Heliyon 10 (2024) e29756

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Development and experimental validation of 3D QSAR models for the screening of thyroid peroxidase inhibitors using integrated methods of computational chemistry

Bharath Basavapattana Rudresh *, Abhishek Kumar Tater, Vaibav Barot, Nitin Patel, Ashita Desai, Sreerupa Mitra, Abhay Deshpande

Jai Research Foundation, Valvada, Vapi, Gujarat, 396105, India

ARTICLE INFO

Keywords:
Endocrine disruptors
TPO
Homology modelling
MD simulation
Molecular docking
QSAR
In vitro rat microsomal assay

ABSTRACT

The intricate network of glands and organs that makes up the endocrine system. Hormones are used to regulate and synchronize the nervous and physiological systems. The agents which perturbate an endocrine system are called endocrine disruptors and they can eventually affect cellular proliferation and differentiation in target tissues. A subclass of endocrine disruptors known as thyroid disruptors (TDs) or thyroid disrupting chemicals (TDCs) influence the hypothalamo-pituitary-thyroid axis or directly interfere with thyroid function by binding to thyroid hormone receptors. Thyroid hormone levels in circulation are now included in more test guidelines (OECD TG 441, 407, 408, 414, 421/422, 443/416). Although these might be adequate to recognize thyroid adversity, they are unable to explain the underlying mechanism of action. Thyroid peroxidase (TPO) and sodium iodide symporter (NIS), two proteins essential in the biosynthesis of thyroid hormones, are well-accepted molecular targets for inhibition. The screening of a large number of molecules using high throughput screening (HTS) requires a minimum quantity of sample, cost, and time consuming. Whereas 3-dimensional quantitative structure-activity relationship (3D-QSAR) analysis can screen the TDCs before synthesizing a compound. In the present study, the human TPO (hTPO) and NIS (hNIS) structures were modelled using homology modeling and the quality of the structures was validated satisfactorily using MD simulation for 100ns. Further, 190 human TPO inhibitors with IC_{50} were curated from Comptox and docked with the modelled structure of TPO using D238, H239 and D240 centric grid. The binding conformation of a molecule with low binding energy was used as a reference and the rest other molecules were aligned after generating the possible conformers. The activity-stratified partition was performed for aligned molecules and training set (139), test set (51) were defined. The machine learning models such as k Nearest Neighbor (kNN) and Random Forest (RF) models were built and validated using external experimental dataset containing 10 molecules. Among the 10 molecules, all 10 molecules were identified as TPO inhibitors and demonstrated 100 % accuracy qualitatively. To confirm the selective TPO inhibition all 10 molecules were docked with the modelled structure of hNIS and the results have demonstrated the selective TPO inhibition.

* Corresponding author.
E-mail address: bharath.rudresh@jrfonline.com (B.B. Rudresh).

https://doi.org/10.1016/j.heliyon.2024.e29756
Received 7 November 2023; Received in revised form 13 April 2024; Accepted 15 April 2024
Available online 16 April 2024

1. Introduction

The endocrine system is a network of secretory glands that secrete hormones and facilitate cellular communication [1]. These hormones are responsible for regulating almost every molecular, cellular and organ function in the body [2]. Hormonal imbalances can disrupt the endocrine system, impairing internal energy levels, metabolism, growth, development, reproduction, and response to injury, stress, and environmental variables [3,4]. There are more than twenty hormones that regulate the endocrine system, and their biosynthesis is regulated by specific enzymes [5]. One of these hormones is thyroid, which plays a key role in metabolism as well as in the growth, development, and activity of the nervous system. Three variables control the biosynthesis and metabolism of thyroid hormone (TH): iodine availability, deiodinases activity, and stimulation induced by thyroid stimulating hormone (TSH) [6]. Iodination of tyrosyl residues on the precursor protein thyroglobulin is another step in the production of TH. High quantities of H_2O_2 and oxidized iodine produced by the enzymes thyroid peroxidase (TPO) and thyroid oxidase (ThOX) 1 and 2 are used in this process [7,8].

The Na+/I− symporter (NIS) mediates active iodide (I−) transport in the thyroid and some extrathyroidal tissues. Thyroid hormone (TH) generation depends critically on NIS-mediated I− absorption [9]. As a result, blocking NIS can impact TPO activity and reduce the amount of iodine available in the follicular colloid [10]. Similarly, there are various reasons which can lead to thyroid dysregulation; nevertheless, the primary causes of this condition are pollutants and xenobiotic interference, which also happens to be the primary TD. Thyroid dysregulation has been linked recently to unfavourable outcomes including cancer, obesity, diabetes, and neurodevelopmental issues. It can also result in diseases like hyper- and hypothyroidism.

The xenobiotic may exert its effects through many mechanisms due to the potential severity of chemical disruption of thyroid homeostasis. Thus, a technique for the screening of thyroid-active chemicals must be developed, particularly in the early stages of drug development. Alternative techniques are being developed since screening xenobiotic libraries in entire animal models is costly and time-consuming. In 2012, a new initiative on the creation of Adverse Outcome Pathways (AOPs) was also introduced by the Organization for Economic Co-operation and Development (OECD). An AOP delineates the progressive sequence of interconnected key events (KEs) across several biological organization levels, commencing with a molecular initiating event (MIE), progressing through many downstream linked KEs, and culminating in an unfavourable impact on health or the environment. In one of these AOPs, the role of TPO inhibition is discussed, which can result in hypothyroxinemia, aberrant neurodevelopment, and neurological impairment in the progeny [11]. This encourages the creation of systems for efficiently evaluating xenobiotics for their ability to inhibit TPO.

Currently multiple *in vitro* assays like high-throughput screening (HTPS) assays to screen TPO inhibitors, the catalyst for TH synthesis, LC-MS based method for the characterization of distinct multiple catalytic activities of TPO are reported [12,13]. However, for the screening of large libraries of novel molecules, it is necessary to synthesize and purify the molecules. Hence, the *in silico* approaches can screen the molecules before synthesis and significantly reduce the cost and time of synthesis. Alongside, the impurity traces with active molecules and formulations can also be screened using *in silico* approaches to save the cost and time on their synthesis [14]. However, the 3D-QSAR can offer better accuracy in predicting the molecules for TPO inhibition activity.

The *in silico* approaches such as molecular docking, molecular dynamics (MD) simulations and QSAR can be used to virtually screen the molecules before synthesis. In this regard, Rosenberg et al. have built a 2D-QSAR model for TPO using TPO inhibition data on 1126 molecules available in ToxCast data. Predictions based on this model have shown that a tiered approach should be used to prioritize potential thyroid modulating chemical for further study [15].

Our knowledge on functional, structural, and antigenic properties of TPO is advanced by the validation of the tenable structural model of a TPO-autoantibody complex [16]. The structure of the extracellular domain of hTPO elucidated by cryo-electron microscopy is described and to be reported to guide the screening of TPO new inhibitors [17]. Since the crystal structure of hTPO is not elucidated, and no structural information is available in protein data bank, the homology modelling of hTPO was performed using Protein Data Bank (PDB) entries 1CXP, 1EMO, and 1VVD as templates [18]. This has provided a detailed insight on enzyme function and autoantigenicity. Habza-Kowalska et al. have also modelled the TPO structure using the PDB entry 1CXP as a template and docked the quercetin, rutin, rosmarinic acid and chlorogenic acid to know their TPO inhibitory potential [14]. However, the 3D-QSAR can offer better accuracy in predicting the molecules for TPO inhibition activity.

In that concern, the current study was aimed at modelling the hTPO structure, validating the model quality by MD simulation and developing 3D QSAR models to screen the potential hTPO inhibitors. So that, the molecules can be screened for the TPO inhibition potential before synthesis and the cost and time of synthesizing the TDCs. Along the side, researchers can also model the hTPO and hNIS structures, using the method proposed in the study and develop new and more robust QSAR models in future.

2. Materials and methods

2.1. Homology modelling of hTPO and hNIS structures

The homology modelling was carried out using SWISS-MODEL for the protein sequences of hTPO (Uniprot ID: P07202) and hNIS (Uniprot ID: Q92911) with 933aa and 643aa length respectively. The crystal structure of myeloperoxidase (PDB ID: 5UZU) and the structure of the rat sodium/iodide symporter (NIS) in complex with iodide and sodium (PDB ID: 7UV0) were used as templates to model hTPO and hNIS, respectively.

2.2. MD simulations

The modelled hTPO and hNIS structures were validated for quality using a Ramachandran plot and Molecular Dynamics (MD)

simulations. The modelled structures were prepared, and molecular dynamics was studied using the Cresset Flare Pro Plus software. The modelled structures were individually solvated with TIP3P solvent system (water molecules) in an orthorhombic box allowing for a 10 Å buffer region between protein atoms and box sides. The temperature of the system was set to 300 K and the total simulation was run up to 120ns using AMBER force field with GAFF2 version. Data was collected every 100ps during the simulation. 3D structures and trajectories were visualized and inspected using the Cresset Flare visualizer [19,20].

2.3. Dataset preparation

A total of 554, containing 466 active and 88 inactive hTPO inhibitors sharing diverse molecular skeletons were collected from the Comptox database with the bioactivities determined using Amplex UltraRed assay [11,21]. The IC_{50} values of these inhibitors distributed in a gradient from 0.001 to 74.60 μM and they were converted to pIC_{50} ($\log_{10} IC_{50}$) values to build the 3D-QSAR models.

