binding energy values. These findings suggest that L22 and L15 are the most promising candidates for effectively inhibiting the enzymatic activity of the aromatase receptor. Furthermore, the variation in inhibitory effectiveness among the five ligands can be attributed to differences in the positioning of their functional groups and the orientation of their molecular structures within the binding pocket, which influence their interactions with critical active site residues. Additionally, it can be concluded that the position and nature of the substituent groups within the molecular structures of the compounds (L22, L15, L19, L7, and L39) play a significant role in influencing their biological activity, particularly their pIC_{50} values.



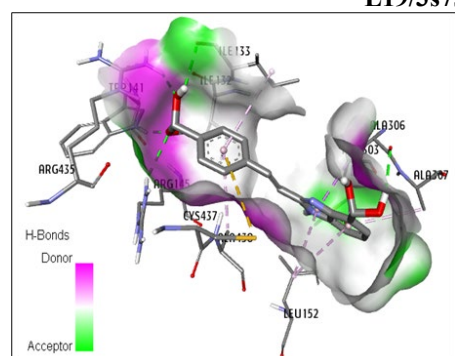
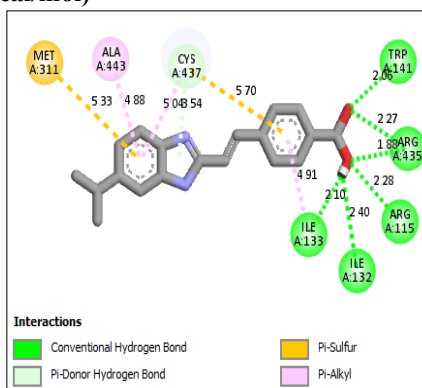
L22/3s7s, (Affinity -10.6 Kcal/mol)



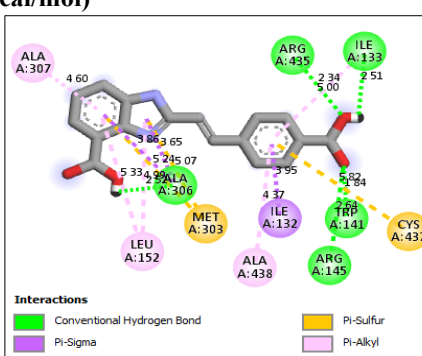
L15/3s7s, (Affinity -10.4 Kcal/mol)

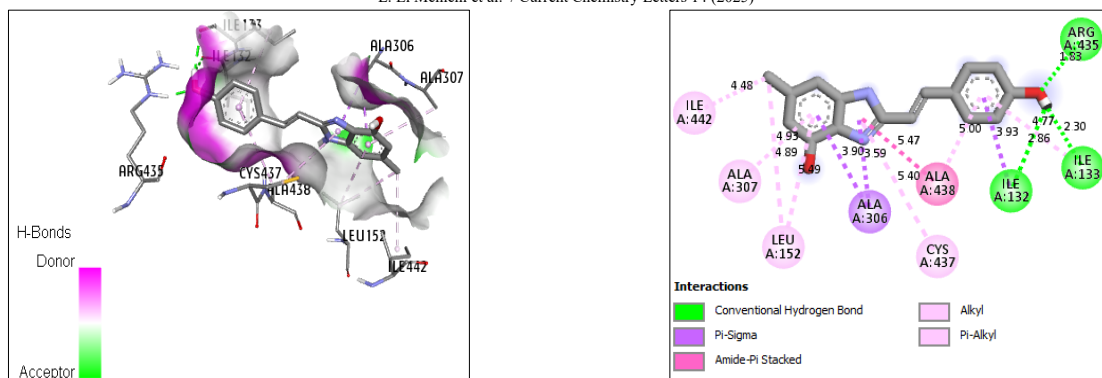


L19/3s7s, (Affinity -10.3 Kcal/mol)



L7/3s7s, (Affinity -10.3 Kcal/mol)





L39/3s7s, (Affinity -9.7 Kcal/mol)

Fig. 13. 3D and 2D docking of designed compounds.

Table 5. Binding affinities and interactions of reference and designed compounds with aromatase receptor.

Protien	Ligands	Affinity Kcal/mol	Hydrogen bonding Interaction		Hydrophobic Interaction			
			Amino acid residue	Distance (Å)	Amino acid residue	Distance (Å)		
Aromatase (3s7s)	L22	-10.6	ARG-115	2.39	Pi-Alkyl (ALA-306)	3.85		
			ARG-145	2.64	Pi-Alkyl (ALA-306)	4.00		
			ARG-435	2.74	Pi-Alkyl (ALA-438)	4.55		
			ILE-132	2.20	Pi-Alkyl (ALA-307)	4.62		
			ILE-132	2.85	Pi-Alkyl (ILE-133)	4.94		
			-	-	Pi-Alkyl (LEU-152)	5.34		
			-	-	Pi-Alkyl (CYS-437)	5.48		
			-	-	Pi-Sulfur (MET-311)	5.99		
			-	-	Pi-Sulfur (CYS-437)	5.80		
	L15	-10.4	TRP-141	1.98	Pi-Sigma (ALA-306)	3.82		
			ARG-435	2.29	Pi-Alkyl (ALA-306)	4.02		
			ILE-133	2.31	Pi-Alkyl (ALAL-307)	4.79		
			ARG-435	2.65	Pi-Alkyl (ALA-438)	4.50		
			ALA-306	2.94	Pi-Stacked (ALA-438)	4.85		
			-	-	Pi-Alkyl (ILE-133)	5.07		
			-	-	Pi-Alkyl (ILE-132)	5.12		
			-	-	Pi-Alkyl (LEU-152)	5.33		
			-	-	Pi-Alkyl (Cys-437)	5.48		
	L19	-10.3	ARG-435	1.88	Pi-Alkyl (ALA-443)	4.88		
			ARG-115	2.28	Pi-Alkyl (CYS-437)	5.04		
			TRP-141	2.06	Pi-Sulfur (MET-311)	5.33		
			ARG-435	2.27	Pi-Sulfur (Cys-437)	5.70		
			ILE-132	2.10	Pi-Alkyl (ILE-133)	4.91		
			ILE-133	2.40	-	-		
			Cys-437	3.54	-	-		
			L7	-10.3	TRP-141	1.84	Pi-Sigma (ALA-306)	3.65
					ARG-145	2.64	Pi-Sigma (ALA-306)	3.86
ARG-435	2.34	Pi-Alkyl (ALA-307)			4.60			
ILE-133	2.51	Pi-Alkyl (LEU-152)			4.99			
ALA-306	2.52	Pi-Sulfur (ILE-132)			3.95			
-	-	Pi-Alkyl (ALA-438)			4.37			
-	-	Pi-Alkyl (ILE-133)			5.00			
-	-	Pi-Alkyl (LEU-152)			5.33			
-	-	Pi-Sulfur (Cys-437)			5.82			
L39	-9.7	ARG-435	1.83	Pi-Sigma (ALA-306)	3.59			
		ILE-133	2.30	Pi-Sigma (ALA-306)	3.90			
		ILE-132	2.86	Alkyl (ALAL-307)	4.89			
		-	-	Pi-Alkyl (LEU-152)	4.93			
		-	-	Pi-Alkyl (LEU-152)	5.49			
		-	-	Pi-Alkyl (ALA-438)	5.00			
		-	-	Alkyl (ILE-132)	5.77			
		-	-	Pi-Alkyl (ILE-442)	4.48			
		-	-	Pi-Alkyl (Cys-437)	5.40			
Reference Ewemestane	-10.2	ARG-115	2.96	Pi-Alkyl (VAL-370)	4.28			
		MET-374	2.85	Alkyl (VAL-370)	4.65			
		ALA-306	2.87	Alkyl (VAL-373)	4.47			
		-	-	Pi-Alkyl (VAL-370)	4.92			
		-	-	Pi-Alkyl (ILE-133)	5.44			
		-	-	Pi-Alkyl (TRP-224)	5.28			
-	-	Pi-Alkyl (PHE-134)	5.46					

The better performance of these designed compounds can be linked to the number of hydrogen bonds formed between important amino acid parts and the changes at the R3 position. In particular, the hydrogen bond donor -OH group at this position has uniform and favourable hydrogen bond interactions. This result indicates that the R3 position is an important position for further modification in the synthesis of new compounds in the future, especially to increase the hydrogen bond potential. Likewise, these findings are also indicative of the good predictive power and the efficiency of the CoMSIA/SHD model in this work.

3.7. ADMET profiling

ADMET predictions of analogues were performed using the pkCSM algorithm. The efficiency of drug absorption is influenced by several biological factors, including the ease with which it crosses cellular barriers such as Caco-2 models, its uptake in the gastrointestinal tract, and its ability to penetrate the skin. In general, compounds with a percentage of intestinal absorption of less than 30 are regarded as low bioavailability agents. However, the new versions L7, L15, L19, L22, and L39 showed a high absorption rate that was above the 30% benchmark, as shown earlier in **Table 6**. Maximizing a drug's efficacy often entails improving its permeation through the skin and is one of the primary aims in the design of transdermal delivery systems. Usually compounds with $\log K_p > -2.5$ cm/h demonstrate negligible skin permeation. The K_p -values of the analogues developed first-order derivative in skin are found to be between -2.735 and -2.972 cm/h (< -2.5) (as seen in **Table 6**). This figure indicates that all analogues should have good skin permeability. High Caco-2 permeability profile: In the predictive model of pkCSM, a compound is defined as having high Caco-2 permeability when its predicted $\log P_{app}$ is higher than 0.90 cm/s (**Table 6**). The $\log P_{app}$ for the designed analogues is within the range of 0.940 – 0.987 cm/s, all of which surpass the limit of 0.9 cm/s. Accordingly, these analogues are expected to have high permeability across the Caco-2 cell model. Water solubility is one of the important parameters for the formation of orally disintegrating tablets, as it allows active ingredients to be delivered effectively in low-volume medications. Solubility values are represented as \log (mol/L) and are considered in the range from insoluble (≤ -10) to very soluble (≥ 0). Reference is made to the **Table 6** data; all the newly tested analogues are soluble (favourable solubility profile).

