

Atom-based 3D-QSAR, molecular docking and molecular dynamics simulation assessment of inhibitors for thyroid hormone receptor α and β

Manish Kumar Gupta · Krishna Misra

Received: 6 February 2014 / Accepted: 1 May 2014 / Published online: 5 June 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract The three-dimensional quantitative structure–activity relationship (3D-QSAR) for inhibitors of thyroid hormone receptors (TR) α and (TR) β was studied. The training set of the TR α model generated a correlation coefficient (R^2) = 0.9535, with standard deviation (SD)=0.3016. From the test set of the TR α model, a Q^2 value for the predicted activities (= 0.4303), squared correlation (random selection R^2 – CV= 0.6929), Pearson- R (= 0.7294) and root mean square error (RMSE=0.6342) were calculated. The P -value for TR α (= 1.411e-96) and TR β (= 2.108e-165) models indicate a high degree of self-reliance. For the TR β model, the training set yielded R^2 =0.9424 with SD = 0.3719. From the test set of TR β , Q^2 value (= 0.5336), the squared correlation (R^2 – CV= 0.7201), the Pearson- R (= 0.7852) and RMSE for test set predictions (=0.8630) all strengthen the good predictive competence of the QSAR model derived. Examination of internal as well as external validation supports the rationality and good predictive ability of the best model. Molecular docking explained the conformations of molecules and important amino acid residues at the docking pocket, and a molecular dynamics simulation study further uncovered the binding process and validated the rationality of docking results. The findings not only lead to a better understanding of interactions between

these antagonists and thyroid hormone receptors α and β , but also provide valuable information about the impact of structure on activity that will be very beneficial in the design of novel antagonists with preferred activity.

Keywords Thyroid hormone receptors α/β · 3D-QSAR · Molecular docking · Pharmacophore modeling · Molecular dynamics simulation · TR antagonist

Introduction

Thyroid hormone receptors (TRs) play an important role in normal development, differentiation, growth and metabolic regulation in humans. TRs can be considered as ligand-regulatable transcription factors, assuming their competencies to interact with ligands and DNA along with their ability to control transcription [1, 2]. TRs, which are members of the nuclear hormone receptor (NR) superfamily, are differentially expressed as many isoforms (TR α 1, TR β 1 and TR β 2) in different tissues of the body and are key targets for commonly used drugs [3–5].

TRs are comprised of functional domains A/B, C, D and E, with a DNA binding domain (DBD), hinge domain and ligand binding domain (LBD) associated with the C, D and E domains, respectively. The sequences of DBD and LBD are extremely highly conserved amongst the TR isoforms. Conversely, no resemblance occurs in the A/B domain between TR α and TR β isoforms [6]. TR α 1 and TR β 1 vary from 400 to 500 amino acids in size, having abundantly homologous DBD and LBD. TR α isoforms α 1, α 2 and α 3, are encoded by the TR α gene. These isoforms vary in their carboxyl termini due to alternative splicing. TR α 1 has a binding capacity for T3, which leads to activation or repression of target genes, whereas TR α 2 and TR α 3 are non-T3 binding products and inhibit T3 function [7]. TR β gene encodes three T3-binding

Electronic supplementary material The online version of this article (doi:10.1007/s00894-014-2286-5) contains supplementary material, which is available to authorized users.

M. K. Gupta
Department of Bioinformatics, University Institute of Engineering and Technology, Chhatrapati Shahu Ji Maharaj University,
Kanpur 208024, Uttar Pradesh, India
e-mail: manish.kumar.gupta@gmail.com

K. Misra (✉)
Centre of Biomedical Research, Sanjay Gandhi Post-Graduate
Institute of Medical Sciences, Campus Raebareli Road,
Lucknow 226014, Uttar Pradesh, India
e-mail: krishnamisra@hotmail.com

splice products isoforms: TR β 1, TR β 2, and TR β 3, which differ in their amino termini. All TR β isoforms are capable of binding to their cognate ligand T3 with high affinity to facilitate target gene transcription [1]. The effects of thyroid hormone have been delineated by observations in investigations on human subject with poor or extreme thyroid hormone levels. Hypothyroidism occurs due to thyroid hormone deficiency, while hyperthyroidism results from oversecretion of thyroid hormone. Resistance to thyroid hormone (RTH) is a condition in which persons have hyposensitivity to thyroid hormone, raised circulating serum levels of T3 and T4, and high or non-suppressed thyroid stimulating hormone (TSH) serum levels. RTH was recognized because of a mutation in TR genes. It is significant that RTH patients can exhibit inconsistent resistance in various tissues and clinical symptoms characteristic of the hypothyroidism and hyperthyroidism [8].

To date, numerous thyroid ligands have been reported [9, 10]. These TR ligands could lead to safe treatments for non-thyroid illness while evading the cardiac side effects and, therefore, could be used as a short-term supplemental therapy to traditional cures [11]. Specifically impeding the consequences of thyroid hormone binding at the receptor level may lead to a substantial improvement in the treatment of hyperthyroid patients [3].

Various quantitative structure–activity relationships (QSAR) from one- to six-dimension models, as well as structure–activity relationships (SARs) have been developed satisfactorily [12–15]. Hormone activities of OH-PBDEs for the TR β and 2D-QSAR model have also been reported [16]. 3D-QSAR studies have been combined with a molecular docking and molecular dynamic (MD) simulation approach [17–19]. These models can assist with finding critical structural features for high binding affinity of novel ligands against TR β . The identification and design of novel TR antagonists could play a significant role in their potential medical applications in the therapy of thyroid disorder [20]. Such antagonists usually interact with hydrophobic and important hydrophilic residues lining the ligand-binding pocket [21].

The aim of the present study was to explore the binding mode of inhibitors with TR α and TR β receptors through molecular docking and MD simulation to gain further insight into their structure–activity relationship, and to obtain robust 3D-QSAR models. Such models could be useful in designing novel antagonists.

Materials and methods

All computational analysis was carried out on a Linux operating system (Cent OS 6.3) on a 64-bit machine running with an Intel quad core (4*2 core) processor and 16 GB RAM, with

nvvidia (Quadro FX 1700) graphic memory cards (512 MB) and a 900-GB SAS hard disk.

Ligand preparation and dataset for analysis

Three dimensional (3D) molecular structural datasets of 357 and 203 inhibitors from the binding database [22] were employed against TR β and TR α , respectively. All molecules were processed with LigPrep 2.5 software [23] in order to assign appropriate protonation states at physiological pH (7.2 ± 0.2), employing the ionizer option. The IC₅₀ values of the biologically active molecules were converted to $pIC_{50} = -\log_{10}(IC_{50}) + 9$. The distribution of pIC_{50} of the entire data set for TR β inhibitors varied from 10.721 to 2.699, whereas for TR α inhibitors fluctuated from 3.67 to 10. Care was taken to definite a uniform distribution for TR α and TR β inhibitors from a varied range of pIC_{50} for both training and test sets.

Generating pharmacophore sites

Six pharmacophore features, such as hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively ionizable (N), positively ionizable (P), and aromatic ring (R) are available in the Phase 3.4 [24] module of Schrodinger. The procedures used to map the locations of pharmacophore sites are acknowledged as feature definitions, and are characterized within the program by a set of SMARTS patterns. Each pharmacophore feature is specified by a set of 3D structural configurations of the compound. Once a feature has been mapped to a precise position in a conformation, it is stated as a pharmacophore site.

Searching common pharmacophores and scoring hypotheses

The desired variants were selected to find common pharmacophores among the active ligands. In this study, with five active compounds in the training set for TR β and ten for TR α , shared pharmacophores were studied using a scoring procedure to recognise the pharmacophore from each surviving n -dimensional box that yields the best alignment of the active-set ligands. This pharmacophore offers hypotheses to describe how the active molecules interact with the receptor. As there are many boxes, there will be many hypotheses. The scoring method offers a ranking of the diverse hypotheses, allowing rational selection of those hypotheses most suitable for further examination. The shared pharmacophore hypotheses were scored by setting the root mean square deviation (RMSD) value $< 1.2 \text{ \AA}$, the vector score value to 0.5 and weighting to include consideration of the alignment of inactive molecules by default constraints.

3D-QSAR model development

Phase—an integrated software module in maestro—was used to produce pharmacophore and 3D-QSAR models for TR α and β inhibitors. 3D-QSAR models of TR α and TR β were developed from varying activities of different compounds, linked with a particular reference molecule, that have all been aligned to a common pharmacophore hypothesis. An atom-based QSAR model was used in this study, which takes all atoms into account. Standard parameters of the program were employed utilizing the partial least-squares (PLS) regression analysis method. The precision of the predicted models increases with rising number of PLS factors until over fitting twitches to transpire.

Model validation

Phase 3D-QSAR models utilise distinct training and test sets rather than internal cross-validation methods. In Phase 3D-QSAR models, leave- n -out models are made and the R^2 value between the leave- n -out estimates is calculated. The estimates come from the model built on the complete training set. This value is designated as the stability value and has a maximum value of the integer one. Models with high stability are preferred because they are not excessively reliant on the peculiarities of any specific training set.

The variance in observed activity (σ_y^2) and coefficient of determination (R^2) are very important statistical parameters used for precision of the training set model where the observed activity for the training set molecule i (v_i), is as shown in Eqs. 1 and 2:

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1} (v_i - \bar{y})^2 \quad (1)$$

$$R^2 = 1 - \frac{\sigma_{err}^2}{\sigma_y^2} \quad (2)$$

The value of R^2 will always be positive, for the reason that the regression coefficients are optimized to have low sum of squared error (sse). The weakest situation is when the independent variables have absolutely no statistical association with activity. The regression coefficients will all be zero, and the predicted model will comprise only an intercept factor, the value of which will be mean observed activity (\bar{y}) as per the above-mentioned conditions. Therefore, all predicted activity will be \bar{y} and variance in errors (σ_{err}^2) will be equivalent to σ_y^2 , resulting in $R^2=0$.

A statistical quantity Q^2 (Eq. 3), equivalent to R^2 is calculated using the experimental and predicted activities for the test set. It is uncommon to obtain negative Q^2 values due to smaller variance in y as the test set does not have as high a series of activity values as the training set and the variance in

the errors is larger because of errors in the test set, which have tendency to be larger than those for the training set. As all values are moved by the sample means, the Pearson correlation coefficient (r) is impervious to logical errors in the estimations; however, Q^2 is not. Therefore, if the rank order of activity computations is basically precise, but there is a key constant shift in the values related to the experiential activities, r may still be moderately high, even if Q^2 is small or negative.

$$Q^2 = R^2(T) \quad (3)$$

$$r = \frac{\sum_{j \in T} (y_j - \bar{y}_T) (y_j - \hat{y}_T)}{\sqrt{\sum_{i \in T} (y_j - \bar{y}_T)^2 (y_j - \hat{y}_T)^2}} \quad (4)$$

The PLS regression was performed by PHASE with all available (i.e., seven) PLS factors. All models were validated by predicting activity for test sets of 47 and 71 molecules for TR α and TR β , respectively.

Pharmacophore screening

For external cross validation of the predicted model, a phase 3D-database was created against both models. A total of 197 inhibitors, consisting of 146 TR β and 51 thyroid oxidase2 antagonists were employed to screen external cross validation of the built 3D QSAR model of TR β . In total, 66 inhibitors, consisting of 15 compounds of TR α and 51 thyroid oxidase 2 inhibitors molecules, were used to cross-validate the constructed 3D QSAR model of TR α .

Molecular docking

Molecular docking was performed to investigate in-depth binding modes of different sets of TR α and TR β inhibitors. The Glide 5.8 [25] module of Schrodinger software was used for molecular docking. Glide applies a hierarchical succession of filters to explore probable sites of the ligand in the active-site region of the target protein. The shape and properties of the target protein are characterized on a grid by numerous diverse sets of fields that make available gradually more precise scoring of the compound poses. Mainly, the degrees of freedom of side chains were sampled, while minor backbone movements were permitted through minimization. The 3D coordinates of the TR α in complex with ligand (PDB ID: 3JZB), and TR β in complex with ligand (PDB ID: 1NAX), were retrieved from the Brookhaven Protein Databank. The four significant amino acids residues Arg 282, Arg 320, Asn 331 and His 435 for TR β and three residues Arg 266, Ser 277, His 381 for TR α , distinct as flexible residues in the docking process, were considered as binding sites. The best docking

score was selected on the basis of the conformation possessing high docking scores as well as maximum activity.