2.4. Molecular docking

The modelled TPO structure was prepared using the protein preparation module available in Cresset Flare for structure-based drug design. During the protein preparation process, the corresponding hydrogens were added at a pH of 7.0, considering the respective ionization states of the acidic and base amino acid residues. The residue gaps were filled and the atoms from residues with incomplete backbone atoms were removed. The active site size was set to 6.00 Å and the prepared protein was copied for molecular docking studies. As per the uniport annotation for Human TPO, the residues Asp238, His239 and Asp240 were considered as the markers of hTPO active site. Further, the residues Met68, Gln72, Val76, Phe87, Trp255, Val293 and Phe417 were identified as the markers of hNIS sodium binding site. The active site was defined with a 10 Å radius around the selected residues and a grid box was generated at the centroid of the active site for docking. To test the docking parameters low energy conformations of the molecule (1H-1,2,4-Triazole-3-amine) with the best IC_{50} 1.76 μM was docked into the active pocket of the modelled TPO using the very accurate but slow method with the extra precision quality available in cresset flare module. Parallelly, the 10 molecules were identified as external dataset to experimentally validate the reliability of QSAR models and they were also docked to hNIS iodine binding site to confirm the selective inhibition of hTPO using the same method and module used for hTPO. The maximum docked conformations were 10, and the final best docked conformation was selected using binding energy (ΔG value). Finally, the lowest energy conformation of amitrole was used as a reference to align the ligands in the dataset.

2.5. Confirmation hunt, and ligand alignment

Ligand alignment is the critical process which determines the quality of the 3D-QSAR model significantly. In the current study, ligand-based alignment was employed to align the ligands in the dataset using the docked conformation of amitrole as the reference molecule. The alignment was performed using Cresset Flare by applying the eXtended Electron Distribution (XED) force field to calculate the field points. To obtain accurate conformations for the generation of 3D-QSAR model, a very accurate and slow calculation method was chosen for the conformation hunt.

2.6. QSAR modeling

The quality of the dataset determines the quality of the QSAR models. The data partitioning method can be implemented by describing the split ratio and the commonly used ratio is 80:20. In practice, other ratios are also used, such as 70: 30, 60: 40, and even 50: 50. There is no specific guidance as to which ratio is best for a particular dataset [22]. Because it is determined based on the relevance and diversity of the data in the dataset. However, the 80: 20 partition is justified by the well-known Pareto principle, but again this is only a rule of thumb used by practitioners. Hence, in the present study, considering the diversity of molecules and activity distribution, the dataset was subjected to activity stratified partition, where 70 % of aligned inhibitors were assigned to the training set and 30 % were assigned to the test set. The regression models like Random Forest (RF), k-nearest neighbor (kNN), Support vector machine (SVM) and field-based field-QSAR models were built for the screening of test samples. The external dataset containing 10 test samples was also screened against the designed model for model validation.

2.7. Reagent and chemicals

L-Tyrosine disodium salt hydrate (CAS 69847-45-6) was obtained from the Tokyo chemical industry, Japan. 3-Iodo-L-tyrosine (MIT, CAS 70-78-0), L- Tyrosine (CAS 60-18-4, Reagent grade), 3,5-Diiodo-L-tyrosine dehydrate (DIT, CAS 18835-59-1), Oxalic acid (CAS 144–62-7, Analytical Reagent grade), Dimethyl sulfoxide (DMSO, CAS 67-68-5, Laboratory Reagent grade), Protease Inhibitor cocktail (Product code P8340), Sodium thiosulfate (CAS 7772-98-7), Potter Elvehjem polytetrafluoroethylene (PTFE) assembly, Cell dissociation sieve - tissue grinder kit (Product code CD1), and Ethylenediaminetetraacetic acid (EDTA, CAS 6381-92-6, Molecular Biology grade) were procured from Sigma-Aldrich, India. Potassium Iodide (KI, CAS 7681-11-0, Excellent Analytical Reagent grade) and Hydrogen peroxide solution (H_2O_2, 30 % w/w, CAS 7722-84-1, Super Quality grade) were obtained from Qualigens, India. Glycerol (Glycerin) Anhydrous (CAS 56-81-5, Analytical Reagent grade) was got from Srlchem. Dulbecco's phosphate buffered Saline (DPBS, 1X, Ref 14191-144) was received from Gibco. Sucrose (Ref RM 3063, Analytical Reagent grade) was purchased from HIMEDIA. Formic acid (98–102 %, CAS 64-18-6, HPLC grade) was supplied by Merck. Acetonitrile (CAS 75-05-8, ULC/MS grade) was taken from

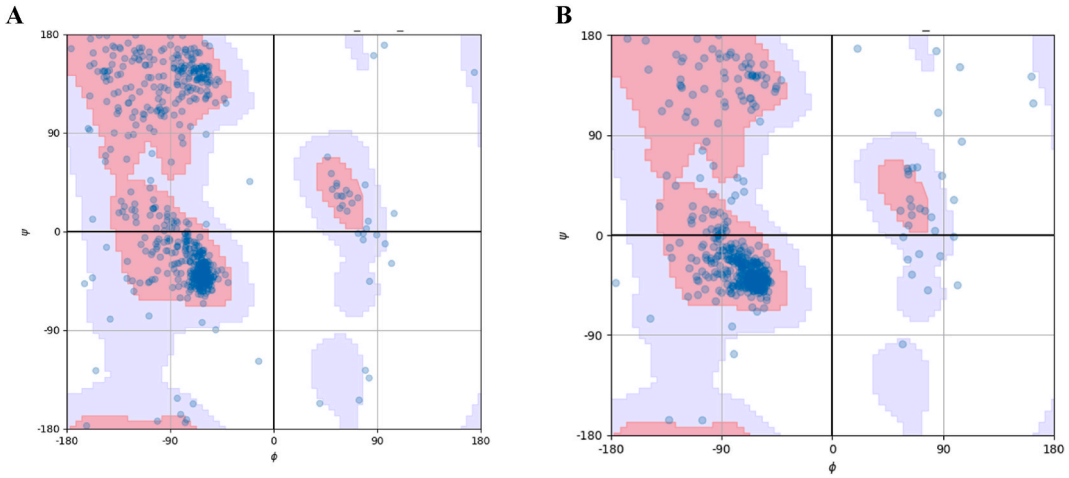


Fig. 1. Ramachandran plot for secondary structure assessment of modelled proteins. **A:** Plot for hTPO with 1.2 % of residues falling in non-allowed region (White) and 98.8 % of residues falling in allowed (gray) and strongly allowed region (Pink). **B:** Plot for hNIS with 1.1 % of residues falling in non-allowed region and 98.9 % of residues falling in allowed and strongly allowed region.

Biosolve. Bicinchoninic acid (BCA) kit for protein estimation was procured from Pierce. Water (Milli-Q, Type-2) was taken from Milli-Q Advantage A10 Merck Laboratory water purification system.

2.8. Isolation of rat thyroid microsomes

The use of rats for thyroid isolation was adopted after obtaining approval from the institutional animal ethics committee (JRF/IAEC/2022/365, Vapi, Gujarat, India). 12-week-old male Wistar rats were perfused with 0.9 % saline before thyroid isolation. The collected thyroid gland was stored at −80 °C. Thyroid glands from 20 rats were pooled for microsome isolation [23]. Pooled thyroid glands were thawed on ice and minced using a sterile surgical blade in DPBS-EDTA containing protease inhibitor cocktail, then homogenized using a Potter Elvehjem PTFE kit. About 3 mL of DPBS-EDTA was used to homogenize the thyroid gland of 20 animals. A cell dissociation sieve-cloth grinder was used to grind and separate the larger remaining fragments. Further, the homogenate was passed through a 23G needle 15 to 20 times and to this homogenate suspension, an equal amount of sucrose (50 %) was added and centrifuged at 600×*g* for 5 min at 4 °C. To the supernatant, an equal volume of sterile water was added. The 250 μL of homogenized tissue suspension was divided equally into multiple tubes and centrifuged at 20,000×*g* for 2 h at 4 °C. The microsomal pellet was pooled and re-suspended in 1 mL of 5 % glycerol-DPBS. Aliquots of microsomes were stored at −80 °C until further use. Microsomal protein content was assessed using a BCA kit.

2.9. Thyroperoxidase inhibition assay using rat thyroid microsomes

Thyroid microsomes were used in the experiment, which was performed in 250 μL DPBS. For every batch of microsomes, the concentrations of KI, H2O2, and the microsome protein were tuned. L-Tyr was used as the starting point for each of the ten concentrations ranging from 0 to 330 μM of Kaempferol, Resveratrol, Acetaminophen, Phenylthiourea, Hexestrol, 4-Benzyloxyphenol, Epigallocathechin gallate, Gallic acid, Curcumin, and Tetrahydrocurcumin in the IC_{50} reaction assays. The concentrations of KI, thyroid microsomes, and H_2O_2 were also included to assess each compound's ability to inhibit the conversion of L-Tyr to MIT. When the concentration of KI was raised to 300 μM, we observed a dose-dependent increase in TPO activity. As a result, the KI concentration was maintained at 300 μM for all subsequent tests. The assay mixture was pre-incubated with microsomes and all other reagents (apart from H_2O_2) for a minimum of 10 min. For a 250 μL assay volume, the ideal range for H_2O_2 and protein concentrations was 40 μM and 10 μg, respectively. The reactions were terminated by adding 1000 μL of acetonitrile containing internal standard (IS) and 50 μL of ice-cold oxalic acid (750 mM). The tubes underwent spinning and centrifugation for 5 min at 8000 rpm and 5 °C. Fresh glass shell vials with a capacity of 1 mL auto-sampler were filled with the supernatant. After loading the samples into an auto-sampler tray, 10 μL of the sample was injected into the LC-MS/MS.

