



ISSN: 2617-6548

URL: www.ijirss.com

Pharmacophore-guided machine learning and molecular simulations of pyrazole-imine ligands targeting estrogen receptor alpha

 Hayder H. Kadhim^{1*},  Ayad M.R. Raauf²,  Shakir Mahmood Alwan³

¹College of Pharmacy, Al-Mustansiriyah University, Baghdad, Iraq.

²College of Pharmacy, Al-Farahidi University, Baghdad, Iraq.

³College of Pharmacy, University of AlShaab, Baghdad, Iraq.

Corresponding author: Hayder H. Kadhim (Email: haider_hassan@uomustansiriyah.edu.iq)

Abstract

Estrogen receptor alpha (ER α) plays a crucial role in breast cancer progression, making it a key target for selective estrogen receptor modulators (SERMs). While Raloxifene has demonstrated therapeutic efficacy, resistance and limited bioactivity in some cases necessitate the development of novel ER α inhibitors with improved pharmacological profiles. A computational drug discovery approach integrating molecular docking, molecular dynamics (MD) simulations, and quantitative structure-activity relationship (QSAR) modeling was employed to design and evaluate new ER α inhibitors. Molecular docking was performed using Glide (XP mode) to predict ligand binding affinity and interaction patterns, while MD simulations over 100 ns assessed the stability and conformational dynamics of the protein-ligand complexes. A QSAR model was developed using a dataset of 1,231 compounds from ChEMBL, incorporating XGBoost regression with optimized hyperparameters for robust predictive performance. Compounds 3b, 3a, and 4a showed notable binding affinities (−9.319, −9.121, and −8.867 kcal/mol, respectively) that are comparable to Raloxifene (−9.791 kcal/mol), suggesting their potential as effective ER α ligands primarily through pi-pi stacking with PHE-404 and hydrogen bonding with Glu 353. MD simulations demonstrated that 3a, 4a, and 4b maintained stable receptor interactions (RMSD < 2.0 Å), while 3e and 4e exhibited higher fluctuations, indicating weaker engagement. The QSAR model achieved high predictive accuracy (RMSE < 0.6, R² > 0.8), identifying NO₂ (3d), OMe (3c), and Cl (3b) substitutions as key structural features enhancing receptor binding. This study identifies 3b, 3d, and 4b as promising lead compounds with strong binding affinity, stability, and predicted estrogen receptor activity. These findings provide a basis for further experimental validation and structural refinement to develop next-generation SERMs for ER α -positive breast cancer therapy.

Keywords: Drug discovery, Breast cancer, Estrogen receptor, Molecular docking, Molecular dynamics, QSAR modeling, Selective estrogen receptor modulators (SERMs).

DOI: 10.53894/ijriss.v8i5.8570**Funding:** This study received no specific financial support.**History: Received:** 22 May 2025 / **Revised:** 26 June 2025 / **Accepted:** 30 June 2025 / **Published:** 16 July 2025**Copyright:** © 2025 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).**Competing Interests:** The authors declare that they have no competing interests.**Authors' Contributions:** All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.**Transparency:** The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.**Acknowledgments:** The authors express their sincere gratitude to the Department of Pharmacy at Al-Mustansiriyah University for providing laboratory facilities and technical support during this study.**Publisher:** Innovative Research Publishing

1. Introduction

Breast cancer is one of the leading causes of cancer-related mortality in women worldwide, with the estrogen receptor (ER) playing a crucial role in its progression. Targeting the alpha-estrogen receptor (ER α) has been a prominent approach in endocrine therapy, with selective estrogen receptor modulators (SERMs) such as Raloxifene demonstrating significant therapeutic potential [1]. However, resistance and limited efficacy in specific patient populations demand the development of innovative ER modulators with higher binding affinity and stability. In this study, a computational drug design strategy integrating molecular docking, molecular dynamics (MD) simulations, and QSAR modeling was applied to generate and assess novel ER α inhibitors with improved pharmacological characteristics [2].

Molecular docking studies were conducted to predict the binding interactions of synthesized drugs with ER α (PDB ID: 1ERR, Chain A) using the Schrödinger software package. The validation of the docking process was confirmed by redocking Raloxifene, which yielded an RMSD of 2 Å, indicating the accuracy of the method. Several developed compounds, particularly 3b, 3a, and 4a, demonstrated significant binding affinities comparable to Raloxifene, mainly through pi-pi stacking interactions with PHE-404 and hydrogen bonding with Glu 353, which are important for receptor stability. Molecular dynamics (MD) simulations provided further insights into the dynamic behavior of the protein-ligand complexes over 100 ns. The results showed that compounds 3a, 4a, and 4b maintained stable interactions, whereas 3e and 4e exhibited larger fluctuations, suggesting weaker receptor engagement [3].

A QSAR model was developed to predict estrogen receptor modulation using a dataset of 1,231 chemicals from ChEMBL. Molecular descriptors, including Lipinski's parameters and Morgan fingerprints, were used to encode structural information, and XGBoost regression was improved through hyperparameter tuning. The model achieved a low RMSE (<0.6) and good R² correlation, indicating robust predictive ability. Notably, the NO₂ (3d), OMe (3c), and Cl (3b) alterations emerged as key factors enhancing receptor binding, emphasizing the importance of electron-withdrawing and electron-donating groups in strengthening ligand affinity [3, 4].

The findings of this study highlight 3b, 3d, and 4b as promising lead compounds due to their high binding affinity, stability in MD simulations, and favorable QSAR-predicted pIC₅₀ values. The results also emphasize the importance of hydrophobic interactions, hydrogen bonding with Glu 353, and ligand-induced conformational changes in optimizing binding efficiency. These insights provide a strong foundation for further experimental validation and structural refinements, aiming to develop next-generation SERMs with enhanced therapeutic efficacy against ER α -positive breast cancer [4].

2. Methodology

2.1. Molecular Docking Studies

The molecular docking simulations were conducted to determine the binding affinity and interaction mode of the synthesized compounds with the alpha-estrogen receptor (PDB ID: 1ERR, chain A). The receptor's crystal structure was obtained from the Protein Data Bank (PDB) and optimized using the Protein Preparation Wizard in Maestro [5]. This preparation required the inclusion of correct protonation states at pH 7.4 using Epik, hydrogen bond network optimization with PROPKA, and minimization using the OPLS4 force field. Water molecules more than 5 Å from the co-crystallized ligand were removed [6].

LigPrep in Maestro generated ligand structures to produce possible tautomeric and ionization forms at pH 7.4, using OPLS4 force field parameters. The grid receptor was a 20 Å cube with the co-crystallized ligand (Raloxifene) binding site at the center of the grid. The Van der Waals scaling factor was 1.0, and the partial charge cutoff was 0.25. Docking was performed using Glide (XP mode) with flexible treatment of the ligands to maximize binding poses [7]. Validation of docking was attained by redocking raloxifene to the receptor to confirm the dependability of the docking protocol using the docked pose and the experimental conformation [1].

2.2. Molecular Dynamics (MD) Simulations

MD simulations were carried out using Desmond [5] to assess the stability and dynamic behavior of the docked protein-ligand complexes. The system was placed in a triclinic simulation box with adequate solvation and boundary conditions. The OPLS4 force field was applied for molecular interactions [8-10].

The system was solvated using the SPC (Simple Point Charge) water model, and counterions were added for charge neutrality. A buffer region was included to prevent interactions with periodic images. The system was then energy-minimized, followed by NPT (constant temperature and pressure) equilibration at 310.15 K, mimicking physiological conditions [11].

Temperature was controlled using the Nose-Hoover chain thermostat, and system pressure was regulated with the Martyna-Tobias-Klein barostat, featuring isotropic coupling and a 2 ps relaxation time. Production was conducted for 100 ns with a 2-fs time step.

Structural and energetic characterization included:

- Root Mean Square Deviation (RMSD)
- Root Mean Square Fluctuation (RMSF)
- Residue contacts
- Binding energy stability ($\Delta G \sim 1.2$ kcal/mol)

2.3. Development of a QSAR Model for Estrogen Receptor Modulation

A Quantitative Structure-Activity Relationship (QSAR) model was developed to predict the activity of estrogen receptor modulators.

1. Dataset Preparation
 - A dataset of 1231 compounds was retrieved from the ChEMBL database, with experimentally determined activity values (pIC₅₀).
 - Missing values were imputed with zero to maintain data integrity.
2. Descriptor Calculation
 - Molecular descriptors were generated using RDKit.
 - Lipinski's descriptors and Morgan fingerprints (radius = 2, 1024-bit) were computed to encode structural features.
3. Feature Selection and Preparation
 - The feature set was normalized using StandardScaler.
 - SelectKBest (f_regression scoring) was utilised to retain the top 200 most informative features.
4. Model Training
 - The dataset was split into 80% train and 20% test sets.
 - XGBRegressor (XGBoost library) was trained.
 - Hyperparameter tuning was performed using RandomizedSearchCV with a 3-fold cross-validation, optimizing:
 - n_estimators: [200, 300, 400]
 - max_depth: [4, 6, 8]
 - learning_rate: [0.005, 0.01, 0.05, 0.1]
 - subsample: [0.6, 0.7, 0.8]
 - colsample_bytree: [0.6, 0.7, 0.8]
 - random_state: 42
 - The best model was selected based on the lowest RMSE (<0.6).
5. Model Evaluation
 - Performance was quantified using RMSE and R².
 - Visualization included:
 - Parity plots
 - Residual plots
 - Learning curves
6. Model Interpretability
 - SHAP (Shapley Additive Explanations) was utilized to recognize features with the most impact on predictions.
7. Validation and Robustness
 - Y-scrambling tests to eliminate the randomness of that model performance.
 - The applicability domain was analyzed to determine limits of prediction.
8. Visualization
 - matplotlib was used for graphical representations of model performance, feature importance, and chemical space distribution.

3. Results and Discussion

Molecular docking was conducted to predict the binding affinity and interaction patterns of the synthesized compounds with the alpha-estrogen receptor (PDB ID: 1ERR, chain A). The docking validation process demonstrated an RMSD of 2 Å

between the original and docked conformations of the reference ligand, Raloxifene, confirming the reliability of the docking protocol (Figure 1) [1].

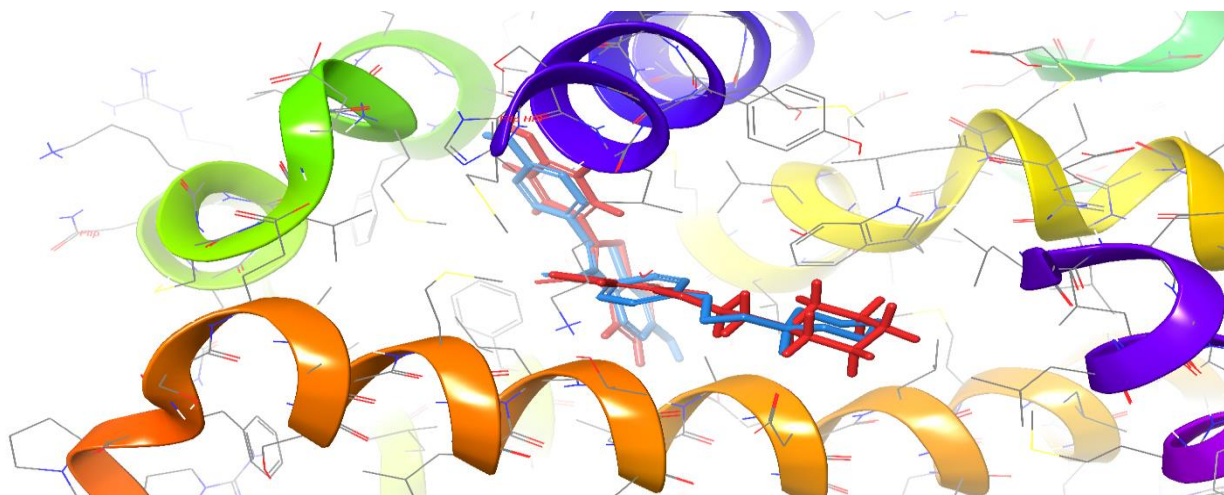
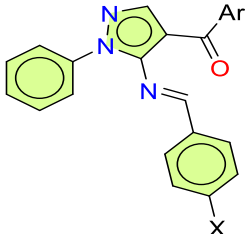


Figure 1.
Validation of Docking process.

Docking scores (Table 1) revealed that most compounds exhibited strong binding affinity, with 3b (-9.319 kcal/mol), 3a (-9.121 kcal/mol), and 4a (-8.867 kcal/mol) ranking among the highest. These values were comparable to Raloxifene (-9.791 kcal/mol), suggesting strong interactions with the receptor binding site. Notably, compound 3e (-4.727 kcal/mol) showed the weakest binding, indicating a suboptimal interaction with the receptor.

Table 1.
Docking results in Kcal/mol for each ligand generated by Glide tool that is a built-in feature in Maestro software.



ID	Docking Score (kcal/mol)	“Ar” group	“X” group
3a	-9.121	Furan	H
3b	-9.319	Furan	Cl
3c	-8.669	Furan	OMe
3d	-7.850	Furan	NO ₂
3e	-4.727	Furan	N(CH ₃) ₂
4a	-8.867	Thiophene	H
4b	-8.316	Thiophene	Cl
4c	-7.924	Thiophene	OMe
4d	-6.822	Thiophene	NO ₂
4e	-4.539	Thiophene	N(CH ₃) ₂
Raloxifene	-9.791	-	-

The docking results were further analyzed using 2D and 3D molecular interaction visualizations Figure 2-11, which highlighted how functional groups influenced ligand conformation and binding strength. Differences in ligand orientation, particularly among substituted benzene rings, directly impacted pi-stacking and hydrogen bonding efficiency, contributing to variations in docking scores. These analyses highlighted pi-pi stacking, salt bridge formation, and hydrogen bonding as crucial determinants of binding strength [7].