The volume of distribution (V_d) is a pharmacokinetic metric that indicates whether a compound tends to remain in the bloodstream or distribute into bodily tissues. As reported by Pires and colleagues, a steady-state volume of distribution (VD_{ss}) is classified as low when it is less than 0.71 L/kg ($\log VD_{ss} < -0.15$) and considered high when it exceeds 2.81 L/kg ($\log VD_{ss} > 0.45$). As shown in **Table 6**, the VD_{ss} values for the new analogues ranged from 0.128 to 2.06 . Among these, only compound L19 exhibited a VD_{ss} value below 0.45 , indicating a lower distribution in tissues. For drugs that need to work in the central nervous system, it's important that they can effectively move from the blood to the brain and spread well in brain tissue. Compounds with a $\log BB < -1$ are generally considered to have limited ability to cross the blood–brain barrier. According to **Table 6**, the new analogues display $\log BB$ values ranging from -0.635 to -1.089 . Based on these values, it can be concluded that L22 and L15, which have $\log BB$ values below -1.0 , probably won't be able to enter the central nervous system well, while the other compounds might be able to enter it to some degree.

The PKCSM pharmacokinetic model estimates total clearance (CL_{tot}) as log-transformed values expressed in mL/min/kg. A higher $\log CL_{tot}$ value indicates more rapid elimination of the compound from the body. As shown in **Table 6**, the benzimidazole analogues exhibit $\log CL_{tot}$ values ranging from 0.854 to 1.085 mL/min/kg, suggesting varying rates of excretion among the analogues based on their predicted clearance profiles.

The digestive system was studied mainly to see how it affects important cytochrome P450 enzymes, especially CYP2C9, CYP1A2, CYP2D6, CYP2C19, and CYP3A4. These enzymes can be inhibited when multiple drugs compete for the same active binding site. Such inhibition can interfere with how the body processes many medications, including some cancer treatments, which may cause higher levels of the drugs in the blood and change how well they work. However, CYP inhibition can still lead to drug toxicity due to altered metabolism. As shown in **Table 6**, compound L19 was predicted to potentially inhibit CYP1A2 and CYP3A4. In contrast, the remaining molecules were not expected to interfere with the major CYP isoforms. Further analysis involving substrates of CYP2D6 and CYP3A4 indicated that none of the tested compounds exhibited inhibitory interactions with these enzymes.

Toxicological assessment is one of the most critical factors in the development of new pharmaceutical compounds. In this study, the potential toxicity of the molecules was evaluated using the pkCSM server, which provides predictive data for various toxicity parameters. These include AMES mutagenicity, human maximum tolerated dose (MTD), inhibition of hERG channels (types I and II), hepatotoxicity, skin sensitization, chronic oral toxicity in rats (LOAEL), *Tetrahymena pyriformis* toxicity, minnow toxicity, and acute oral toxicity in rats (LD50), as summarized in **Table 6**. The Ames test is widely used for assessing the mutagenic potential of chemical compounds using bacterial models. A positive result indicates that the compound may be mutagenic and potentially carcinogenic. In this study, only compound L7 was predicted to be Ames-positive, suggesting possible mutagenicity, while the remaining compounds were considered non-mutagenic. Additionally, none of the compounds showed signs of hepatotoxicity or skin sensitization, except for compound L7, which was predicted to have a positive skin sensitization response.

The human ether-a-go-go-related gene (hERG) encodes potassium channels that play a crucial role in cardiac repolarization. Blocking these channels is a major cause of acquired long QT syndrome, which can lead to serious heart rhythm problems. Several drugs have been withdrawn from the market due to their adverse effects on hERG channel activity. In this study, it was predicted that compounds L15 and L22 could be harmful to the heart because they block hERG II.

The study also predicted the values for *Tetrahymena pyriformis* toxicity, minnow toxicity (MT), oral rat acute toxicity (LD₅₀), and human maximum tolerated dose (MTD) for the candidate compounds and compared them with established reference values. Notably, *T. pyriformis* toxicity is shown as the negative logarithm of the amount needed to stop 50% of growth (pIGC₅₀), measured in log µg/L. A higher pIGC₅₀ value indicates greater toxicity. The predicted LD₅₀ values showed that all compounds have an estimated oral LD₅₀ between 500 and 5000 mg/kg in rats, which means they have low immediate toxicity and are considered slightly toxic.

In the drug-likeness evaluation, all compounds complied fully with Lipinski's Rule of Five, showing no violations (Table 7). Additionally, we applied the pan-assay interference compounds (PAINS) filter to identify potential structural alerts that are commonly associated with false positives, chemical reactivity, or instability. This approach is widely used in pharmaceutical chemistry to identify substructures that may contribute to undesirable pharmacological properties or toxicity. None of the compounds L7, L15, L19, L22, and L39 raised any alarms according to the PAINS filter, indicating that they do not exhibit properties commonly associated with undesirable reactivity or toxicity and thus hold promise as potential therapeutic candidates. The synthetic accessibility (SA) score, which quantitatively assesses the ease of synthesizing drug-like compounds, revealed that all the compounds have favourable SA scores, suggesting they can be synthesized with relative ease (Table 7).

In conclusion, the comprehensive ADMET studies, drug-likeness evaluation, and medicinal chemistry analysis indicated that only two compounds, L22 and L15, demonstrated favourable properties.

Table 6. ADMET profile of the designed compounds.

Adsorption										
Entry	Water solubility (log mol/l)	caco-2 permeability	intestinal absorption		skin permeability					
L22	-2.926	0.94	81.637		-2.735					
L15	-2.92	0.949	82.401		-2.735					
L19	-2.883	0.953	78.225		-2.735					
L7	-1.577	-1.577	66.888		-2.972					
L39	-1.698	0.987	86.542		-2.919					
Distribution				Excretion						
Entry	Vdss	BBB permeability	CNS permeability	Total clearance	Renal OCT2 substrate					
L22	0.667	-0.927	-2.532	1.021	Yes					
L15	0.614	-1.089	-2.536	0.957	Yes					
L19	0.128	-0.635	-2.016	1.085	Yes					
L7	1.757	-0.878	-2.722	1.008	Yes					
L39	2.065	-0.657	-1.91	0.854	Yes					
Metabolism										
Entry	CYP substrates inhibition			CYP enzymes inhibition						
	CYP2D6	CYP3A4	CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4			
L22	No	No	No	No	No	No	No			
L15	No	No	No	No	No	No	No			
L19	No	No	Yes	No	No	No	Yes			
L7	No	No	No	No	No	No	No			
L39	No	No	No	No	No	No	No			
Toxicity										
Entry	Ames	MTDD log mg/kg/day	hERG		LD ₅₀ 10 ⁻³ mg/kg	LOAEL log mg/kg bw/day	HT	SS	T.Pyriformistoxicity log µg/L	MTLC ₅₀ logmM
			I	II						
L22	No	0.370	No	No	2286	2.366	No	No	0.285	1.127
L15	No	0.361	No	No	2325	2.241	No	No	0.285	1.33
L19	No	0.29	No	No	2466	0.78	No	No	0.285	0.702
L7	Yes	0.682	No	Yes	2231	2.265	Yes	No	0.354	2.142
L39	No	0.691	No	Yes	2644	2.411	No	No	0.704	1.278

Table 7. Drug likeness and medicinal chemistry properties of the designed compounds.

Entry	Lipinski's parameters						Viol.	Water Solubility		Medicinal Chemistry	
	MW	Log P	RB	HBA	HBD	PSA (Å ²)		Log S	Class	PAINS	Synthetic accessibility
L22	282.29	2.04	3	4	4	89.37	0	-3.53	Soluble	0 alert	2.34
L15	282.29	2.66	3	4	4	89.37	0	-3.18	Soluble	0 alert	2.48
L19	308.37	2.71	4	3	3	93.33	0	-4.16	Moderately	0 alert	2.59
L7	312.32	1.50	4	5	5	109.6	1	-2.57	Soluble	0 alert	2.53
L39	266.29	1.96	2	3	3	69.14	0	-4.09	Soluble	0 alert	2.39
Threshold	< 500	< 5	≤ 10	≤ 10	≤ 5	< 140	Viol ≤ 1	> - 4	Soluble	0 alert	≤ 4

3.8. Assessing the stability of protein–ligand complexes throughout MD simulations

The selected compounds, L15 and L22, which exhibit favorable pharmacokinetic properties and high predicted activity, were subjected to molecular dynamics (MD) simulations to assess their stability in complex with the aromatase target.

3.8.1. Temperature and total energy

Throughout the 100 ns MD simulation, both complexes, L15-3s7s and L22-3s7s maintained a constant system temperature of 310.15 K and 310.16 K, respectively (Fig. 14). Similarly, the total energy of both systems remained consistent at -8.3×10^5 throughout the simulation. This demonstrates the stability of the systems throughout the simulation.