2.10. Chromatography and LC-MS/MS conditions

Analytes were separated chromatographically utilizing an Exion-LC Sciex HPLC system outfitted with an X-bridge-C18 column having 3.0 × 150 mm (i.d. × L) specification and 3.5 μm particle size. The oven was kept at 40 °C. Each sample was run for a total of 4 min at a steady flow rate of 0.550 mL/min. The gradient of the mobile phase was as follows: A: Acetonitrile B: 0.1 % formic acid (v/v) from 85 % B for 0.1 min. After that, the gradient was changed to 5 % B for 1 min, kept at that level for 3 min, and then returned to 85 % B for 4 min.

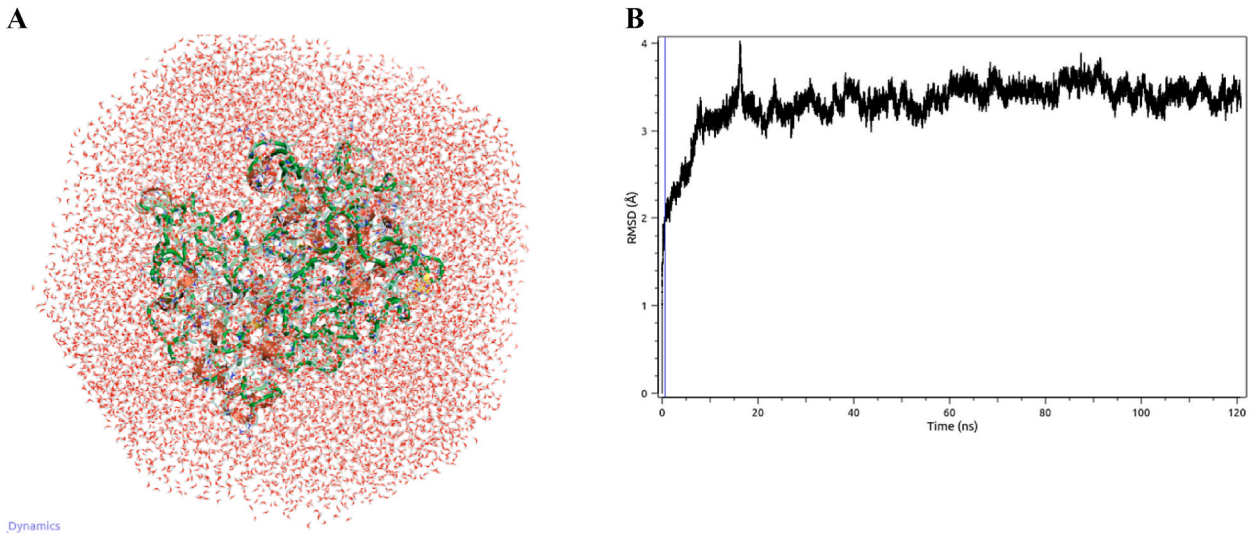


Fig. 2. MD simulation to assess the model quality at tertiary level. **A:** The modelled hTPO solvated in TIP3P water system. **B:** The RMSD plot for modelled hTPO simulated for 120ns.

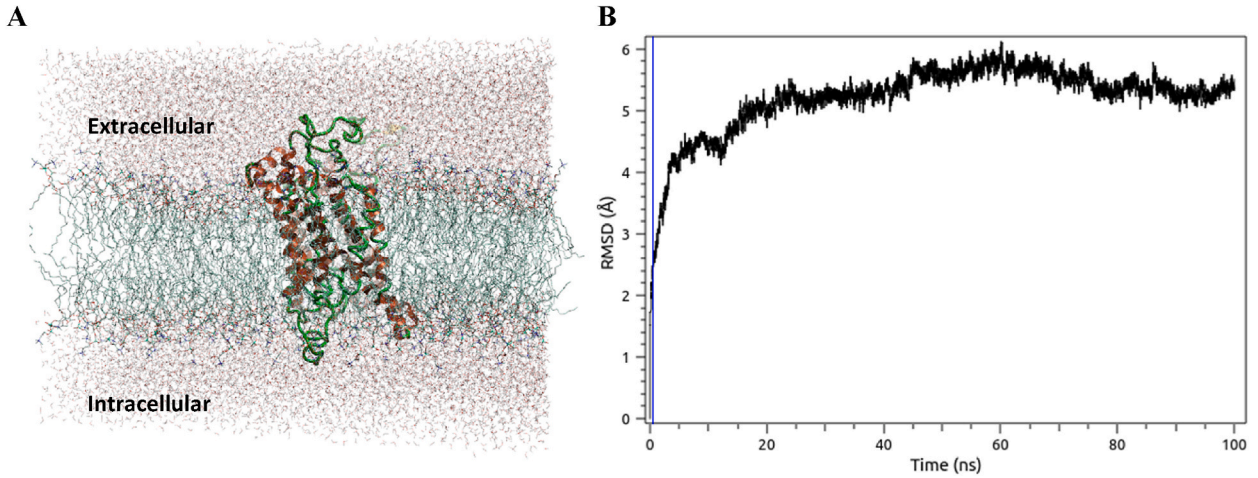


Fig. 3. MD simulation to assess the model quality at tertiary level. **A:** The modelled hNIS solvated in TIP3P water and POPC membrane system. **B:** The RMSD plot for modelled hNIS simulated for 100ns.

For mass spectrometric analysis, AB Sciex's QTRAP 5500 LC-MS/MS (linear ion trap triple quadrupole) was utilized. To detect MIT and IS, rolipram, electrospray ionization (ESI) method was applied in positive polarity employing multiple reaction monitoring (MRM scan) [19]. In the Q1 quadrupole, the mass transition acquired for MIT quantification was recognized and measured as 308 Da for the parent ion, and 262 Da for the fragment mass in the Q3 quadrupole. Similarly, the mass transition was found to be 276/208.3 Da in Q1/Q3 as parent/fragment mass ions, and Rolipram was chosen as the IS for quantifying MIT. Ion source temperature was kept at 550 °C, ion spray voltage at 5500 V, and pressures of 65 and 55 psi for sheath and auxiliary gases, respectively. The CAD gas was adjusted to medium, and curtain gas was utilized at 35 psi. The collision exit potential (CXP) was 15 V, while the entrance potential (EP) was 10 V.

3. Results and discussion

3.1. Human TPO protein structure modelling and validation

The structure for hTPO was modelled using the crystal structure of myeloperoxidase (PDB ID: 5UZU) as a template. The template was 48.10 % identical to hTPO with 61 % coverage between 148 and 735th residue. Similarly, the structure for hNIS was modelled using the crystal structure of the sodium/iodide symporter (NIS) in complex with iodide and sodium (PDB ID: 7UV0) as a template. The

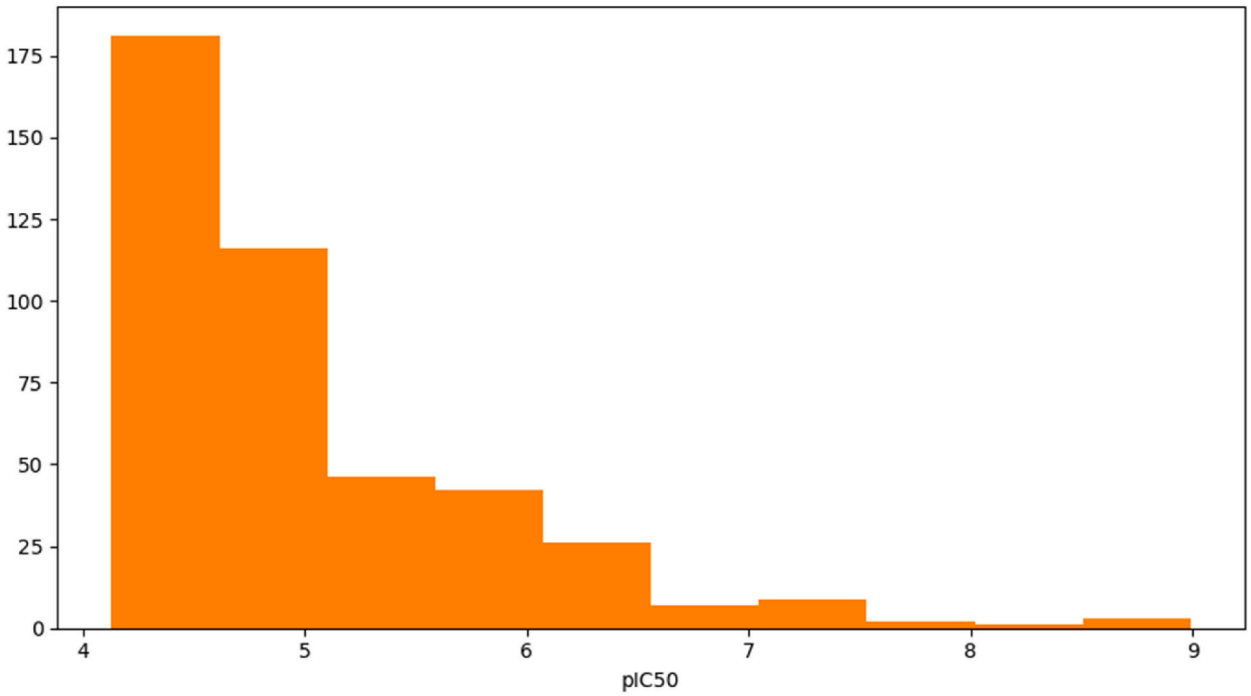


Fig. 4. Histogram showing the pIC50 distribution of hTPO inhibitors in dataset, the 30 % of dataset was assigned to test set.

