Targeting TFF3 in obstructive airway diseases: a computational approach to novel therapeutics

Alireza Shahriary
Baqiyatallah University of Medical Sciences
Mohsen Sisakht
Shiraz University of Medical Sciences
Masoud Arabfard
Baqiyatallah University of Medical Sciences
Esmaeil Behmard
Fasa University of Medical Sciences
Ali Najafi (najafi74@bmsu.ac.ir)
Baqiyatallah University of Medical Sciences

Research Article

Keywords: Trefoil Factor 3 (TFF3), Airway remodeling, Obstructive pulmonary disease, Toxicity Assessment, Genistein

Posted Date: January 31st, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3907985/v1

Additional Declarations: No competing interests reported.

Abstract

Background

Airway remodeling, a hallmark of chronic obstructive pulmonary disease (COPD) and Mustard lung disease, is influenced by the Trefoil Factor 3 (TFF3). This study sought to pinpoint a compound with minimal toxicity that can effectively suppress TFF3 expression and activity.

Methods and Results

We employed an integrative approach, combining gene expression analysis, molecular docking, and molecular dynamics simulations, to identify potential TFF3 inhibitors. The biological safety of these compounds was ascertained using a sophisticated deep neural network model. Of the compounds assessed, eight manifested a significant reduction in TFF3 expression, with binding affinities (ΔG) ranging from − 7 to -9.4 kcal/mol. Notably, Genistein emerged as the frontrunner, showcasing potent TFF3 downregulation, minimal toxicity, and a robust inhibitory profile as evidenced by molecular dynamics simulations.

Conclusion

Genistein holds promise as a therapeutic agent for TFF3-mediated conditions, including mustard lung disease. Its potential to address the current therapeutic gaps is evident, but its clinical utility necessitates further *in vitro* and *in vivo* validation.

1. Introduction

Airway remodeling is a defining feature of respiratory diseases, notably Chronic Obstructive Pulmonary Disease (COPD) [1] and Mustard Lung (ML) [2]. This remodeling, characterized by airway wall thickening, mucous gland hypertrophy, and increased smooth muscle mass, leads to airflow obstruction and compromised respiratory function. The molecular underpinnings of these diseases are complex, with disruptions in specific pathways causing pathological changes in the airway [3, 4]. Recent studies highlight Trefoil Factor 3 (TFF3) as a central figure in airway remodeling and associated pathologies [5, 6]. Elevated TFF3 levels correlate with chronic inflammatory respiratory diseases, including COPD, ML, and various adenocarcinomas [7–10]. Beyond respiratory contexts, TFF3 is implicated in diverse conditions such as Type 2 Diabetes Mellitus (T2DM), Non-Alcoholic Fatty Liver Disease (NAFLD), neurodegeneration, gastric ulcers, colitis, and several malignancies [11, 12]. TFF3, part of the human TFF peptide family, is primarily found in mucosal environments [13]. Its dimer form, more potent than its monomer counterpart, is particularly significant for its anti-apoptotic properties [14, 15]. This dimerization enhances mucus thickness by interacting with soluble mucins like MUC5AC and MUC6, emphasizing its role in mucus overproduction and airway obstruction [13]. All human TFFs, including TFF1, TFF2, and TFF3, are dual-capacity lectins recognizing the GlcNAc-α-1,4-Gal disaccharide, crucial for their mucus-thickening ability [13, 16]. Given TFF3's broad roles, targeted therapeutic strategies are essential. Advanced techniques, including the Library of Integrated Network-Based Cellular Signatures (LINCS) [17], molecular docking, molecular dynamics simulations, and AI algorithms, offer avenues for drug discovery. Our research, leveraging these tools, sought a low-toxicity compound to down-regulate TFF3 expression, targeting TFF3 to inhibit its dimerization. From our analyses, eight compounds emerged, reducing TFF3 expression by at least two-fold, with binding affinities (ΔG values) between − 7 and − 9.4 kcal/mol. Notably, Genistein emerged as a potential TFF3 antagonist, introducing new therapeutic possibilities for obstructive airway diseases.

2. Materials and methods

2.1. TFF3 gene expression analysis

We aimed to identify compounds that suppress TFF3 gene expression, a critical element in airway remodeling observed in diseases such as COPD, asthma, and ML. The LINCS database [17], which captures changes in expression for ~ 12,000 genes across various human cell lines post chemical exposure, served as our primary resource. We focused on compounds that reduced TFF3 expression by more than twofold. All experiments utilized the A549 cell line, a standard *in vitro* model for airway remodeling, derived from human lung carcinoma.

2.2. Preparation of 3D structures

Compound 3D structures were initiated by extracting PubChem IDs (CID) from the LINCS dataset API service [18]. These structures were then sourced from the PubChem database (https://pubchem.ncbi.nlm.nih.gov). Concurrently, the 3D structure of the TFF3 protein was retrieved from the Protein Data Bank (https://www.rcsb.org).

2.3. Phytochemical compound analysis

We sourced 2,845 phytochemical compounds from the PubChem database. Their druglike properties were assessed using the SwissADME service (http://www.swissadme.ch) that operates with the SMILES format.

2.4. Molecular docking analysis

We employed molecular docking to estimate the binding affinity between ligands and the TFF3 protein, specifically at the interface between the two monomers of its dimeric form (1PE3) structure. Preparation involved removing water molecules and the co-crystal ligand. Both TFF3 and compound structures were converted to PDBQT format, with added Gasteiger partial charges [19] for docking analyses. Docking simulations utilized AutoDock Vina [20] on an 192-core processor.

2.5. Toxicity assessment of compounds

Using Python (v3.10), TensorFlow (v2.21), and DeepChem (v2.7.1) [21], we assessed the toxicity of identified compounds. The Tox21 dataset guided the training, validation, and testing of the Graph Convolutional Network (GCN) architecture.

2.5.1. Dataset

Tox21 encompasses 7,831 chemicals and 12 toxicological endpoints. These include five stress responses (SR-ARE, SR-ATAD5, SR-HSE, SR-MMP, SR-p53) and seven nuclear receptor signals (e.g., NR-AR, NR-ER). Chemicals, in SMILES format, have binary toxicity labels. To prevent overfitting, scaffold splitting clustered compounds by molecular fingerprints. DeepChem's butina splitter method partitioned the dataset.

2.5.2. Graph convolutional network

While CNNs excel with Euclidean data, they falter with non-Euclidean data like chemical structures [22, 23]. GCNs, designed for non-Euclidean data [24], represent compounds as graphs: atoms as nodes and bonds as edges. Convolutional and pooling layers extract molecular patterns. DeepChem's GraphConv featurizer prepared chemical features. Training parameters included 7 hidden layer units, a 0.4 dropout rate, a 0.0007 learning rate, and 100 epochs. Given dataset imbalance, ROC-AUC was the primary metric. Training used a Linux OS with GPU support, with SMILES representations as input.

2.5.3. Feature importance in Tox21

We gauged the contribution of each Tox21 task to toxicity predictions using feature importance. Both Random Forest and Permutation methods calculated scores. By contrasting both methods' importances, we assessed each task's AUC impact.

2.6. Comprehensive bio-evaluation through integrated analyses

We amalgamated data from gene expression, molecular docking, and toxicity assessments to pinpoint chemicals that: i) reduce TFF3 gene expression (Z-score < -2), ii) hinder TFF3 dimerization (ΔG < -7), and iii) exhibit low toxicity. While many chemicals affecting TFF3 also influence other genes, certain ones show heightened selectivity. Using the Signature Strength (SS) of the LINCS dataset, we discerned molecules with pronounced specificity to the TFF3 gene. SS is gauged by the tally of genes with an absolute Z-score ≥ 2.