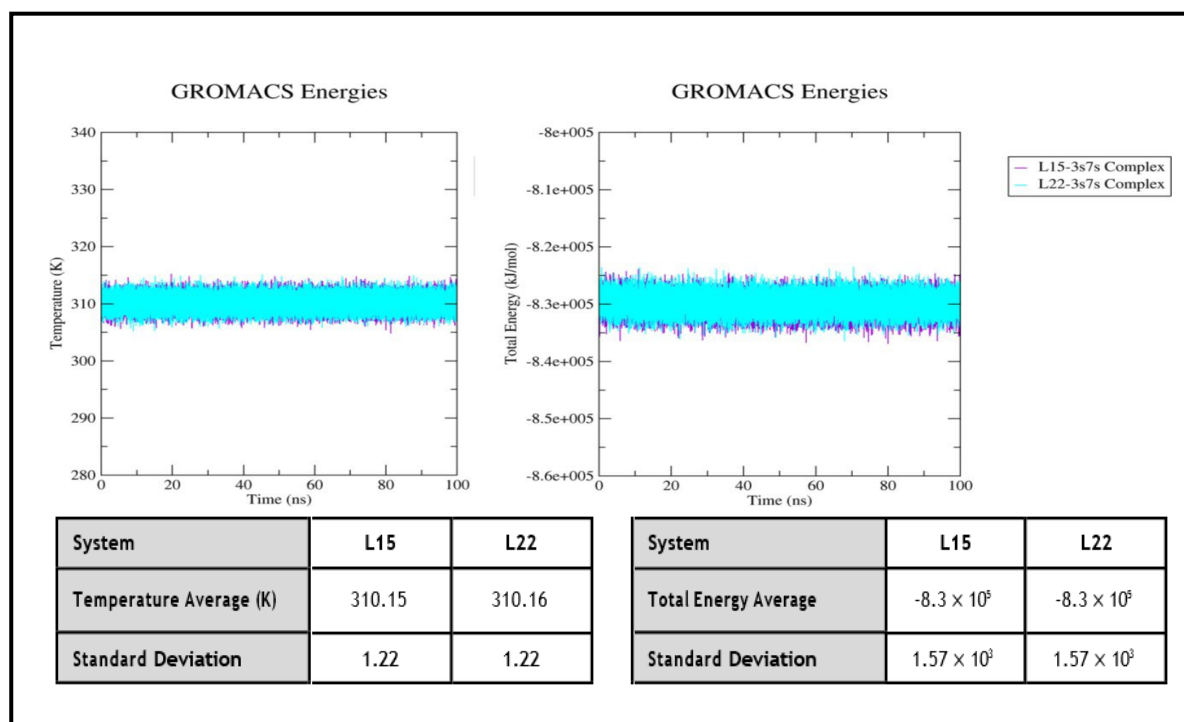


Fig. 14. Temperature and Total energy of both complex L15-3s7s and L22-3s7s.

3.8.2. Dynamic Structural Metrics: RMSD, RMSF, Rg and SASA.

RMSD, RMSF, Rg, and SASA are very important factors for examining the protein stability in its different structural conditions under the MD simulations. RMSD provides valuable information about the global stability of protein-ligand complexes through time. For the levamisole complexes, L15-3s7s and L22-3s7s, they both maintained their stability during the 100 ns MD simulations (the RMSD graph became stationary with only slight spikes) throughout. The 3s7s protein was stabilized 30 ns after binding to L15. In L22 the complex (and the protein) were less stable than in L15 but still stabilized from 70 ns to the end of the 100 ns MD simulation. The RMSD plots of both the complexes showed no significant spikes and were quite stable (Fig. 15).

RMSF Conversely, RMSF is supplemented to the measure of flexibility of individual residues and reflects regions of dynamic behaviour in the protein⁷³. The aromatase protein 3s7s demonstrates flexibility in specific residues with RMSF value more than 0.2 nm when bound to different ligands. Analysis revealed significant flexibility at residues 45, 54, 154-156, 237, 270-271, 273-275, 465, and 475-477 when the protein was bound to L15. In contrast, flexibility was observed at residues 45, 52-56, 233, 236-240, 270-271, 464, and 496 when the protein was bound to L22. This suggests that ligand binding induces conformational changes in the aromatase protein 3s7s, with distinct regions exhibiting flexibility depending on the specific ligand present. When the aromatase protein 3s7s is combined with different ligands, we see some flexible parts with RMSF greater than 0.2 nm. When complexed to L15, the analysis revealed high flexibility at positions 45, 54, 154-156, 237, 270-271, 273-275, 465, and 475-477. On the other hand, flexibility appeared at positions 45, 52-56, 233, 236-240, 270-271, 464, and 496 in the presence of the protein subunit bound to L22. This implies that binding of the ligand perturbs, to different extents, the conformation of the aromatase protein 3s7s at various regions.

The radius of gyration is particularly useful as a measure of the overall folding, distance, and compactness of proteins. A reduction in Rg throughout an MD simulation algorithmically corresponds to proteins becoming more compact, which is frequently used as a marker of enhancement in stability⁷⁴. When complexed with L15 and L22, the 3s7s protein became more compact than in its free state, with average Rg values calculated at 2.29 nm in both complexes.

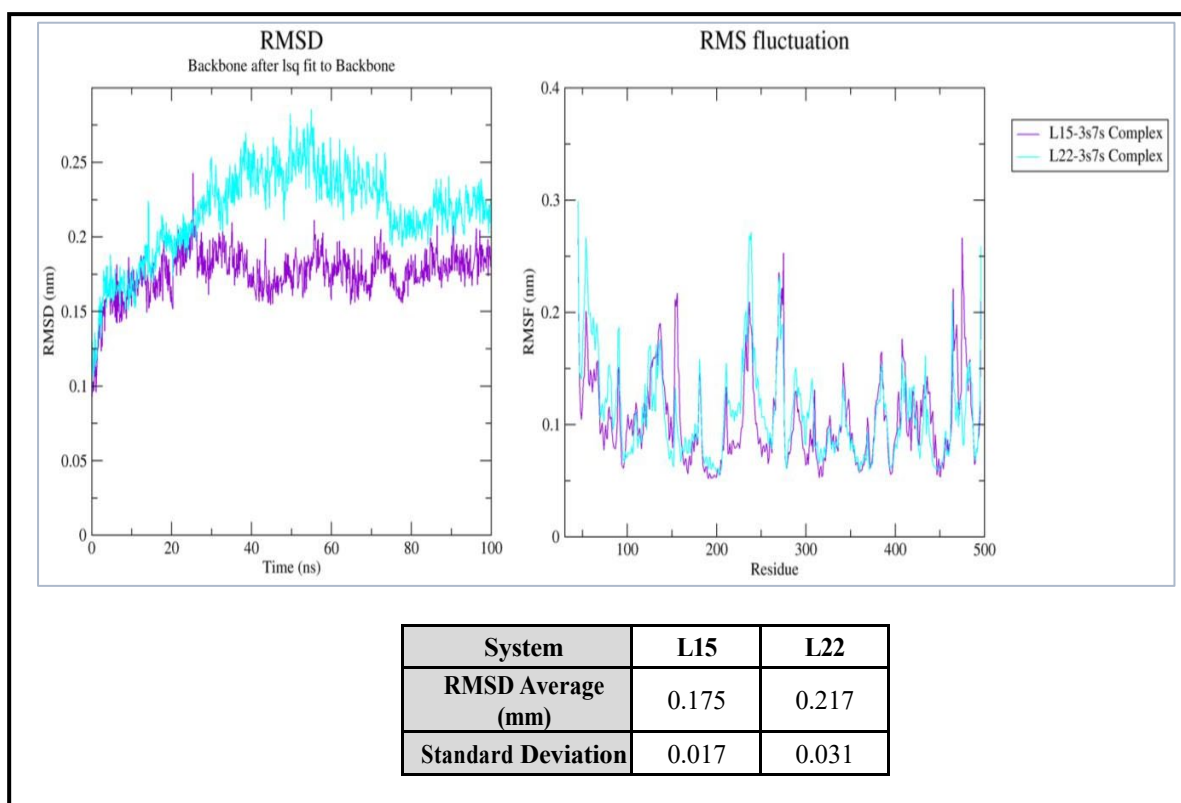


Fig. 15. RMSD and RMSF of both complex L15-3s7s and L22-3s7s.

In contrast, SASA can offer information about how the protein interacts with the environment and may reflect on protein folding states. A drop in the SASA value means a protein structure becomes tighter and more folded in the course of a simulation. Like the Rg value, the SASA of the protein also follows the same trend during the binding of L15 and L22, and it is observed to be decreased in value, which is an indication of the compact condensation. The 3s7s protein seems more compact in the L22-bound state, with the average SASA value equal to 214.74 nm² (**Fig. 16**). In contrast, the aromatase protein 3S7S shows an average SASA of 217.42 nm² when complexed with L15.

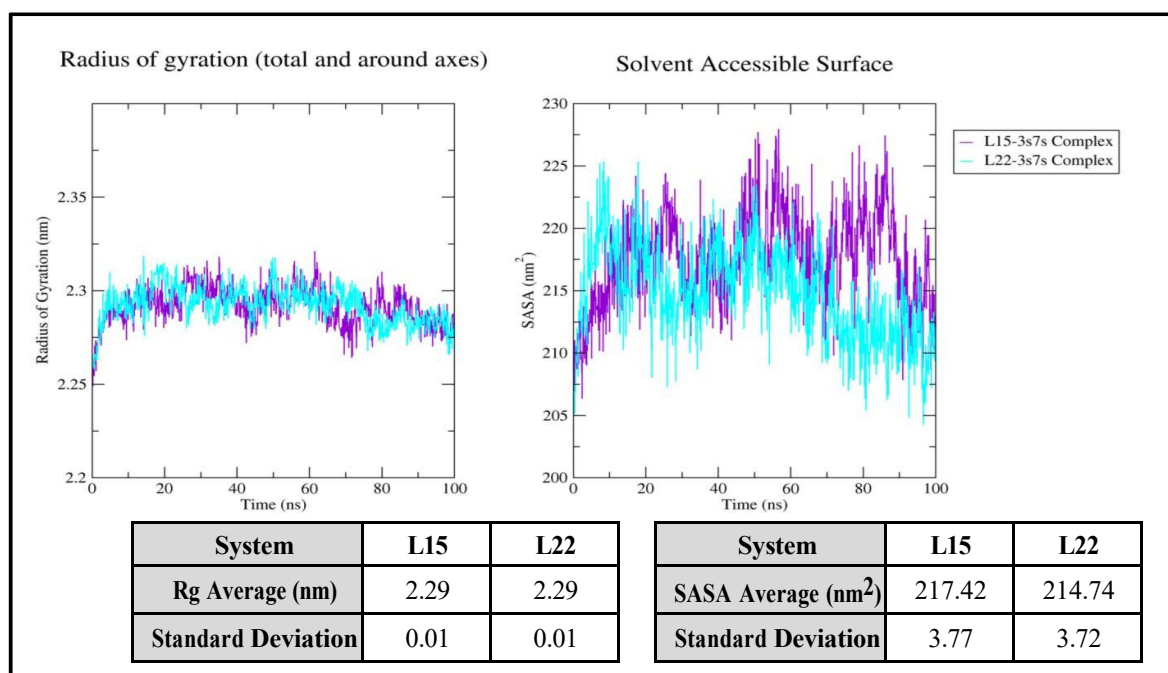


Fig. 16. Radius of Gyration (Rg) and Solvent Accessible Surface Area (SASA) of both complex L15-3s7s and L22-3s7s.