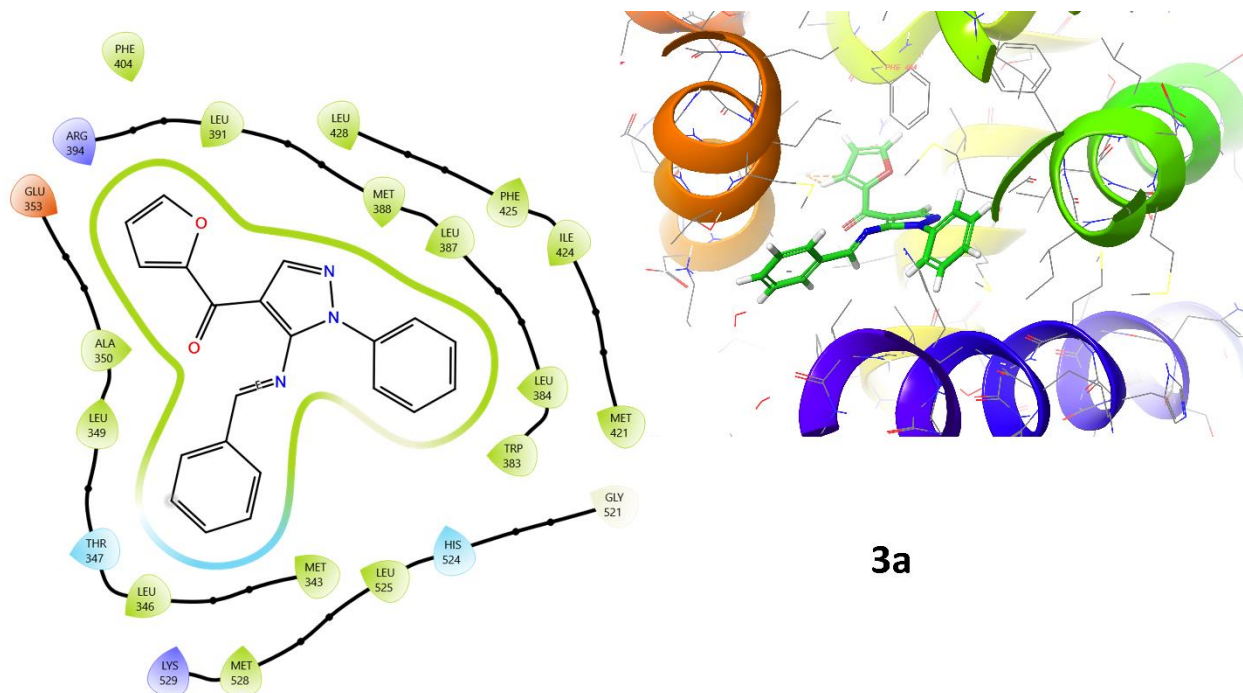


Figure 2.
Compound 3a within the binding site of ER- α .

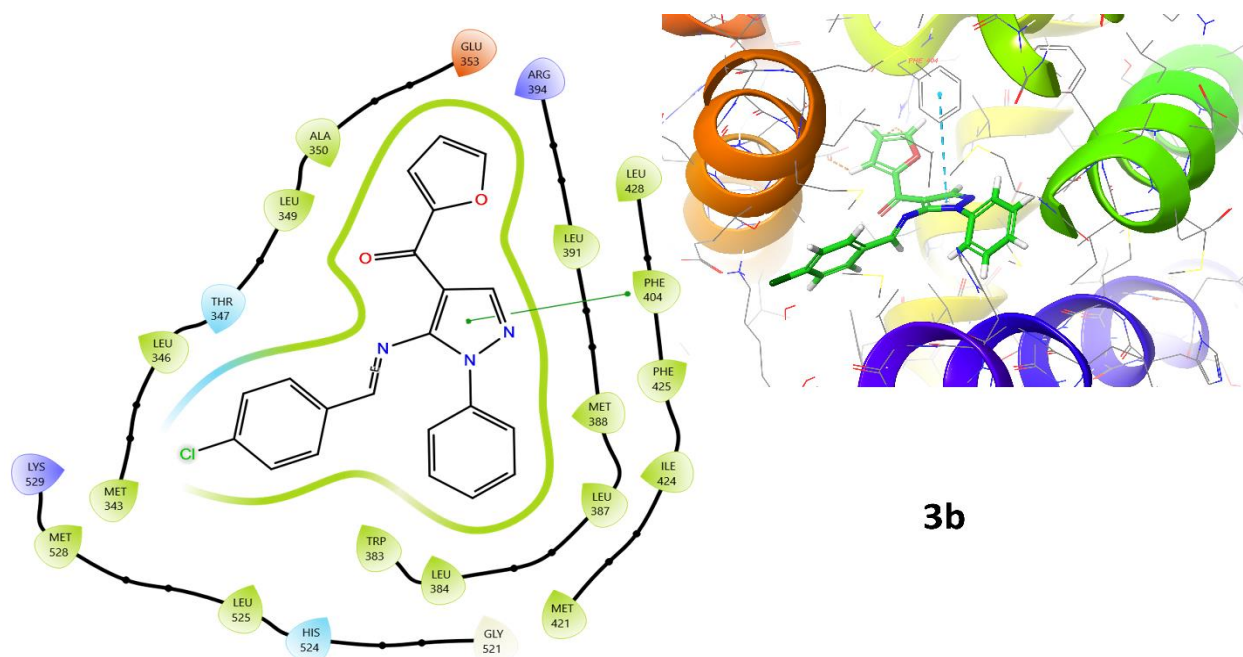


Figure 3.
Compound 3b within the binding site of ER- α .

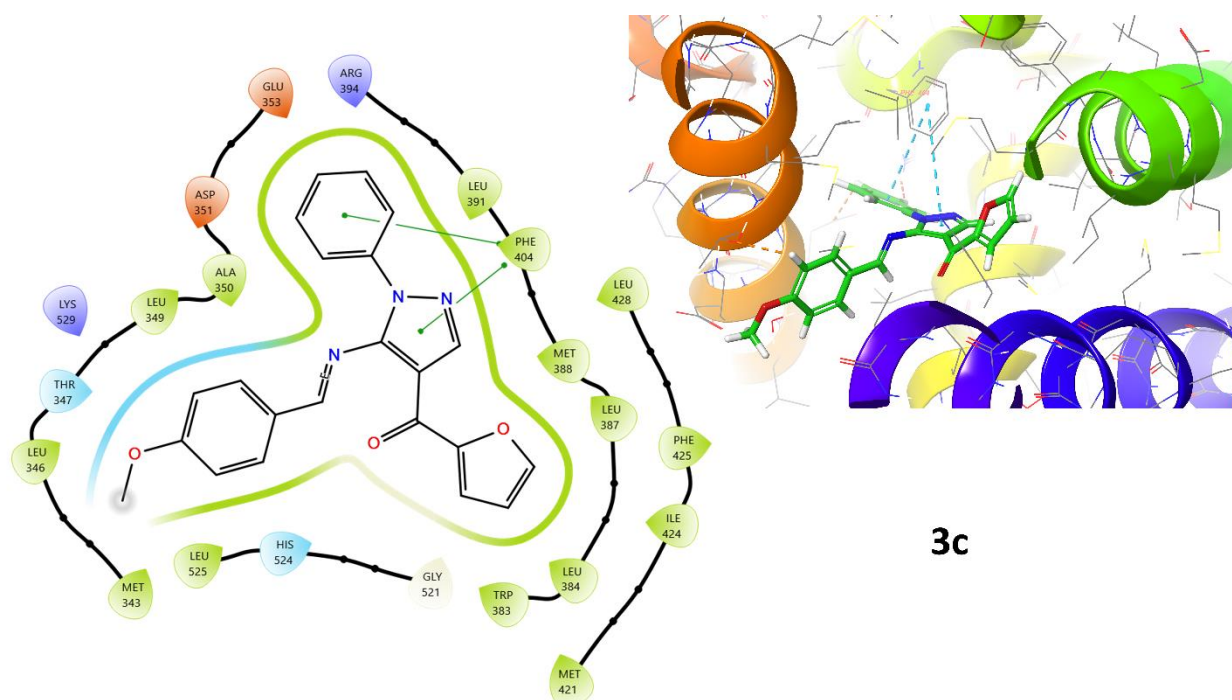


Figure 4.
Compound 3c within the binding site of ER- α .

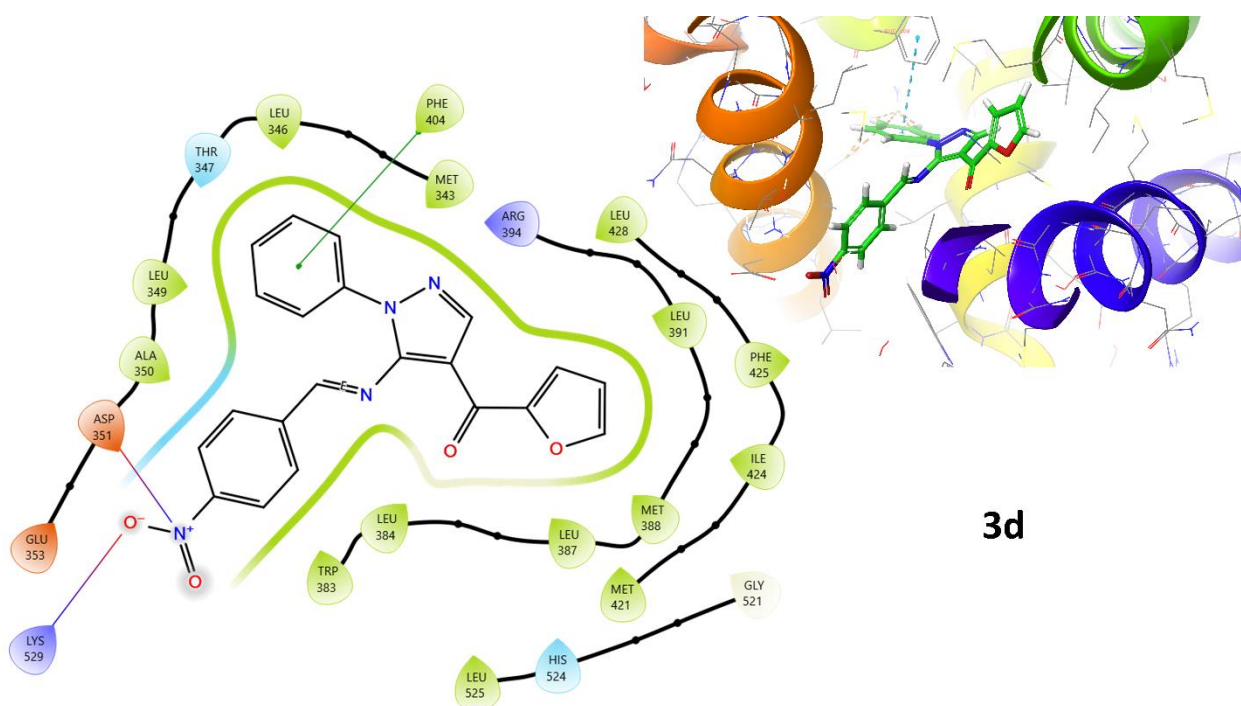


Figure 5.
Compound 3d within the binding site of ER- α .

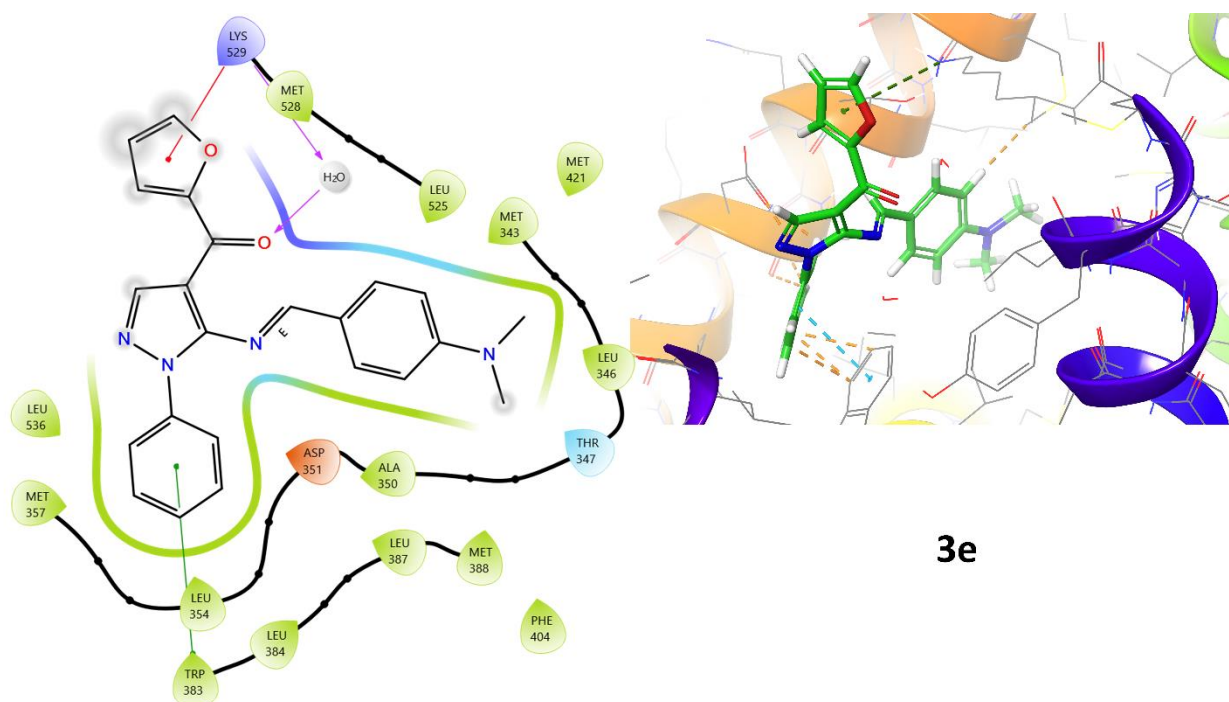


Figure 6.
Compound 3e within the binding site of ER- α .