2.7. Molecular dynamics simulation

To determine the optimal binding conformation of the compound with 1PE3, we employed molecular dynamics simulations using GROMACS (v5.1) on a Linux server. Force field parameters were defined using the Charmm27 force field. SwissParam (https://www.swissparam.ch/) generated coordinate and topology files. The complex was solvated in a TIP3P water cubic box, ensuring a 1.0 nm distance from each edge, and neutralized with sodium chloride. Energy minimization and system equilibration were achieved using appropriate algorithms, NVT and NPT ensembles, over 100 ps at 300 Kelvin and 1 atm. Simulations used the Ewald (PME) method and lincs algorithm constraints, running for 100 ns. RMSD assessed protein atom stability. Ligand binding conformation was analyzed using Chimera, and binding energies were computed via the MM-PBSA method.

3. Results

3.1. TFF3 Gene expression modulation by chemicals

From the LINCS database, 15,300 experimental conditions indicated TFF3 downregulation. Of these, 6,001 matched compounds in the PubChem database, fitting our criteria. Focusing on the A549 cell line, 798 experiments were pertinent. These formed our analysis foundation. Table 1 lists the top 10 experiments.

The Z-score in Table 1 quantifies gene expression deviation from a reference group mean:

$$Z = \frac{\mathrm{x} - \mu}{\sigma}$$

Where x is the gene expression, μ is the reference mean, and σ is the standard deviation. In LINCS, the Z-score standardizes gene expression comparisons, minimizing technical and batch effect variations.

Table 1
Leading chemicals down-regulating TFF3 in A549 cells. The dose is presented in μM and the time is indicated in hours (h).

#	Signature ID	CID	Name	Dose	Time	Z-score
1	DOS041_A549_24H:BRD-K36747900-001-01-8:5.01	54638598	BRD-K36747900	5	24	-3.9
2	CPC016_A549_6H:BRD-K82731415-001-05-4:10	4592	Olomoucine	10	6	-3.8
3	PCLB002_A549_24H:BRD-A70731303:10	73707610	Avrainvillamide-analog-5	10	24	-3.8
4	CPC010_A549_24H:BRD-K39462424-050-07-2:10	33036	Dexchlorpheniramine	10	24	-2.9
5	CPC010_A549_6H:BRD-A72180425-001-10-6:10	3689416	K784-3188	10	6	-2.9
6	CPC015_A549_6H:BRD-A84174393-236-03-0:10	10291556	Meloxicam	10	6	-2.8
7	CPC016_A549_6H:BRD-K33396764-001-02-0:10	5280934	Alpha-linolenic-acid	10	6	-2.8
8	CPC015_A549_6H:BRD-A78391468-001-01-0:10	9847023	Prednisolone-hemisuccinate	10	6	-2.7
9	CPC010_A549_24H:BRD-A66435872-050-04-0:10	124846	HTMT	10	24	-2.6
10	DOS042_A549_24H:BRD-K49434056-001-01-0:4.94	54645999	BRD-K49434056	5	24	-2.6

3.2. Phytochemicals as potential drug candidates

Following Lipinski's rule via SWISSADME (http://www.swissadme.ch), 2,223 phytochemicals exhibited drug-like characteristics. Cross-referencing with PubChem IDs identified 101 as genuine phytochemicals. Notably, 13 from this subset were relevant to A549 cell line experiments.

3.3. Compound interactions with TFF3 dimer

Docking studies targeting the TFF3 dimer interface assessed compound binding affinities. The top 10 compounds exhibited ΔG values between − 8.6 to -9.4 (Table 2). A negative ΔG suggests spontaneous, energetically favorable binding, with larger negative values indicating stronger affinities.

Table 2
Leading molecular docking scores for TFF3, expressed as ΔG values (kcal/mol).

#	CID	Name	ΔG
1	124846	HTMT	-9.4
2	118221163	BRD-K26510616	-9
3	54654640	SA-1472514	-9
4	409805	NSC-23766	-8.8
5	44142121	BRD-K11611839	-8.8
6	73707542	BG-1024	-8.7
7	54654197	BRD-K18511213	-8.7
8	15301607	VU-0415012	-8.7
9	73707610	Avrainvillamide-analog-5	-8.6
10	7217941	BRD-K55186349	-8.6

3.4. Model performance

Figure 2 Evaluation of model performance using AUC metrics. (A) PR-AUC curve for each of the 12 toxicological tasks, sorted by their respective AUC values with the highest at the top. The curve illustrates the trade-off between precision and recall for different threshold values. (B) ROC-AUC curve for each task, depicting the true positive rate against the false positive rate. Tasks are ranked by their AUC values, with the most predictive ones at the top. The overall performance of the model across all tasks is represented by the green dotted line.

3.5. Feature importance in Tox21 dataset: Random Forest vs. Permutation

We assessed toxicological endpoint contributions using Random Forest and Permutation feature importance. Both methods highlighted SR-MMP, NR-AhR, and SR-ARE as key contributors. While there was agreement between the techniques, the Permutation method, through value permutation, provided a more robust evaluation. Figure 3 visualizes the relative importance from both methods.

3.6. Compound toxicity assessment

The established model was employed to ascertain the toxicity profiles of compounds across 12 distinct toxicological endpoints. Table 3 showcases the ten foremost compounds exhibiting the most favorable toxicity profiles.

Table 3

Top ten compounds with favorable toxicity profiles. The table enumerates compounds based on their mean toxicity scores, derived from evaluations across 12 distinct toxicological endpoints. The mean represents the average score across all tasks for each chemical.

#	CID	Name	NR-AR	NR-AR-LBD	NR-AhR	NR-Aromatase	NR-ER	NR-ER-LBD	NR-PPAR-gamma	SR-ARE	SR-ATAD5	SR-HSE	SR-MMP	SR-p53	Mean
1	125519	Aminogenistein	0.05	0.15	0.00	0.15	0.07	0.01	0.40	0.02	0.01	0.33	0.01	0.13	0.11
2	230748	INCA-6	0.44	0.07	0.14	0.05	0.06	0.24	0.10	0.01	0.26	0.06	0.02	0.04	0.12
3	68296	3H-1,2-Dithiole-3-thione	0.59	0.03	0.00	0.30	0.49	0.05	0.01	0.02	0.01	0.00	0.00	0.04	0.13
4	2326992	GNF-PF-254	0.36	0.07	0.02	0.05	0.15	0.11	0.24	0.08	0.37	0.05	0.11	0.05	0.14
5	5281707	COUMESTROL	0.04	0.02	0.01	0.42	0.01	0.01	0.72	0.00	0.00	0.58	0.00	0.02	0.15
6	5280443	Apigenin	0.10	0.44	0.04	0.50	0.10	0.04	0.48	0.00	0.03	0.10	0.00	0.01	0.16
7	5035	Raloxifene	0.78	0.08	0.02	0.03	0.00	0.01	0.21	0.02	0.03	0.73	0.01	0.01	0.16
8	222515	Brazilin	0.11	0.08	0.37	0.95	0.03	0.02	0.01	0.04	0.33	0.00	0.01	0.00	0.16
9	824226	Ro 90-7501	0.53	0.13	0.00	0.32	0.06	0.48	0.11	0.00	0.00	0.44	0.01	0.02	0.18
10	5741425	BRD-K30715099-001-01-2	0.58	0.47	0.12	0.08	0.19	0.14	0.31	0.03	0.11	0.15	0.01	0.07	0.19

3.7. Integrative analysis of gene expression, molecular docking, and toxicity

Our comprehensive assessment of compounds on TFF3 gene expression, dimerization inhibition, and toxicity identified Genistein as a prime candidate. As detailed in Table 4, Genistein demonstrated significant TFF3 downregulation (Z-score of -2.02), a favorable binding affinity (ΔG of -7.6) for TFF3 dimerization inhibition, and a low toxicity score (0.19) (see Fig. 3). A graphical representation of the 2D structure of the final compounds from Table 4 is depicted in Fig. 4. Additionally, a 2D image of Genistein and its associated toxicities across the 12 tasks is illustrated in Fig. 5. Given its phytochemical origin, Genistein's potential for reduced side effects underscores its promise for further exploration.