3.9. Protein-ligand distance, hydrogen bond and binding free energy (MMPBSA)

Protein-ligand distance, hydrogen bonds, and Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) are used for analysis of ligands binding to the target protein. The distance between a ligand and a protein is based on one of the most important parameters to illustrate the ligand is close to the active site of the protein. Ligands closer than typical to the residues are also a favourable indicator of tight binding. It is also important to note that a distance of 3.5 Å between hydrogen bond donors and acceptors shows a strong interaction that is needed for the stability of the protein-ligand complex. At the end of the protein-L15/L22 simulation, the average protein-ligand distances for L15 and L22 are both ≈ 0.188 and ≈ 0.183 nm, respectively (Fig. 17).

Hydrogen bonding is important for sustaining any protein-ligand complex. The hydrogen bond occupancy of MD simulations may imply the strength of the binding; a larger occupancy correlates with stronger binding. Hydrogen bond analysis suggests that L22 has higher hydrogen bond occupancy than L15 when bound to the 3s7s protein. In particular, for L15, a high hydrogen-bond occupancy was conserved during the entire MD simulation from 60 ns to 100 ns.

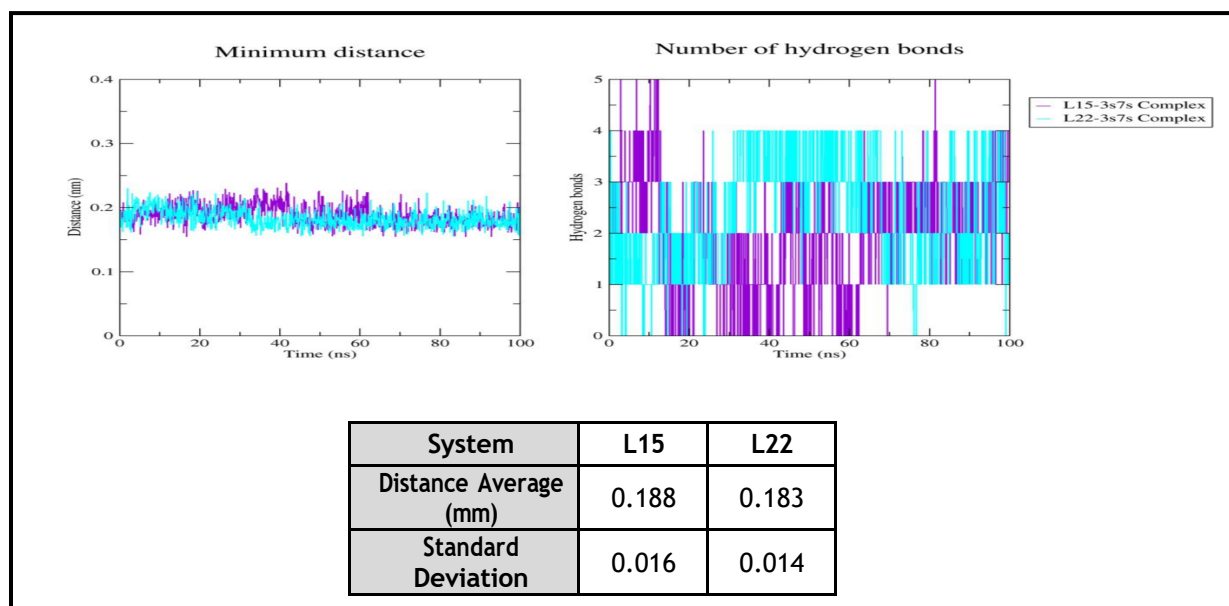


Fig. 17. Protein-ligand distance, hydrogen bonding of both complex L15-3s7s and L22-3s7s.

The `gmx_MMPBSA` tool is a computational protocol for calculating the binding free energy of protein-ligand complexes in GROMACS molecular dynamics software. This server uses the MM/PBSA method, which combines molecular mechanics with models that consider how molecules interact with their environment, to calculate how strongly proteins and ligands bind together. The MM/PBSA methodology accounts for energy contributed by both the protein and ligand to the computation of the free energy of binding, which provides a kinetic connection between molecular dynamics simulations and thermodynamic considerations 63. Both compound L15 and L22 had small MMPBSA values of -24.46 kcal/mol when they bound to the 3s7s protein, and both had small differences when they bound to 3c57 during the 100 ns MD simulation (Fig. 18). This finding reveals that both ligands are closely associated with the protein at a similar binding strength. Table 8 summarizes the contributions to the binding free energy of the L15-3s7s and L22-3s7s complexes.

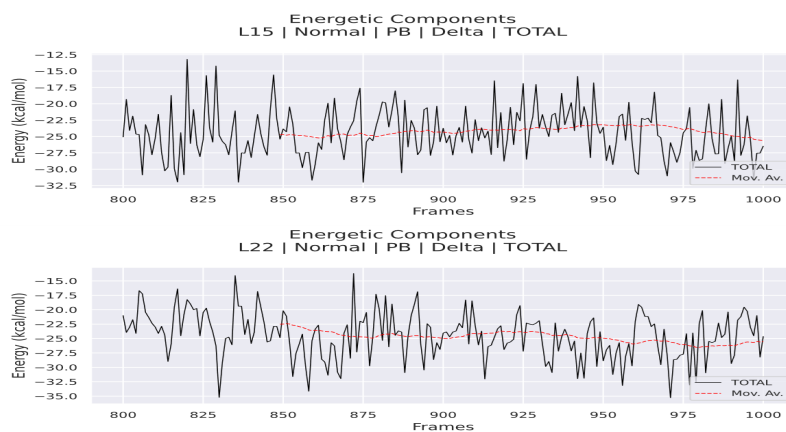


Fig. 18. The binding free energy of both complex L15-3s7s and L22-3s7s.

Table 8. Estimated binding affinities for the studied complexes, reported in kcal/mol.

Complex	The binding affinities (kcal/mol)							
	VDWAALS	EEL	EPB	ENPOLAR	GGAS	GSOLV	TOTAL	
L15-3s7s Complex	Average	-34.27	-46.08	59.46	-3.58	-80.34	55.88	-24.46
	SD	2.78	5.68	4.27	0.09	5.51	4.26	3.75
	SEM	0.2	0.4	0.3	0.01	0.39	0.3	0.26
L22-3s7s Complex	Average	-41.57	-40.17	60.93	-3.65	-81.73	57.28	-24.46
	SD	2.59	7.89	5.43	0.09	7.34	5.46	4
	SEM	0.18	0.56	0.38	0.01	0.52	0.38	0.28

3.10. Retrosynthesis analysis

We used the ASKCOS platform to forecast the synthetic pathways for A498 kidney cancer cells as possible inhibitors of the synthesis of the two molecule candidates L15 and L22 to support. In this program, both reactants and products are represented using Morgan circular fingerprints and encoded through their SMILES notation, as shown in **Figure 19**. A retrosynthetic analysis led to suitable routes starting with commercial material. The obtained multi-step syntheses have moderate reaction conditions and high degrees of synthetic accessibility. This approach serves to improve the experimental planning while saving time and resources for laboratory synthesis.

ASKCOS has created different retrosynthesis frameworks that can work together to help plan experiments interactively, allowing us to include single-step results into full synthetic pathways, and it also has a feature that automatically builds multi-step retrosynthetic strategies. In this study, we employed both functions to design efficient synthetic routes for the compounds L15 and L22.

3.10.1. Interactive path planning

In the current study, the one-step retrosynthesis module gave rise to five excellent synthetic routes to access L22 and L15. We listed these pathways in priority order 1-5 (**Figs. 20 and 22**). ASKCOS provided the best experimental conditions for each suggested reaction, including the necessary chemicals, solvents, temperature, and catalysts when applicable. The results indicate that the reaction conditions in L22 and L15 are mild and suitable for standard lab work, which makes these reaction systems easy to use in regular synthesis without any noticeable issues (**Figs. 21 and 23**). The information emphasizes the utility of the retrosynthetic routes and potential for scale-up.

3.10.2. Tree-Builder module

The Tree-Builder module automatically generated multi-step retrosynthetic routes for the target compounds L22 and L15. This software builds the reaction tree by successively performing one-step retrosynthetic transformations, which thereby enables exploration from the final product to readily available starting materials via prospective synthetic routes.

In our application, the Tree-Builder proposed a number of complete synthetic routes with different lengths and complexity. The pathways were evaluated based on factors like how many steps they had, how easy they were to make, whether the reactions could happen in a regular lab, and how easy it was to get the starting materials. Favourable routes often encompassed 3–5 steps and employed classic organic reactions.

Presentation of retrosynthetic trees (see **Figs. 24 and 25**) shows important intermediates and junctions where other disconnections were probable. This overarching perspective serves as a guide for the most effective practical synthetic approaches, minimizing ambiguity in the planning of experiments.

In the end, the Tree-Builder module presented itself as a valuable tool for retrosynthetic explorations, providing an orderly and automated means to evaluate various possible roads and thus aiding in the selection of the best synthetic strategies for both inhibitors L22 and L15.