The XP glide docking process is very precise and produces 10,000 poses for every ligand during docking and provides the top pose on the basis of the energy term E_{model} . The top poses of each compound were further given preference on the basis of $\text{XPG}_{\text{score}}$ [26]. Lower $\text{XPG}_{\text{score}}$ for a ligand specifies improved binding affinity with the protein. The limit of the $\text{XPG}_{\text{score}}$ constraint for XP glide docking was kept to 0.0 kcal mol⁻¹, a restraint set to reject compounds with positive $\text{XPG}_{\text{score}}$ after docking yield.

Molecular dynamics simulation

Desmond 3.1 and 3.5 [27]—a suite of computer program for executing classical MD simulation systems—was used to evaluate the stability and conformational changes, and to generate high-quality simulation trajectories of moderate timescales in solution along with thermodynamic measures of docking complexes [28]. The MD simulations were studied through using the OPLS2005 force field with the predefined [29] with the docked complex of TR β protein with second lowest value docking score of ligand PCID 9933119 and TR α protein with lowest docking score of ligand PCID 10456672.

The original coordinates for the MD simulations were pocketed from the docking results. Prior to the simulation, an energy minimization process was performed on the full system without constraints using the steepest descent integrator for maximum 2,000 iterations, and the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithms, with a convergence threshold of 20.0 kcal mol⁻¹ Å⁻¹. The system was rooted with simple point charge (SPC) water model along with 10 Å × 10 Å × 10 Å ortho-rhombic water box, and the systems were neutralized by replacing solvent molecules with counter ions (Na⁺) to balance the net charges of the systems. This ensures that the major surface of both complexes (TR α and TR β) were covered by the solvent model. In all, 64,009 atoms of TR α and 59,141 atoms of TR β protein and ligand complex were simulated through a multi sim step procedure of the final system.

Prior to equilibration and extensive MD simulations, the structures were minimized and pre-equilibrated using a default relaxation routine. To perform this, the program ran six phases composed of minimizations and small (12 and 24 ps) MD simulations to relax the model system before accomplishment of the final long simulations. A 5-ns production MD simulation was then performed for each system. The shake algorithm was applied to all hydrogen atoms and the van der Waals (VDW) cut off was set to 9 Å [30]. The temperature of the system was maintained at 300 K, using the Nose-Hoover thermostat approach with a relaxation time of 1 ps, and pressure was maintained by employing the Martyna-Tobias-Kleinbarostat approach using coupling style isotropic with

relaxation time of 2 ps. Long-range electrostatic forces were considered by means of the particle-mesh Ewald (PME) method [31]. Trajectory was recorded at 4.8 ps during the MD runs and the recording interval energy was set at 1.2 ps. The dynamics equilibration was observed by studying the stability of temperature, energy, pressure and density of the system along with RMSD of the backbone atoms. RMSD and energy fluctuations of the complex in every trajectory were examined with respect to simulation time. The root mean square fluctuations (RMSF) of overall atoms, backbone and side chains of TR α and TR β were studied for each residue. The docking complex was explored and examined for consistency of hydrogen bonding interactions.

Results and discussion

Analysis of the atom-based 3D-QSAR model

To explore the common pharmacophore hypotheses, the complete dataset was separated into an active set and an inactive set, underlying the attribute of pharm set. Compounds with $\text{pIC}_{50} > 8.60$ for TR α and $\text{pIC}_{50} > 10.00$ for TR β were considered to be active and those with $\text{pIC}_{50} < 5.00$ for both (TR α and TR β) were set to be inactive, whereas those in-between were considered as moderately active (supplementary Tables 1 and 2). Total 439 (8 variant) and 33 (5 variant) hypotheses were identified for TR α and TR β , respectively (Table 1).

These hypotheses yielded a scoring process comprising of three scores: survival (solely on the basis of the active set), surv-inactive (on the basis of the active and inactive sets) and post-hoc score; calculated on the basis of active and inactive sets, with a reward assigned based on the pIC_{50} of each compound of the data set. The default parameter of the post-hoc score was taken as cut-off value. A total of 80 TR α (Table 2) and 20 TR β (Table 3) hypotheses survived, and only hypotheses with the highest survival-inactive scores were considered for building 3D-QSAR models. All molecules in the dataset were then aligned, matching with at least three pharmacophore features. The selected test set members were considered by the minimum distance from the centroid of each cell in the top map. After analyzing the alignment between active ligands and the generated hypothesis, the best model was DHHNRR.251 of TR α , in which five active molecules in the active set matched with the hypotheses and were selected for further study. Likewise, the built QSAR model generated by hypothesis AHNNRR.25 for TR β had two active molecules in the active set out of five that matched hypotheses. Hypothesis 79 for TR α (Table 2) and 19 for TR β (Table 3) are also shown to allow the reader to draw comparisons. The QSAR model was constructed by an atom-based QSAR modeling method.

Table 1 Pharmacophore hypotheses with variant of thyroid hormone receptors (TR) TR α and TR β

Target protein	Variant no.	Variant	Maximum no. of hypotheses
TR α	1	ADHHNRR	75
	2	AAHHHNR	42
	3	DHHHNRR	64
	4	AAHHNRR	44
	5	AHHHNRR	72
	6	AAHHHRR	30
	7	ADHHHRR	52
	8	ADHHHNR	60
	Total		439
TR β	1	ADHHHRR	6
	2	AAHHNRR	1
	3	AAHHHRR	20
	4	DHHHNRR	1
	5	AHHHNRR	5
	Total		33

The best models for TR α and TR β were selected on the basis of the PLS factor model's minimum SD (standard deviation) and RMSE (root-mean-square error) value for both sets. The PLS regression was achieved by Phase with a maximum of seven PLS factors at a minimum of ten ligands per PLS factor. The legitimacy of each model was tested based on the calculated correlation coefficient (R^2) and variance (F) for the test set. Training and test set ligands were respected in a 4:1 ratio to obtain the best model. " F " enlightens the proportion of the model variance to the observed activity variance. Higher values of " F " specify a more statistically significant regression. Pearson- R value indicates the correlation between the experimental and observed activity for the test set. The " P " value indicates probability, i.e., correlation could occur by chance and signify a level of " F " when considered as a ratio of Chi-squared distributions. Lower values indicate a higher degree of confidence. Cross-validated R^2 (R^2 -CV) value, computed from predictions is obtained by a leave- n -out approach.

It is advocated that Phase works well when $R^2 > 0.4$ [32]. The statistical constraints related to the built atom-based 3D-QSAR model for TR α were as follows: the training set (= 156 compounds) attained an R^2 value of 0.9535, with SD of 0.3016; the test set (= 47 compounds) obtained Q^2 (value of Q^2 for the predicted activities) of 0.4303, squared correlation (random selection R^2 -CV=0.6929), Pearson- R of 0.7294 and RMSE of 0.6342. The P value of 1.411e-96 stated a high degree of self-reliance. The anticipated activities of the training and test set molecules were given in Table 4. The regression line for predicted activity of TR α (cross validated test data set) is displayed in Fig. 1.

A total of 357 molecules was employed for the TR β model, of which 336 were retrieved from the binding database and 21 from the published literature [3].

The entire data set for developing the TR β QSAR model was distributed, randomly choosing 279 compounds to be in the QSAR training set and 71 compounds for the test set. Training and test set were maintained at a one to four ratio in an automatic process. Thus, total 350 molecules were engaged for making the model and seven molecules were discarded because they did not fulfill the criteria to fit the model. The training set yielded an R^2 value 0.9424 with SD (= 0.3719). The squared correlation (random selection R^2 -CV=0.7201), the Pearson- R (= 0.7852), and RMSE for test set predictions (= 0.8630) all strengthen the good predictive competences of the concluding QSAR model for the test set of TR β . The P value, 2.108e-165 specified a high degree of self-reliance for TR β model. The predicted activities of the training and test set molecules are given in Table 5. The regression line for predicted activity of TR β (cross validated test data set) is shown in Fig. 2. Stability of the model predictions to modifications in training set composition have a maximum value of 1.

The Q^2 value of TR α at PLS factor 3 yielded better at the cost of R^2 , so it was omitted. The optimum model of TR β was found at PLS factor 6. The RMSE value of PLS factor 6 was found to be lower (= 0.8630) compared to PLS factor 7 (= 0.9067) and there was very little difference between R^2 values.

The established 3D QSAR model satisfies model requirements and has good correlation with biological activity. After analyzing the statistical values, the QSAR model for TR β was found to be much improved as compared to that of TR α .

QSAR visualization of TR α and TR β models

In the present study, the best QSAR model for inhibitors of both TR α and TR β were developed at PLS factor 6. The selected hypothesis DHHHNRR.251 of TR α contains one hydrogen-bond donor (light blue), three hydrophobic sites (green), one negatively ionizable atom (red) and two aromatic rings (dusky saffron) in Fig. 3a. The 3-D arrangement of pharmacophore sites reveals that the main section of the model is engaged by a hydrophobic pocket. The hydrogen bond donor site is located on one aromatic ring and the negatively ionizable atom on the other.

The pharmacophore hypothesis displaying distance site and angle mapping between the pharmacophoric sites of TR α is shown in Fig. 3b,c. The distances between sites D3-H4 and D3-R8 were 2.169 Å and 3.266 Å, respectively. The details of distances and angles between different groups of TR α are listed in supplementary Table 3.

The chosen hypothesis AHHHNRR.25 was found as the best among 20 hypotheses of the TR β model; this hypothesis is comprised of one hydrogen-bond acceptor (light red), three

Table 2 All 80 pharmacophore hypotheses of TR α with parameters and their values for all 10 matching sites