Table 4

Assessment of compounds for TFF3 modulation and inhibition. Summary of compounds' effects on TFF3 gene expression and inhibitory potential, tested on the A549 cell line over 6 or 24-hour durations at varied dosages. Toxicity, averaged from 12 scenarios, ranges from 0 (least toxic) to 1 (most toxic). Genistein, emphasized, is chosen for further study. "Signature Strength" (SS) indicates the specificity of a compound's gene expression impact; "NA" denotes broad gene effects. Gene expression is represented by Z-scores, with dosages in μM and time in hours.

#	CID	Name	Dose	Time	Expression	ΔG	Toxicity	SS
1	**5280961**	**Genistein**	**0.04**	**24**	**-2.02**	**-7.6**	**0.20**	**410**
2	44142059	BRD-K88269385	10	6	-2.17	-7.2	0.58	538
3	73707610	Avrainvillamide-analog-5	10	24	-3.84	-8.6	0.63	652
4	54645999	BRD-K49434056	5	24	-2.63	-7.2	0.75	NA
5	44489706	BRD-K16046246	10	6	-2.14	-8.1	0.78	NA
6	54654234	BRD-K07660364	5	24	-2.12	-7	0.84	NA
7	969516	Curcumin	100	24	-2.58	-8.1	0.85	NA
8	124846	HTMT	10	24	-2.69	-9.4	0.97	NA

3.8. Molecular dynamics insights into ligand-receptor interactions

Through molecular dynamics simulations, we assessed the Genistein-TFF3 interaction dynamics. Binding energies, inclusive of electrostatic and van der Waals forces, were derived using the MM-PBSA approach (Table 5). RMSF analysis (Fig. 6B) highlighted significant fluctuations, indicative of TFF3 dimer dissociation upon Genistein binding. This is further supported by the Gyration trajectory (Fig. 7B), where an increased gyration radius suggests TFF3 monomer separation. Figure 8 offers a 3D depiction of the TFF3 dimer at the simulation's start and end, with Genistein distinctly marked in yellow.

Table 5
Binding energy components of the Genistein-TFF3 complex derived from MM-PBSA calculations. The table presents the electrostatic, van der Waals, and total binding energies computed throughout the molecular dynamic simulation.

Compound	Binding Energy	Electrostatic Energy	Van der Waals Energy
Genistein	-19.62 ± 3.32	-18.05 ± 6.14	-34.26 ± 2.55

4. Discussion

The pivotal role of TFF3 in the pathophysiology of airway remodeling, especially in obstructive airway diseases, has been previously highlighted [6]. This understanding set the stage for our comprehensive quest to pinpoint a compound with the capability to effectively down-regulate TFF3 gene expression, inhibit its function, and simultaneously exhibit minimal toxicity.

Our initial approach was rooted in the gene expression analysis. Utilizing the expansive LINCS database, we meticulously gauged the impact of a myriad of compounds on TFF3 gene expression. This foundational analysis was instrumental in shortlisting candidates for further evaluation. Building on this, we recognized the significance of TFF3's dimerization for its functional stability [13, 25]. The potential to disrupt TFF3's activity by targeting its dimerization interface became a focal point of our investigation. Our molecular docking studies, tailored specifically to TFF3's dimeric form, furnished critical insights into the binding affinities of the shortlisted compounds. This dual-faceted strategy led to the identification of eight compounds that showcased pronounced downregulation of TFF3 expression and also manifested favorable ΔG values, signaling strong binding affinities. To ensure the holistic evaluation of these compounds, it was imperative to assess their biological safety. We embarked on a rigorous toxicity evaluation, harnessing the cutting-edge GCN method within the DeepChem library. This assessment was pivotal in ensuring that the identified compounds, while effective, also adhered to safety benchmarks. Among the evaluated compounds, Genistein emerged as a standout candidate, ranking 13th in terms of low toxicity. Intriguingly, its derivative, amino-genistein, took precedence as the compound with the lowest toxicity (Table 3). This naturally occurring phytochemical not only demonstrated prowess in down-regulating TFF3 expression and inhibiting its dimerization but also showcased minimal toxicity. The inherent nature of Genistein, being a phytochemical, augments its appeal, suggesting a potential for enhanced biocompatibility and a reduced spectrum of side effects compared to synthetic counterparts. Our molecular dynamics simulations further endorsed Genistein's potential, revealing a stable Genistein-TFF3 complex and suggesting Genistein's capability to disrupt TFF3 dimerization.

Beyond our findings, Genistein's broader pharmacological profile is noteworthy. Derived from soybeans, this isoflavone has been extensively researched for its myriad health benefits, ranging from its antioxidant and anti-inflammatory properties to its demonstrated antiproliferative effects on diverse cancer cell lines [26–29]. Additionally, Genistein has been shown to have a potential role in the prevention and treatment of osteoporosis by increasing bone mineral density [30]. Genistein has also been shown to have potential benefits in the treatment of cardiovascular disease, diabetes, and neurodegenerative disorders [31–33]. Furthermore, research suggests that Genistein may have a positive impact in preventing obesity and enhancing insulin sensitivity [34, 35]. Additionally, Genistein has been found to have potential therapeutic effects in the treatment of liver injury and inflammation [36, 37]. The exact mechanisms by which Genistein exerts its biological effects are still under investigation, but it is thought to modulate the activity of various signaling pathways, including tyrosine kinase, NF-κB, Nrf2, and oxidative stress [38–41]. These findings suggest that Genistein may also have potential therapeutic benefits in the treatment of obstructive airway diseases such as ML, COPD, and asthma. In fact, studies have demonstrated that Genistein can inhibit the activity of various inflammatory pathways associated with these diseases. For example, one study found that Genistein was able to inhibit the NF-κB, TNF-α, and MMP-9-associated pathways in lymphocytes from patients with COPD, leading to a reduction in inflammatory markers [42]. Similarly, in an in vivo guinea pig model of asthma, Genistein was shown to significantly reduce airway hyperresponsiveness, inflammation, and eosinophilia [43]. Additionally, in a murine model of asthma, soy isoflavones, of which Genistein is a major constituent, were found to reduce airway hyperresponsiveness, eosinophil counts, and inflammatory biomarkers, as well as inhibiting lung tissue eosinophil infiltration, airway mucus production, and collagen deposition [44]. So, by inhibiting TFF3, the main regulator in airway remodeling, Genistein may have potential therapeutic benefits in the treatment of obstructive airway diseases such as COPD, asthma, and ML. However, further studies are needed to confirm its efficacy and safety in human subjects.

In conclusion, our multidimensional investigation, encompassing gene expression analysis, molecular docking, toxicity assessment, and molecular dynamics simulations, has illuminated the potential of Genistein as a promising therapeutic candidate for obstructive airway diseases. By targeting the pivotal TFF3 molecule, Genistein not only showcases the ability to down-regulate its expression but also to inhibit its dimerization, a key aspect of its functionality.

Declarations

Acknowledgements

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of interest

The authors have stated explicitly that there is no conflict of interest in connection with this article.

Author contributions

Alireza Shahriary: Led project administration and workflow design; contributed to drafting of the original manuscript. Mohsen Sisakht: Responsible for software development and methodology; co-authored the original manuscript. Masoud Arabfard: Key in project conceptualization and major contributor to the original manuscript. Esmaeil Behmard: Assisted in project conceptualization and co-authored the original manuscript. Ali Najafi: Central to project conceptualization; pivotal in manuscript preparation.

Declarations

Ethical Approval

Not applicable. This study did not involve human participants, human data or human tissue, nor did it involve the use of animals.