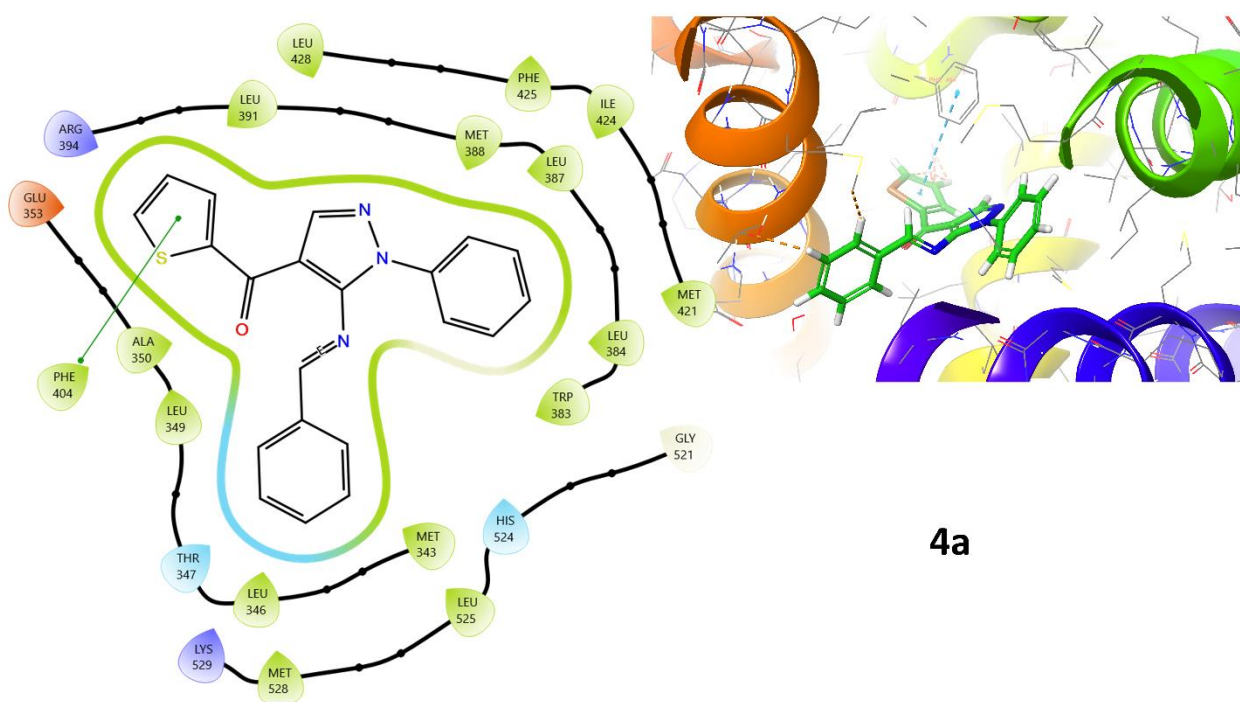


Figure 7.
Compound 4a within the binding site of ER- α .

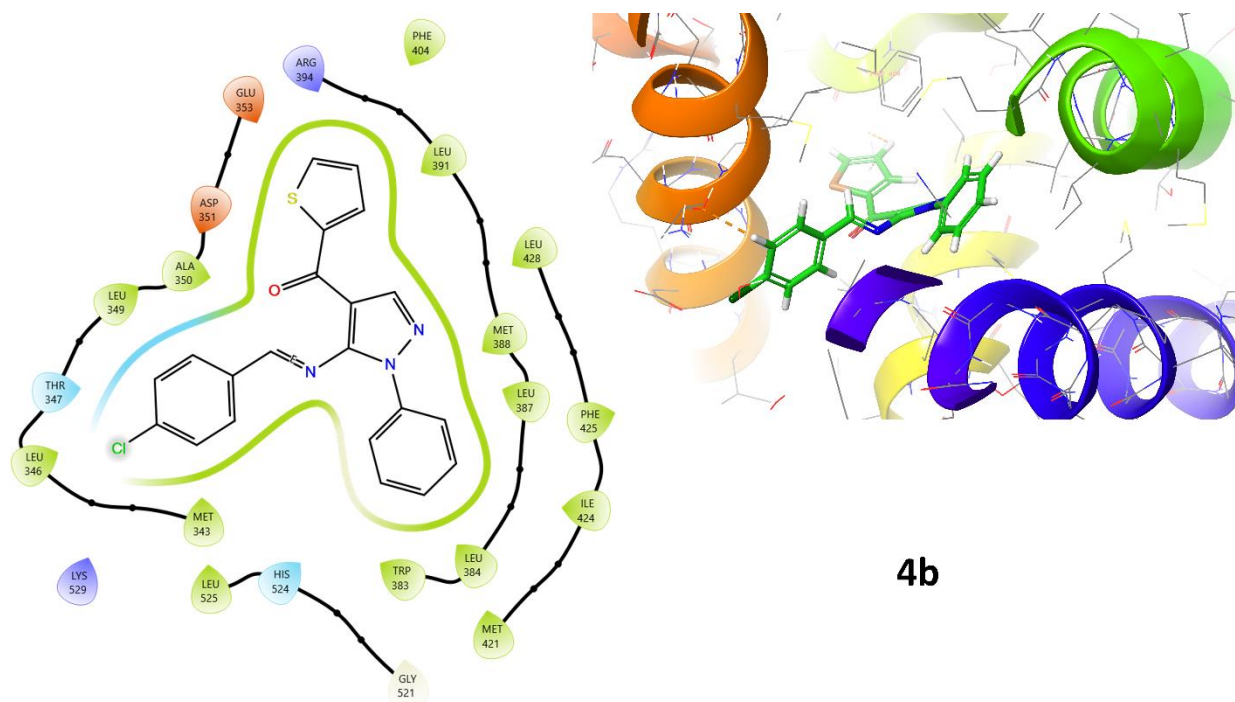


Figure 8.
Compound 4b within the binding site of ER- α .

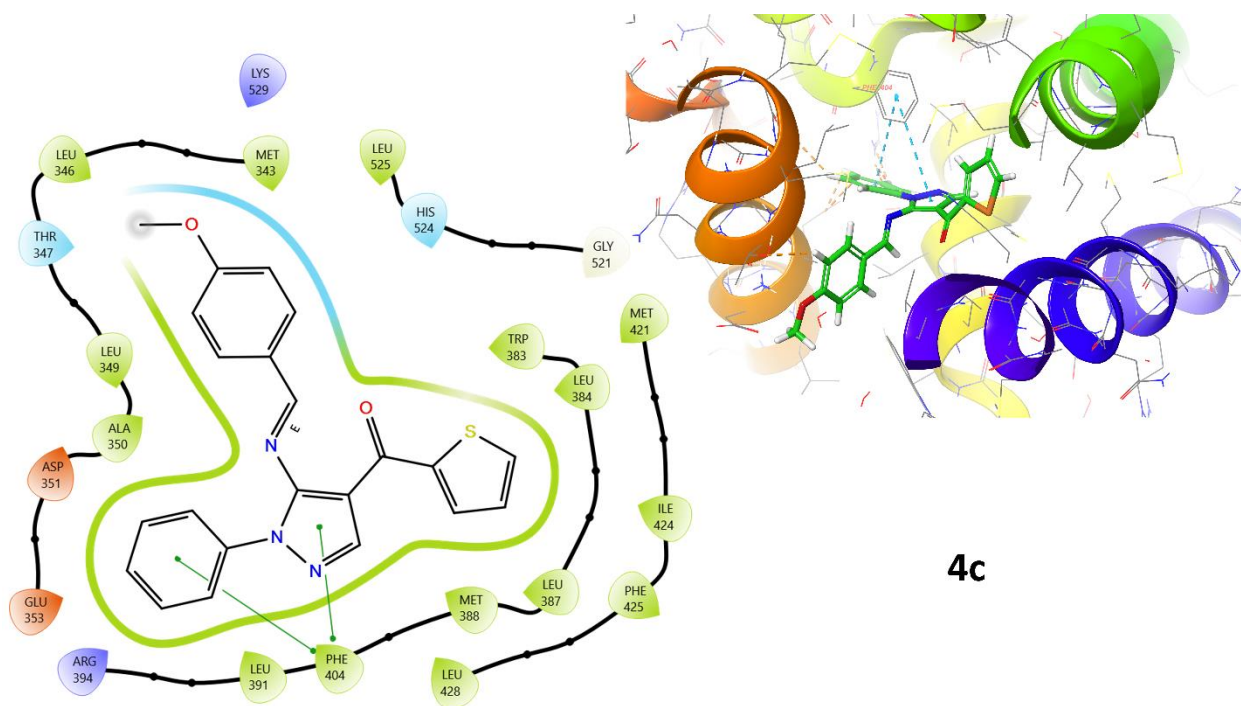


Figure 9.
Compound 4c within the binding site of ER- α .

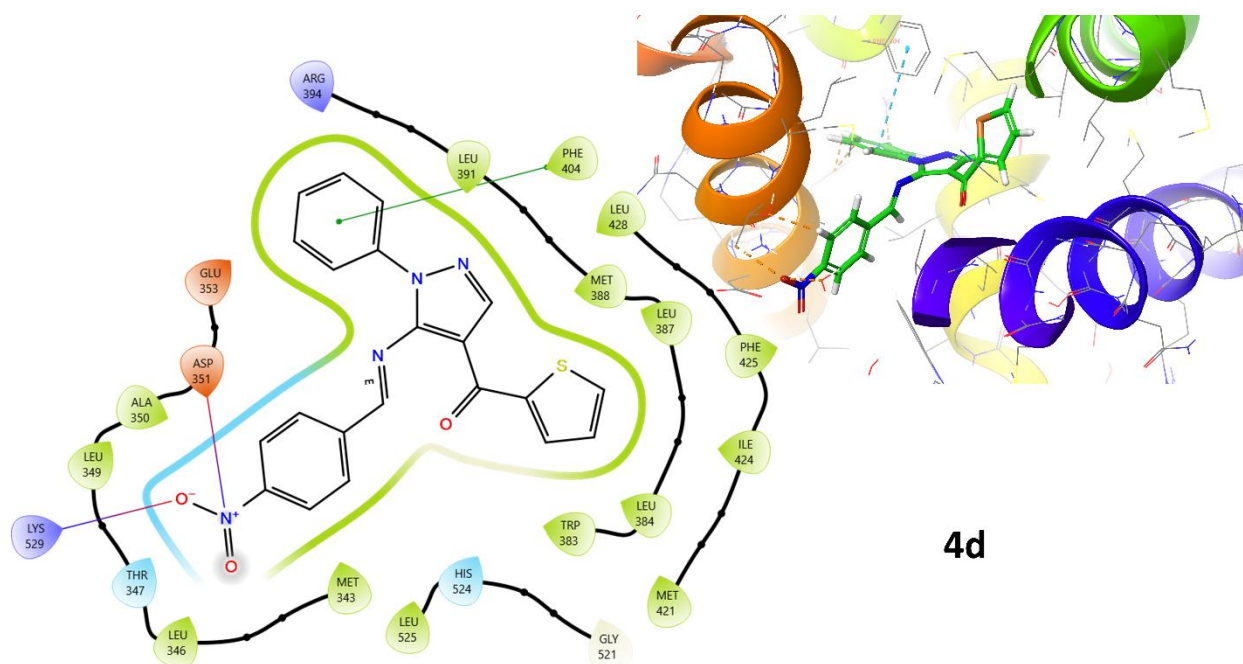


Figure 10.
Compound 4d within the binding site of ER- α .

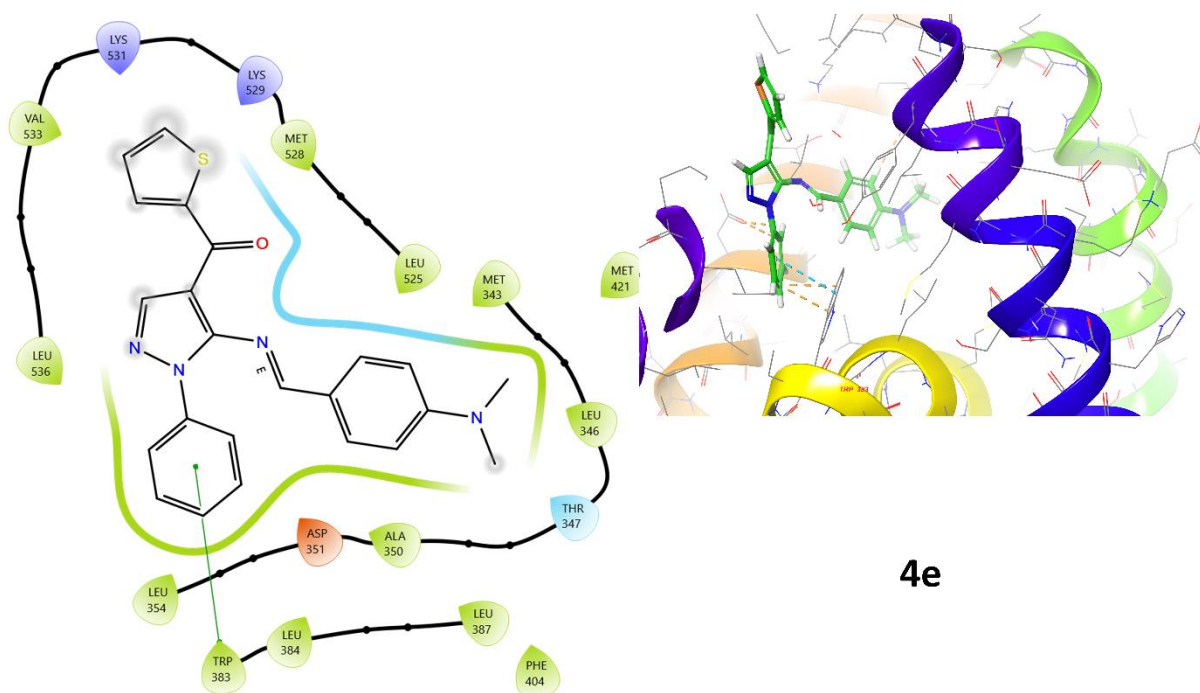


Figure 11.
Compound 4e within the binding site of ER- α .