Sample no.	Hypotheses ID	Survival	Inactive	Survival -inactive	Post-hoc	Site	Vector	Volume	Selectivity	Energy	Activity (pIC ₅₀)
1	DHHHNRR.251	8.693	0.53	8.163	3.511	0.68	0.995	0.837	3.823	2.16	10.39
2	DHHHNRR.246	8.693	0.53	8.163	3.511	0.68	0.995	0.837	3.823	2.16	10.39
3	DHHHNRR.277	8.663	0.704	7.959	3.491	0.67	0.995	0.83	3.814	1.364	10
4	DHHHNRR.284	8.663	0.704	7.959	3.491	0.67	0.995	0.83	3.814	1.364	10
5	DHHHNRR.279	8.65	0.747	7.903	3.485	0.67	0.996	0.82	3.808	2.844	9.85
6	DHHHNRR.288	8.65	0.747	7.903	3.485	0.67	0.996	0.82	3.808	2.844	9.85
7	DHHHNRR.415	8.656	0.794	7.862	3.496	0.68	0.994	0.825	3.802	4.376	10.39
8	DHHHNRR.418	8.657	0.81	7.848	3.486	0.67	0.994	0.821	3.814	2.421	10.39
9	DHHHNRR.416	8.626	0.795	7.831	3.475	0.67	0.994	0.812	3.793	2.519	9.62
10	DHHHNRR.420	8.639	0.813	7.827	3.478	0.67	0.994	0.817	3.804	3.066	9.62
11	AHHHNRR.101	8.58	0.843	7.737	3.48	0.68	0.995	0.809	3.742	2.728	9.85
12	AHHHNRR.94	8.58	0.843	7.737	3.48	0.68	0.995	0.809	3.742	2.728	9.85
13	ADHHNRR.44	8.535	0.869	7.665	3.545	0.72	0.995	0.825	3.632	2.844	9.85
14	ADHHNRR.45	8.535	0.869	7.665	3.545	0.72	0.995	0.825	3.632	2.844	9.85
15	ADHHNRR.48	8.529	0.871	7.658	3.539	0.72	0.994	0.823	3.632	3.066	9.62
16	ADHHNRR.49	8.529	0.871	7.658	3.539	0.72	0.994	0.823	3.632	3.066	9.62
17	ADHHNRR.51	8.513	0.858	7.655	3.523	0.71	0.995	0.818	3.632	2.728	9.85
18	ADHHNRR.52	8.513	0.858	7.655	3.523	0.71	0.995	0.818	3.632	2.728	9.85
19	ADHHNRR.47	8.508	0.867	7.642	3.514	0.71	0.995	0.811	3.636	2.311	9.62
20	ADHHNRR.50	8.508	0.867	7.642	3.514	0.71	0.995	0.811	3.636	2.311	9.62
21	ADHHNRR.240	8.517	0.879	7.638	3.533	0.72	0.996	0.82	3.626	2.948	9.85
22	ADHHNRR.234	8.517	0.879	7.638	3.533	0.72	0.996	0.82	3.626	2.948	9.85
23	AHHHNRR.96	8.578	1.011	7.567	3.474	0.68	0.993	0.802	3.745	2.838	9.62
24	AHHHNRR.102	8.578	1.011	7.567	3.474	0.68	0.993	0.802	3.745	2.838	9.62
25	AAHHNRR.141	8.491	0.947	7.544	3.481	0.67	0.997	0.815	3.651	2.519	9.62
26	ADHHNRR.130	8.46	0.941	7.52	3.482	0.66	0.996	0.823	3.62	2.844	9.85
27	ADHHNRR.124	8.46	0.941	7.52	3.482	0.66	0.996	0.823	3.62	2.844	9.85
28	AHHHNRR.104	8.588	1.084	7.504	3.496	0.69	0.995	0.814	3.734	2.844	9.85
29	AHHHNRR.97	8.588	1.084	7.504	3.496	0.69	0.995	0.814	3.734	2.844	9.85
30	AHHHNRR.144	8.592	1.112	7.48	3.499	0.69	0.994	0.813	3.735	3.066	9.62
31	AHHHNRR.477	8.652	1.184	7.468	3.494	0.67	0.994	0.829	3.8	4.328	10.39
32	AHHHNRR.143	8.556	1.099	7.457	3.464	0.68	0.99	0.799	3.734	2.969	9.62
33	AHHHNRR.131	8.556	1.099	7.457	3.464	0.68	0.99	0.799	3.734	2.969	9.62
34	AAHHNRR.28	8.595	1.168	7.426	3.547	0.73	0.996	0.822	3.69	2.597	9.85
35	AAHHNRR.27	8.595	1.168	7.426	3.547	0.73	0.996	0.822	3.69	2.597	9.85
36	ADHHNRR.186	8.458	1.034	7.423	3.724	0.87	0.999	0.856	3.376	2.203	10
37	ADHHNRR.193	8.458	1.034	7.423	3.724	0.87	0.999	0.856	3.376	2.203	10
38	ADHHNRR.156	8.366	0.953	7.413	3.496	0.69	0.996	0.813	3.513	2.844	9.85
39	ADHHNRR.150	8.366	0.953	7.413	3.496	0.69	0.996	0.813	3.513	2.844	9.85
40	ADHHNRR.153	8.331	0.947	7.384	3.474	0.66	0.996	0.814	3.499	2.597	9.85
41	ADHHNRR.149	8.331	0.947	7.384	3.474	0.66	0.996	0.814	3.499	2.597	9.85
42	AAHHNRR.129	8.526	1.169	7.357	3.485	0.67	0.996	0.815	3.683	2.597	9.85
43	AAHHNRR.128	8.515	1.168	7.348	3.475	0.67	0.995	0.814	3.682	2.928	9.62
44	AAHHNRR.148	8.521	1.176	7.345	3.493	0.7	0.992	0.804	3.67	2.519	9.62
45	ADHHNRR.216	8.362	1.019	7.342	3.489	0.68	0.995	0.81	3.514	3.066	9.62
46	AAHHNRR.131	8.534	1.193	7.341	3.502	0.7	0.993	0.81	3.673	2.844	9.85
47	AAHHNRR.146	8.502	1.167	7.335	3.468	0.68	0.99	0.798	3.677	2.43	9.62
48	AAHHNRR.130	8.513	1.183	7.331	3.477	0.67	0.996	0.813	3.678	2.928	9.62
49	AAHHNRR.127	8.542	1.216	7.326	3.498	0.68	0.997	0.817	3.686	2.844	9.85

Table 2 (continued)

Sample no.	Hypotheses ID	Survival	Inactive	Survival -inactive	Post-hoc	Site	Vector	Volume	Selectivity	Energy	Activity (pIC50)
50	ADHHHNR.139	8.346	1.032	7.313	3.476	0.68	0.996	0.804	3.512	2.311	9.62
51	ADHHHNR.137	8.346	1.032	7.313	3.476	0.68	0.996	0.804	3.512	2.311	9.62
52	ADHHHNR.214	8.346	1.036	7.31	3.486	0.69	0.994	0.805	3.502	2.519	9.62
53	ADHHHRR.164	8.419	1.111	7.307	3.697	0.86	0.999	0.838	3.364	4.634	10.39
54	ADHHHRR.163	8.419	1.111	7.307	3.697	0.86	0.999	0.838	3.364	4.634	10.39
55	AAHHNRR.132	8.531	1.226	7.305	3.492	0.69	0.996	0.808	3.681	2.519	9.62
56	ADHHHRR.151	8.432	1.13	7.302	3.71	0.86	0.999	0.853	3.364	2.16	10.39
57	ADHHHRR.158	8.432	1.13	7.302	3.71	0.86	0.999	0.853	3.364	2.16	10.39
58	ADHHHRR.188	8.438	1.161	7.277	3.726	0.88	0.998	0.847	3.355	0.237	10
59	ADHHHRR.195	8.438	1.161	7.277	3.726	0.88	0.998	0.847	3.355	0.237	10
60	AAHHHNR.108	8.411	1.141	7.269	3.499	0.69	0.997	0.812	3.554	2.844	9.854
61	AAHHHNR.104	8.411	1.141	7.269	3.499	0.69	0.997	0.812	3.554	2.844	9.854
62	ADHHHRR.275	8.412	1.164	7.248	3.705	0.86	0.998	0.846	3.349	1.801	10.39
63	ADHHHRR.251	8.412	1.164	7.248	3.705	0.86	0.998	0.846	3.349	1.801	10.39
64	AAHHHNR.144	8.403	1.156	7.247	3.495	0.69	0.997	0.804	3.55	2.519	9.62
65	AAHHHNR.107	8.381	1.157	7.224	3.477	0.67	0.998	0.813	3.546	2.928	9.62
66	AAHHHNR.105	8.381	1.157	7.224	3.477	0.67	0.998	0.813	3.546	2.928	9.62
67	AAHHHNR.131	8.369	1.15	7.219	3.46	0.67	0.994	0.797	3.551	2.43	9.62
68	AAHHHNR.143	8.369	1.15	7.219	3.46	0.67	0.994	0.797	3.551	2.43	9.62
69	AAHHHNR.94	8.384	1.176	7.208	3.475	0.68	0.995	0.803	3.551	2.311	9.62
70	AAHHHNR.98	8.384	1.176	7.208	3.475	0.68	0.995	0.803	3.551	2.311	9.62
71	AAHHHRR.122	8.287	1.236	7.051	3.727	0.87	0.999	0.856	3.202	2.203	10
72	AAHHHRR.129	8.287	1.236	7.051	3.727	0.87	0.999	0.856	3.202	2.203	10
73	AAHHHRR.127	8.259	1.224	7.035	3.71	0.86	0.999	0.85	3.191	2.16	10.39
74	AAHHHRR.133	8.259	1.224	7.035	3.71	0.86	0.999	0.85	3.191	2.16	10.39
75	AAHHHRR.139	8.246	1.222	7.024	3.697	0.86	0.999	0.837	3.191	4.634	10.39
76	AAHHHRR.140	8.246	1.222	7.024	3.697	0.86	0.999	0.837	3.191	4.634	10.39
77	AAHHHRR.124	8.269	1.245	7.024	3.727	0.88	0.998	0.848	3.183	0.237	10
78	AAHHHRR.131	8.269	1.245	7.024	3.727	0.88	0.998	0.848	3.183	0.237	10
79	AAHHHRR.298	8.238	1.215	7.023	3.706	0.86	0.999	0.845	3.174	1.801	10.39
80	AAHHHRR.338	8.238	1.215	7.023	3.706	0.86	0.999	0.845	3.174	1.801	10.39

hydrophobic (green), one negatively ionizable atom (red) and two aromatic rings (dusky saffron) (Fig. 3d). The pharmacophore hypothesis displaying distance site and angle mapping between the pharmacophoric sites of TR β is shown in Fig. 3e,f. The distance between sites A2–R9 and A2–H6 were 2.77 Å and 3.238 Å, respectively. The details of different distances and angles between different groups of TR β are given in supplementary Table 4. The spatial arrangement of the pharmacophore sites shows a major area of the model being occupied by a hydrophobic pocket. The hydrogen bond acceptor site is situated on one aromatic ring and a negatively ionizable atom at the other ring with a COOH⁻ group.

All the active compounds (ten active compounds for TR α and five for TR β) used for building the model fitted and occupied into the similar spatial arrangement is illustrated in Fig. 4a,b. The mapping reveals both that the model satisfied

the hypothesis and that all active molecules taken for the training set are aligned.

Pictorial representations of volume occupied maps (hydrophobic, donor, aromatic ring and electron-withdrawing features) for TR α hypothesis DHHHNR.251 and TR β hypothesis AHHHNR.25 are shown in Fig. 5. The reference compounds for TR α and TR β are 3-[4-(4-hydroxy-3-iodophenoxy)-3, 5-diiodophenyl] propanoic acid (PCID 5804), 2-[3, 5-dibromo-4-(4-hydroxy-3-propan-2-ylphenoxy) phenyl] acetic acid (PCID 9933119), respectively. Favorable and unfavorable contacts are displayed in blue and red cubes in the contour map. The volume occupied map is helpful to identify significant sites that necessitate a convinced physico-chemical property for developing an effective antagonist drug for hyper level thyroid hormone secretion as discussed in the following section.

Table 3 Twenty hypotheses generated to build three dimensional quantitative structure–activity relationship (3D-QSAR) model of TR β for five matching sites with their parameters

Sample no.	Hypotheses ID	Survival	Inactive	Survival -inactive	Post-hoc	Site	Vector	Volume	Selectivity	Energy	Activity
1	AHHHNRR.25	7.615	1.079	6.536	3.351	0.6	0.957	0.796	3.8	2.266	10.721
2	AHHHNRR.26	7.598	1.075	6.522	3.335	0.58	0.964	0.79	3.799	2.279	10.721
3	AHHHNRR.1	7.585	1.108	6.477	3.309	0.56	0.964	0.789	3.812	4.588	10.602
4	ADHHHRR.3	7.239	0.975	6.264	3.404	0.72	0.899	0.788	3.371	2.337	10.602
5	ADHHHRR.4	7.239	0.975	6.264	3.404	0.72	0.899	0.788	3.371	2.337	10.602
6	AAHHHRR.20	7.352	1.094	6.258	3.686	0.88	0.975	0.833	3.202	0	10.022
7	AAHHHRR.16	7.352	1.094	6.258	3.686	0.88	0.975	0.833	3.202	0	10.022
8	AAHHHRR.22	7.341	1.132	6.209	3.693	0.88	0.976	0.839	3.183	0.284	10.022
9	AAHHHRR.18	7.341	1.132	6.209	3.693	0.88	0.976	0.839	3.183	0.284	10.022
10	AHHHNRR.4	7.282	1.092	6.19	3.022	0.38	0.919	0.724	3.796	0.237	10.602
11	AHHHNRR.3	7.282	1.092	6.19	3.022	0.38	0.919	0.724	3.796	0.237	10.602
12	AHHHRR.172	7.293	1.136	6.157	3.655	0.84	0.976	0.837	3.174	0.151	10.319
13	AHHHRR.196	7.293	1.136	6.157	3.655	0.84	0.976	0.837	3.174	0.151	10.319
14	ADHHHRR.42	7.231	1.087	6.143	3.411	0.71	0.897	0.8	3.355	4.588	10.602
15	ADHHHRR.46	7.231	1.087	6.143	3.411	0.71	0.897	0.8	3.355	4.588	10.602
16	AHHHRR.211	7.272	1.149	6.123	3.638	0.84	0.973	0.829	3.17	3.561	10.022
17	AHHHRR.187	7.272	1.149	6.123	3.638	0.84	0.973	0.829	3.17	3.561	10.022
18	AAHHNRR.1	7.23	1.176	6.054	3.084	0.42	0.947	0.713	3.682	3.493	10.721
19	AAHHHRR.21	7.171	1.122	6.049	3.522	0.83	0.976	0.711	3.185	0.242	10.222
20	AAHHHRR.17	7.171	1.122	6.049	3.522	0.83	0.976	0.711	3.185	0.242	10.222

Hydrogen bond donor

The volume occlusion map for the hydrogen bond donor (HBD) of TR α explains the favorable 3D arrangement of hydrogen bonding or non-covalent interactions with acceptor groups of the protein. The structure of the HBD acquired from the QSAR model could be applied to the most active ligands (Fig. 5a). HBD favorable blue cubes are present near the fourth position of the benzene (R8) group of the model. All highly active molecules contain the HBD group (OH) at their fourth position in molecules PCID 5804, 10456830, 23648079, 5803, 71212, 7048703, 10904700, 9933119, 9951790 and 10479779. The unfavorable red contours are not found in this volume occupied region. These HBD groups

also show H-bonding contacts with His381 in docking (Fig. 6) with the most active ligand PCID (5804) but the docking score with (2-[4-(5-bromo-6-hydroxynaphthalen-1-yl)-3, 5-dichloroanilino]-2-oxoacetic acid) PCID(10456672) was the best compared to that mentioned above, which is a moderately active ligand, thereby indicating the importance of this interaction for the activity of all molecules. This result thus justifies the occurrence of HBD blue regions at this site.