Consent to Participate

Not applicable. This study did not involve human participants.

Consent to Publish

Not applicable. This study did not involve human participants.

Funding

Not applicable. No funding was received for conducting this study.

Data Availability Statement

The data supporting the findings of this study are available within the article. The datasets analyzed during the current study are publicly available: the LINCS dataset can be accessed at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92742, and the Tox21 dataset can be accessed at https://github.com/deepchem/deepchem/tree/master/datasets.

References

1. Hirota N, Martin JG (2013) Mechanisms of airway remodeling. Chest 144:1026–1032
2. Shahriary A, Seyedzadeh MH, Ahmadi A, Salimian J (2015) The footprint of TGF-β in airway remodeling of the mustard lung. Inhalation Toxicol 27:745–753
3. Doherty T, Broide D (2007) Cytokines and growth factors in airway remodeling in asthma. Curr Opin Immunol 19:676–680
4. Wang Y, Xu J, Meng Y, Adcock IM, Yao X (2018) Role of inflammatory cells in airway remodeling in COPD. Int J Chronic Obstr Pulm Dis. 3341–3348
5. Marzouni ET, Dorcheh SP, Nejad-Moghaddam A, Ghanei M, Goodarzi H, Hosseini SE et al (2020) Adipose-derived mesenchymal stem cells ameliorate lung epithelial injury through mitigating of oxidative stress in mustard lung. Regen Med 15:1861–1876
6. Najafi A, Masoudi-Nejad A, Ghanei M, Nourani M-R, Moeini A (2014) Pathway reconstruction of airway remodeling in chronic lung diseases: a systems biology approach. PLoS ONE. 9, e100094
7. Weste J, Houben T, Harder S, Schlüter H, Lücke E, Schreiber J et al (2022) Different Molecular Forms of TFF3 in the Human Respiratory Tract: Heterodimerization with IgG Fc Binding Protein (FCGBP) and Proteolytic Cleavage in Bronchial Secretions. Int J Mol Sci 23:15359
8. Viby NE, Pedersen L, Lund TK, Kissow H, Backer V, Nexø E et al (2015) Trefoil factor peptides in serum and sputum from subjects with asthma and COPD. The clinical respiratory journal. 9:322–329
9. Mihalj M, Bujak M, Butković J, Zubčić Ž, Tolušić Levak M, Čes J et al (2019) Differential Expression of TFF1 and TFF3 in Patients Suffering from Chronic Rhinosinusitis with Nasal Polyposis. Int J Mol Sci. 20
10. Doubková M, Karpisek M, Mazoch J, Skrickova J, Doubek M (2016) Prognostic significance of surfactant protein A, surfactant protein D, Clara cell protein 16, S100 protein, trefoil factor 3, and prostatic secretory protein 94 in idiopathic pulmonary fibrosis, sarcoidosis, and chronic pulmonary obstructive disease. Sarcoidosis Vasc Diffuse Lung Dis 33:224–234
11. Yang Y, Lin Z, Lin Q, Bei W, Guo J (2022) Pathological and therapeutic roles of bioactive peptide trefoil factor 3 in diverse diseases: recent progress and perspective. Cell Death Dis 13:62
12. Xue Y, Shen L, Cui Y, Zhang H, Chen Q, Cui A et al (2013) Tff3, as a novel peptide, regulates hepatic glucose metabolism. PLoS ONE. 8, e75240
13. Järvå MA, Lingford JP, John A, Soler NM, Scott NE, Goddard-Borger ED (2020) Trefoil factors share a lectin activity that defines their role in mucus. Nat Commun 11:2265
14. Poulsen SS, Kissow H, Hare K, Hartmann B, Thim L (2005) Luminal and parenteral TFF2 and TFF3 dimer and monomer in two models of experimental colitis in the rat. Regul Pept 126:163–171
15. Kinoshita K, Taupin DR, Itoh H, Podolsky DK (2000) Distinct pathways of cell migration and antiapoptotic response to epithelial injury: structure-function analysis of human intestinal trefoil factor. Mol Cell Biol 20:4680–4690
16. Matsuda K, Yamauchi K, Matsumoto T, Sano K, Yamaoka Y, Ota H (2008) Quantitative analysis of the effect of Helicobacter pylori on the expressions of SOX2, CDX2, MUC2, MUC5AC, MUC6, TFF1, TFF2, and TFF3 mRNAs in human gastric carcinoma cells. Scand J Gastroenterol 43:25–33

17. Keenan AB, Jenkins SL, Jagodnik KM, Koplev S, He E, Torre D et al (2018) The library of integrated network-based cellular signatures NIH program: system-level cataloging of human cells response to perturbations. Cell Syst 6:13–24
18. Pilarczyk M, Fazel-Najafabadi M, Kouril M, Shamsaei B, Vasiliauskas J, Niu W et al (2022) Connecting omics signatures and revealing biological mechanisms with iLINCS. Nat Commun 13:4678
19. Morris G, Huey R, Olson A (2008) Using AutoDock for Ligand-Receptor Docking. Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Malden, MA, p 14
20. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455–461
21. Ramsundar B (2018) Molecular machine learning with DeepChem. Stanford University
22. Rao B, Zhang L, Zhang G (2020) Acp-gcn: the identification of anticancer peptides based on graph convolution networks. IEEE Access 8:176005–176011
23. Altae-Tran H, Ramsundar B, Pappu AS, Pande V (2017) Low data drug discovery with one-shot learning. ACS Cent Sci 3:283–293
24. Li G, Muller M, Thabet A, Ghanem B, Deepgcns (2019) : Can gcns go as deep as cnns? In: Proceedings of the IEEE/CVF international conference on computer vision, pp. 9267-76
25. Thim L, Madsen F, Poulsen S (2002) Effect of trefoil factors on the viscoelastic properties of mucus gels. Eur J Clin Invest 32:519–527
26. Sohel M, Biswas P, Al Amin M, Hossain MA, Sultana H, Dey D et al (2022) Genistein, a potential phytochemical against breast cancer treatment-insight into the molecular mechanisms. Processes 10:415
27. Yu X, Yan J, Li Y, Chen J, Zheng L, Fu T et al (2023) Inhibition of castration-resistant prostate cancer growth by genistein through suppression of AKR1C3. Food & Nutrition Research
28. Alorda-Clara M, Torrens-Mas M, Morla-Barcelo PM, Roca P, Sastre-Serra J, Pons DG et al (2022) High Concentrations of Genistein Decrease Cell Viability Depending on Oxidative Stress and Inflammation in Colon Cancer Cell Lines. Int J Mol Sci 23:7526
29. Zhao L, Wang Y, Liu J, Wang K, Guo X, Ji B et al (2016) Protective effects of genistein and puerarin against chronic alcohol-induced liver injury in mice via antioxidant, anti-inflammatory, and anti-apoptotic mechanisms. J Agric Food Chem 64:7291–7297
30. Kushwaha AC, Mohanbhai SJ, Sardoiwala MN, Jaganathan M, Karmakar S, Choudhury R (2022) Nanoemulsified genistein and vitamin D mediated epigenetic regulation to inhibit osteoporosis. ACS Biomaterials Science & Engineering 8:3810–3818
31. Li R, Robinson M, Ding X, Geetha T, Al-Nakkash L, Broderick TL et al (2022) Genistein: A focus on several neurodegenerative diseases. J Food Biochem. 46, e14155
32. Li R, Ding X, Geetha T, Fadamiro M, Aubin S, Shim CR et al (2022) M.,. Effects of Genistein and Exercise Training on Brain Damage Induced by a High-Fat High-Sucrose Diet in Female C57BL/6 Mice. Oxidative Medicine and Cellular Longevity. 2022
33. Rasheed S, Rehman K, Shahid M, Suhail S, Akash (2022) M.S.H. Therapeutic potentials of genistein: New insights and perspectives. J Food Biochem. 46, e14228
34. Gan M, Chen X, Chen Z, Chen L, Zhang S, Zhao Y et al (2022) Genistein Alleviates High-Fat Diet-Induced Obesity by Inhibiting the Process of Gluconeogenesis in Mice. Nutrients 14:1551
35. Azgomi RND, Jazani AM, Karimi A, Pourreza S (2022) Potential roles of genistein in polycystic ovary syndrome: A comprehensive systematic review. Eur J Pharmacol. 175275
36. Ding Q, Pi A, Hao L, Xu T, Zhu Q, Shu L et al (2023) Genistein Protects against Acetaldehyde-Induced Oxidative Stress and Hepatocyte Injury in Chronic Alcohol-Fed Mice. Journal of Agricultural and Food Chemistry
37. Goh YX, Jalil J, Lam KW, Husain K, Premakumar CM (2022) Genistein: a review on its anti-inflammatory properties. Front Pharmacol 13:820969
38. Li Y, Zhang J-J, Chen R-J, Chen L, Chen S, Yang X-F et al (2022) Genistein mitigates oxidative stress and inflammation by regulating Nrf2/HO-1 and NF-κB signaling pathways in hypoxic-ischemic brain damage in neonatal mice. Annals of Translational Medicine. 10
39. Jafari A, Esmaeilzadeh Z, Khezri MR, Ghasemnejad-Berenji H, Pashapour S, Sadeghpour S et al (2022) An overview of possible pivotal mechanisms of Genistein as a potential phytochemical against SARS-CoV-2 infection: A hypothesis. J Food Biochem. 46, e14345
40. Li Y, Ou S, Liu Q, Gan L, Zhang L, Wang Y et al (2022) Genistein improves mitochondrial function and inflammatory in rats with diabetic nephropathy via inhibiting MAPK/NF-κB pathway. Acta Cirúrgica Brasileira. 37
41. Shirvanian K, Vali R, Farkhondeh T, Abderam A, Aschner M, Samarghandian S (2023) Genistein Effects on Various Human Disorders Mediated via Nrf2 Signaling. Current Molecular Medicine
42. Liu XJ, Bao HR, Zeng XL, Wei JM (2016) Effects of resveratrol and genistein on nuclear factor–κB, tumor necrosis factor–α and matrix metalloproteinase–9 in patients with chronic obstructive pulmonary disease. Mol Med Rep 13:4266–4272
43. Duan W, Kuo IC, Selvarajan S, Chua KY, Bay BH, Wong WF (2003) Antiinflammatory effects of genistein, a tyrosine kinase inhibitor, on a guinea pig model of asthma. Am J Respir Crit Care Med 167:185–192
44. Bao Z-S, Hong L, Guan Y, Dong X-W, Zheng H-S, Tan G-L et al (2011) Inhibition of airway inflammation, hyperresponsiveness and remodeling by soy isoflavone in a murine model of allergic asthma. Int Immunopharmacol 11:899–906