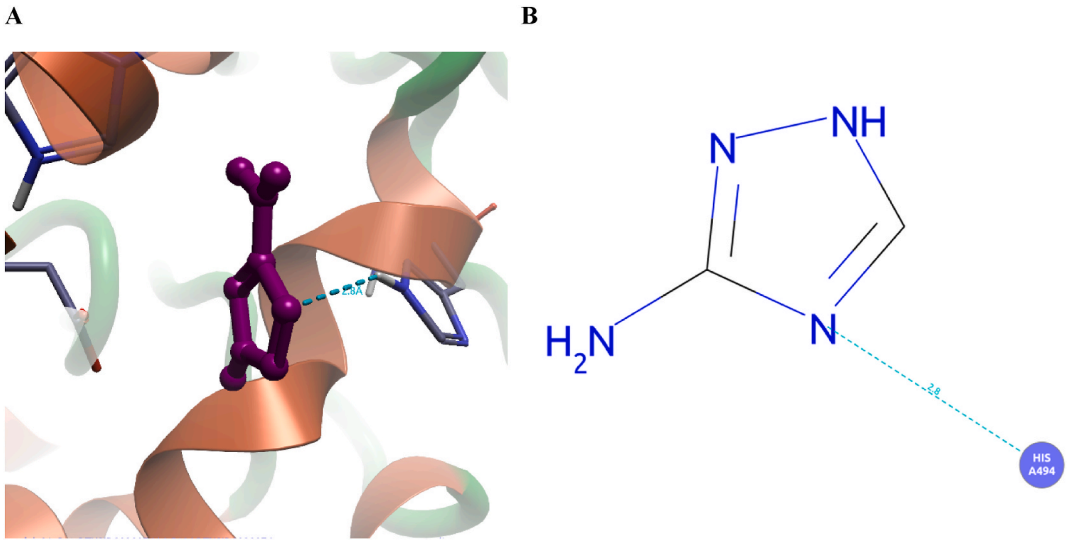


Fig. 5. Molecular docking of amitrole in the active pocket of modelled hTPO. **A**: The 3D illustration of amitrole interacting with His494. The molecule colored in violet color is an amitrole and it is interacting with His494 by weak H-Bond. **B:** The 2D illustration of amitrole interacting with His494, the blue dotted line corresponds to weak H-Bond.

template was 86.32 % identical to hNIS with 89 % coverage between 10 and 566th residue. The quality of the modelled structures was validated at secondary and tertiary levels of protein structure. The Ramachandran plot was built for both the hTPO, and hNIS models as shown in Fig. 1A and B respectively. The φ-ψ angles cluster into distinct regions in the Ramachandran plot where each region corresponds to a particular secondary structure. The hTPO model structure was seen with both right-handed alpha helices and beta sheets and the residues plotted at the left top and left bottom quadrants have complemented the same. In the Ramachandran plot, many of the residues (black dots) were observed in allowed (yellow) and strongly allowed (red) regions, whereas few glycine residues (triangles) were observed in non-allowed (white) regions. Glycine with only hydrogen in sidechain, does not have the Cβ atom, which is likely to induce many steric clashes in the generic Ramachandran plot [24]. Hence, the glycine residues in non-allowed regions have no

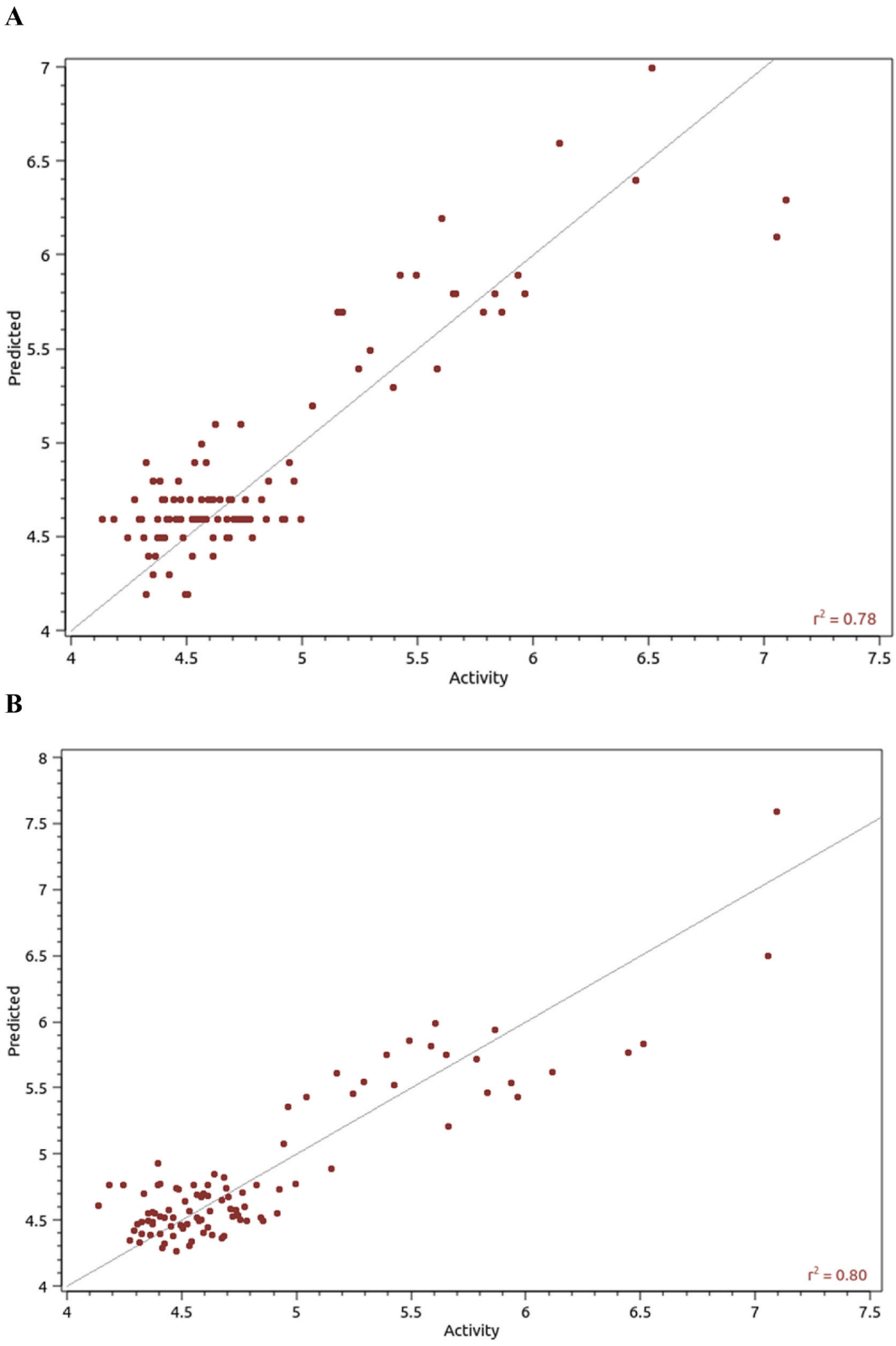
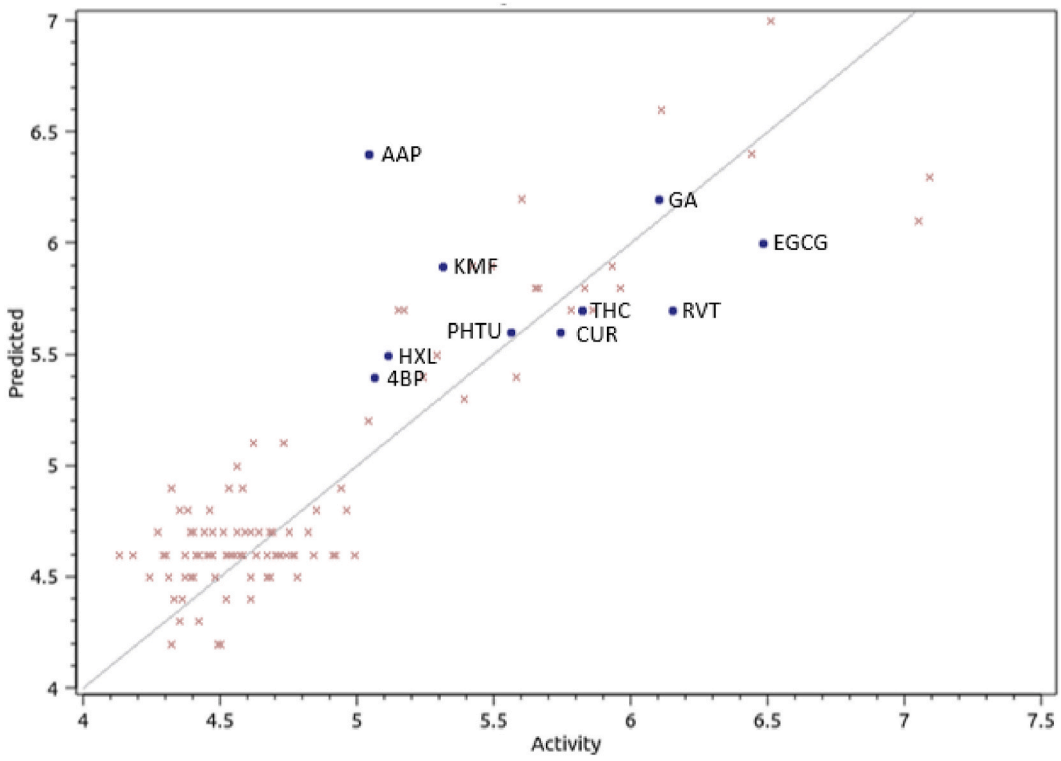


Fig. 6. 3D-QSAR models for the screening of hTPO inhibitors. **A:** The RF regression model built using training and test datasets, dots on the plot corresponds to the molecules in test set. **B:** The kNN regression model built using training and test datasets, dots on the plot corresponds to the molecules in test set.