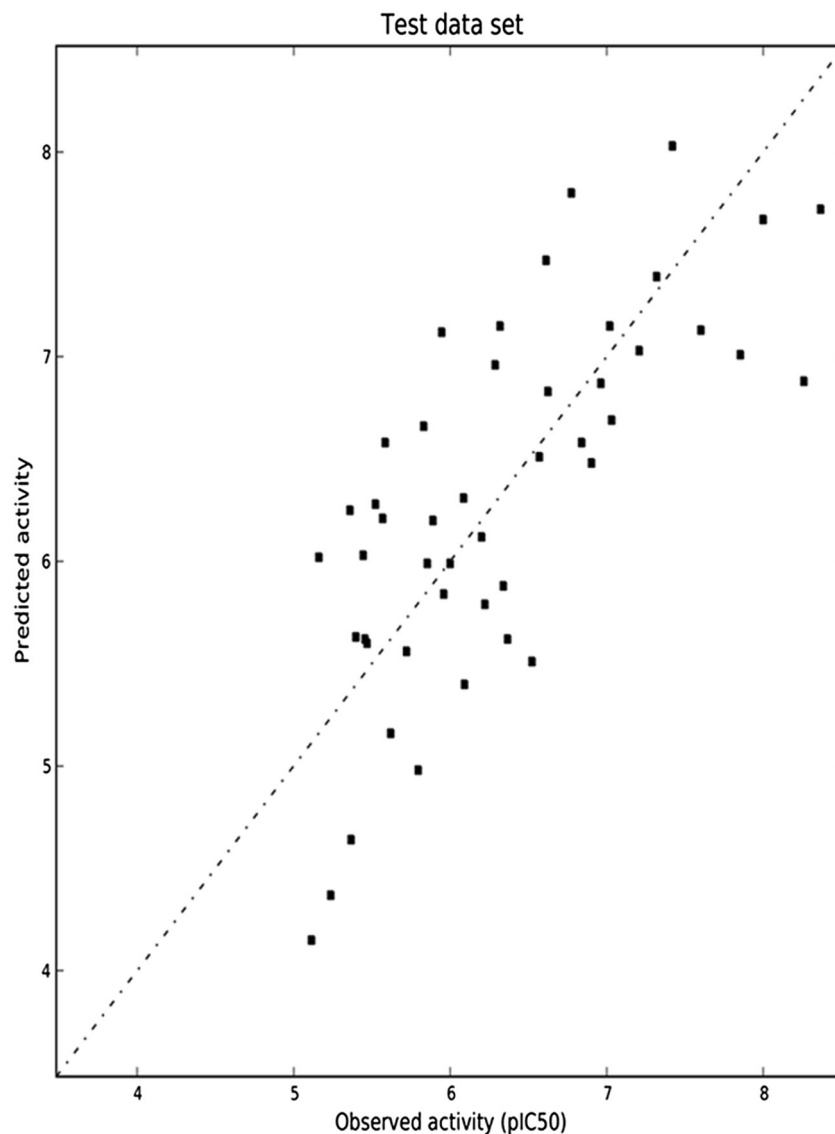
Hydrophobic/non-polar property

The hydrophobic property of the TR α 3D-QSAR model was applied to the most active compound PCID (5804) as shown in Fig. 5b. In this figure, favorable blue maps are seen near the

Table 4 Statistical analysis of 3D-QSAR model for TR α . PLS Partial least squares, SD standard deviation, RMSE root-mean-square error

PLS factor	SD	R ²	F	P	R ² -CV	Stability	RMSE	Q ²	R -Pearson	Opt model
1	0.9493	0.5238	169.4	1.36E-26	0.4703	0.9918	0.7511	0.2009	0.5729	
2	0.7268	0.7227	199.3	2.46E-43	0.5845	0.9532	0.589	0.5085	0.7772	
3	0.6128	0.8041	208	1.37E-53	0.6034	0.9126	0.5555	0.5629	0.7977	
4	0.4571	0.8917	310.9	8.65E-72	0.6601	0.8329	0.6298	0.4382	0.7307	
5	0.3756	0.9274	383.1	1.70E-83	0.6819	0.8038	0.6274	0.4425	0.7223	
6	0.3016	0.9535	509.1	1.41E-96	0.6929	0.7917	0.6342	0.4303	0.7294	√
7	0.2675	0.9637	560.6	4.07E-103	0.6949	0.7759	0.6247	0.4471	0.7344	

Fig. 1 Scatter plot of cross validated predicted values of thyroid hormone receptor α (TR α)



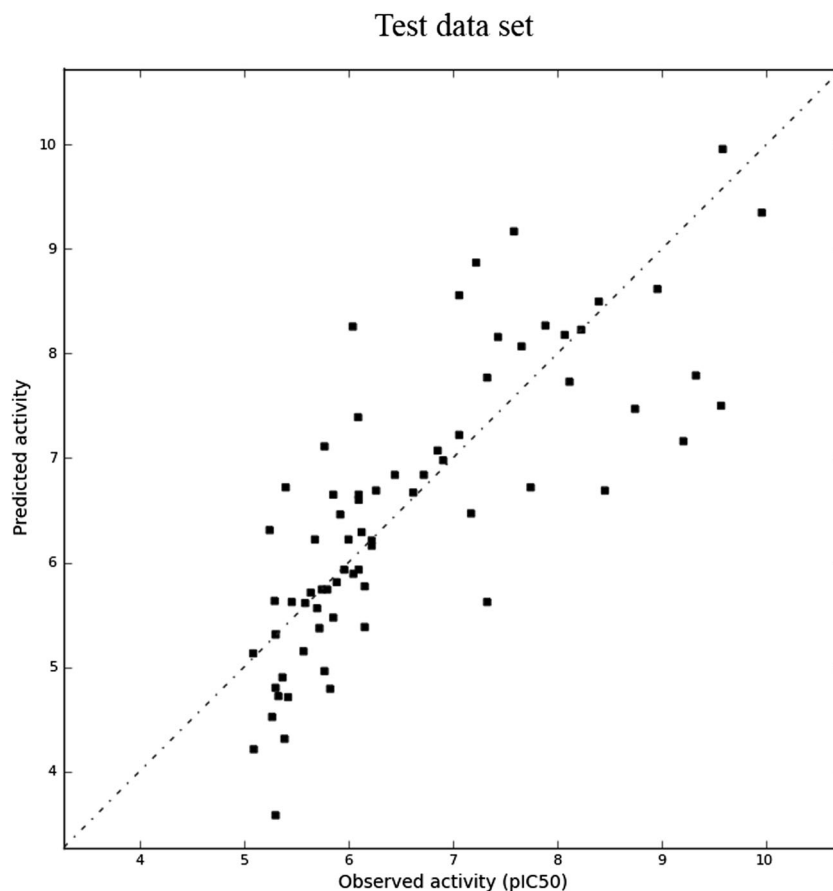
third position of the benzene ring (R8), and the third and fifth positions of the phenyl ring (R9) in active molecules. On the other hand, hydrophobic unfavorable red contours are visualized near the ortho- position of the benzene ring (R8) in most of the inactive molecules. Hydrophobic favorable blue

contours are also shown near the first position of the diiodophenyl ring in active molecules. This location is especially favorable for hydrophobic aliphatic chains having a carboxylic group at their terminus. This noticeably specifies that hydrophobic substitutions can be acknowledged at this

Table 5 Statistical analysis of 3D-QSAR for TR β

No. of factors	SD	R^2	F	P	R^2 -CV	Stability	RMSE	Q^2	R -Pearson	Opt model
1	0.9669	0.6037	422	1.30E-57	0.5743	0.9971	0.925	0.4642	0.7036	
2	0.738	0.7699	461.9	8.55E-89	0.6639	0.9678	0.8443	0.5536	0.7682	
3	0.6283	0.8339	460.1	7.80E-107	0.6852	0.9443	0.8671	0.5292	0.7743	
4	0.5017	0.8944	580.4	2.04E-132	0.6977	0.8943	0.864	0.5325	0.7881	
5	0.4139	0.9284	708.2	5.23E-154	0.7031	0.8445	0.8714	0.5244	0.7737	
6	0.3719	0.9424	741.9	2.11E-165	0.7201	0.833	0.863	0.5336	0.7852	√
7	0.3194	0.9577	876.4	4.47E-182	0.7172	0.8077	0.9067	0.4852	0.7717	

Fig. 2 Scatter plot of cross validated predicted values of TR β



position to increase activity. The appearance of unfavorable red cubes at approximately similar sites in inactive molecules revealed that massive hydrophobic assemblies at this position are harmful for activity. Favorable blue contours are present at the second position of the diiodophenyl ring. Thus, hydrophobic groups affixed through a linker are favorable at this site, whereas the non-appearance of hydrophobic groups at the same site is likely to reduce the biological activity. A red contour located not too close to any of the atoms of the compounds suggests that occupancy of this spatial region by a hydrogen bond acceptor group would cause a decrease in activity.

Figure 5e shows the significant favorable regions and unfavorable hydrophobic/non-polar contacts that arise with the TR β 3D-QSAR model applied to the most active ligand of PCID (9933119). Favorable blue regions are seen near the third position (H4) and fifth (H6) position of the benzene ring (R8), and the third position (H9) of the phenyl ring (R9) in active molecules. On the other hand, hydrophobic unfavorable red regions are visualized near the aliphatic chain associated with the phenyl ring (R9) in most inactive molecules. Hydrophobic favorable blue contours are shown more closer to the dibromophenyl ring in active molecules. This place is mainly favorable for hydrophobic aliphatic chains containing a carboxylic group at the terminus. This favors the view that

hydrophobic substitutions can be agreed at this position to enhance the activity of the molecule. The presence of red cubes at a nearly comparable position in inactive compounds suggests that vast hydrophobic assemblies at this site are detrimental for activity.

Electron withdrawing and hydrogen bond acceptor

An electron withdrawing (EW) property contour map for TR α inhibitors is shown in Fig. 5c. Electronegative atoms such as N, O, S and halogen along with hydrogen bond acceptor (HBA) groups were considered under the EW property map using atom-based QSAR. EW group blue regions are seen close to atoms associated with N7 groups. In highly active molecules, favorable blue cubes are present near the 'O' atom of the 3-propanoic acid group attached to the position of the phenyl ring (R9), and 'O' atoms form H-bonds with Arg266 in docking analysis acting as HBA (Fig. 6). Favorable blue cubes are also present near the fourth position of the benzene ring (R8) in all active molecules, thus suggesting the importance of this interaction for a molecule's inhibitory activity for TR α .