Figures

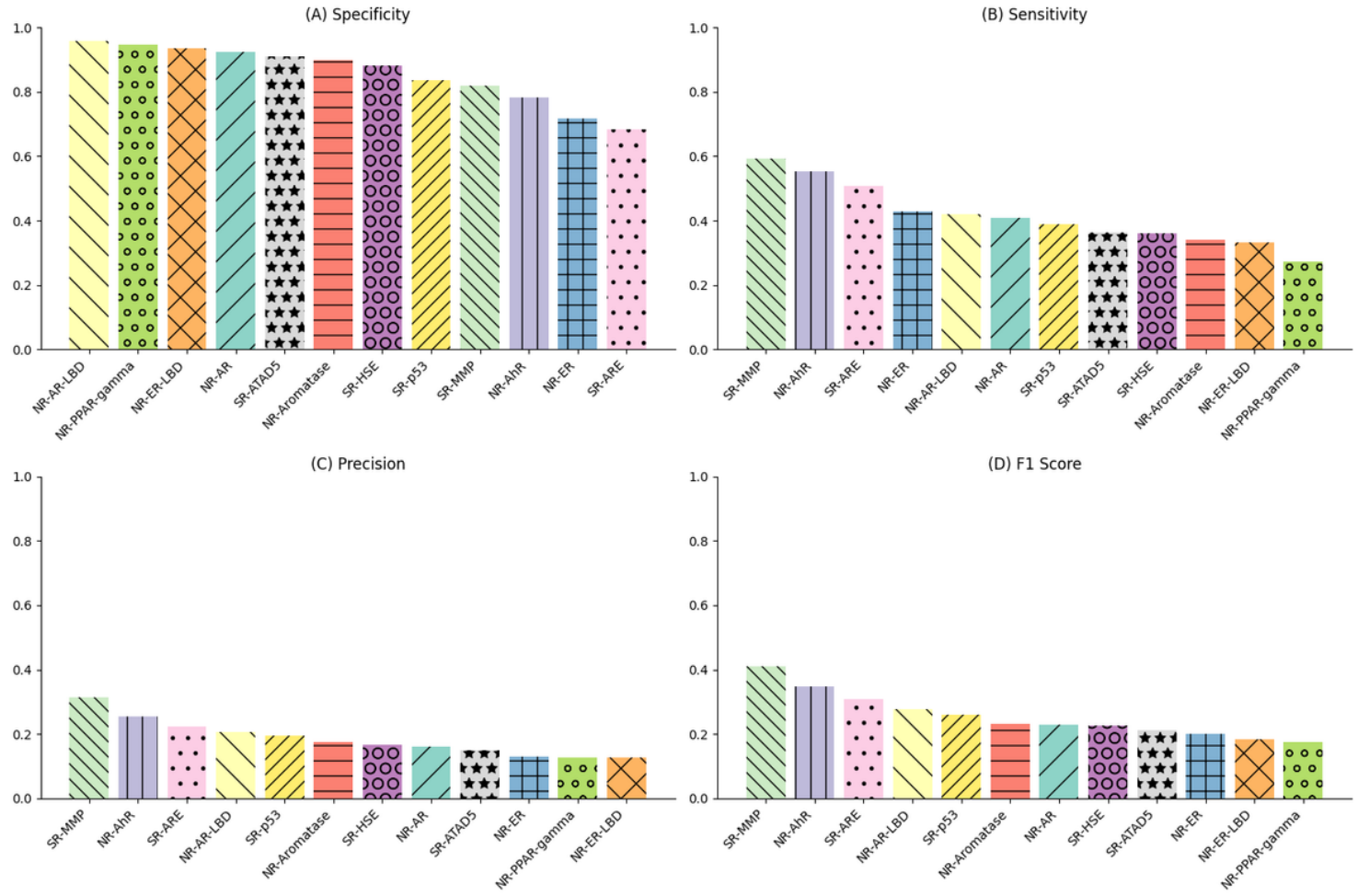

Figure 1

Model Performance metrics for predicted toxicities. (A) Specificity of the model across different toxicities, indicating the true negative rate. (B) Sensitivity, showcasing the true positive rate for each toxicity type. (C) Precision, representing the proportion of true positive predictions among all positive predictions. (D) F1 Score, a harmonic mean of precision and sensitivity, providing a balanced measure of the model's accuracy for each toxicity.