Compounds 3b and 3c displayed broad interactions within the binding pocket, particularly engaging in pi-pi stacking with PHE-404, contributing significantly to their hydrophobic stabilization. The presence of furan rings appears to play a role in enhancing these interactions. Compound 3d exhibited moderate binding affinity (-7.850 kcal/mol), with the NO₂ group forming two salt bridges with ASP-351 and LYS-529. The 2D and 3D molecular interaction analyses (Figure 5) revealed that the NO₂ group induced a distinct ligand reorientation compared to unsubstituted analogs, optimizing electrostatic interactions but reducing overall hydrophobic engagement. The altered conformation may explain its lower docking score relative to other high-affinity compounds. These interactions highlight the significant role of electrostatic contributions in ligand stability. This interaction likely contributed to moderate binding affinity (-7.850 kcal/mol) despite the presence of strong electrostatic interactions. Compound 3e showed a pi-cation interaction with LYS-529; however, its weaker binding score (-4.727 kcal/mol) suggests limited interaction. Additionally, the furan carbonyl oxygen formed a hydrogen bond with a bridging water molecule, which likely contributed to some degree of affinity but was not sufficient to enhance overall binding activity. The thiophene-containing compounds (4-series) exhibited similar interactions to their

furan counterparts with minimal impact on docking scores. Figures 7-11 suggest that thiophene substitution led to minor conformational adjustments without significantly altering the key pi-stacking and hydrogen bonding interactions within the receptor binding site. Compound 4a maintained strong pi-pi stacking with PHE-404, contributing to its robust docking score (-8.867 kcal/mol). Compounds 4c and 4d displayed multiple pi-pi stacking interactions with PHE-404, with 4d also forming salt bridges with ASP-351 and LYS-529, similar to 3d. In contrast, compound 4e, which formed pi-pi stacking with TRP-383, exhibited a lower docking score (-4.539 kcal/mol), indicating weaker receptor engagement. The low scores of 3e and 4e are likely due to steric hindrance, a lack of significant electrostatic or hydrogen bonding interactions, and an unfavorable ligand orientation within the binding site. Unlike electron-withdrawing groups (NO₂, Cl), which enhance receptor interactions, the bulky diethylamine group may disrupt binding efficiency by interfering with key stabilizing forces.

In relation to molecular dynamics (MD) simulations, we conducted a root-mean-square deviation (RMSD) analysis to evaluate the structural stability of the protein in its interactions with the designed ligands (3a–4e) and the reference compound, Raloxifene (Ralo). The RMSD plots (Figure12) provide insights into the conformational changes of the protein and ligands over the 100 ns molecular dynamics (MD) simulations.

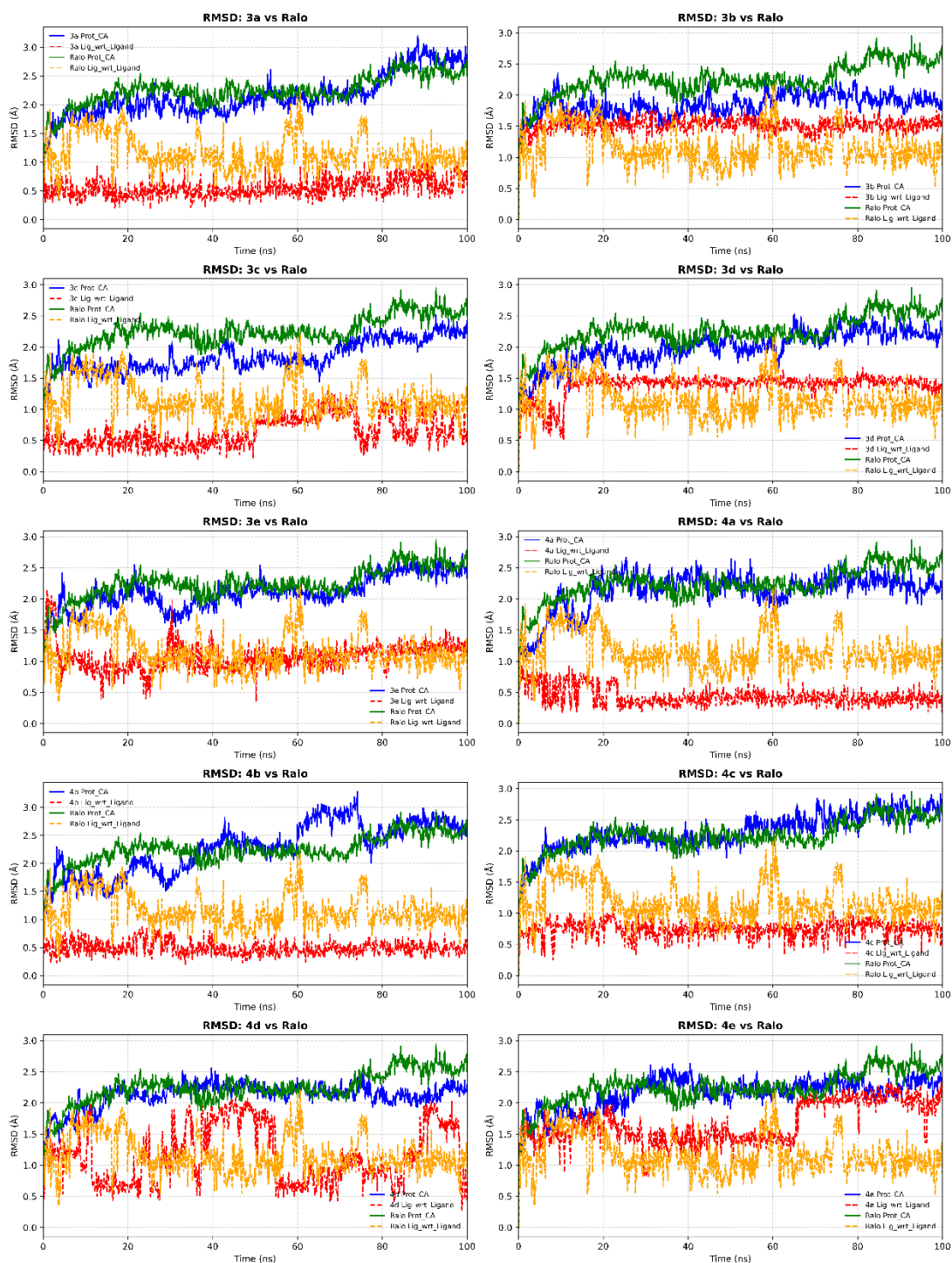


Figure 12.

Root Mean Square Deviation (RMSD) plots comparing protein and ligand stability over 100 ns molecular dynamics simulations for Raloxifene (Ralo) and the compounds (3a–4e). Each subplot shows RMSD profiles of protein alpha carbon (Prot_CA) and ligand (Lig_wrt_Ligand) for the tested compound alongside Ralo for direct comparison. Stable RMSD values suggest consistent ligand binding and structural integrity of the protein-ligand complex over time.

The protein's RMSD upon binding to the selected ligands (3a–4e), illustrated by the blue line, exhibits mild variations, often stabilizing within a range of 1.5–3.0 Å. This indicates that the protein maintains its structural integrity with only slight conformational changes while interacting with these novel ligands. Comparatively, the protein's RMSD upon binding to Raloxifene, illustrated by the green line, follows a similar trend, suggesting that the overall protein stability is comparable between the developed compounds and the reference ligand.

The RMSD of Raloxifene itself, illustrated by the orange line, is very steady throughout the simulation, often varying between 1.0–2.0 Å. This stability indicates that Raloxifene maintains a consistent binding mode with few variations. In contrast, the RMSD of the proposed ligands (red line) varies among different compounds. Some ligands, such as 3a, 4a, and 4b, have low RMSD values, remaining below 2.0 Å, indicating robust and stable binding within the active site. Other ligands, such as 3c, 3d, and 4e, display mild variations (1.5–2.5 Å), suggesting some degree of flexibility but retained

binding. However, ligands such as 3e and 4e show higher variations, exceeding 2.5 Å, which may imply weaker interactions, conformational rearrangements, or possibly partial dissociation from the binding site.

A key observation is the comparative action of the proposed ligands against Raloxifene. While some ligands demonstrate equivalent or better stability within the binding pocket, others exhibit larger fluctuations, suggesting that certain structural alterations may affect the binding mechanism and stability. The results indicate that ligands 3a, 4a, and 4b show promise as potential lead candidates due to their stability; however, ligands such as 3e and 4e may need further modification to enhance their binding affinity and eliminate structural variations.

Moreover, the root-mean-square fluctuation (RMSF) difference plots (Figure 13) offer insights into the dynamic behavior of protein residues upon the binding of the proposed ligands (3a–4e) compared to Raloxifene (Ralo). These differences in flexibility highlight how distinct ligands alter the conformational landscape of the binding pocket. The y-axis displays the Δ RMSF, where positive values imply higher flexibility upon binding of the proposed ligands, and negative values suggest that the residues are more flexible when bound to Raloxifene. Residues with values approaching 0 remain stiff and are not significantly altered by ligand binding.

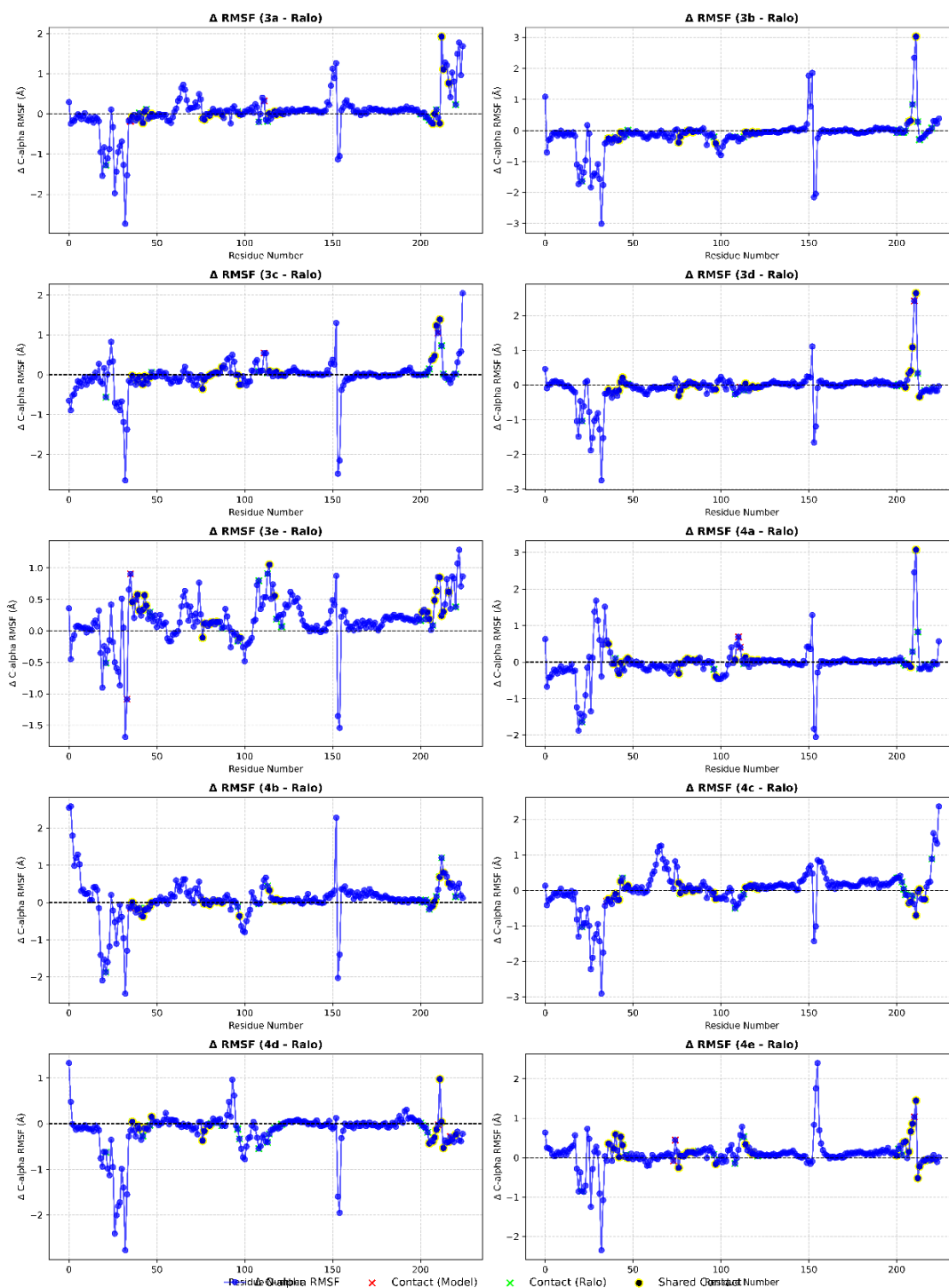


Figure 13.