EW/HBA properties of TR β inhibitors are displayed in Fig. 5c. HBA group favorable regions are observed close to atoms associated with the carboxylic group of the

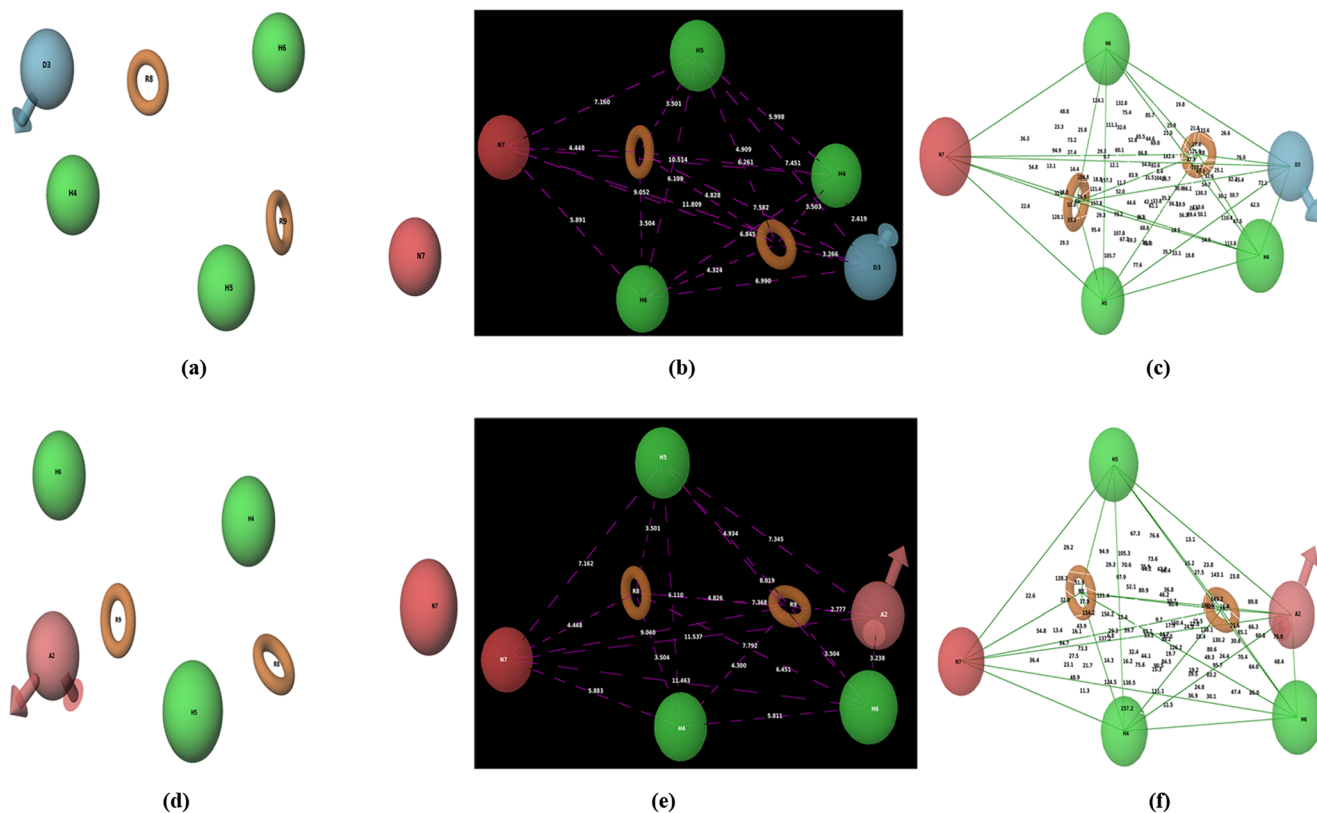


Fig. 3 (a) Pharmacophore mapping of hypothesis DHHNRR.251 for TR α (b) Distance site mapping (Å) for TR α (c) Angle mapping for TR α (d) Pharmacophore mapping of hypothesis AHHNRR.25 for TR β (e) Distance mapping of TR β (Å) (f) Angle site mapping for TR β model

dibromophenyl ring (R8). In highly active ligands, blue cubes are seen near the ‘O’ atom of the acetic acid group attached to the dibromophenyl ring (R8) and forming H-bonds with amino acid Asn331 in docking analysis acting as HBA (Fig. 7d). EW property favorable regions are seen near the aliphatic chain associated with the phenyl ring (R9) in all active molecules. Analysis of this interaction is important for a molecule’s inhibitory activity for TR β .

Negative ionizable property

The negative ionizable property contour map of TR β is shown in Fig. 5d. Negative ionizable group favorable blue cubes are found adjacent to the carboxylic group attached to the benzene ring (R8). All highly active compounds examined in this study bear a negative ionizable ‘O’ atom at this position. The negative ionizable property contour map of TR α is shown in

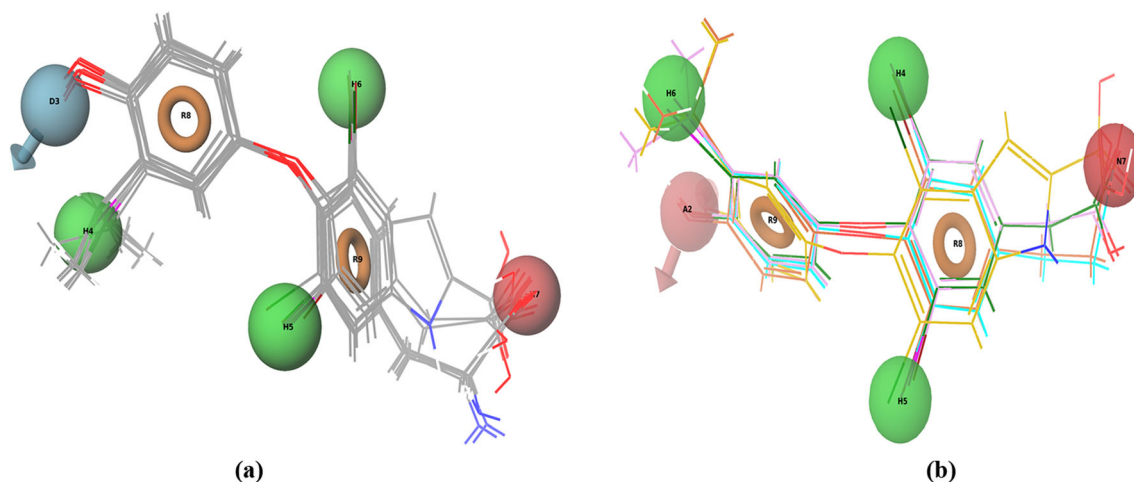


Fig. 4 (a) Alignment of the ten active compounds for the TR α model (b) Superposition of five active compounds for the TR β model

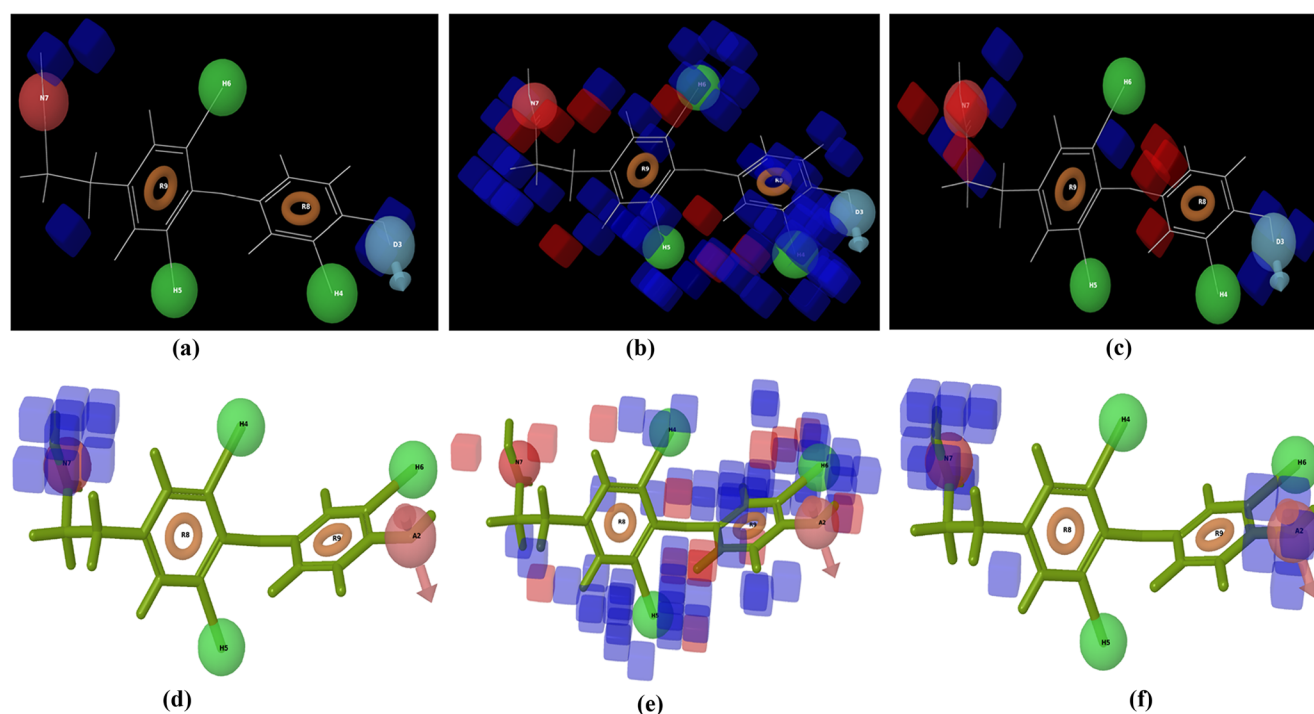


Fig. 5 3D-QSAR visualization of contour map for TR α and TR β . (a) Hydrogen bond donor, (b) hydrophobic donor/non-polar (c) electron-withdrawing features, of TR α hypothesis (DHHNRR), and (d)

Negative ionic effect (e) hydrophobic donor/non-polar (f) electron-withdrawing properties, of TR β hypothesis (AHHNRR)

supplementary Fig. 1. Negative ionizable group favorable blue cubes are seen adjacent to the phenyl ring (R9) and aliphatic chain.

For TR β , the active ligands are quite sensitive to steric bulky groups or expansions to the phenyl ring, which lead to a

decrease in activity of the molecules. However, a propyl group attached to the carboxylic group maintained the activity. The expansion of ring size, i.e., *n*-propyl to cyclo-pentyl, reduces the activity. Halogen atoms along with two benzene rings are important for maintaining the activity of the molecules.

These results provide perceptions of the structural requirements of the most active compound in our dataset, which in turn can help in the rational drug design of novel TR antagonist derivatives. This analysis opens up the possibility of predicting the biological activity of new compounds and their active sites inside their receptors through pharmacophore-based virtual screening and interaction analysis.

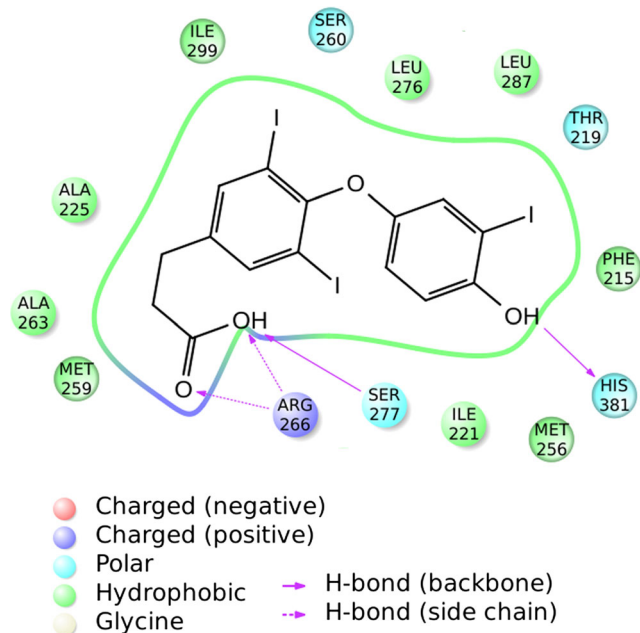


Fig. 6 Two-dimensional (2D) docked view of ligand 3-[4-(4-hydroxy-3-iodophenoxy)-3, 5-diiodophenyl] propanoic acid (PCID 5804) with TR α receptor

External validation of 3D-QSAR model

For external validation of the TR α hypothesis DHHNRR, a 3D-phase database of 66 compounds was created that includes 15 molecules of TR α inhibitors and 51 known thyroid oxidase 2 inhibitors, which are provided in the supplementary Table 5. The structures of known inhibitors were retrieved from the binding database. The source organism taken for TR α was *Rattus norvegicus* (mammal) due to the non-availability of a human data set in the binding database; *R. norvegicus* was considered for our study because it belongs to the mammalian class. Only one molecule (T3) made a good hit, with a fitness value of 2.39 (supplementary Table 6), giving a good predictive value of being an active molecule for TR α . A fitness value range of 1.89 (minimum) to 3 (maximum) was used in