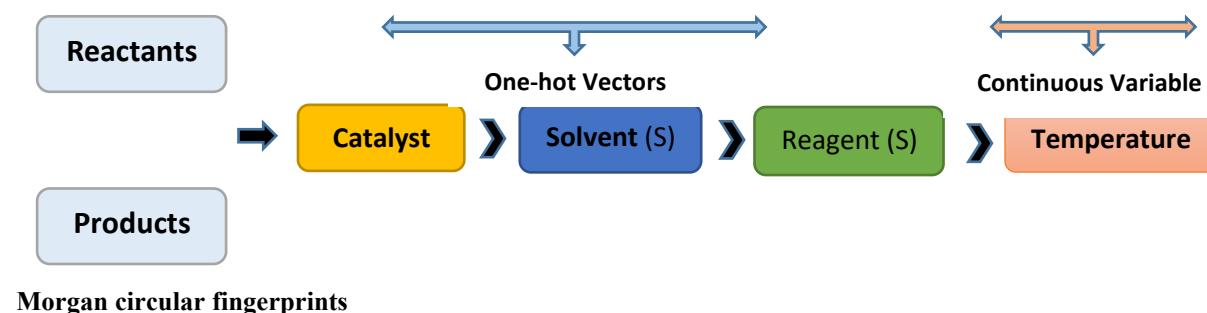


Fig. 19. Condition prediction flowchart using a neural model in ASKCOS.

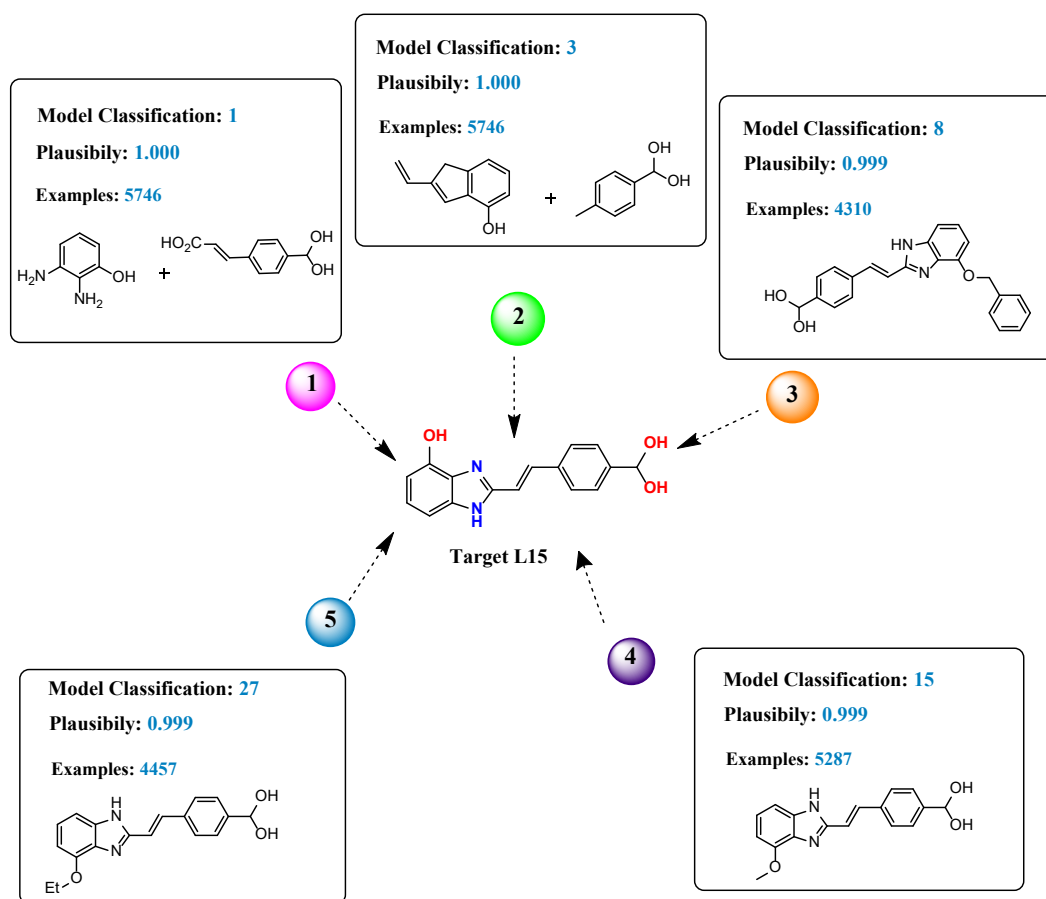


Fig. 20. One-step retrosynthetic prediction results for target L15 using interactive path-planning module.

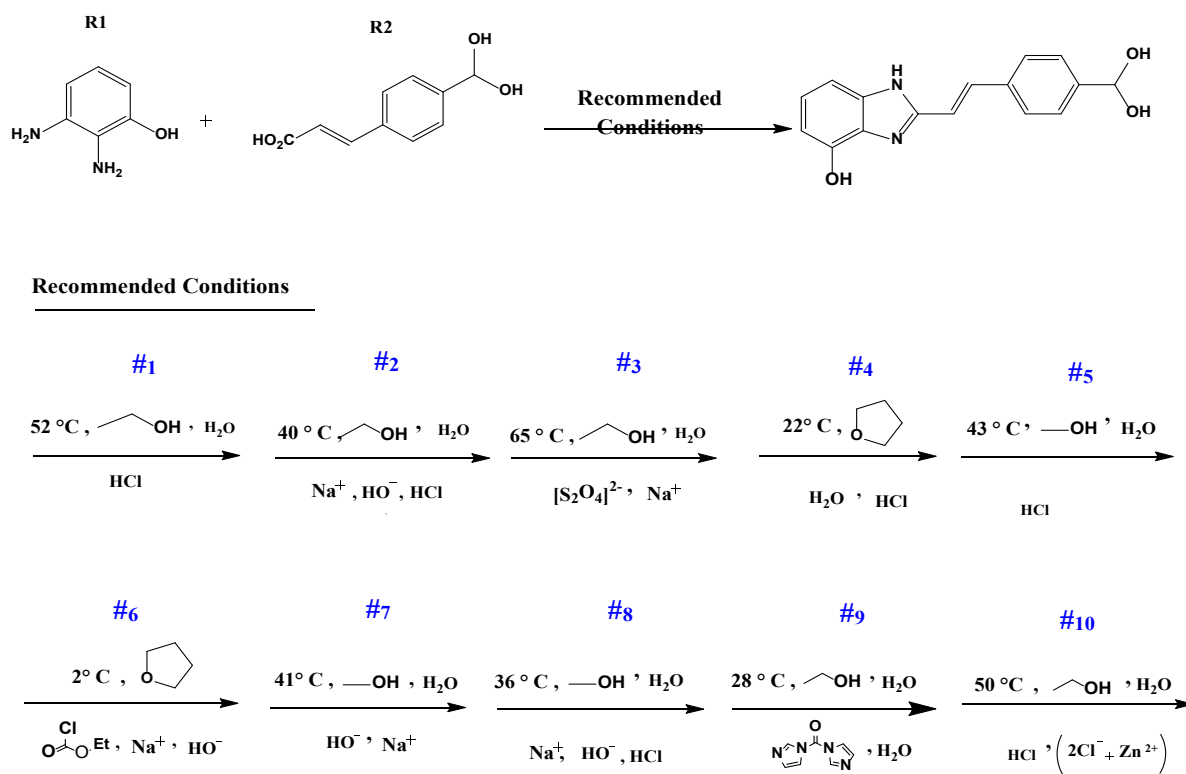


Fig. 21. The ten optimal reaction conditions recommended by the model for the target L15.

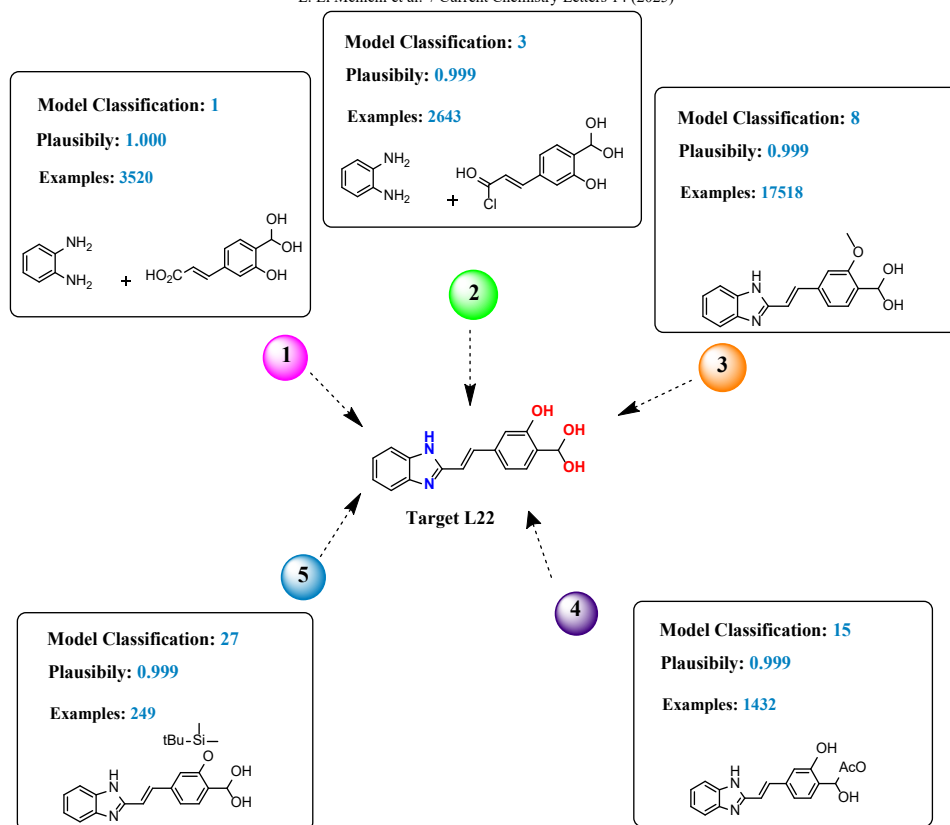


Fig. 22. One-step retrosynthetic prediction results for target L22 using interactive path-planning module.