B

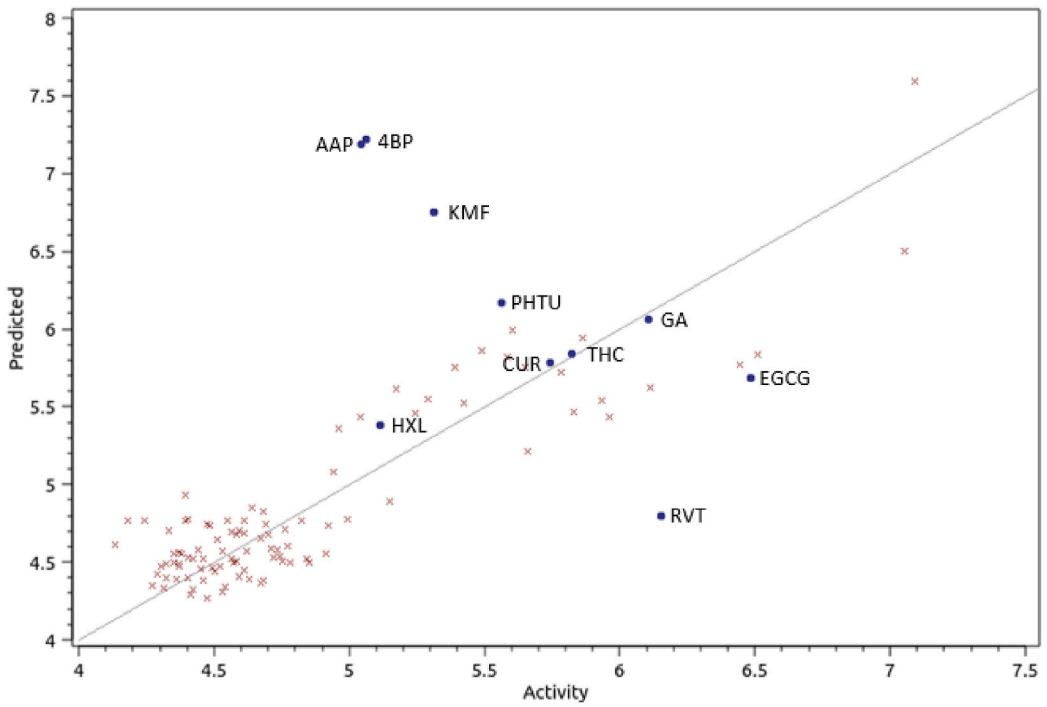


Fig. 7. 3D-QSAR models for the screening of hTPO inhibitors. **A:** The RF regression model is overlayed with external dataset, dots on the plot correspond to the molecules in external dataset and crosses corresponds to test set. **B:** The kNN regression model overlayed with external dataset, dots on the plot corresponds to the molecules in external dataset and crosses correspond to test set.

Table 1
IC_{50} of molecules in external dataset obtained by LC-MS/MS based *in vitro* assay.

S.No.	Molecules in external dataset (0.015–330 μM)	Abb.	L-Tyr to MIT				
			N = 1	N = 2	N = 3	AVG IC50 ± SD	Log10 Value
1	Kaemferol	**KMF**	6.58	5.45	2.74	**4.93 ± 1.97**	5.30
2	Phenylthiourea	**PHTU**	2.51	2.17	3.51	**2.73 ± 0.69**	5.56
3	Epigallocatechin gallate	**EGCG**	0.27	0.39	0.31	**0.32 ± 0.06**	6.48
4	Hexestrol	**HXL**	5.38	10.73	7.19	**7.77 ± 2.71**	5.10
5	Reservatrol	**RVT**	0.64	0.73	0.76	**0.71 ± 0.05**	6.14
6	Acetaminophen	**AAP**	9.69	10.48	7.25	**9.14 ± 1.68**	5.03
7	Gallic acid hydrate	**GA**	0.96	1.01	0.39	**0.79 ± 0.34**	6.09
8	4-Benzyl oxy phenol	**4BP**	8.55	10.22	7.52	**8.76 ± 1.36**	5.05
9	Curcumin	**CUR**	2.31	1.59	1.53	**1.81 ± 0.43**	5.74
10	Tetrahydro curcumin	**THC**	1.55	1.67	1.32	**1.51 ± 0.17**	5.81

significance in determining the quality of secondary structures in modelled proteins. The NIS model structure was seen with only alpha helices in the transmembrane region and the corresponding residues were observed in the left bottom quadrant of the plot which represents alpha helices, but the residues mapping to extracellular helices were observed in the left top quadrant of the plot which represents beta sheets. To derive the consensus, the modelled structure for hTPO and hNIS were observed with 98.80 % and 98.90 % of Ramachandran favoured residues and ensured the good quality of secondary structures in the models.

The quality of the models was also assessed at the tertiary structure level to understand the folding quality and stability of the structures when solvated in the water system. The MD simulation for 120ns in the TIP3P solvent system has demonstrated the stability of TPO as shown in Fig. 2A. Since the NIS is a transmembrane transporter, the phosphatidylcholine (POPC) membrane for NIS model structure was built using OPM structure of 7UV0 as shown in Fig. 3A and the MD simulation was carried out for 100ns. During the simulation, the root mean square deviation (RMSD) of structure coordinates were not more than 1 Å as shown in Fig. 2 B and 3B whereas for the globular proteins, the RMSD between 1 and 3 Å is acceptable [20]. The free-energy landscape of proteins is highly complex because of their complex topology and potential to establish nonbonded interactions. This has imperative results since the roughness of the landscape affects the ease of chain folds and influences the dynamic behaviour of the folded structure, thus influencing its functional and stability properties [25]. The potential energy minima of the protein are the major indicator of the folded structure and in the current study, the potential energy of hTPO and hNIS modelled structures was maintained at a minimum throughout simulation.

3.2. Molecular docking and QSAR modeling

The hTPO inhibitors dataset used for the QSAR modeling consisted of 432 molecules after removing the molecules with duplication and ambiguous activity information. The activity (pIC50) of the small molecules in the dataset was distributed between 4 and 9 as shown in the histogram (Fig. 4). The amitrole was the potential hTPO inhibitor in the dataset with IC_{50} of 1.75 μM and it was docked to modelled hTPO structure as shown in Fig. 5A and B with the ΔG -2.65 kJ/mol and virtual screening score (VSscore) −5.51. ΔG is relatively higher because of the very small size of the molecule, however it is sterically complimentary in the hTPO active pocket. Further, the docked conformation of amitrole was used as a reference to align the remaining molecules in the dataset.

Further, three regression 3D-QSAR models were developed using diverse hTPO inhibitors dataset to predict the hTPO inhibition potential of the small molecules. Random Forest is an aggregation of unpruned classification or regression trees created by bootstrap sampling of training data and random selection of features in tree induction, among the three models that have been developed [26]. Here, predictions are generated by combining majority votes or averaging together ensembles' forecasts. We built a predictive model for our hTPO inhibitors data set. In the current study, the RF model was built with the coefficient of determination R^2 0.78 as shown in Fig. 6A to predict the TPO inhibition potential of a target molecule. Similarly, to build a correlation model between chemical descriptors and biological activities of compounds kNN is known to be a popular nonlinear method. Here, the models are built by finding a subspace of an original descriptor space where each compound's activity in the data set is most accurately predicted based on its average activity with its nearest neighbours within that subspace [27]. Our analysis demonstrates that kNN is a powerful and suitable model with R^2 0.80 as shown in Fig. 6B to predict the TPO inhibition potential of a given molecule.

The molecules in the external dataset were aligned with both RF and kNN 3D-QSAR models to predict their hTPO inhibition potential and the IC_{50} was predicted as shown in Fig. 7A and **B** respectively. The predicted hTPO inhibition activity of molecules in the external dataset was experimentally evaluated.