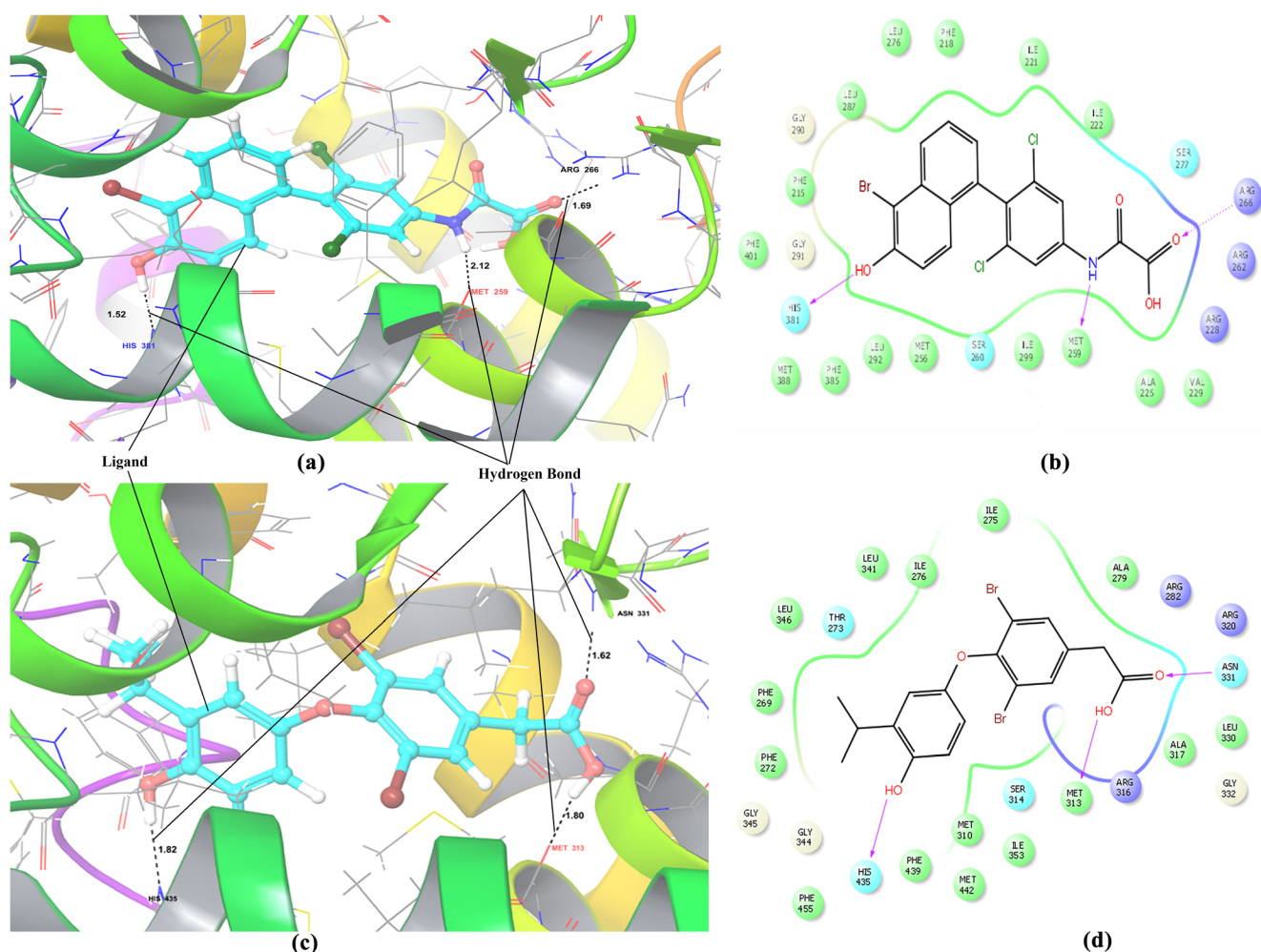


Fig. 7 Docking analysis of TR α and TR β . **(a)** Ligand and protein interaction hydrogen bond interaction view (black dotted lines hydrogen bonds) **(b)** 2D docked view of ligand 2-[4-(5-bromo-6-hydroxynaphthalen-1-yl)-3,5-dichloroanilino]-2-oxoacetic acid (PCID

10456672) with TR α receptor **(c)** Ligand and protein interaction view with hydrogen bond **(d)** 2D docked view of 2-[3,5-dibromo-4-(4-hydroxy-3-propan-2-ylphenoxy) phenyl] acetic acid (PCID9933119) ligand with TR β receptor

our study to build the 3D-QSAR model (supplementary Table 1).

For external validation, a 3D-phase database of 197 compounds was created including 146 known TR β inhibitors (138 from *Homo sapiens* and 08 from *R. norvegicus*), the dataset for which was retrieved from the binding database (given in supplementary Table 7). The structures of known inhibitors were taken from the binding database. The other 51 compounds, thyroid oxidase 2, employed in the test set used for external validation were selected from the binding database, with care taken to ensure that none of the molecules was structurally correlated with those in the original training and test sets. When the AHHNRR.25 hypothesis was used as the query to build the 3D phase database, 18 molecules (supplementary Table 8) were found as hits that target human TR β , but only 11 of them gave good predictive value as the range of fitness values in the built model was predicted to lie in the range of 1.78 (minimum) to 3 (maximum) (supplementary

Table 2). None of the thyroid oxidase 2 inhibitors made hits. The biological activity of these molecules was also predicted for all seven PLS factors. Hits with the highest fitness score were retained. The score assortment for the vector score is dispersed between -1.0 and 1.0 , for the alignment score 0.0 to 1.0 , and for the volume score 0.0 to 1.0 . For all three terms, the weight range is 0.0 to 1.0 , with a default weight of 1.0 . The fitness score varied from -1.0 to 3.0 and is a linear combination of the site and vector alignment scores, in addition to the volume score. The fitness score is a parameter that explains how well the aligned compound conformer coordinates with the hypothesis built on RMSD site matching, vector alignments and volume terms.

Molecular docking analysis

Flexible molecular docking of all the ligand molecules was performed, but only 139 molecules were successful in

molecular docking. Docking on the whole pharma-set was carried out at the active site Arg 266, Ser 277 and His 381 of TR α (Pdb Id: 3JZB). The most active compound (PCID 5804) was found to have a gscore of -8.731492 . As can be seen clearly in Fig. 6, there are two hydrogen bond interactions between the carboxylic group ('O' atoms) of inhibitors and the NH of Arg266. It is also observed that NH of Ser 277 forms a hydrogen bond with the oxygen atom of the ligand group. The 'N' atom of imidazole group of His 381 makes a hydrogen bond at fourth position of the OH group with the ligand. The top gscore of compound PCID 10456672 was obtained as -9.89261 at the cost of activity. The best possible pose view of ligand 10456672 in the TR α binding site is shown in Fig. 7a and the corresponding 2D-view of the ligand-protein interaction is displayed in Fig. 7b. The hydrophobic surrounding is observed at the distal position of the ligand, shown in green circles labeled with the three-letter code of the amino acid. Note that three polar residues, viz. Ser 260, Ser 277 and His 381 (circled in cyan), two glycine residues at positions 290–291 and charged positive residue Arg 266 play an important role in the interaction. The binding contacts arrangements noted during docking analyses are in agreement with that of HBD and HBA contour maps. Docking gscores of TR α inhibitors with receptors are given in supplementary Table 9.

Figure 7c shows the best possible pose view for ligand PCID (9933119). The most active compound among the whole data set was docked into the binding site of the TR β receptor (PDB Id: 1NAX). Figure 7d shows the 2D representation of the ligand-protein interaction of the same. The predicted gscore of ligand PCID 9933119 was -11.52 . It can be seen clearly from Fig. 7d that the ligand molecule is surrounded by hydrophobic residues Ile 275, Ile 276, Ala 279, Met 310, Met 313, and Ala 317, mainly through hydrophobic interactions. The charged positive residues Arg 282 and Arg 316 also surround the ligand molecule. In addition, the two oxygen atoms of the terminus of the ring form three hydrogen bonds to Asn 331 as HBAs and with Met 313 and His 435 as HBD. Therefore, the inhibitor is stabilized in the binding pocket by H-bond interactions. It should be emphasized that analysis of the docking modes of the inhibitors PCID (9933119) in terms of gscore as well as activity specifies the establishment of more extensive contact forms.

Thus, it can be concluded that a ligand containing a halogen or a more electronegative atom can result in increased activity. This is in agreement with our contour maps as more favorable regions (blue cubes) are observed at that location. Another hydrophobic interaction was formed considering mainly Ile 275, Ile 276, Ala 279, Met 310 Met 313, and Ala317 amino acid residues, so a benzene ring in this position can increase activity. Out of 357 ligands, only 97 succeeded in docking. The succeeded docking gscores of these compounds with 1NAX protein are given in supplementary Table 10.

Comparative docking studies suggest that the TR β receptor is a better potential target compared to TR α because of the binding energy and biological activity of inhibitors towards this target.

Molecular dynamics simulation analysis of TR α and TR β inhibitor complexes

The MD simulation of docking complexes, compound 10456672-TR α and compound9933119-TR β receptor, was performed for 5 ns to confirm the dynamic stability and binding modes to uncover the conformational changes occurring. Figures 8 and 9 show the RMSD trajectory, RMSF, protein-ligand contacts (PLC) and ligand-protein contacts (LPC) of both complexes. Fluctuations within 1–3 Å RMSD values are perfectly acceptable for small proteins [33–35].

Ligand 10456672-TR α -complex reaches 1.5 to 3 Å RMSD from initial to 1.8 ns and undergoes a conformational change during the MD simulation as reflected by its RMSD value (Fig. 8a). During MD simulation from 1.8 to 5 ns, the RMSD was found to be relatively stable about a 3-Å RMSD value. The mean value of the TR α -complex of the backbone atom was found to be 2.212 Å. A small perturbation in the initial state was observed. Figure 9a displays the RMSD evolution of a protein (left y-axis) and with ligand added (right y-axis). All protein frames are aligned on the backbone. From initial to 3.8 ns, the compound 9933119-TR β -complex trajectory was observed to have an average RMSD of about 1.5 Å. During 3.80–3.90 ns MD simulation the backbone protein of TR β complex (red) underwent a slightly higher conformational variation due to ligand and protein contacts and later attained a second plateau after 3.9 ns simulation at about 2.5 Å RMSD. The range of backbone TR β complexes was found between 0 to 2.708 Å and the mean value was observed at 1.742 Å. The average RMSD of heavy atoms and side chain atoms were found to be 2.028 and 2.582 Å, respectively. It is observed from the above plot that the ligand RMSD is lower than that of the protein backbone atoms. Ligand RMSD attains two plateaus, first from initial to 1 Å and later after 3.9 ns simulation at about 1.75 Å. Lig fit Lig RMSD value calculates the internal fluctuations of the ligand atoms and is aligned on its reference conformation. Ligand RMSD specifies that the ligand is stable with respect to the protein and its binding pocket during the course of simulation.

Figures 8b and 9b depict the analysis of RMSF versus the residue number for TR α (aa 144–407) and TR β (211–460), respectively. RMSF is beneficial for describing local changes along the protein chain. The analysis of RMSF is important for the explicit relationship between the inhibitory capability of a ligand and the folding pattern of a helix. Alpha-helical and beta-strand regions are given in red and blue backgrounds, respectively. These regions are explained by helices or strands that continue over 70 % of the entire simulation. In Figs. 8b