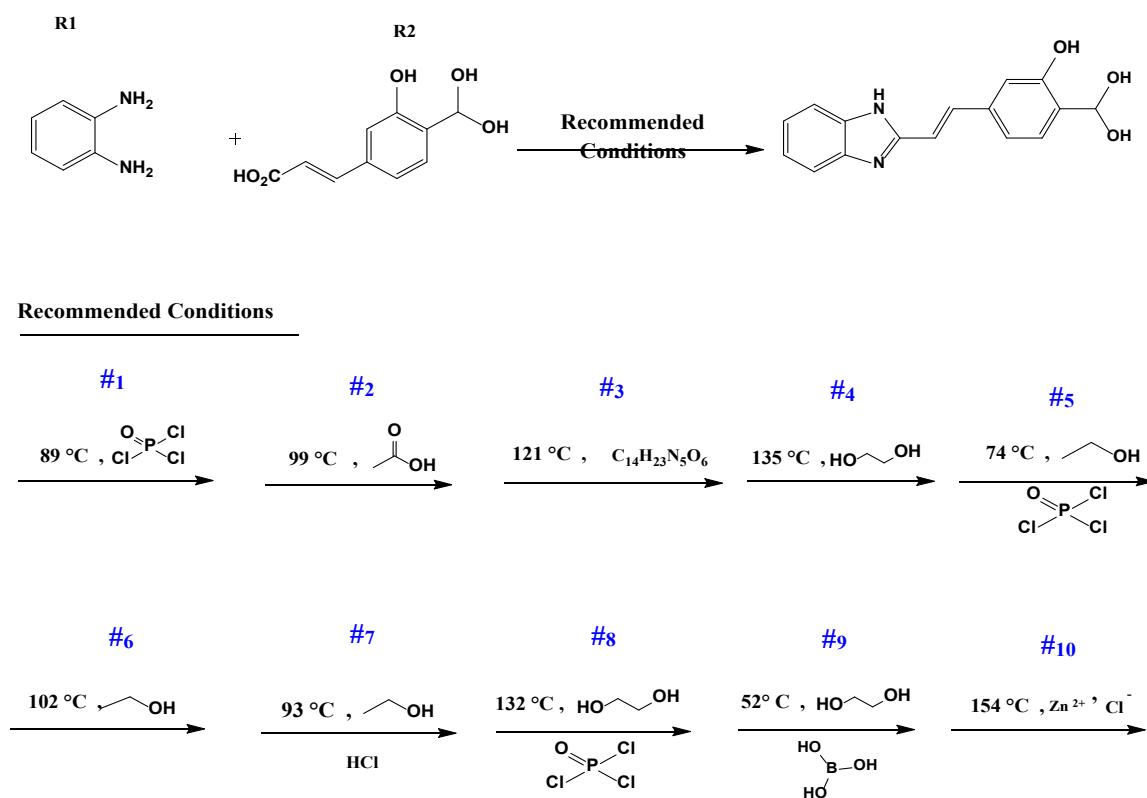


Fig. 23. The ten optimal reaction conditions recommended by the model for the target L22.

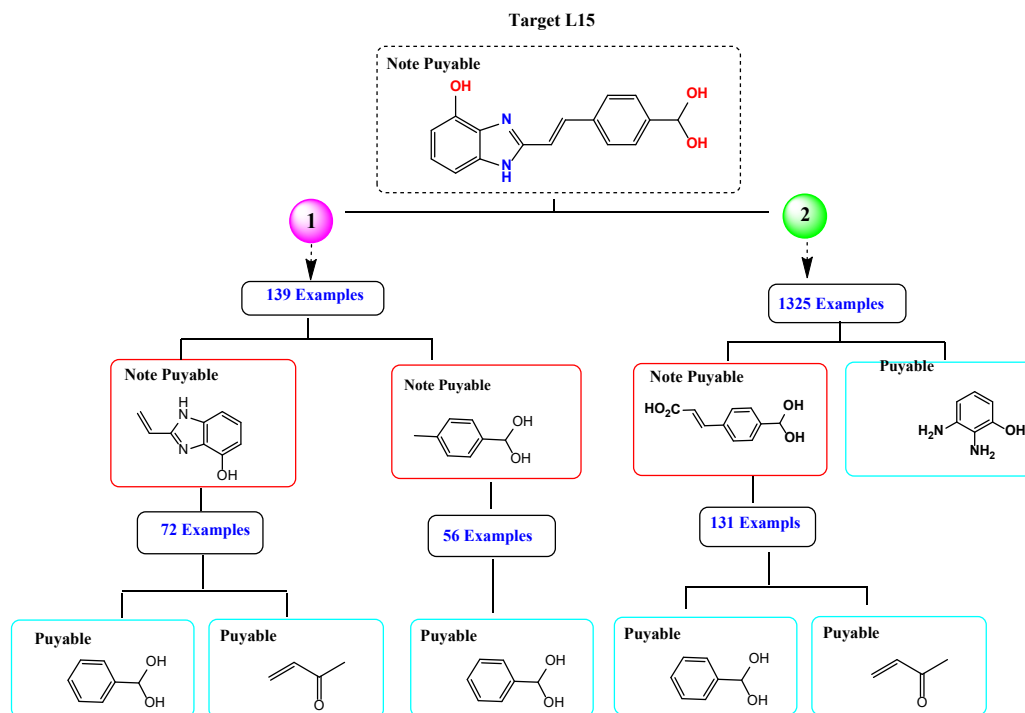
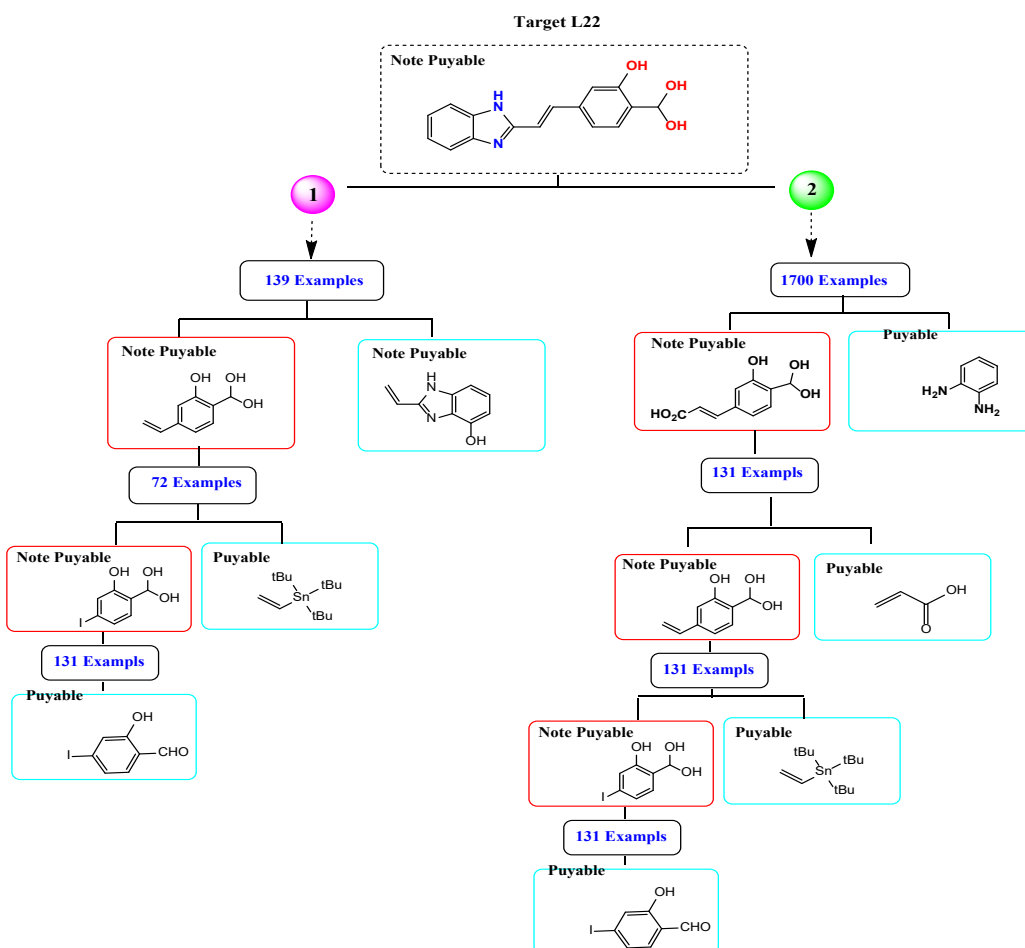


Fig. 24. Proposed synthetic route for the target L15 using the tree builder module implemented in ASKCOS.



4. Conclusion

To date, benzimidazole and its congeners are still among the most significant heterocyclic structures due to their wide range of pharmacological fields of application. Nevertheless, finding and developing new courses is challenging, costly, and time-consuming. Being performed in this way, computational methods have been powerful for different tasks in drug discovery, such as analyzing protein interaction networks, predicting targets, identifying binding sites, virtual screening, and so on. In the current research, we were dedicated to creating new benzimidazole compounds for the potential treatment of RCC. We designed five new derivatives with higher molecular biological activity based on the 3D-QSAR CoMSIA/SHD model, specifically the characteristic contour analysis. Molecular docking investigation was used to study the binding interaction between the developed compounds and aromatase. Compounds L22 and L15 exhibited reasonable ADMET profiles, scoring well both computationally in terms of binding affinities and in vitro for predicted activities. To verify the stability of their binding in the aromatase's active site, 100 ns MD simulations were carried out on the novel ligands L15 and L22. The simulation results supported that these two compounds were still binding to the ligand-binding pocket of aromatase during the simulation. Moreover, retrosynthetic analysis was carried out to suggest the feasible synthetic routes for L15 and L22 to serve tomorrow in experimental work. In conclusion, our data demonstrated the potential of benzimidazole derivatives as good candidates for RCC therapy and potential aromatase inhibitors. These compounds could be useful starting points in the search for effective oral drugs.

Acknowledgements

Acknowledgements We are grateful to the "Association Marocaine des Chimistes Théoriciens" (AMCT) for its pertinent help concerning the programs.