Δ RMSF plots showing the difference in Ca root mean square fluctuation (RMSF) between Raloxifene (Ralo) and the compounds (3a–4e) across protein residues. Positive or negative values indicate increased or decreased flexibility compared to Ralo, respectively. Key residue contacts are marked: red crosses for contacts unique to the 3a–4e, green Xs for Ralo-specific contacts, and yellow circles for shared interactions. These plots highlight structural perturbations and dynamic differences induced by ligand binding.

An important observation in these RMSF plots is the presence of unique contact residues (indicated with red X) that interact with the intended ligands but not with Raloxifene. These residues correspond to locations where the proposed ligands explore previously vacant areas within the binding pocket, potentially revealing additional binding mechanisms or allosteric interactions. This suggests that the novel chemicals may induce local conformational changes that enable deeper or broader engagement within the active site, which could enhance binding affinity or introduce new stabilizing interactions. The flexibility observed in these locations (positive Δ RMSF) further supports this notion, as increased mobility generally correlates with ligand-induced reconfiguration of the binding environment.

In contrast, residues exclusively engaged by Raloxifene (shown by green X) are typically linked to negative Δ RMSF values, signifying increased flexibility upon Raloxifene binding while maintaining greater stability when interacting with the designed ligands. This indicates that Raloxifene can promote a unique dynamic equilibrium within the pocket,

potentially stabilizing alternative conformations compared to the intended ligands. This varied flexibility profile suggests that whereas Raloxifene adheres to a recognized binding mode, certain designed ligands may be interacting with alternate subpockets or eliciting minor allosteric effects.

The conserved contact residues (yellow circles) shared by Raloxifene and the proposed ligands are predominantly linked to areas that exhibit minimal flexibility (Δ RMSF near zero). This indicates that these fundamental interactions are essential for binding stability and are preserved regardless of ligand identity. The presence of flexible sections adjacent to unique contact residues suggests that the engineered ligands may be enlarging the available binding pocket volume, which could be utilized for future enhancements in drug design.

The interaction examination of Raloxifene (Ralo) and the synthesized ligands (3a–4e) with the target protein indicates significant variations in binding processes. Raloxifene predominantly depends on a synergy of hydrophobic contacts, hydrogen bonds, and water bridges for stable binding. The most significant interaction is a robust hydrogen bond with Glu 353 Figure 14a-c, which is essential for maintaining the ligand in the active site. Moreover, hydrophobic interactions with leucine, phenylalanine, and methionine residues substantially enhance the ligand's stable binding conformation. Water bridge connections improve their affinity by providing additional stabilization through solvent-mediated contacts.

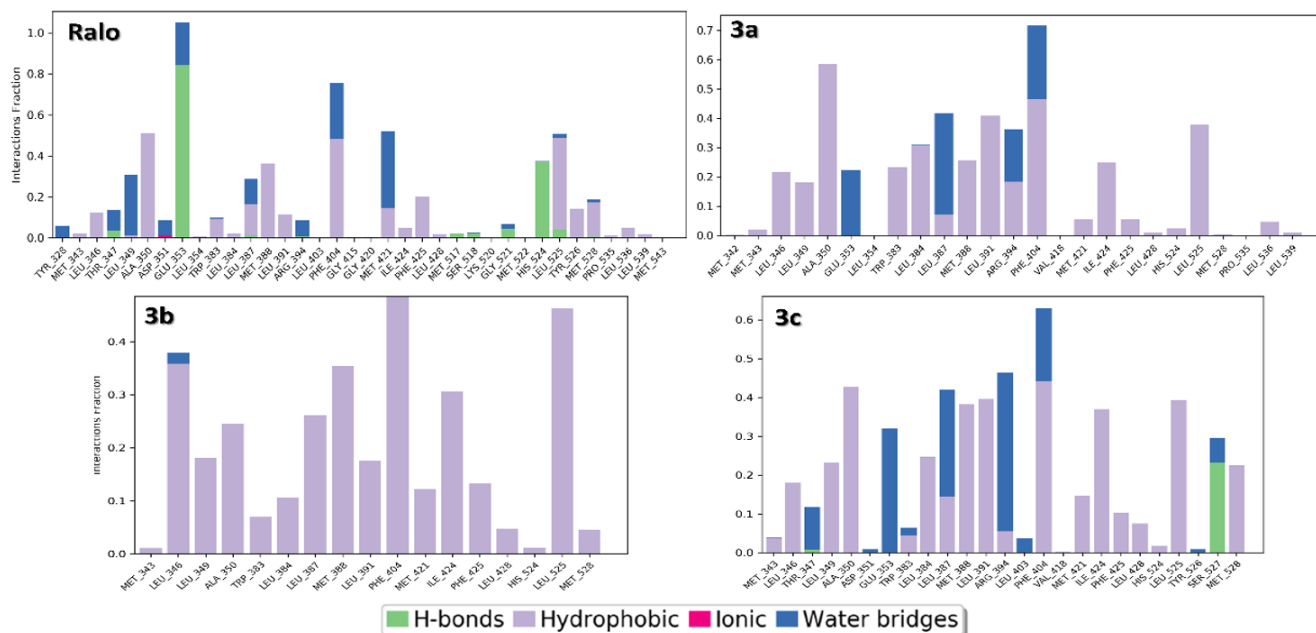


Figure 14a.

Interaction profiles of Raloxifene (Ralo), 3a, 3b, and 3c with the target protein during MD simulations. Bars represent interaction fractions with key residues, categorized as hydrogen bonds (green), hydrophobic (purple), ionic (pink), and water bridges (blue).

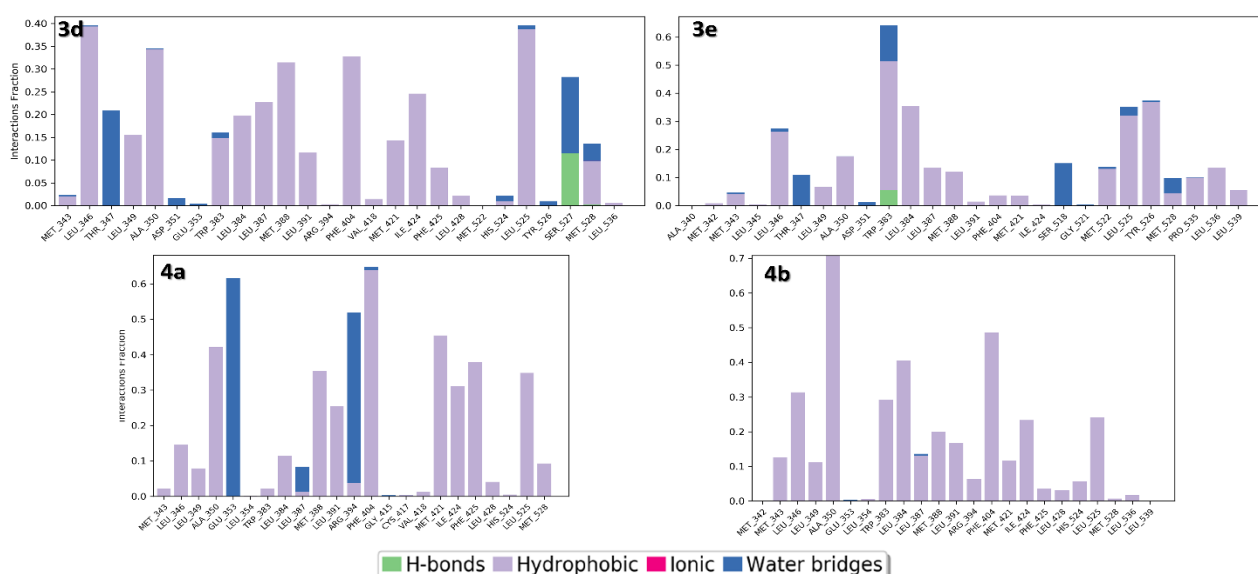
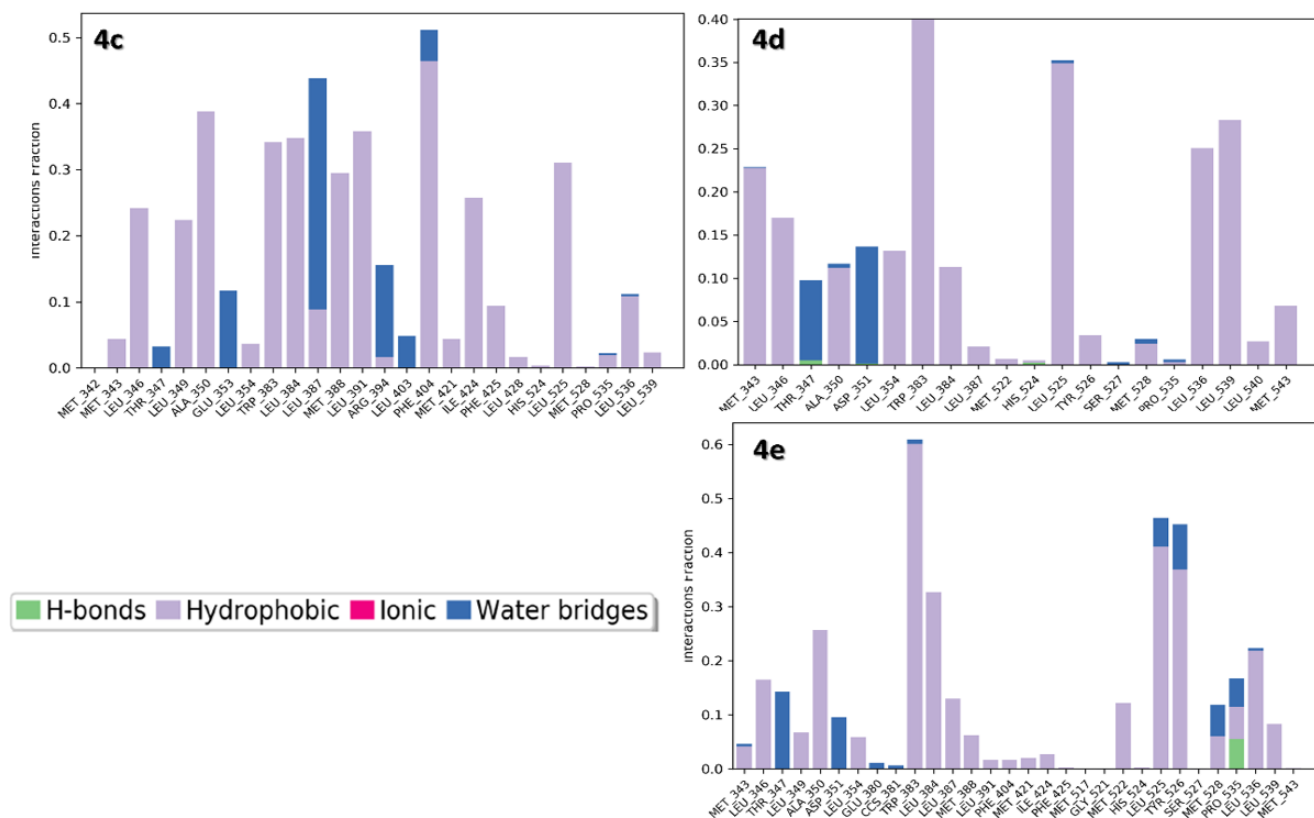


Figure 14b.

Interaction profiles of 3d, 3e, 4a, and 4b with the target protein during MD simulations. Bars represent interaction fractions with key residues, categorized as hydrogen bonds (green), hydrophobic (purple), ionic (pink), and water bridges (blue).

**Figure 14c.**

Interaction profiles of 4c, 4d, 4e with the target protein during MD simulations. Bars represent interaction fractions with key residues, categorized as hydrogen bonds (green), hydrophobic (purple), ionic (pink), and water bridges (blue).

In comparison, the designed ligands exhibit diverse interaction profiles, suggesting variations in their binding mechanisms. Many of the new ligands, such as 3a, 3b, 4a, and 4b, maintain extensive hydrophobic interactions similar to Raloxifene, particularly with Leu, Phe, and Met residues, indicating strong van der Waals contributions to their stability. However, differences emerge in hydrogen bonding patterns. Some ligands, such as 3c, 3e, and 4e, introduce additional hydrogen bonds and water bridge interactions, potentially exploring alternative binding regions within the pocket. While this may enhance ligand stability, the absence of the critical Glu 353 interaction in some cases could impact overall binding affinity.

Another significant difference is observed in the role of water bridge interactions. Ligands such as 3c, 4a, and 4e show an increased fraction of water bridges compared to Raloxifene, indicating a stronger dependence on solvent-mediated stabilization. Conversely, ionic interactions appear to play a minimal role across all designed ligands, such as Raloxifene, suggesting that charge-based interactions are not a major factor in ligand binding.

Among the designed compounds, 3b and 4b exhibit the strongest hydrophobic interactions, making them promising candidates for maintaining stable binding. On the other hand, 3c and 4e explore new binding regions through additional water bridge interactions, which may offer alternative stabilization mechanisms but require further evaluation to confirm their effectiveness. The variations in interaction patterns suggest that certain modifications to the designed ligands could enhance their binding properties, particularly by reinforcing hydrogen bonding with Glu 353 while maintaining strong hydrophobic interactions.

Moreover, the QSAR model developed for predicting estrogen receptor activity was comprehensively evaluated through multiple validation techniques, including applicability domain analysis, molecular similarity assessment, model learning performance, and prediction reliability. The findings provide insights into the molecular interactions driving receptor activity and highlight the predictive capability of the model for novel compounds.