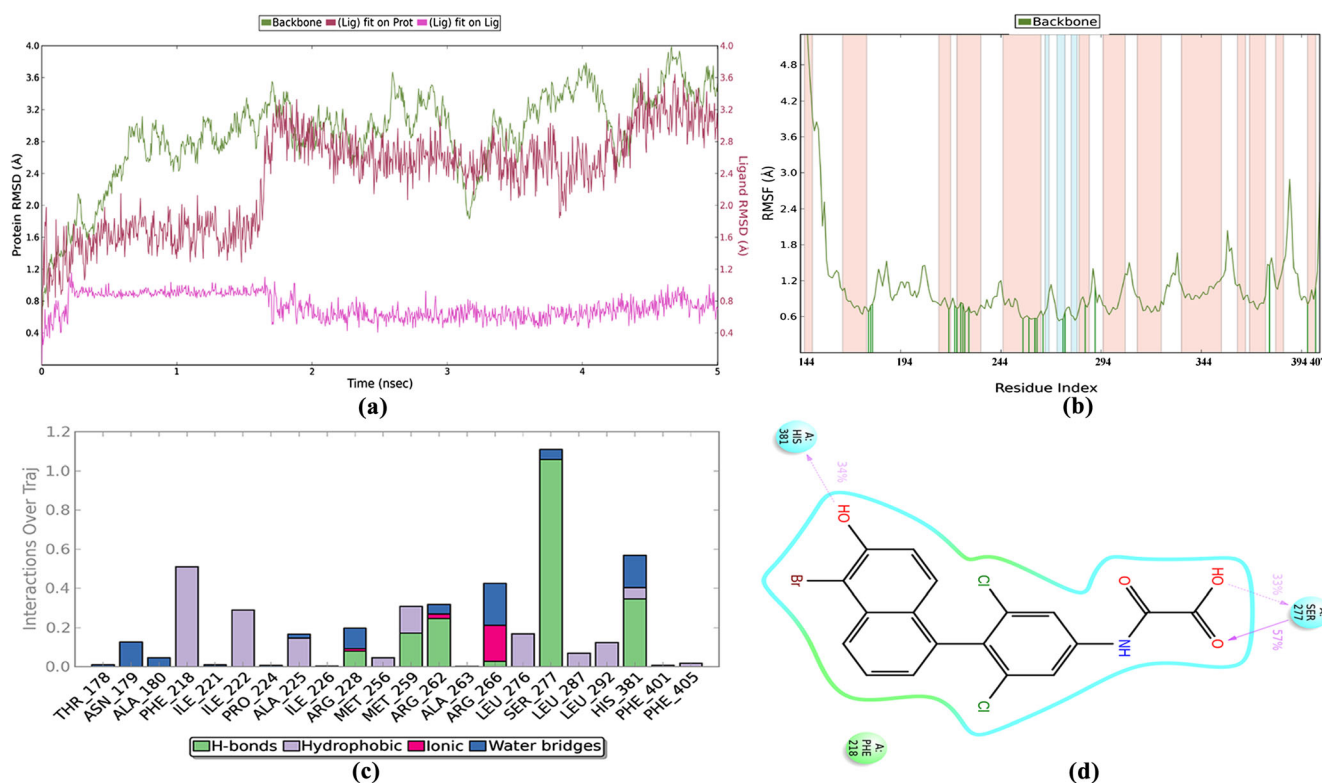


Fig. 8 Results of molecular dynamics (MD) simulation analysis. For TR α complex (a) MD simulation time vs RMSD in Å for all residues shown in different colours: *light green* backbone, *red* along with ligand fit on protein, and *pink* ligand fit on ligand (b) RMSF vs residues index for

TR α (c) Protein ligand contact interaction over trajectory (d) Average conformation of the binding pocket of ligand 10456672-TR α complex throughout the simulation

and 9b, peaks show the segments of protein that fluctuate the most during the simulation. Typically, the tails (N- and C-terminal) fluctuate more than any other part of the protein. It is seen from the plot that alpha helices (red) and beta (blue) strands are more rigid and have less fluctuation than the unstructured part loop regions (white region) for both TR α and TR β complexes. Protein residues that interact with the ligand are manifested as green vertical bars. It can be observed that both TR α and TR β proteins alter the conformation, which agrees with the result of RMSF. Two major residue segments, 357–362 and 391–400 and flexible protein regions of the backbone atom (green) can be seen in TR α protein but only one is seen in TR β protein (251–263), which indicates that TR β protein is more stable than TR α . The fluctuation of TR β protein is seen at Asn 257 with a backbone RMSF value of 4.17 Å, which could be due to a bend or turn.

PLCs for TR α are classified into four types: viz. hydrogen bonds, and hydrophobic, ionic and water bridges (Fig. 8c). The binding site includes Thr 178, Asn 179, Ala 180, Phe 218, Ile221, Ile 222, Pro 224, Ala 225, Ile 226, Arg 228, Met 256, Met 259, Arg 262, Ala 263, Arg 266, Leu 276, Ser 277, Leu 287, Leu 292, His 381, Phe 401, Phe 405 residues, with the TR α complex coming mainly under the alpha-helix region. The y-axis is normalized over the course of the trajectory. Phe 218 with 0.5 values, Ile 222 with 0.3 values and Leu 276 with

0.1 interactions over trajectory values are found, which suggests that 50, 30 and 10 % of the simulation time, the specific interaction of Phe 218, Ile 222 and Leu 276, are maintained with ligand 10456672. The geometric criteria for TR α protein-10456672 ligand, H-bond is 2.5 Å between the donor and acceptor atoms (D–H...A, 2.5 Å, $\geq 120^\circ$). Ser 277 values over 1.0 suggests that multiple contacts (water-bridge and H-bond) with the ligand take place. Ser 277 (one side chain and one backbone H-bond) and His 381(side chain) residues participate mainly in intermolecular hydrogen bonding. Arg 266, with an interaction trajectory value of about 0.4, shows that it participates in ionic- and water-bridges with very little interaction of hydrogen bonding with the ligand, which differs from the docking results (Fig. 7b). The significance of the hydrogen-bonding interactions in drug design is important because of their deep impact on drug specificity, metabolization and adsorption. Three H-bonds appear in both docking and MD simulation for the Ligand 10456672-TR α complex, but distinct variations are observed in both results. In docking His381, a backbone H-bond is observed, whereas in MD simulation, a side chain H-bond is noted for same residue. Met 259 makes a backbone H-bond with the NH atom of the ligand in docking but shows only 20 % of interaction over trajectory in the initial phase of the MD simulation and later forms 10 % hydrophobic interaction.

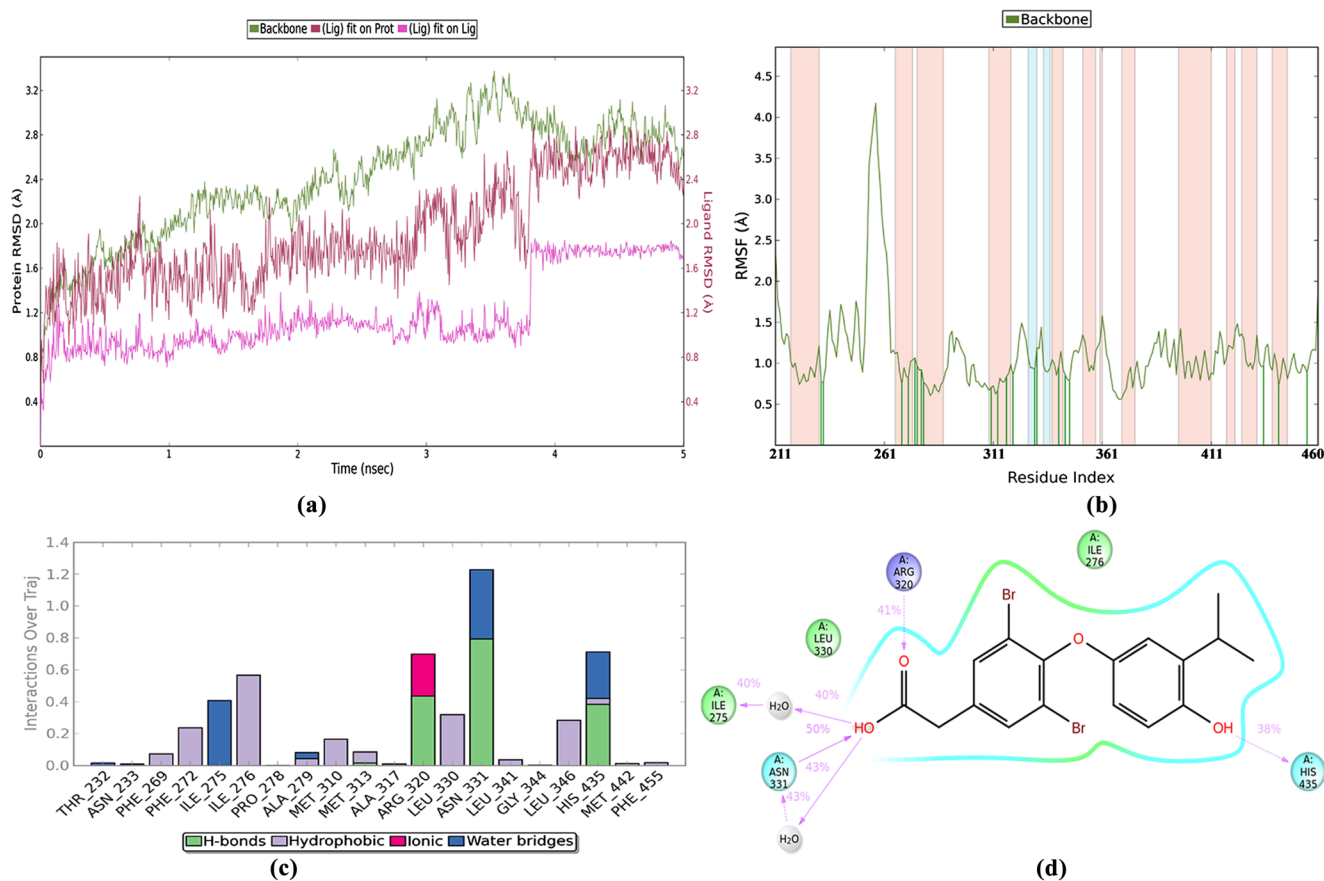


Fig. 9 Results of MD simulation for the TRβ complex. **(a)** MD simulation time vs RMSD in Å for all residues **(b)** RMSF vs residues index **(c)** Protein ligand contact interaction over trajectory **(d)** Average 2D-docking view of ligand 9933119-TRβ complex throughout MD simulation

An ionic interaction is seen between TRα protein residue Arg 266 and oxygen charged atoms of ligands that are within 3.7 Å of each other in the docking complex and 20 % of simulation time with specific interaction over the trajectory. Arg 266 forms a side chain H-bond with the oxygen atom of the –COOH group of ligand 10456672 in docking but has a negligible interaction during MD simulation. Ser 277 plays a significant role in MD simulation as it forms one side chain and one backbone H-bond with the ligand and stabilizes the interaction with TRα protein. In this analysis it is evident that the conformational change in ligand 10456672 is due mainly to fluctuations of the –NH–CO–COOH group attached to the dichlorobenzene ring, which can be explained by both docking and MD simulation study.

It is evident from the analysis that aromatic residues, mainly Phe 218 of TRα protein, can make π – π interactions with the aromatic ring of ligand 10456672. The geometric criteria for hydrophobic contacts are measured with π -cation- aromatic and charged groups within 4.5 Å; π – π , i.e., two aromatic groups stacked face-to-face or face-to-edge and others, i.e., a non-specific hydrophobic sidechain within 3.6 Å of a ligand's aromatic or aliphatic carbons.

The binding site, including mainly Phe 269, Phe 272, Ile 275, Ile 276, Ala 279, Met 310, Met 313, Arg 320, Leu 330, Asn 331, Leu 346 and His 435 for TRβ protein-ligand 9933119 participates in PLC (Fig. 9c). The binding site comes mainly under the region of alpha-helix (red). The interactions of key residues Ile 276, Leu 330, Leu 346 over trajectory values are about 0.3 or greater, which suggests that, for ≥ 30 % of the simulation time, the specific interaction of these residues is maintained with ligand atoms that make hydrophobic contacts. Asn 331, His 435 and Arg 320 residues participate mainly in hydrogen bonding (Fig. 9c,d). Asn 331 with about >1.2 interaction trajectory value indicates that it makes multiple contacts (H-bond side chain and backbone) with the ligand in the form of hydrogen bonding through a water-bridge. Ionic interaction is observed between TRβ protein residue Arg 320 and the oxygen atom of the terminal –COOH group of ligand 9933119 by 25 % of simulation time with specific interaction over trajectory. Further, the new H-bond formed between the ketonic oxygen of the ligand and the side chain of Arg 320, is altered from docking results. Three H-bonds participate in both docking and MD simulation, of which two H-bonds with residues Asn 331 and His 435 of TRβ protein are consistent in