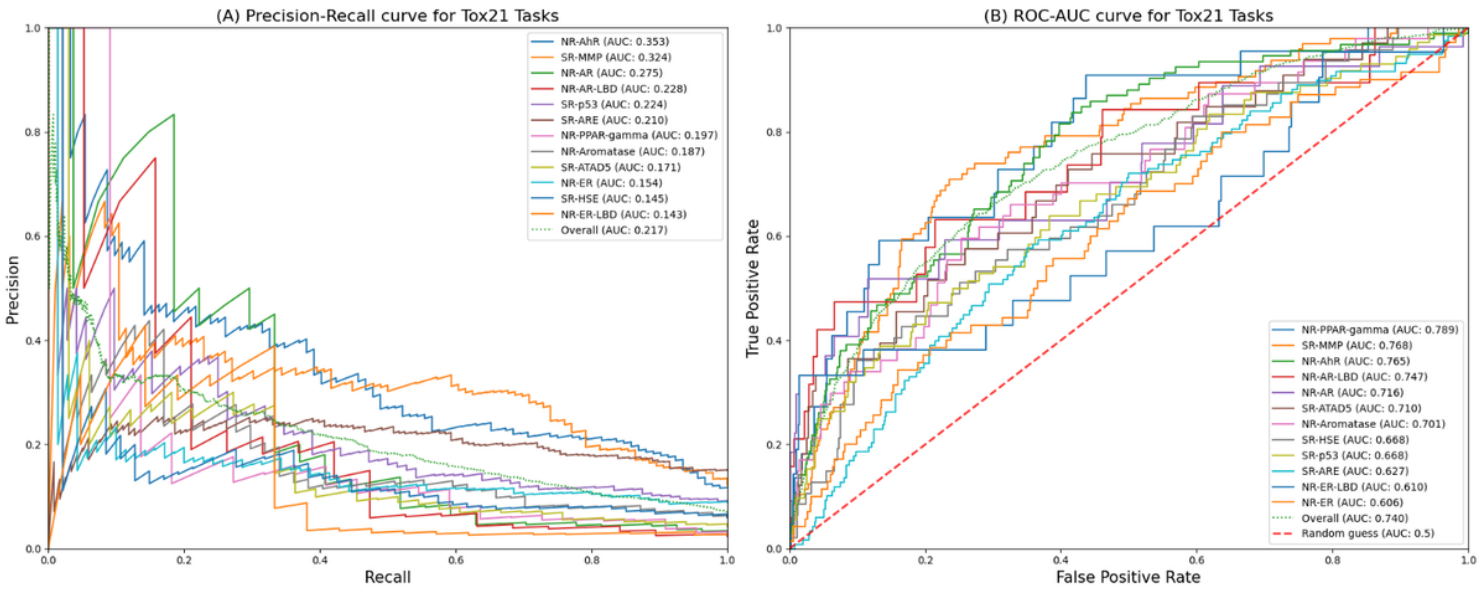

Figure 2

Evaluation of model performance using AUC metrics. (A) PR-AUC curve for each of the 12 toxicological tasks, sorted by their respective AUC values with the highest at the top. The curve illustrates the trade-off between precision and recall for different threshold values. (B) ROC-AUC curve for each task, depicting the true positive rate against the false positive rate. Tasks are ranked by their AUC values, with the most predictive ones at the top. The overall performance of the model across all tasks is represented by the green dotted line.

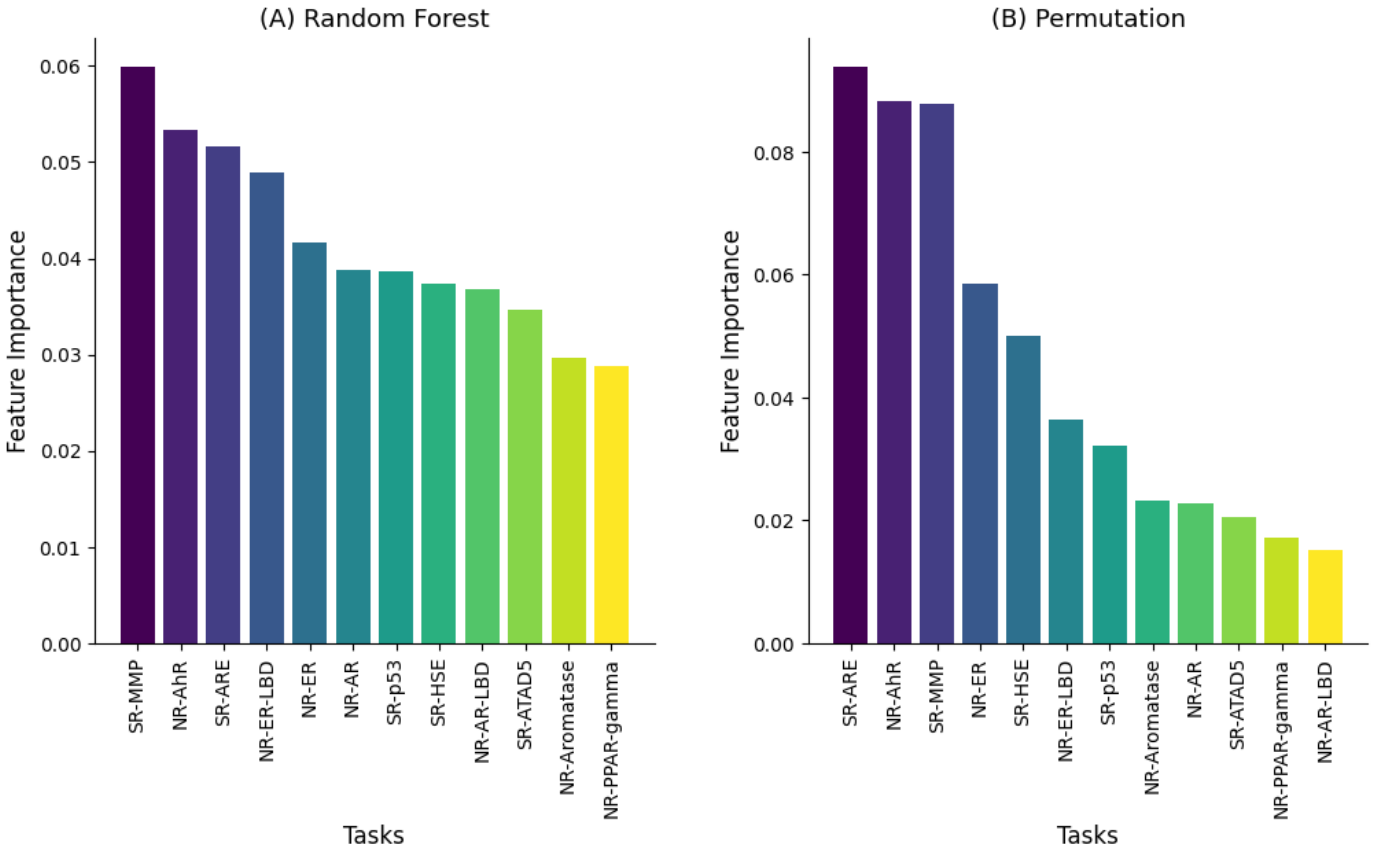

Figure 3

Feature importance assessment. A juxtaposition of feature importance derived from Random Forest (A) and Permutation (B) methods across multiple tasks. Predominant features for each task are denoted by the most elevated bars.