The applicability domain analysis (Figure 15) was conducted using a leverage-based approach to assess the reliability of predictions. The leverage threshold ($h^* = 0.773$) defines the space where predictions are expected to be reliable. Most compounds fall within this threshold, indicating high confidence in their predicted values. However, a few molecules exceed this boundary, suggesting they may represent outliers with unique structural characteristics. These high-leverage compounds may require further experimental validation or additional training data to improve model generalization.

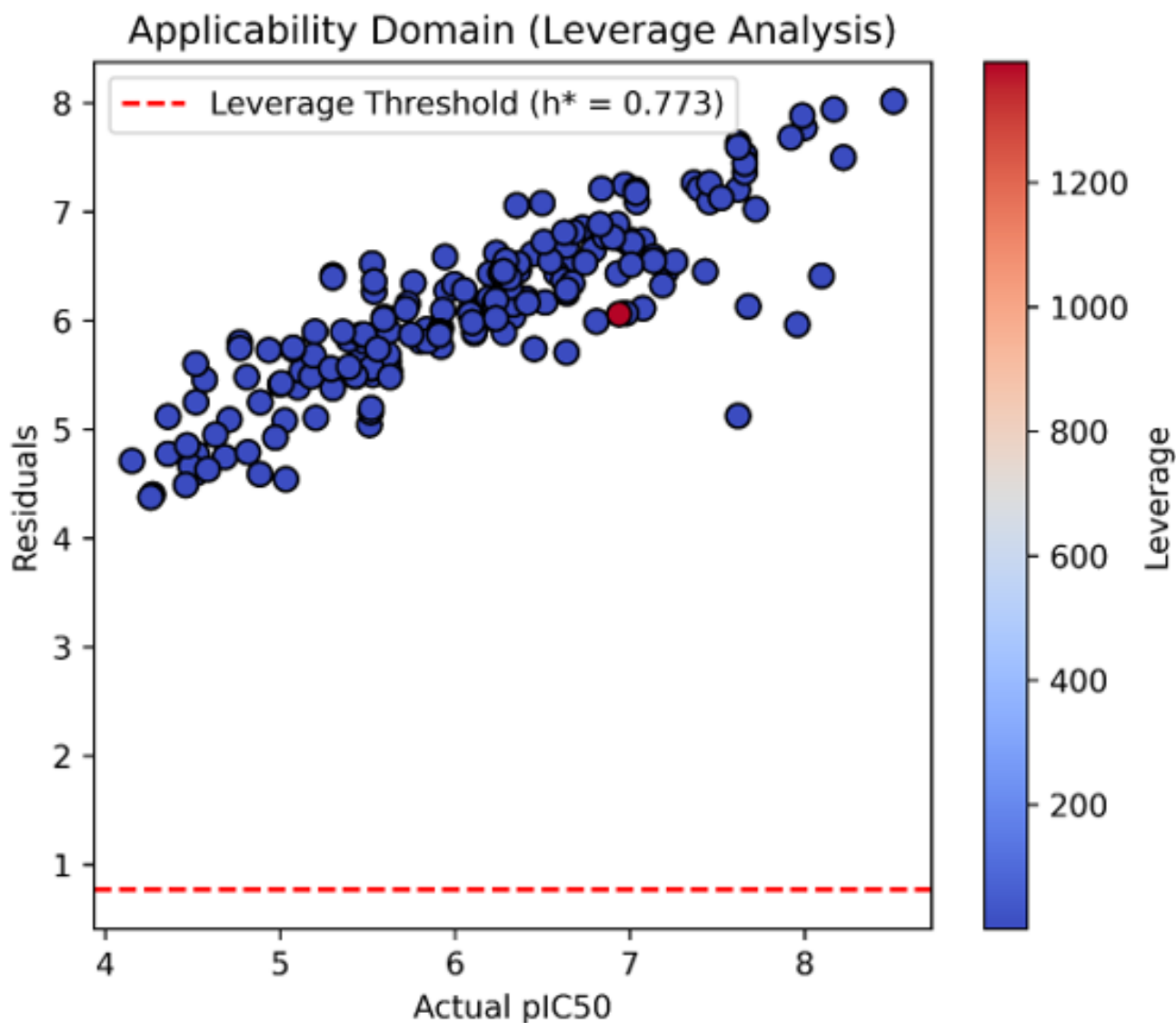


Figure 15.

Applicability domain plot based on leverage analysis. The x-axis shows actual pIC_{50} values, and the y-axis represents model residuals. Each point corresponds to a compound, colored by leverage score. The red dashed line indicates the leverage threshold ($h^* = 0.773$); compounds beyond this limit may be outside the model's reliable prediction domain.

The chemical space visualization (Figure 16) using principal component analysis (PCA) provides an overview of the molecular diversity in the dataset. The distribution of compounds suggests a well-balanced chemical space with no significant clustering of highly active molecules, which is advantageous for broad applicability in virtual screening. The molecular similarity map (Figure 17) using t-SNE further explores compound relationships based on molecular fingerprints. Structurally similar molecules with similar pIC_{50} values are grouped, validating the model's ability to differentiate active from inactive compounds. This clustering pattern also highlights structurally diverse molecules with potential novel interactions, which may serve as lead candidates for further optimization.

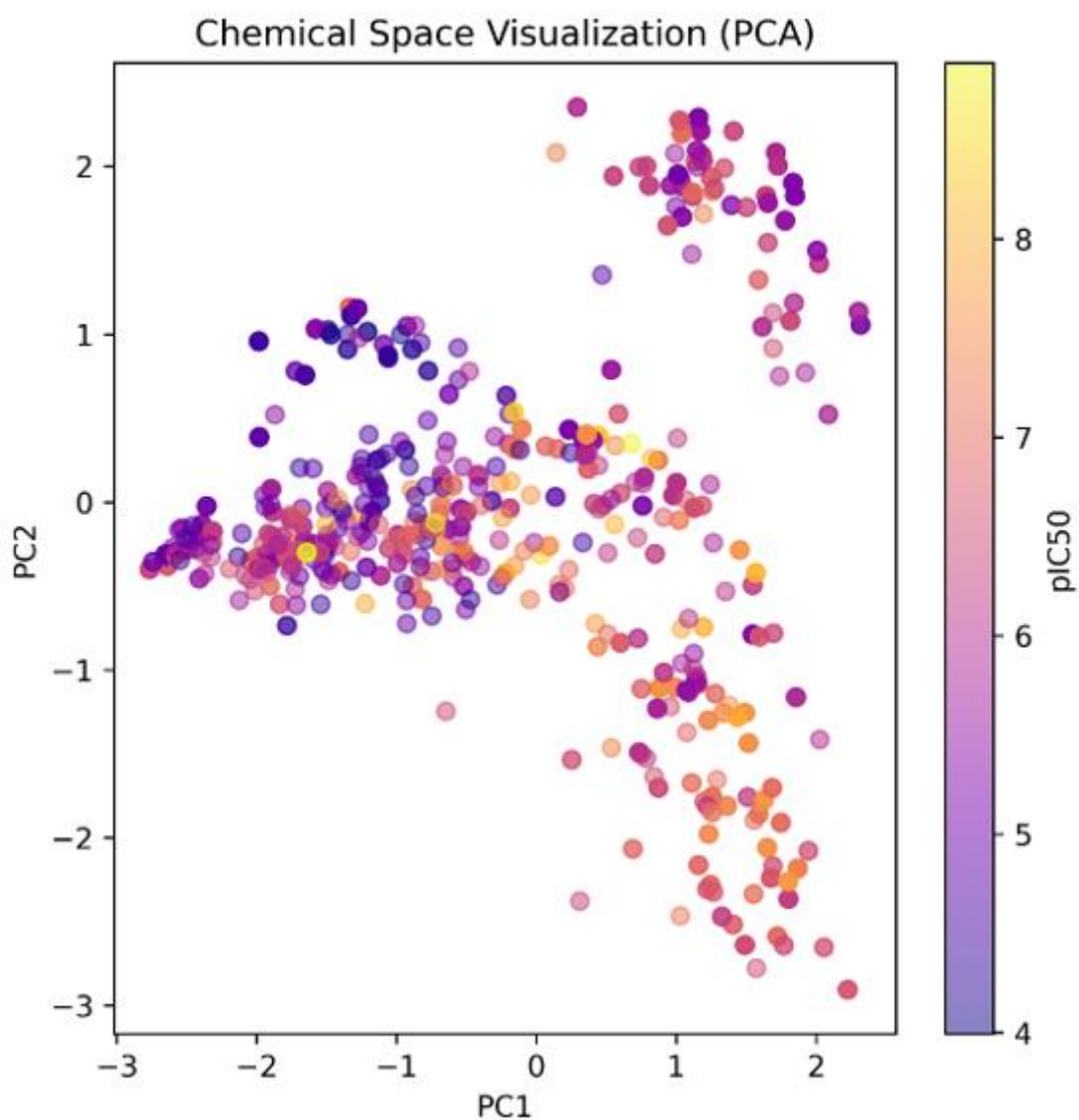


Figure 16.

Chemical space visualization using Principal Component Analysis (PCA). Compounds are projected onto the first two principal components (PC1 and PC2), with colors representing pIC₅₀ values. This plot illustrates the structural diversity and distribution.

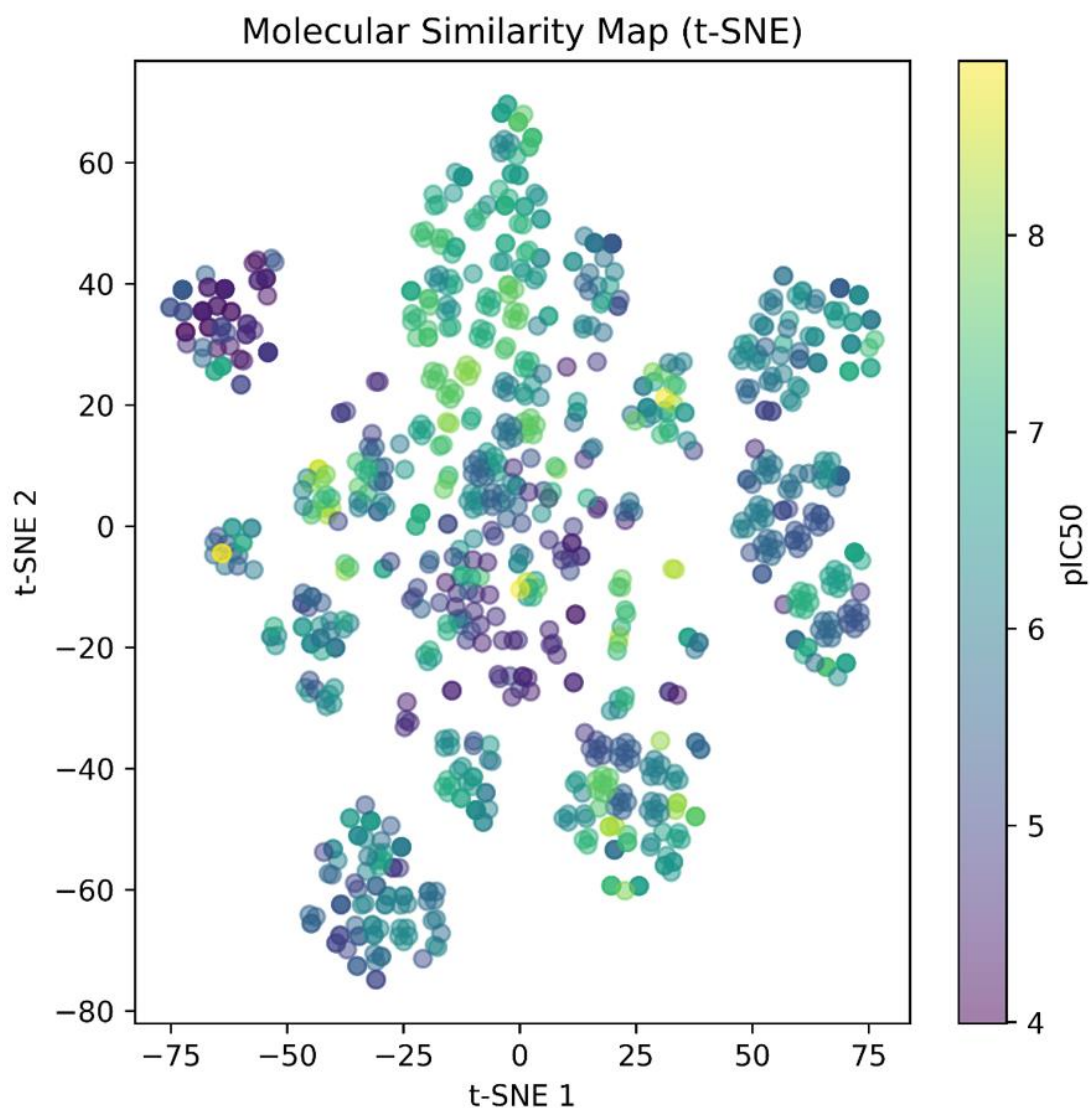


Figure 17.

Molecular similarity map generated using t-distributed stochastic neighbor embedding (t-SNE). Each point represents a compound projected into a 2D space based on structural features, with color indicating pIC₅₀ values. Clustering reflects molecular similarity, revealing structure–activity relationships within the dataset.

The learning curve analysis (Figure 18) evaluates the model’s generalization performance. As the training set size increases, the test RMSE steadily decreases and stabilizes, confirming that the model benefits from additional data without overfitting. The small gap between training and test RMSE further indicates that the model maintains a balance between accuracy and robustness.