3.3. Experimental validation of QSAR models

The well optimized *in vitro* assay by Tater et al. was used to evaluate the TPO inhibitory potential of molecules in the external dataset and the L-Tyr-to-Methyl methionine conversion was observed by LC-MS/MS [13]. Half maximal inhibitory concentration (IC_{50}), a numerical indicator of test chemical inhibitor potency, was calculated using GraphPad Prism software for N = 3 and reported as a measure of variability for the IC_{50} values as tabulated in Table 1. Fig. 8 displays four parametric sigmoidal IC_{50} curves of molecules from an external dataset for the TPO-catalyzed iodination from L-Tyro to MIT. The corroboration of 3D-QSAR based prediction of

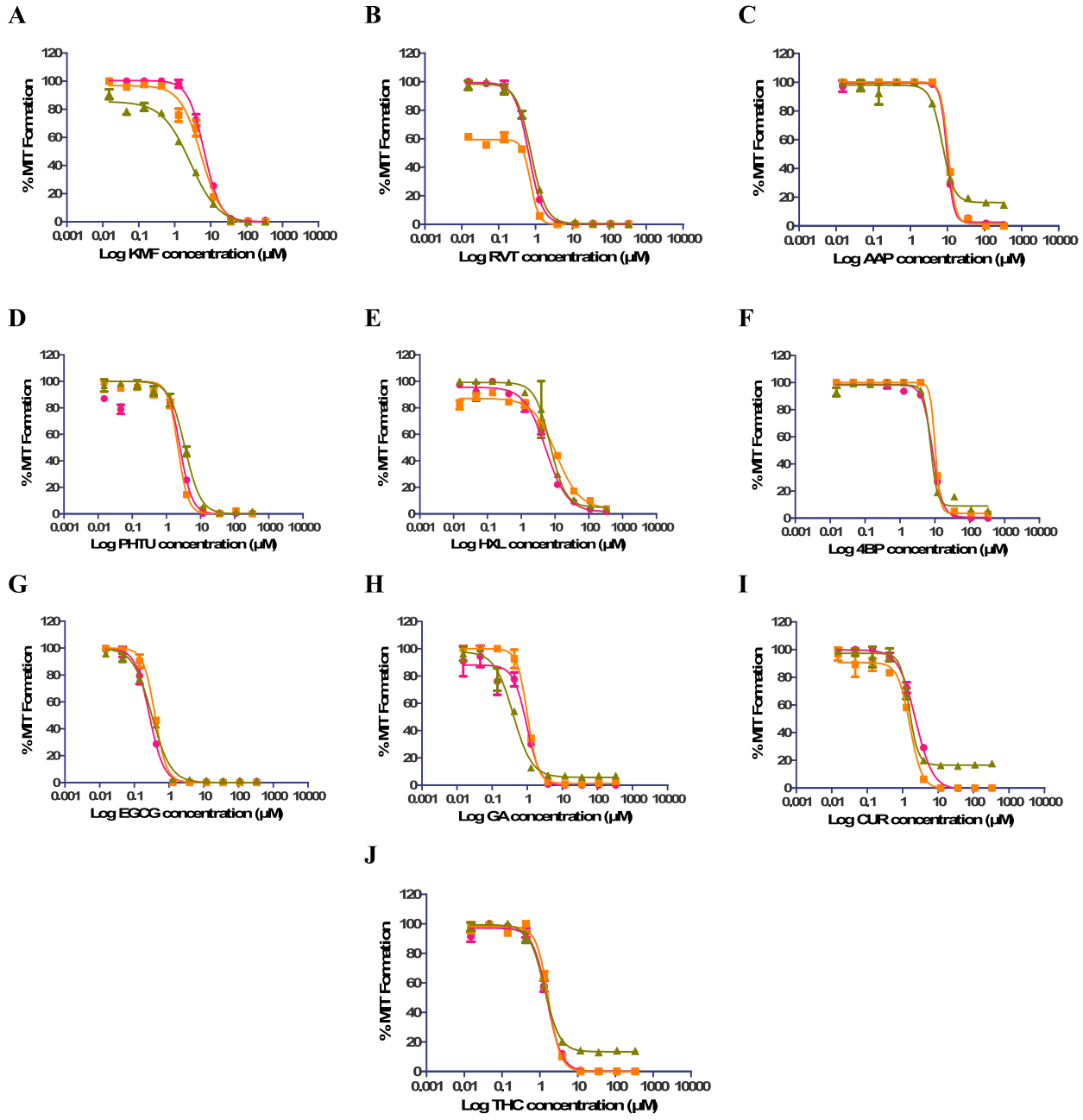


Fig. 8. Four parametric sigmoidal IC_{50} curves of molecules in external dataset for respective TPO reactions. **A:** Kaempferol, **B:** Resveratrol, **C:** Acetaminophen, **D:** Phenylthiourea, **E:** Hexesterol, **F:** 4-Benzyloxyphenol, **G:** Epigallocatechin, **H:** Gallic acid, **I:** Curcumin, **J:** Tetrahydrocurcumin.

hTPO inhibition with the experimental data tabulated in Table 2 was assessed by calculating the correlation coefficient. Though the kNN 3D-QSAR model was seen with a higher R^2 value than the RF model, the prediction for molecules in the external dataset has shown a better correlation with the RF 3D-QSAR model as shown in Fig. 9A and B.

3.4. NIS and thyroid biosynthesis

The biosynthesis of thyroid hormone requires various molecules in thyroid follicular cells, including NIS, which transports iodide into the cells, and TPO, which oxidizes iodide into iodinium (I+) that in turn iodinates thyroglobulin (Tg) [28]. Since iodine availability is one of the key factors regulating hormone biosynthesis, the inhibitors may bind to hNIS and induce the scarcity of iodide which is essential for thyroid biosynthesis. Hence, the study has demanded the screening of hNIS inhibition potential of the hTPO inhibitors. Since there is no information available on NIS inhibitors, the 3D-QSAR model was not developed to predict the hNIS

Table 2
Comparison of *in vitro* and *in silico* IC_{50} values for molecules in external dataset to validate the RF and kNN 3D-QSAR models.

Sl. No	Abbreviation	Actual pIC_{50}	Predicted pIC_{50}	
			RF	kNN
1	**KMF**	5.30	5.90	6.80
2	**PHTU**	5.56	5.60	6.20
3	**EGCG**	6.48	6.00	5.70
4	**HXL**	5.10	5.50	5.40
5	**RVT**	6.14	5.70	4.80
6	**AAP**	5.03	6.40	7.20
7	**GA**	6.09	6.20	6.10
8	**4BP**	5.05	5.40	7.20
9	**CUR**	5.74	5.60	5.80
10	**THC**	5.81	5.70	5.80

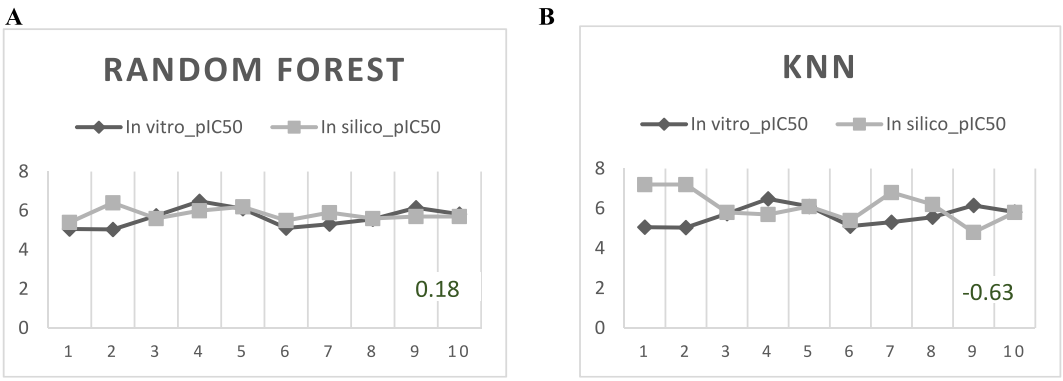


Fig. 9. Correlation between 3D-QSAR and *in vitro* IC_{50}. **A:** The RF model. **B:** The kNN model.

inhibition potential. However, the structure-based screening of NIS inhibition potential of molecules in an external dataset was carried out using molecular docking. All 10 molecules were docked to the iodide binding site of hNIS (Met 68, Gln 72, Val 76, Phe 87, Trp 255, Val 293, and Phe 417) and only 3 molecules such as Kaempferol, Curcumin, and Tetrahydro curcumin have shown interaction with NIS at iodide binding site (Fig. 10 A-10F). On the other hand, Kaempferol has shown interaction with the key residue Asp 238 and Amitrole binding site His 494 as shown in Fig. 11A and B, but Curcumin and Tetrahydrocurcumin have neither shown interaction with key residues nor with Amitrole binding site His 494 as shown in Fig. 11C-F. This interaction suggests that the Kaempferol, Curcumin and Tetrahydrocurcumin are inhibiting the hNIS and reducing the availability of iodide for thyroid biosynthesis. The docking energy of all the molecules in external dataset is tabulated in Table 3. All the ten molecules in external dataset have shown higher docking energy for hNIS than the docking energy for hTPO. Because the hTPO active site is in the internal gaseous (IG) tunnel. The IG tunnels are likely to be seen in the enzymes anchored in the lipid bilayer. The gaseous tunnels facing the lipid membrane are the O2 and NO transporting class of cytochrome *ba*3, cytochrome *aa*3, and cytochrome *c* dependent nitric oxide reductase cNOR [29]. These enzymes belong to the heme-copper oxidoreductase superfamily, containing multiple heme centers. hTPO also possesses the same architecture, hence the docking energy is good in the case of hTPO. On the other hand, the iodine binding site of hNIS is in the pore region enriched with non-polar residues, hence the interaction is observed with high docking energy. Therefore, the docking energy alone cannot be considered as a parameter to understand the binding efficiency of molecules, and the active site residues interacting with the molecules and their steric, electrostatic complementarities are also a determining factor. In that regard, Kaempferol, Curcumin and Tetrahyrdocurcumin are found to be good inhibitors of hNIS. To conclude, in the present study, the integration of more than one method to predict the hTPO inhibition potential of target molecules is demonstrated and three out of ten molecules are predicted as hNIS inhibitors.