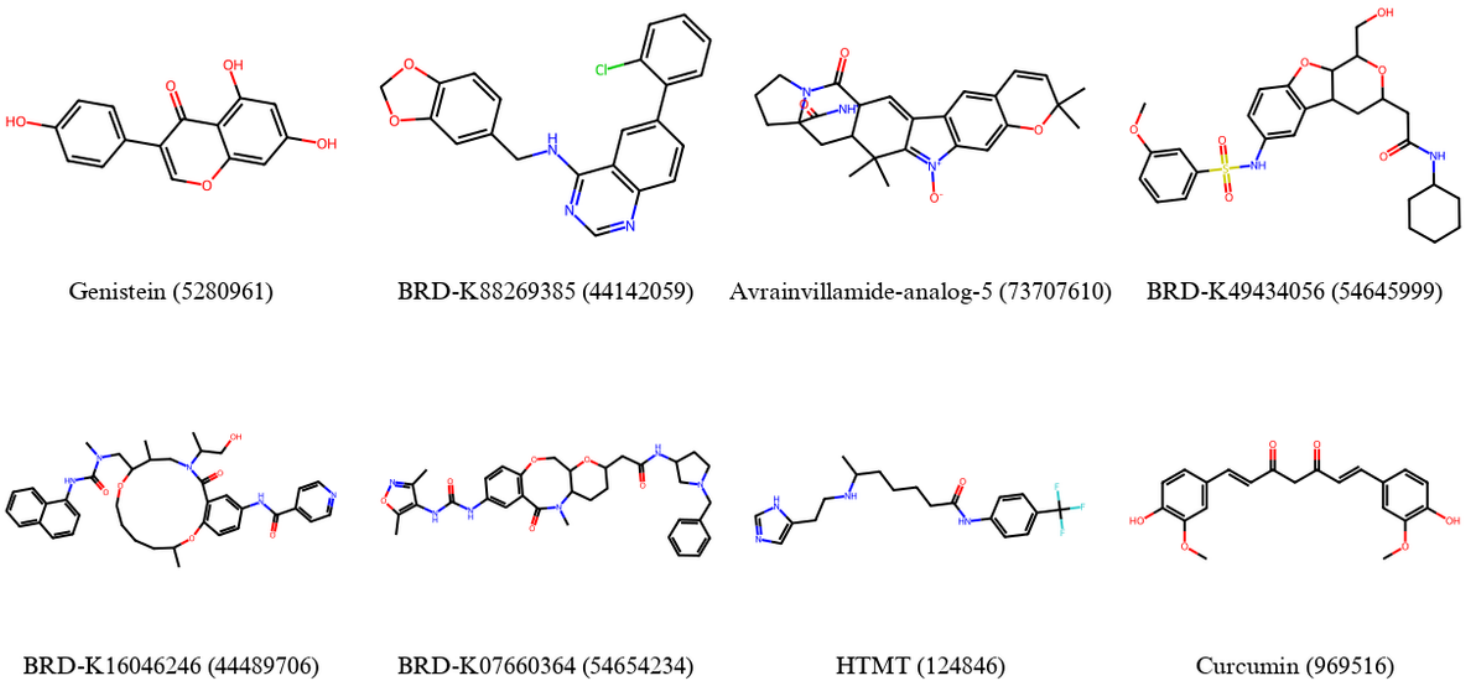

Figure 4

2D structural representations of the final compounds from Table 4. Each structure visually depicts the molecular arrangement with reactive atoms distinctly colored. The compounds are presented in the order based on their toxicity as detailed in Table 4.

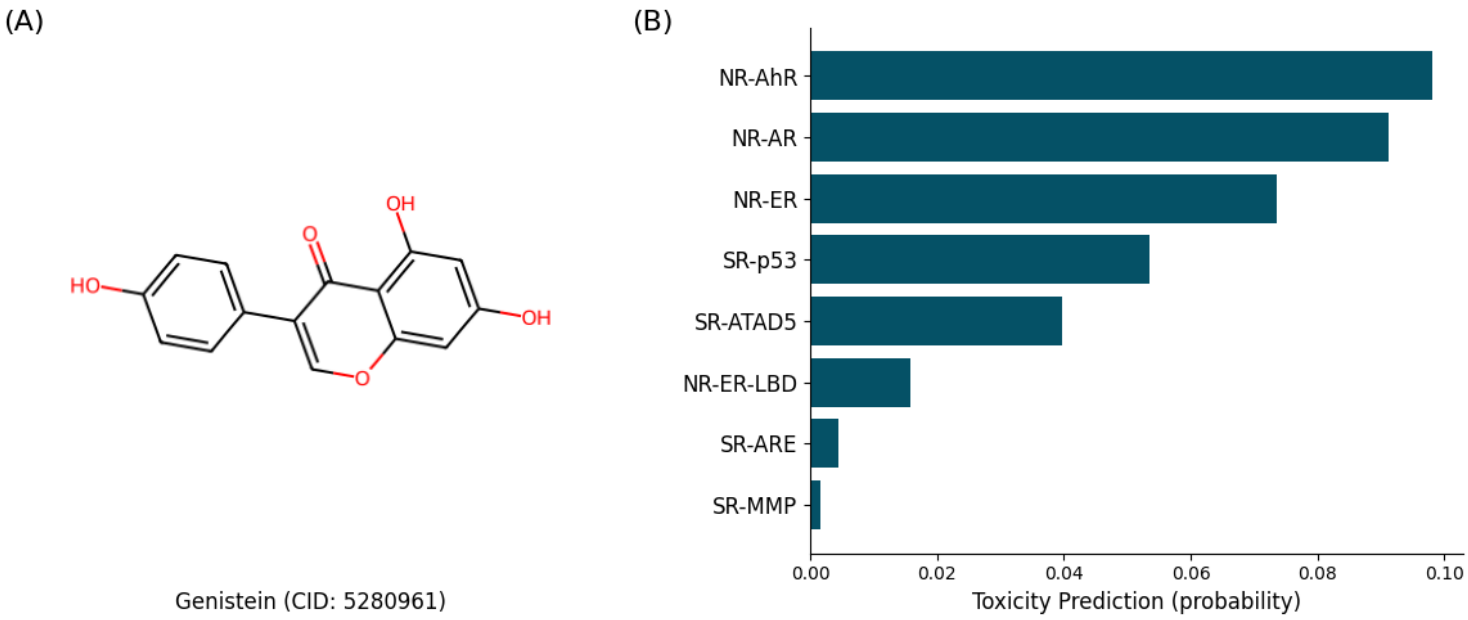

Figure 5

Molecular structure and predicted toxicity of the Genistein compound (CID: 5280961). (A) Graphical representation of Genistein's molecular structure with reactive atoms distinctly colored. (B) Bar chart illustrating the predicted toxicity levels of Genistein across various biological tasks.

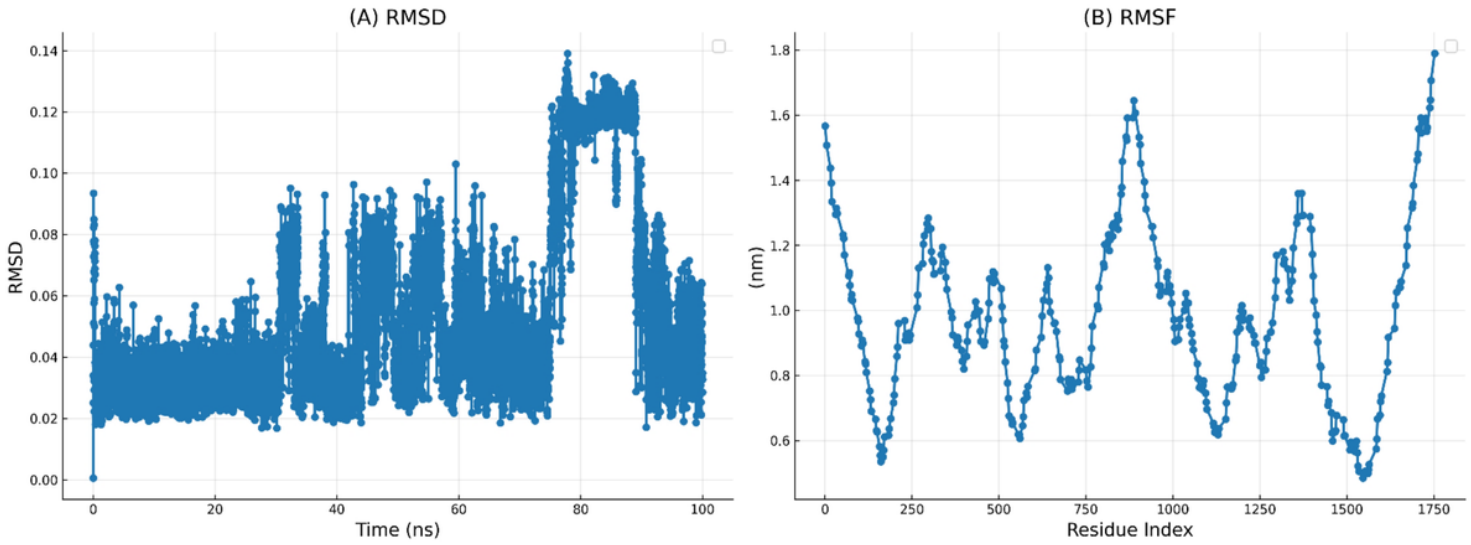

Figure 6

Dynamics and Stability of the Genistein-TFF3 Complex. (A) Root Mean Square Deviation (RMSD) trajectory of the protein backbone atoms, reflecting the temporal stability of the Genistein-TFF3 complex. (B) Root Mean Square Fluctuation (RMSF) of individual protein residues, showcasing residue-specific flexibility within the complex. Notable fluctuations in the RMSF plot hint at Genistein's potential role in modulating TFF3 dimerization.