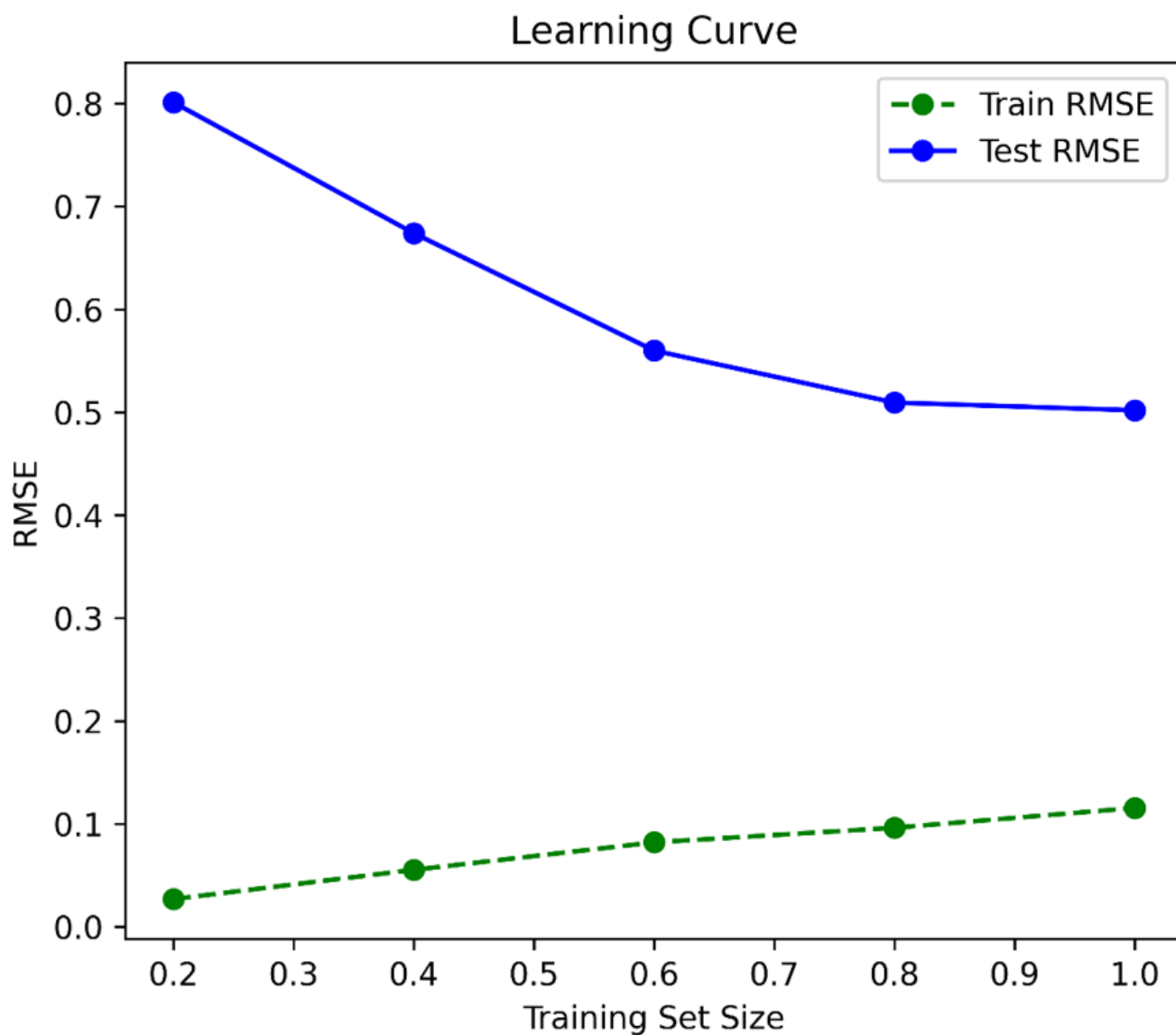


Figure 18.

Learning curve showing the root mean squared error (RMSE) for training and test sets as a function of training set size. The decrease in test RMSE with increasing training data indicates improved model generalization, while the slight increase in train.

The parity plot (Figure 19) compares actual versus predicted pIC₅₀ values, showing that most compounds align closely with the diagonal reference line, indicating strong predictive performance. However, minor deviations exist, particularly in highly active compounds, which suggests some degree of bias in extreme pIC₅₀ values. The residuals plot (Figure 20) supports this observation by displaying a near-zero mean residual distribution, confirming that systematic errors are minimal.

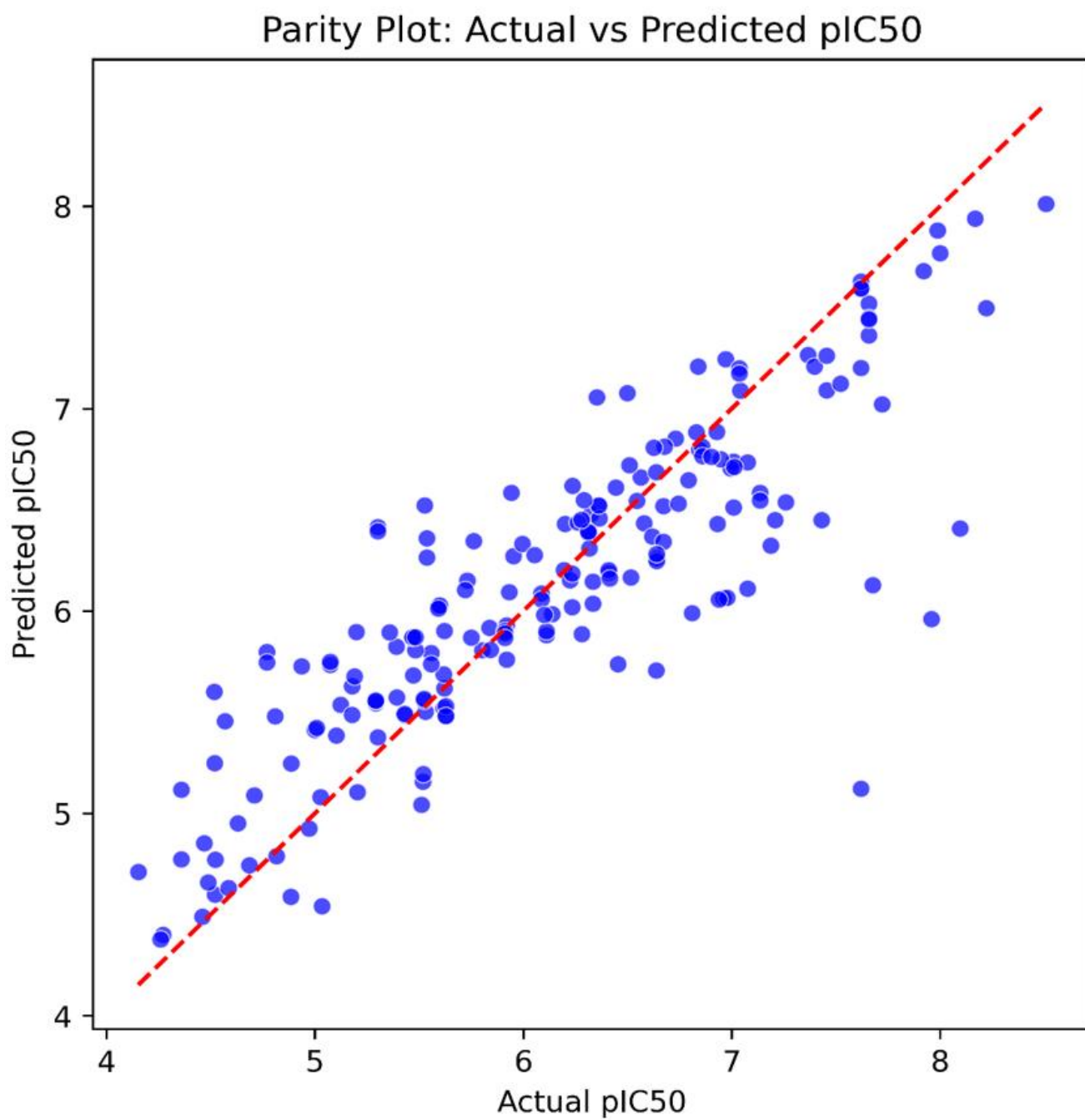


Figure 19.

Parity plot comparing actual versus predicted pIC₅₀ values. Each point represents a compound, with the red dashed line indicating the ideal correlation ($y = x$). The clustering of points around the diagonal suggests good predictive performance of the QSAR model.

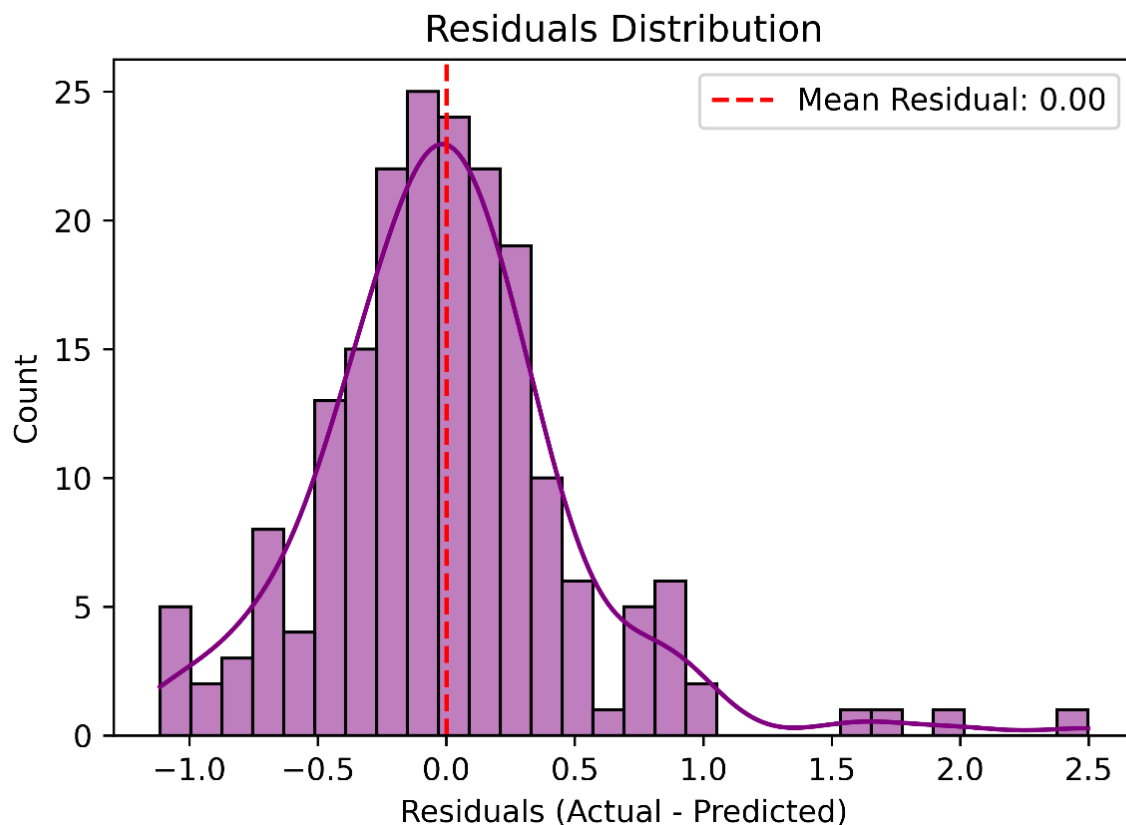


Figure 20.

Distribution of residuals (actual – predicted pIC_{50}) for the QSAR model. The histogram shows a near-normal distribution centered around zero (red dashed line), indicating minimal bias and good model accuracy. The residual spread reflects the variability in prediction error across the dataset.

The Y-scrambling validation further tests the model's reliability. The distribution of R^2 values for shuffled target labels (Figure 21) is significantly lower than the original model's R^2 , demonstrating that the model's predictive ability is not due to random correlations. Similarly, the RMSE distribution for scrambled targets (Figure 22) is significantly higher, reinforcing the validity of the learned structure-activity relationships.