4. Conclusion

The universal RF 3D-QSAR model developed during the study was found reliable to predict the hTPO inhibition potential of small molecules. Integrated modeling approaches often allow us to attain the highest external predictive ability, compared to individual approaches. This is especially the case when the individual models are created using different methods and models, as in the case of this study, both 3D-QSAR and docking techniques are integrated to compensate for and address the limitations of each method. However, due to the limited experimental data available, the data set used in the 3D-QSAR model is very small. Therefore, the robustness of the model can be improved by continuously updating the data set by adding new molecules along with experimental data in the future. Hence, molecular docking of the target molecule with hNIS is also required to confirm the selectivity of the target molecule.

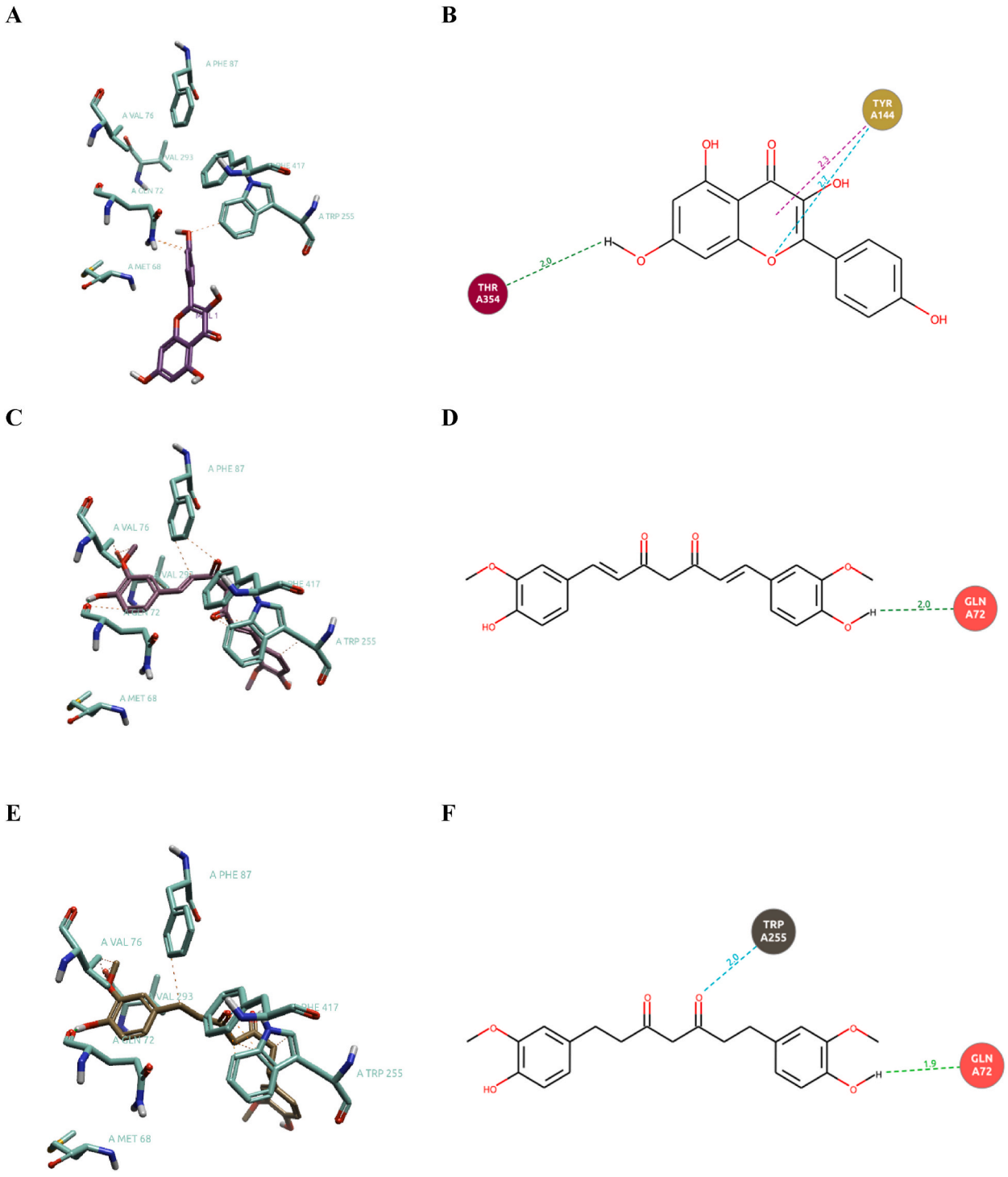


Fig. 10. Molecular docking of external dataset with hNIS iodine binding site. **A:** 3D illustration of Kaemferol interaction with active site residues. **B:** 2D illustration of Kaemferol interaction with active site residues. **C:** 3D illustration of Curcumin interaction with active site residues. **D:** 3D illustration of Curcumin interaction with active site residues. **E:** 3D illustration of Tetrahydro curcumin interaction with active site residues. F: 3D illustration of Tetrahydro curcumin interaction with active site residues.

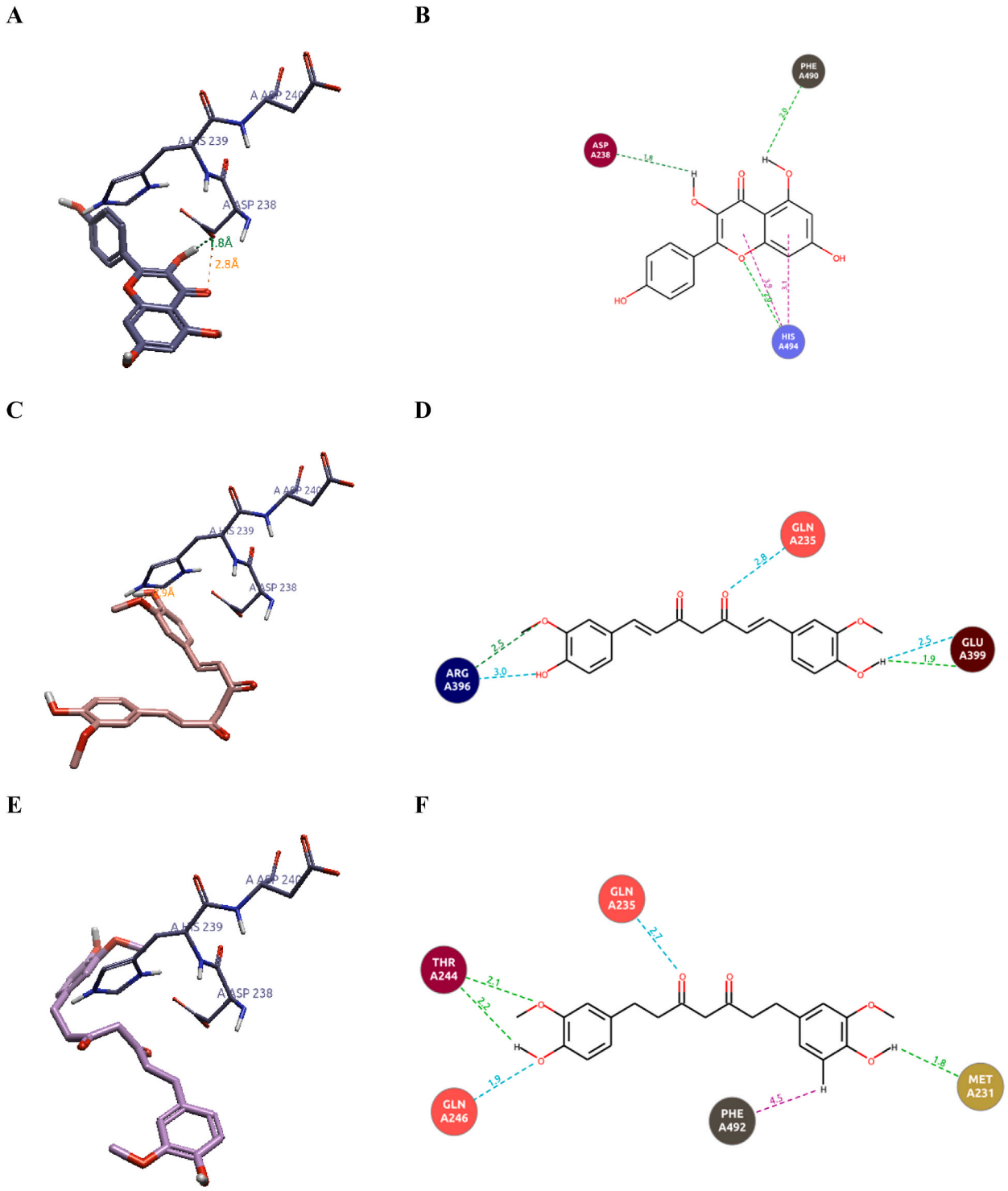


Fig. 11. Molecular docking of external dataset with hTPO considering the protonation site (D238, H 239, and D240) as active site. **A:** 3D illustration of Kaemferol interaction with active site residues. **B:** 2D illustration of Kaemferol interaction with active site residues. **C:** 3D illustration of Curcumin interaction with active site residues. **D:** 3D illustration of Curcumin interaction with active site residues. **E:** 3D illustration of Tetrahydro curcumin interaction with active site residues. F: 3D illustration of Tetrahydro curcumin interaction with active site residues.

Table 3
The docking energy of external dataset molecules docked with hTPO and hNIS.