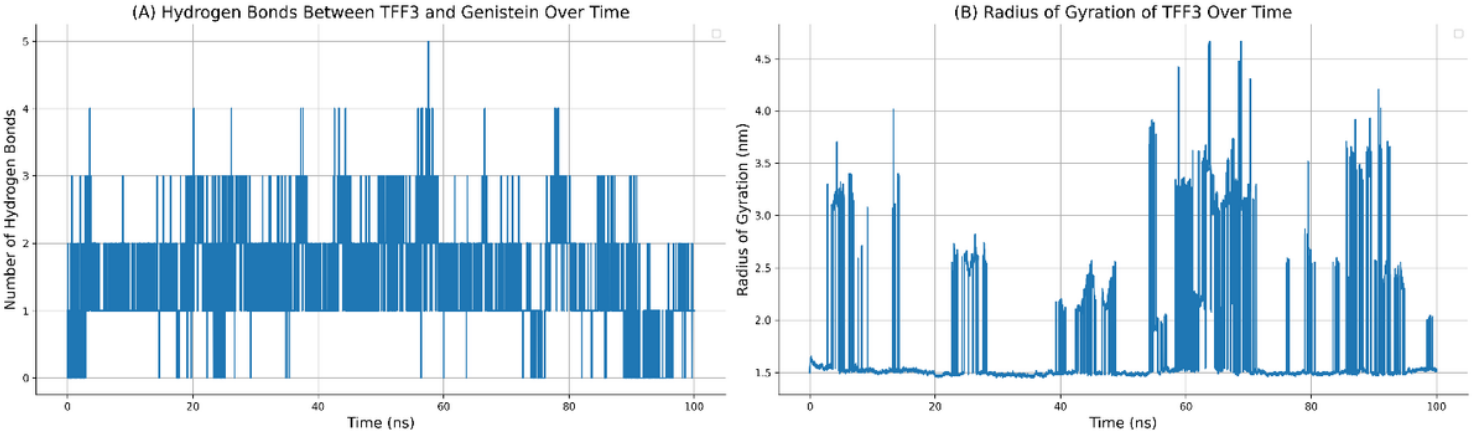

Figure 7

Interaction Dynamics and Structural Changes in the Genistein-TFF3 Complex. (A) Temporal evolution of hydrogen bonds between TFF3 and Genistein throughout the 100 ns simulation, highlighting the stability of their interaction. (B) Radius of Gyration trajectory for TFF3, with an observed increase suggesting the dissociation of TFF3 monomers upon Genistein binding.

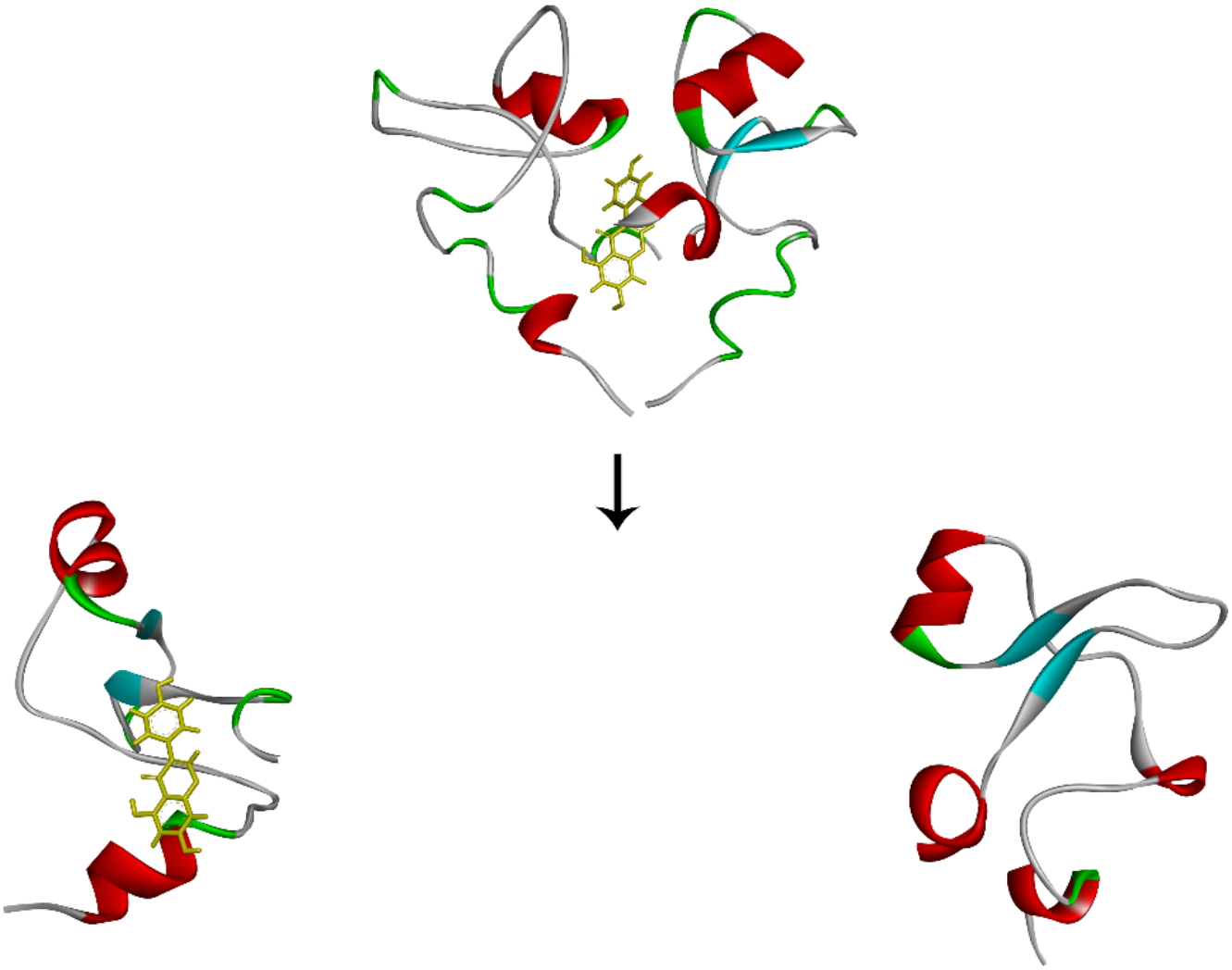

Figure 8

Three-dimensional visualization of the Genistein-TFF3 complex over simulation time. Depicted are two distinct frames from the molecular dynamics simulation: The initial frame (up), capturing the TFF3 dimer in the presence of Genistein, and the concluding frame (down), illustrating the dissociation of TFF3 monomers subsequent to Genistein interaction. Genistein is delineated in yellow, underscoring its instrumental role in mediating the observed dimeric separation.