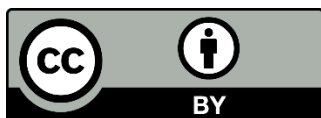
References

- Siegel R. L., Giaquinto A. N., and Jemal A. (2024) Cancer statistics, 2024. *CA: CA Cancer J Clin.*, 74 (1) 12–49.
- Fares J., Fares M. Y., Khachfe H. H., Salhab H. A., and Fares Y. (2020) Molecular principles of metastasis: A hallmark of cancer revisited. *Sig Transduct Target Ther.*, 5 (1) 28.
- Bahadoram S., Davoodi M., Hassanzadeh S., Bahadoram M., Barahman M., and Mafakher L. (2022) Renal cell carcinoma: An overview of the epidemiology, diagnosis, and treatment. *G Ital Nefrol.*, 39 (3) 2022.
- Motzer R. J., Jonasch E., Agarwal N., Alva A., Bagshaw H., Baine M., Beckermann K., Carlo M. I., Choueiri T. K., and Costello B. A. (2024) Nccn guidelines® insights: Kidney cancer, version 2.2024: Featured updates to the nccn guidelines. *J. Natl. Compr. Cancer Netw.*, 22 (1) 4–16.
- Yeh S., Xu H., Hu Y., Yang G., Joseph J., and Messing E. M. (2024) Pd51-10 estrogen receptor signals promote autophagy and m2 macrophage polarization in the tki-resistant rcc. *J. Urol.*, 211 (5S) e1069.
- Ding J., Cui X.-G., Chen H.-J., Sun Y., Yu W.-W., Luo J., Xiao G.-Q., Chang C., Qi J., and Yeh S. (2022) Targeting circdgdgd intercepts tki's effects on up-regulation of estrogen receptor β and vasculogenic mimicry in renal cell carcinoma. *Cancers*, 14 (7) 1639.
- Correia De Sousa M., Delangre E., Türkal M., Foti M., and Gjorgjieva M. (2023) Endoplasmic reticulum stress in renal cell carcinoma. *Int. J. Mol. Sci.*, 24 (5) 4914.
- Pontes O., Oliveira-Pinto S., Baltazar F., and Costa M. (2022) Renal cell carcinoma therapy: Current and new drug candidates. *Drug Discov. Today*, 27 (1) 304–314.
- Bouzian Y., El Hafii M., Parvizi N., Kim W., Subaşıoğlu M., Ozcan M., Turkez H., and Mardinoglu A. (2024) Design and evaluation of novel inhibitors for the treatment of clear cell renal cell carcinoma. *Bioorg. Chem.*, 151 (2024) 107597.
- Song M. (2017) Recent developments in small molecule therapies for renal cell carcinoma. *Eur. J. Med. Chem.*, 142 (2017) 383–392.
- Ashaq A., Maqbool M. F., Maryam A., Khan M., Shakir H. A., Irfan M., Qazi J. I., Li Y., and Ma T. (2021) Hispidulin: A novel natural compound with therapeutic potential against human cancers. *Phytother. Res.*, 35 (2) 771–789.
- Kousar R., Naeem M., Jamaludin M. I., Arshad A., Shamsuri A. N., Ansari N., Akhtar S., Hazafa A., Uddin J., and Khan A. (2022) Exploring the anticancer activities of novel bioactive compounds derived from endophytic fungi: Mechanisms of action, current challenges and future perspectives. *Am. J. Cancer Res.*, 12 (7) 2897.
- El-Deek H. E., Ahmed A. M., Hassan T. S., and Morsy A. M. (2018) Expression and localization of estrogen receptors in human renal cell carcinoma and their clinical significance. *Int. J. Clin. Exp. Pathol.*, 11 (6) 3176.
- El Mchichi L., Tabti K., Kasmi R., El-Mernissi R., El Aissouq A., En-Nahli F., Belhassan A., Lakhlifi T., and Bouachrine M. (2022) 3d-qsar study, docking molecular and simulation dynamic on series of benzimidazole derivatives as anti-cancer agents. *J. Indian Chem. Soc.*, 99 (9) 100582.
- Gaba M., and Mohan C. (2016) Development of drugs based on imidazole and benzimidazole bioactive heterocycles: Recent advances and future directions. *Med. Chem. Res.*, 25, 173–210.
- Ebenezer O., Oyetunde-Joshua F., Omotoso O. D., and Shapi M. (2023) Benzimidazole and its derivatives: Recent advances (2020–2022). *Results Chem.*, 5 (2023) 100925.
- Sadowski M., Dresler E., Zawadzińska K., Wróblewska A., and Jasiński R. (2024) Syn-propanethial s-oxide as an available natural building block for the preparation of nitro-functionalized, sulfur-containing five-membered heterocycles: An medt study. *Molecules*, 29 (20) 4892.

18. Sadowski M., and Kula K. (2024) Unexpected course of reaction between (1e, 3e)-1, 4-dinitro-1, 3-butadiene and n-methyl azomethine ylide—a comprehensive experimental and quantum-chemical study. *Molecules*, 29 (21) 5066.
19. Yadav G., and Ganguly S. (2015) Structure activity relationship (sar) study of benzimidazole scaffold for different biological activities: A mini-review. *Eur. J. Med. Chem.*, 97, 419–443.
20. Solmaz B., Oguz A., Oguz M., Ozturk B., and Yilmaz M. (2025) Synthesis, anticancer activity, and mitochondria-targeted bioimaging applications of novel fluorescent calix [4] arenes-benzimidazole derivatives. *Curr Med Chem.*
21. Kabi A. K., Sravani S., Gujjarappa R., Garg A., Vodnala N., Tyagi U., Kaldhi D., Singh V., Gupta S., and Malakar C. C. (2022) An overview on biological activity of benzimidazole derivatives. *Nanostruct. Biomater.: Basic Struct. and App.*, 351–378.
22. El Mchichi L., Aouidate A., Chokrafi F. Z., Ghaleb A., Khalil F., Lakhliifi T., and Bouachrine M. (2018) Prediction of biological activity of pyrazolo [3, 4-b] quinoliny acitamide by qsar results. *RHAZES: Green and Appl. Chem.*, 3 (3) 79–93.
23. Tabti K., Elmchichi L., Sbai A., Maghat H., Bouachrine M., Lakhliifi T., and Ghosh A. (2022) In silico design of novel pin1 inhibitors by combined of 3d-qsar, molecular docking, molecular dynamic simulation and admet studies. *J. Mol. Struct.*, 1253, 132291.
24. Zrinej J., Elmchichi L., Alaqrbeh M., Lakhliifi T., and Bouachrine M. (2023) Computational approach: 3d-qsar, molecular docking, admet, molecular dynamics simulation investigations, and retrosynthesis of some curcumin analogues as parp-1 inhibitors targeting colon cancer. *New J. Chem.*, 47 (45) 20987–21009.
25. Kasmi R., Elmchichi L., El Aissouq A., Bouachrine M., and Ouammou A. (2021) In silico drug design: Development of new pyrimidine-based benzothiazole derivatives, selective for cdk2. *Lett. Drug Des. Discov.*, 18 (10) 961–975.
26. El Mchichi L., Bouachrine M., and Lakhliifi T. (2025) Anticancer activity in heterocyclic organic structures: A pathway to novel drug development part 1. 1 (1st ed.). CRC Press. <https://doi.org/10.1201/9781003653134>.
27. Zrinej J., Ouabane M., El Mchichi L., Qara A., Sekkate C., Lakhliifi T., and Bouachrine M. (2025) Chemical and biological evaluation of the plumbago zeylanica plant against prostate cancer. A unified approach combining molecular docking, dynamic simulations, pharmacophore modeling, dft and admet studies. *Vietn. J. Chem.*
28. Soudani W., Zaki H., Alaqrbeh M., Elmchichi L., Bouachrine M., and Hadjadj-Aoul F. Z. (2023) Discover the medication potential of algerian medicinal plants against sars-cov-2 main protease (mpro): Molecular docking, molecular dynamic simulation, and admet analysis. *Chem. Afr.*, 6 (6) 2879–2895.
29. Zaki R. M., El-Dean A. M. K., Radwan S. M., and Saber A. F. (2018) A convenient synthesis, reactions and biological activity of some new 6h-pyrazolo [4', 3': 4, 5] thieno [3, 2-d][1, 2, 3] triazine compounds as antibacterial, anti-fungal and anti-inflammatory agents. *J. Braz. Chem. Soc.*, 29 (12) 2482–2495.
30. Zaki R. M., Saber A. F., El-Dean A. M. K., and Radwan S. M. (2020) A concise review on synthesis, reactions and biological importance of thienopyrazoles. *Arkivoc*, 2020 (1) 20–60.
31. Saber A. F., Zaki R. M., Kamal El-Dean A. M., and Radwan S. M. (2020) Synthesis, reactions, and spectral characterization of some new biologically active compounds derived from thieno [2, 3-c] pyrazole-5-carboxamide. *J. Heter. Chem.*, 57 (1) 238–247.
32. El-Dean A. M. K., Zaki R. M., Radwan S., and Saber A. F. (2017) Synthesis, reactions and spectral characterization of novel thienopyrazole derivatives. *Eur. Chem. Bull*, 6 (12) 550–553.
33. Fouada M. R., El-Aswada A. F., Alya M. I., and Badawya M. E. (2023) *Curr. chem. lett.*, 13 (2024) 503–514.
34. Abdel-Raheem S. A., Fouad M. R., Gad M. A., El-Dean A. M. K., and Tolba M. S. (2023) Environmentally green synthesis and characterization of some novel bioactive pyrimidines with excellent bioefficacy and safety profile towards soil organisms. *J. Environ. Chem. Eng.*, 11 (5) 110839.
35. Ersan R. H., Kuzu B., Yetkin D., Alagoz M. A., Dogen A., Burmaoglu S., and Algul O. (2022) 2-phenyl substituted benzimidazole derivatives: Design, synthesis, and evaluation of their antiproliferative and antimicrobial activities. *Med. Chem. Res.*, 31 (7) 1192–1208.
36. Elmchichi L., Belhassan A., Aouidate A., Ghaleb A., Lakhliifi T., and Bouachrine M. (2020) Qsar study of new compounds based on 1, 2, 4-triazole as potential anticancer agents. *Ph. Chem. Res.*, 8 (1) 125–137.
37. Nepali K., Sharma S., Sharma M., Bedi P., and Dhar K. (2014) Rational approaches, design strategies, structure activity relationship and mechanistic insights for anticancer hybrids. *Eur. j. med. chem.*, 77, 422–487.
38. Elmchichi L., Belhassan A., Lakhliifi T., and Bouachrine M. (2020) 3d-qsar study of the chalcone derivatives as anticancer agents. *J. Chem.*, 2020 (1) 5268985.
39. Alaqrbeh M., El Mchichi L., Abouzied A. S., Bouzzine S. M., Huwaimel B., and Bouachrine M. (2023) Computational investigation of structural-biological inhibitory activity for au (iii) porphyrin complexes against mcf-7 human breast cancer. *Chem. Data Collec.*, 48, 101094.
40. Chtita S., Aouidate A., Belhassan A., Ousaa A., Taurati A. I., Elidrissi B., Ghamali M., Bouachrine M., and Lakhliifi T. (2020) Qsar study of n-substituted oseltamivir derivatives as potent avian influenza virus h5n1 inhibitors using quantum chemical descriptors and statistical methods. *New J. Chem.*, 44 (5) 1747–1760.
41. Clark M., Cramer Iii R. D., and Van Opdenbosch N. (1989) Validation of the general purpose tripos 5.2 force field. *J. comput. chem.*, 10 (8) 982–1012.
42. Soukaina E., Al-Zaqri N., Warad I., Ichou H., Yassine K., Guenoun F., and Bouachrine M. (2023) Novel antiproliferative inhibitors from salicylamide derivatives with dipeptide moieties using 3d-qsar, molecular docking, molecular dynamic simulation and admet studies. *J. Mol. Struct.*, 1282, 135219.