Sl. No	Abbreviation	ΔG for hTPO kj/mol	ΔG for hNIS kj/mol
1	**KMF**	−8.75	−6.30
2	**PHTU**	−6.26	−4.43
3	**EGCG**	−12.16	Did not dock
4	**HXL**	−7.25	−2.41
5	**RVT**	−7.47	−5.93
6	**AAP**	−5.32	−4.44
7	**GA**	−6.90	−4.91
8	**4BP**	−6.27	−5.05
9	**CUR**	−10.34	−6.06
10	**THC**	−9.06	−7.09

Data availability statement

The data on molecules used for the development of 3D-QSAR is not specified because of commercial constrains.

CRediT authorship contribution statement

Bharath Basavapattana Rudresh: Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Abhishek Kumar Tater:** Writing – original draft, Methodology. **Vaibav Barot:** Writing – original draft, Software, Methodology. **Nitin Patel:** Methodology. **Ashita Desai:** Methodology. **Sreerupa Mitra:** Writing – original draft, Methodology. **Abhay Deshpande:** Project administration, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Bharath BR reports administrative support, article publishing charges, equipment, drugs, or supplies, statistical analysis, travel, and writing assistance were provided by Jai Research Foundation. Bharath BR reports a relationship with Jai Research Foundation that includes: employment. All the authors of this publication are employed by Jai Research Foundation If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Anders, N. Malin, K. Kirstine, The Endocrine System, Compendium of Histology, 2017, pp. 517–539, https://doi.org/10.1007/978-3-319-41873-5_24.
[2] W. Anika, M. Ralf, New insights into the astonishing diversity of hormone functions, Acta Physiol. 224 (4) (2018) e13188, https://doi.org/10.1111/apha.13188.
[3] H. Dong, M. Godlewska, M.G. Wade, A rapid assay of human thyroid peroxidase activity, Toxicol. Vitro 62 (2020) 104662, https://doi.org/10.1016/j.tiv.2019.104662.
[4] T. Filomena, Nutritional influences on hormonal health, Integrative and Functional Medical Nutrition Therapy (2020) 517–532, https://doi.org/10.1007/978-3-030-30730-1_31.
[5] M. Dario, J.A. Richard, Androgens, estrogens, and hydroxysteroid dehydrogenases, Mol. Cell. Endocrinol. 301 (10) (2009) 37–42, https://doi.org/10.1016/j.mce.2008.08.029.
[6] L. Patrick, Thyroid disruption: mechanism and clinical implications in human health, Altern Med Rev 14 (4) (2009) 326–346. PMID: 20030460.
[7] M. Nakamura, S. Ohtaki, Molecular mechanism of thyroid hormone synthesis, Jpn. J. Clin. Med. 52 (4) (1994) 857–863. PMID: 8196170.
[8] M. Godlewska, P.J. Banga, Thyroid peroxidase as a dual active site enzyme: focus on biosynthesis, hormonogenesis and thyroid disorders of autoimmunity and cancer, Biochimie 160 (2019) 34–45, https://doi.org/10.1016/j.biochi.2019.02.003.
[9] S. Ravera, A. Reyna-Neyra, G. Ferrandino, L.M. Amzel, N. Carrasco, The sodium/iodide symporter (NIS): molecular physiology and preclinical and clinical applications, Annu. Rev. Physiol. 79 (2017) 261–289, https://doi.org/10.1146/annurev-physiol-022516-034125.
[10] S. Christine, John C. Morris, Sodium iodide symporter (NIS) and thyroid, Hormones (Basel) 1 (1) (2002) 22–34, https://doi.org/10.1016/B978-0-12-816386-3.00064-8.
[11] F.K. Paul, E.D. Watt, M.W. Hornung, J.M. Hedge, R.S. Judson, K.M. Crofton, K.A. Houck, S.O. Simmons, Tiered high-throughput screening approach to identify thyroperoxidase inhibitors within the ToxCast phase I and II chemical libraries, Toxicol. Sci. 151 (1) (2016) 160–180, https://doi.org/10.1093/toxsci/kfw034.
[12] B.P. Katie, M.H. Joan, M.R. Daniel, W.H. Michael, M.C. Kevin, O.S. Steven, Development of a thyroperoxidase inhibition assay for high-throughput screening, Chem. Res. Toxicol. 27 (3) (2014) 387–399, https://doi.org/10.1021/tx400310w.
[13] A. Tater, A. Gupta, G. Upadhyay, A. Deshpande, R. Date, I.Y. Tamboli, In vitro assays for characterization of distinct multiple catalytic activities of thyroid peroxidase using LC-MS/MS, Curr Res Toxicol 2 (2021) 19–29, https://doi.org/10.1016/j.crtox.2021.01.00.
[14] E. Habza-Kowalska, A.A. Kaczor, J. Zuk, D. Matosiuk, U. Gawlik-Dziki, Thyroid peroxidase activity is inhibited by phenolic compounds—impact of interaction, Molecules 24 (15) (2019) 2766, https://doi.org/10.3390/molecules24152766.
[15] S.A. Rosenberg, E.D. Watt, R.S. Judson, S.O. Simmons, F.K. Paul, M. Dybdahl, N.G. Nikolov, E.B. Wedebye, QSAR models for thyroperoxidase inhibition and screening of U.S. and EU chemical inventories, Comp Tox. 4 (2017) 11–21, https://doi.org/10.1016/j.comtox.2017.07.006.
[16] D.E. Williams, S.N. Le, D.E. Hoke, P.G. Chandler, M. Gora, M. Godlewska, J.P. Banga, A.M. Buckle, Structural studies of thyroid peroxidase show the monomer interacting with autoantibodies in thyroid autoimmune disease, Endocrinology 161 (2) (2020) bqaa016, https://doi.org/10.1210/endocr/bqaa016.
[17] S. Baker, R.N. Miguel, D. Thomas, M. Powell, J. Furmaniak, B.R. Smith, Cryo-electron microscopy structures of human thyroid peroxidase (TPO) in complex with TPO antibodies, J. Mol. Endocrinol. 70 (3) (2023) e220149, https://doi.org/10.1530/JME-22-0149.
[18] S.N. Le, B.T. Porebski, J. McCoey, J. Fodor, B. Riley, M. Godlewska, et al., Modelling of thyroid peroxidase reveals insights into its enzyme function and autoantigenicity, PLoS One 10 (12) (2015) e0142615, https://doi.org/10.1371/journal.pone.0142615.

[19] Z. Guo, U. Mohanty, J. Noehre, T.K. Sawyer, W. Sherman, G. Krilov, Probing the α-helical structural stability of stapled p53 peptides: molecular dynamics simulations and analysis, Chem. Biol. Drug Des. 75 (2010) 348–359, https://doi.org/10.1111/j.1747-0285.2010.00951.x.
[20] B.R. Bharath, H. Damle, S. Ganju, L. Damle, In silico screening of known small molecules to bind ACE2 specifi c RBD on Spike glycoprotein of SARS-CoV-2 for repurposing against COVID-19, F1000 Research 9 (2020) 663, https://doi.org/10.12688/f1000research.24143.1.
[21] K.B. Paul, J.M. Hedge, D.M. Rotroff, M.W. Hornung, K.M. Crofton, S.O. Simmons, Development of a thyroperoxidase inhibition assay for high-throughput screening, Chem. Res. Toxicol. 27 (3) (2014) 387–399, https://doi.org/10.1021/tx400310w.
[22] V.R. Joseph, Optimal ratio for data splitting, Stat. Anal. Data Min.: The ASA Data Science Journal (2022) 1–19, https://doi.org/10.1002/sam.11583.
[23] L. Abas, C. Luschnig, Maximum yields of microsomal-type membranes from small amounts of plant material without requiring ultracentrifugation, Anal. Biochem. 401 (2) (2010) 217–227, https://doi.org/10.1016/j.ab.2010.02.030.
[24] B.K. Ho, R. Brasseur, The Ramachandran plots of glycine and pre-proline, BMC Struct. Biol. 16 (5) (2005) 14, https://doi.org/10.1186/1472-6807-5-1.
[25] I. Tavernelli, S. Cotesta, E.E. Di Iorio, Protein dynamics, thermal stability, and free-energy landscapes: a molecular dynamics investigation, Biophys. J. 85 (4) (2003) 2641–2649, https://doi.org/10.1016/S0006-3495(03)74687-6.
[26] S. Vladimir, L. Andy, T. Christopher, C. Christopher, P.S. Robert, P.F. Bradley, Random forest: a classification and regression tool for compound classification and QSAR modeling, J. Chem. Inf. Comput. Sci. 43 (6) (2003) 1947–1958, https://doi.org/10.1021/ci034160g.
[27] P. Itskowitz, A. Tropsha, Kappa Nearest neighbors QSAR modeling as a variational problem: theory and applications, J. Chem. Inf. Model. 45 (3) (2005) 777–785, https://doi.org/10.1021/ci049628+.
[28] B. Salvatore, E. Giusy, R. Francesca, R.P. Sabrina, M.S. Mattia, M.F. Silvia, A. Alessandro, Fallahi Poupak, Endocrine disruptors and thyroid autoimmunity, Best Pract. Res. Clin. Endocrinol. Metabol. 34 (1) (2020) 101377, https://doi.org/10.1016/j.beem.2020.101377.
[29] S. Singh, R. Anand, Tunnel architectures in enzyme systems that transport gaseous substrates, ACS Omega 6 (49) (2021) 33274–33283, https://doi.org/10.1021/acsomega.1c05430.