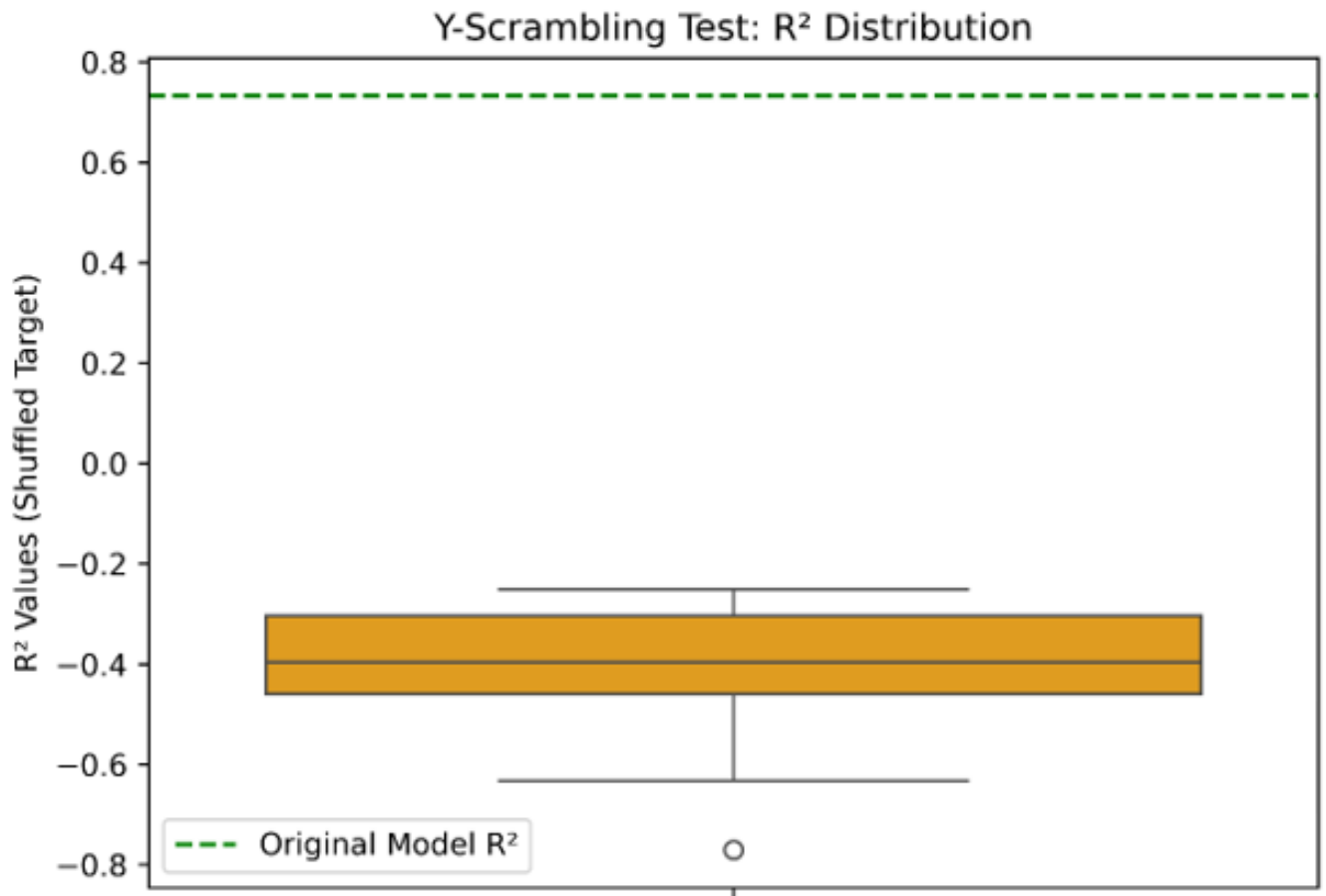


Figure 21. Y-scrambling validation test showing the distribution of R^2 values for models trained on randomly shuffled target values. The boxplot reflects the performance of these random models, while the green dashed line indicates the R^2 of the original (unshuffled) model. The significant difference confirms that the original model's performance is not due to chance, supporting its robustness and reliability.

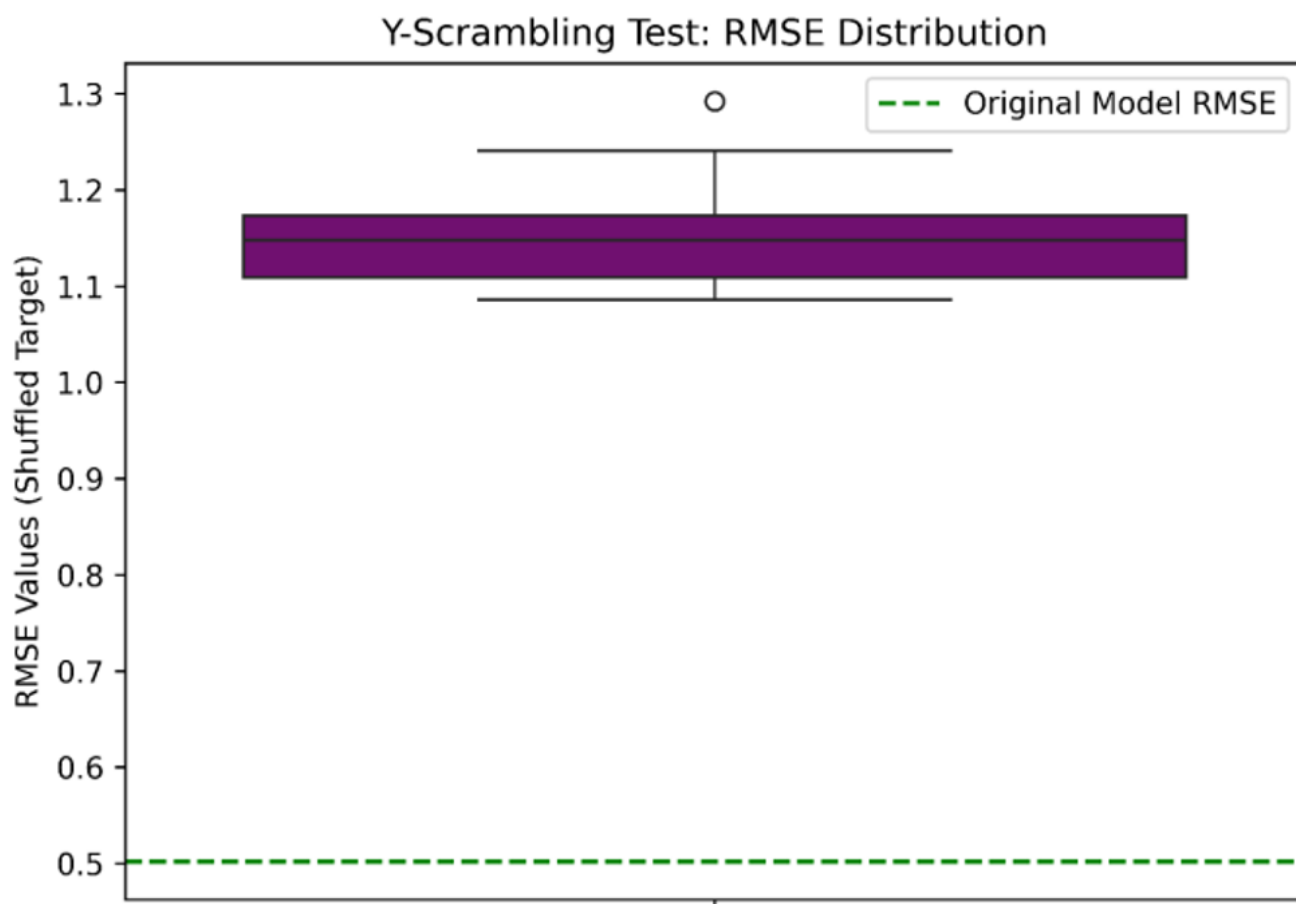


Figure 22.

Y-scrambling validation test showing the distribution of RMSE values for models built on randomly shuffled target data. The boxplot demonstrates the poor predictive performance of these randomized models. The green dashed line indicates the RMSE of the original model, which is significantly lower, confirming that the model's accuracy is not due to chance and supports its robustness.

3.1. Model Predictions for Designed Compounds

The predicted pIC₅₀ values for the designed compounds, summarized in Table 2, highlight the potential of these molecules as estrogen receptor modulators. All designed ligands exhibit higher predicted pIC₅₀ values compared to Raloxifene, indicating improved inhibitory potential.

Table 2.

Model Predictions of pIC₅₀ for the Designed Compounds.

Molecule	Key Group	Predicted pIC ₅₀	Δ (vs Raloxifene)
Raloxifene	Reference	5.3936	0.0000
3b	Cl (Chloro)	5.8097	+0.4161
3a	Phenyl	5.8097	+0.4161
4a	Thiophene	5.7538	+0.3602
3c	OMe (Methoxy)	5.8133	+0.4197
4b	Cl (Thiophene)	5.7538	+0.3602
4c	OMe (Thiophene)	5.7574	+0.3638
3d	NO ₂ (Nitro)	5.8387	+0.4451
4d	NO ₂ (Thiophene)	5.7828	+0.3892
3e	N(CH ₃) ₂ (Dimethylamine)	5.8058	+0.4122
4e	N(CH ₃) ₂ (Thiophene)	5.7499	+0.3563

Among the designed compounds, 3d (NO₂ substitution) has the highest predicted pIC₅₀ value (5.8387), surpassing Raloxifene by +0.4451. This suggests that the nitro group enhances receptor binding, possibly through electronic effects or additional hydrogen bonding interactions. Similarly, 3c (OMe substitution, 5.8133) and 3b (Cl substitution, 5.8097) also exhibit substantial increases in activity, implying that methoxy and chloro groups contribute positively to ligand affinity.

The introduction of a thiophene scaffold (4a–4e) leads to slightly lower pIC₅₀ values than their benzyl counterparts (3a–3e), although they still demonstrate improved activity over Raloxifene. For example, 4a (Thiophene-Phenyl, 5.7538) has a lower predicted pIC₅₀ compared to 3a (Phenyl, 5.8097), suggesting that the benzyl core may be more favorable for

receptor interactions. This trend is also observed in 4c (OMe-Thiophene, 5.7574) and 4e (N(CH₃)₂-Thiophene, 5.7499), indicating that thiophene modifications maintain activity but do not necessarily enhance receptor binding.

Electron-donating groups (OMe, N(CH₃)₂) and electron-withdrawing groups (NO₂, Cl) consistently contribute to increased pIC₅₀ values, reinforcing their importance in modulating receptor interactions. The dimethylamine derivative (3e, 5.8058) exhibits a notable increase over Raloxifene (+0.4122), suggesting that this functional group enhances binding affinity through steric or electronic effects.

Overall, the QSAR model predicts that the designed compounds possess improved estrogen receptor activity compared to Raloxifene. The NO₂, Cl, and OMe substitutions appear particularly promising, while the thiophene core, although effective, does not significantly enhance binding over the benzyl-based derivatives. Future experimental validation, molecular docking, and molecular dynamics simulations will be necessary to confirm these predictions and further refine lead candidates for optimization.

4. Conclusion

This study combined molecular docking, molecular dynamics (MD) simulations, and quantitative structure-activity relationship (QSAR) modeling to evaluate the binding affinity, stability, and predicted activity of novel estrogen receptor modulators. Molecular docking results demonstrated that several designed compounds, particularly 3b, 3a, and 4a, exhibited strong binding affinities comparable to Raloxifene, with key interactions such as pi-pi stacking with PHE-404 and hydrogen bonding with Glu-353 contributing to stability.

MD simulations provided further insights into the structural dynamics of the protein-ligand complexes, showing that 3a, 4a, and 4b maintained stable binding with low RMSD values comparable to Raloxifene, while 3e and 4e exhibited higher fluctuations, suggesting weaker interactions and potential partial dissociation. Root-mean-square fluctuation (RMSF) analysis revealed ligand-induced conformational changes in the binding pocket, with certain compounds engaging previously unoccupied regions, which may offer opportunities for further optimization.

The QSAR model demonstrated strong predictive capability, with 3d (NO₂), 3c (OMe), and 3b (Cl) exhibiting the highest predicted pIC₅₀ values, reinforcing the role of electron-withdrawing and electron-donating groups in enhancing receptor binding. The results also indicated that benzyl-based derivatives generally performed better than their thiophene counterparts, suggesting that the choice of scaffold plays a crucial role in activity modulation.

Overall, this study identifies 3b, 3d, and 4b as the most promising lead compounds due to their strong binding affinity, stability in MD simulations, and high predicted pIC₅₀ values. Further experimental validation and structural refinements will be necessary to optimize their pharmacological properties and enhance their therapeutic potential as estrogen receptor modulators.

References

- [1] F. Gentile, M. A. Deriu, G. Licandro, A. Prunotto, A. Danani, and J. A. Tuszynski, "Structure based modeling of small molecules binding to the TLR7 by atomistic level simulations," *Molecules*, vol. 20, no. 5, pp. 8316-8340, 2015.
- [2] J. Min *et al.*, "Dual-mechanism estrogen receptor inhibitors," *Proceedings of the National Academy of Sciences*, vol. 118, no. 35, p. e2101657118, 2021. <https://doi.org/10.1073/pnas.2101657118>
- [3] E. Gazzillo *et al.*, "Exploring the chemical space of functionalized [1,2,4]triazolo[4,3-a]quinoxaline-based compounds targeting the bromodomain of BRD9," *Bioorganic Chemistry*, vol. 139, p. 106677, 2023. <https://doi.org/10.1016/j.bioorg.2023.106677>
- [4] L. Suigo *et al.*, "Benzodioxane-benzamides as promising inhibitors of Escherichia coli FtsZ," *International Journal of Biological Macromolecules*, vol. 253, p. 126398, 2023.
- [5] Schrödinger, *Maestro, protein preparation wizard, Epik, and OPLS4*. (2023). New York, NY, USA: Schrödinger, LLC, 2023.
- [6] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular docking: A powerful approach for structure-based drug discovery," *Current Computer-Aided Drug Design*, vol. 7, no. 2, pp. 146-157, 2011.
- [7] A. Shirali, V. Stebliankin, U. Karki, J. Shi, P. Chapagain, and G. Narasimhan, "A comprehensive survey of scoring functions for protein docking models," *BMC Bioinformatics*, vol. 26, no. 1, p. 25, 2025. <https://doi.org/10.1186/s12859-024-05991-4>
- [8] Q. Lu *et al.*, "Molecular docking and molecular dynamics studies on the interactions of hydroxylated polybrominated diphenyl ethers to estrogen receptor alpha," *Ecotoxicology and Environmental Safety*, vol. 101, pp. 83-89, 2014. <https://doi.org/10.1016/j.ecoenv.2013.12.018>
- [9] S. A. Hollingsworth and R. O. Dror, "Molecular dynamics simulation for all," *Neuron*, vol. 99, no. 6, pp. 1129-1143, 2018. <https://doi.org/10.1016/j.neuron.2018.08.011>
- [10] M. Hernández-Rodríguez, M. C. Rosales-Hernández, J. E. Mendieta-Wejebe, M. Martínez-Archundia, and J. Correa Basurto, "Current tools and methods in molecular dynamics (MD) simulations for drug design," *Current Medicinal Chemistry*, vol. 23, no. 34, pp. 3909-3924, 2016.
- [11] L. Pinzi and G. Rastelli, "Molecular docking: Shifting paradigms in drug discovery," *International Journal of Molecular Sciences*, vol. 20, no. 18, p. 4331, 2019. <https://doi.org/10.3390/ijms20184331>