43. Wold S., (1991). Validation of QSAR's. *Quant Struct Relatsh* 10: 191–193.
44. Cruciani G., Baroni M., Clementi S., Costantino G., Riganelli D., and Skagerberg B. (1992) Predictive ability of regression models. Part i: Standard deviation of prediction errors (sdep). *J. Chemometr.*, 6 (6) 335–346.
45. Roy K., Das R. N., Ambure P., and Aher R. B. (2016) Be aware of error measures. Further studies on validation of predictive qsar models. *Chemometr. Intell. Lab. Syst.*, 152, 18–33.
46. Tropsha A., Gramatica P., and Gombar V. K. (2003) The importance of being earnest: Validation is the absolute essential for successful application and interpretation of qspr models. *QSAR & Combina. Sc.*, 22 (1) 69–77.
47. Reda E.-M., Khatabi K. E., Khaldan A., Bouamrane S., Elmchichi L., Ajana M. A., Lakhlifi T., and Bouachrine M. (2022) 3d-qsar, admet and docking studies for design new 5, 5-diphenylimidazolidine-2, 4-dione derivatives agents against cervical cancer. *Orbital: Electron. J. Chem.*, 24–32.
48. El Mchichi L., Alaqrbeh M., Lakhlifi T., and Bouachrine M. (2024) Discovery of a new isatin scaffold for bcr-abl tyrosine kinase inhibitors using a comprehensive computational approach. *ChemistrySelect*, 9 (43) e202402053.
49. Golbraikh A., and Tropsha A. (2002) Beware of q²! *J. Mol. Graph. Model.*, 20 (4) 269–276.
50. Rossafi B., Outahar F., Hammoudan I., Moumou M., Touti R., Benharref A., and Chtita S. (2025) Mechanism and stereoselectivity of epoxidation reaction of a β -himachalene derivative: Insights from dft, molecular docking, admet, and molecular dynamics investigations. *J. Mol. Struct.*, 1320, 139664.
51. Kar S., and Roy K. (2010) Predictive toxicology using qsar: A perspective. *J. Indian Chem. Soc.*, 87 (12) 1455.
52. Roy K., and Mitra I. (2012) On the use of the metric rm^2 as an effective tool for validation of qsar models in computational drug design and predictive toxicology. *Mini Rev. Med. Chem.*, 12 (6) 491–504.
53. Roy, K., Das, R., Ambure, P., and Aher, R. (2016) Be aware of error measures Further. *Chemometr. Intell. Lab. Syst.*, 152 (2016) 18–33
54. Dearden J. C. (2017) The history and development of quantitative structure-activity relationships (qsars): Addendum. *Int. J. Quantit. Struct. Proper. Relation. (IJQSPR)*, 2 (2) 36–46.
55. El Rhabori, S., Alaqrbeh, M., Allouche, Y.E., Naanaai, L., El Aissouq, A., Bouachrine, M., Chtita, S., and Khalil, F. (2024). Exploring innovative strategies for identifying anti-breast cancer compounds by integrating 2D/3D-QSAR, molecular docking analyses, ADMET predictions, molecular dynamics simulations, and MM-PBSA approaches. *J. Mol. Struct.*, 8 (2024) 100455.
56. Kanzouai, Y., Bouribab, A., El Moussaoui, A., Touhtouh, J., Laghmari, M., Bouzammit, R., Aflak, N., Boujdi, K., Chtita, S., and Benali, T. (2025). Design, synthesis and characterization of new amide-linked Chromone-Isoxazole Hybrids: In Vitro anti-bacterial and antioxidant evaluation, DFT calculations, ADMET profiling, docking and molecular dynamics simulation. *J. Mol. Struct.* 1325, 140972.
57. Zrinej J. (2022) Curcumin and derivatives against human colon cancer (hct-116). *Computational. RHAZES: Green and Appl. Chem.*, 16, 83–102.
58. Chalkha M., Akhazzane M., Moussaid F. Z., Daoui O., Nakkabi A., Bakhouch M., Chtita S., Elkhatabi S., Housseini A. I., and El Yazidi M. (2022) Design, synthesis, characterization, in vitro screening, molecular docking, 3d-qsar, and adme-tox investigations of novel pyrazole derivatives as antimicrobial agents. *New J. Chem.*, 46 (6) 2747–2760.
59. Bouamrane S., Khaldan A., Hajji H., El-Mernissi R., Alaqrbeh M., Alsakhen N., Maghat H., Ajana M. A., Sbai A., and Bouachrine M. (2023) In silico identification of 1, 2, 4-triazoles as potential candida albicans inhibitors using 3d-qsar, molecular docking, molecular dynamics simulations, and admet profiling. *Mol. Divers.*, 27 (5) 2111–2132.
60. Daina A., Michielin O., and Zoete V. (2017) Swissadme: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. repor.*, 7 (1) 42717.
61. Dulsat J., López-Nieto B., Estrada-Tejedor R., and Borrell J. I. (2023) Evaluation of free online admet tools for academic or small biotech environments. *Molecules*, 28 (2) 776.
62. Pires D. E., Blundell T. L., and Ascher D. B. (2015) PkcsM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. med. chem.*, 58 (9) 4066–4072.
63. Lipinski C. A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods*, 44 (1) 235–249.
64. Trott O., and Olson A. J. (2010) Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. comput. chem.*, 31 (2) 455–461.
65. Cao H., Sun Y., Wang L., Zhao C., Fu J., and Zhang A. (2017) Understanding the microscopic binding mechanism of hydroxylated and sulfated polybrominated diphenyl ethers with transthyretin by molecular docking, molecular dynamics simulations and binding free energy calculations. *Mol. Biosyst.*, 13 (4) 736–749.
66. Bouachrine M., Elmchichi L., El Aissouq A., Belhassan A., Zaki H., Ouammou A., and Lakhlifi T. (2021) Molecular docking, drug likeness studies and admet prediction of flavonoids as platelet-activating factor (paf) receptor binding. *Chem. Rev. Lett.*, 4 (3) 145–152.
67. Fattouche M., Belaidi S., Abchir O., Al-Shaar W., Younes K., Al-Mogren M. M., Chtita S., Soualmia F., and Hochlaf M. (2024) Ann-qsar, molecular docking, admet predictions, and molecular dynamics studies of isothiazole derivatives to design new and selective inhibitors of hcv polymerase ns5b. *Pharmaceuticals*, 17 (12) 1712.
68. Fattouche M., Belaidi S., Ouassaf M., Chtita S., Al-Mogren M. M., and Hochlaf M. (2024) Computational studies of pyrimidine derivatives as inhibitors of human σ_1 receptor using 3d-qsar analysis, molecular docking, admet properties and dft investigation. *Chem. Phys. Impact.*, 8, 100463.
69. Jo S., Kim T., Iyer V. G., and Im W. (2008) Charmm-gui: A web-based graphical user interface for charmm. *J. comput. chem.*, 29 (11) 1859–1865.

70. Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A. E., and Berendsen H. J. (2005) Gromacs: Fast, flexible, and free. *J. comput. chem.*, 26 (16) 1701–1718.
71. Valdés-Tresanco M. S., Valdés-Tresanco M. E., Valiente P. A., and Moreno E. (2021) Gmx_mmpbsa: A new tool to perform end-state free energy calculations with gromacs. *J. Chem. Theory Comput.*, 17 (10), 6281–6291.
72. Coley C. W., Thomas Iii D. A., Lummiss J. A., Jaworski J. N., Breen C. P., Schultz V., Hart T., Fishman J. S., Rogers L., and Gao H. (2019) A robotic platform for flow synthesis of organic compounds informed by ai planning. *Science*, 365 (6453) eaax1566.
73. Choi K., Lee Y., and Kim C. (2023) An in silico study for expanding the utility of cannabidiol in alzheimer’s disease therapeutic development. *Int. J. Mol. Sci.*, 24 (21) 16013.
74. Ahmed F. Z., Ahmad A. A., and Kharazian B. (2023) A molecular dynamics study of the interaction between graphene as a carrier and gemcitabine as a chemotherapy. *Zanco J. Pure Appl. Sci.*, 35 (1) 23–29.



© 2025 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).