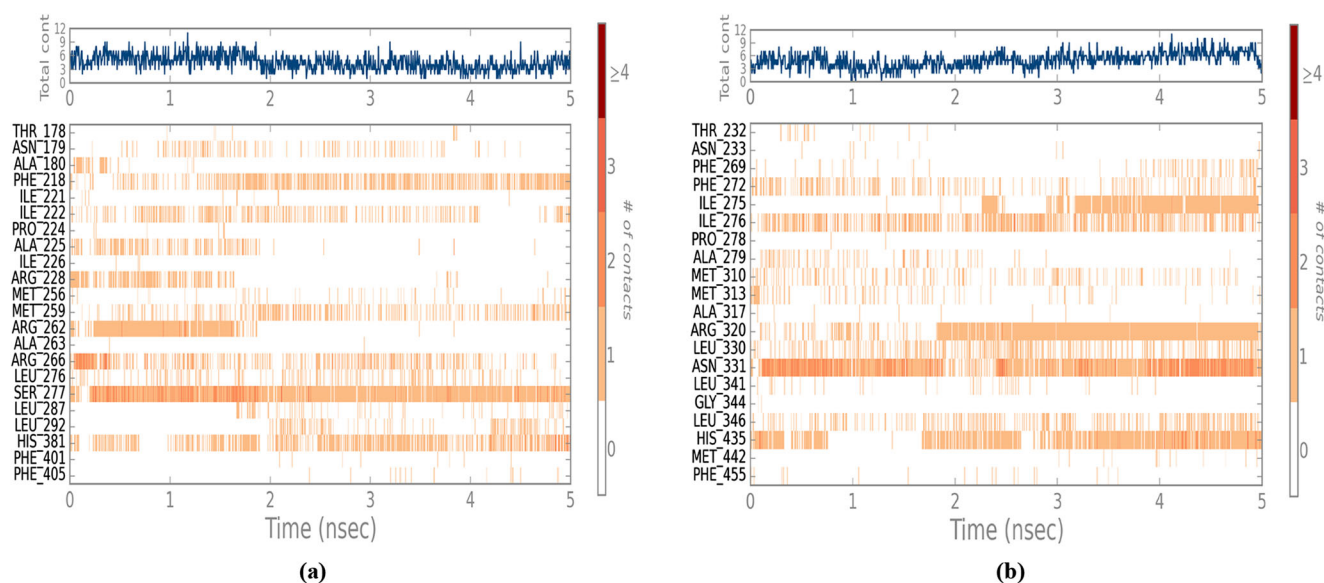


Fig. 10 Protein ligand contact analysis (a) Ligand 10456672-TR α complex (b) Ligand 9933119-TR β complex

both events but only Met 313 is altered to Arg 320 during MD simulation. This study suggests that these H-bonds play a significant role in stabilizing interactions between ligand 9933119 and TR β protein. Thus, the conformational change of ligand 9933119 results mainly from movement of the terminal $-\text{COOH}-\text{CH}_2$ group attached to the dibromobenzene ring, as could be elucidated by binding mode from docking and MD simulation analysis.

Interactions that take place more than 30 % of the simulation time in the selected trajectory (0.0 through 5.0 ns), are shown in Figs. 8d and 9d for TR α protein and the TR β protein–ligand complex, respectively. The current geometric criteria for a protein–water or water–ligand H-bond are: ($\text{D}\cdots\text{A}$, $2.7 \text{ \AA} \geq 110^\circ$); and an acceptor angle of $\geq 80^\circ$ between the hydrogen-acceptor-bonded atom, i.e., atoms ($\text{H}\cdots\text{A}-\text{X}$). Additionally, two water bridges, hydrogen-bonded protein–ligand interactions mediated by a water molecule, are found in TR β protein–ligand 9933119 but not in TR α –10456672. In TR β protein–ligand 9933119, one water bridge is formed between Asn 331 (H-bond side chain) and Ile 275 (H-bond backbone).

The upper section of Fig. 10a,b displays the total number of definite interactions for the 10456672-TR α complex and ligand 9933119-TR β complex, respectively, over the course of the trajectory. The lower section depicts the residues that make contact with the ligand in each trajectory frame. Ser 277 residues of TR α protein and Asn 331 of TR β protein form more than one contact with ligands 10456672 and 9933119, respectively, as denoted by a darker shade of orange, according to the scale on the y-axis (right) of the plot. Ile 275, Arg 320 and His 435 of TR β protein can be seen as more consistent after 2 ns during MD simulation, which makes a single contact and signifies the stability of the ligand 9933119-TR β

complex. A different pattern is observed in ligand 10456672-TR α complex as only Ser 277 is steady during the MD simulation. These results suggest that the average structure and backbone atoms of TR β protein and its ligands are relatively more stable than those of the TR α -complex during simulation time, suggesting the acceptability of the model.

These outcomes suggest that the TR β -complex is relatively more stable during the course of simulation as compared to the TR α -complex, and established the good predictive results achieved from the molecular docking. Though both complexes experienced numerous changes throughout the MD simulation, the binding pocket and the conformation of 9933119-TR β complex are more constant compared to those of the 10456672-TR α complex, imparting confidence in the rationality and validity of the docking result. The regions achieved from 3D contour maps are consistent with the contacts between ligands and amino acid residues of receptor TR α and TR β recognised by molecular docking.

These findings validate the notion that computational 3D-QSAR approaches can be an effective tool with which to predict the biological activity of compounds, to comprehend the toxicity process, and to propose novel molecules with precise biological activity. The benefit of the molecular docking methodology is that it takes the characteristics of the binding pocket of both TR α and TR β receptors into consideration, to obtain the active conformation of compounds. Due to both large numbers, i.e., 64,009 atoms of TR α and 59,141 atoms of TR β protein ligand complex, and computational constraints, MD simulations were performed at 5 ns. A study by Genheden and Ryde [36] showed that simulations of protein–ligand complexes with several approaches do not converge to equilibria even up to 500 ns, meaning that not all available states have been touched.

Hence, MD simulation will not reach equilibrium with the current computation and algorithms.

Conclusions

The objective of this study was to establish a robust relationship between structural properties and antagonist activity of human TR α and TR β by using 3D-QSAR, molecular docking and MD simulation methods to explore receptor-ligand contacts, which can be useful to minimize the adverse effect of current drugs. The built model produces high Q^2 values with small standard errors of estimation. The TR β model was found to be more optimized and robust as compared to TR α after analyzing the results. The model was validated internally and externally both in order to confirm consistency and to maintain high predictive ability. Contour maps revealed that inhibitory activity can be improved by modulating the donor abilities of nitrogen or oxygen atoms in the fused aromatic rings that are involved in the H-bond interactions with the binding site of the receptor. The built model can be used to design more effective antagonists than existing drugs (PTU and MMI), which have side effects such as agranulocytosis, aplastic anemia, and the ability to cross the placenta [37–39]. The model built in this work could be useful to recognize substantial essential features that affect thyroid hormone function and are favorable to deliver certain statistics about the interaction of ligands with receptor. The coherence of the outcome found from 3D-QSAR, molecular docking and MD simulation further specifies the robustness of the best 3D-QSAR model.

References

- Brent GA (2012) Mechanism of thyroid hormone action. *J Clin Invest* 122(9):3035–3043
- Oetting A, Yen PM (2007) New insights into thyroid hormone action. *Best Pract Res Clin Endocrinol Metab* 21(2):193–208
- Schapiro M, Raaka BM, Das S, Fan L, Totrov M, Zhou ZG, Wilson S, Abagyan R, Samuels HH (2003) Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. *Proc Natl Acad Sci USA* 100(12):7354–7359
- Yen PM (2001) Physiological and molecular basis of thyroid hormone action. *Physiol Rev* 81(3):1097–142
- Flamant F, Baxter JD, Forrest D, Refetoff S, Samuels H, Scanlan TS, Vennstrom B, Samarut J (2006) International Union of Pharmacology. LIX. The pharmacology and classification of the nuclear receptor superfamily: thyroid hormone receptors. *Pharmacol Rev* 58(4):705–11
- Cheng SY (2000) Multiple mechanisms for regulation of the transcriptional activity of thyroid hormone receptors. *Rev Endocr Metab Disord* 1(1–2):9–18
- Macchia PE, Takeuchi Y, Kawai T, Cua K, Gauthier K, Chassande O, Seo H, Hayashi Y, Samarut J, Murata Y, Weiss RE, Refetoff S (2001) Increased sensitivity to thyroid hormone in mice with complete deficiency of thyroid hormone receptor alpha. *Proc Natl Acad Sci USA* 98(1):349–54
- Song Y, Yao X, Ying H (2011) Thyroid hormone action in metabolic regulation. *Protein Cell* 2(5):358–368
- Hedfors A, Appelqvist T, Carlsson B, Bladh LG, Litten C, Agback P, Grynfarb M, Koehler KF, Malm J (2005) Thyroid receptor ligands. 3. Design and synthesis of 3, 5- Dihalo-4-alkoxyphenylalkanoic acids as indirect antagonists of the thyroid hormone receptor. *J Med Chem* 48(9):3114–3117
- Koehler K, Gordon G, Brandt P, Carlsson B, Backsbro-Saeidi A, Appelqvist T, Agback P, Grover GJ, Nelson W, Grynfarb M, Fa'megardh M, Rehnmark S, Malm J (2006) Thyroid receptor ligands. 6. A high affinity “direct antagonist” selective for the thyroid hormone receptor. *Bioorg Med Chem Lett* 49(23):6635–6637
- Grover GJ, Malm J (2005) Selective thyroid hormone agonists: a strategy for treating metabolic syndrome. *Drug Discov Today: Ther Strateg* 2(2):137–142
- Ren Y, Liu H, Li S, Yao X, Liu M (2007) Prediction of binding affinities to $\beta 1$ isoform of human thyroid hormone receptor by genetic algorithm and projection pursuit regression. *Bioorg Med Chem Lett* 17(9):2474–2482
- Liu HX, Gramatica P (2007) QSAR study of selective ligands for the thyroid hormone receptor β . *Bioorg Med Chem* 15(15):5251–5261
- Vedani A, Dobler M, Lill MA (2006) The challenge of predicting drug toxicity in silico. *Basic Clin Pharmacol* 99(3):195–208
- Vedani A, Zumstein M, Lill MA, Ernst B (2007) Simulating alpha/beta selectivity at the human thyroid hormone receptor: consensus scoring using multidimensional QSAR. *Chem Med Chem* 2(1):78–87
- Li F, Xie Q, Li XH, Li N, Chi P, Chen JW, Wang ZJ, Hao C (2010) Hormone activity of hydroxylated polybrominated diphenyl ethers on human thyroid receptor-beta: in vitro and in silico investigations. *Environ Health Perspect* 118(5):602–606
- Du J, Qin J, Liu H, Yao X (2008) 3D-QSAR and molecular docking studies of selective agonists for the thyroid hormone receptor beta. *J Mol Graph Model* 27(2):95–104
- Li X, Ye L, Wang X, Wang X, Liu H, Zhu Y, Yu H (2012) Combined 3D-QSAR, molecular docking and molecular dynamics study on thyroid hormone activity of hydroxylated polybrominated diphenyl ethers to thyroid receptors β . *Toxicol Appl Pharmacol* 265(3):300–307
- Hao M, Li Y, Wang Y, Yan Y, Zhang S, Li G, Yang L (2011) Combined 3D-QSAR, molecular docking, and molecular dynamics study on piperazinyll-glutamatepyridines/ pyrimidines as potent P2Y12 antagonists for inhibition of platelet aggregation. *J Chem Inf Model* 51(10):2560–2572
- Malm J, Färnegårdh M, Grover GJ, Ladenson PW (2009) Thyroid hormone antagonists: potential medical applications and structure activity relationships. *Curr Med Chem* 16(25):3258–66
- Martínez L, Webb P, Polikarpov I, SkafM S (2006) Molecular dynamics simulations of ligand dissociation from thyroid hormone receptors: evidence of the likeliest escape pathway and its implications for the design of novel ligands. *J Med Chem* 49(1):23–26
- Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* 35(Database issue):D198–201
- (2012) LigPrep, version 2.5, Schrödinger, LLC, New York, NY
- (2012) Phase, version 3.4, Schrödinger, LLC, New York, NY
- (2012) Glide, version 5.8, Schrödinger, LLC, New York, NY
- Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic

- enclosure for protein-ligand complexes. *J Med Chem* 49(21): 6177–96
27. Shaw D E Research (2008) Desmond molecular dynamics system, version 3.5, New York, NY
 28. Bowers KJ, Chow E, Xu H, Dror R O, Eastwood M P, Gregerson B A, Klepeis J L, Kolossvary I, Moraes M A, Sacerdoti F D, Salmon J K, Shan Y, Shaw D E (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In Proc. ACM/IEEE conf. on supercomputing (Tampa, FL, 2006). http://sc06.supercomputing.org/schedule/event_detail.php?evid=9088
 29. Kaminski GA, Friesner RA (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem* 105:6474–6487
 30. Andersen HC (1983) Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *J Comput Phys* 52(1):24–34
 31. Strahan GD, Keniry MA, Shafer RH (1998) NMR structure refinement and dynamics of the K⁺-[d(G3T4G3)]₂ quadruplex via particle mesh Ewald molecular dynamics simulations. *Biophys J* 75(2):968–981
 32. Evans DA, Doman TN, Thomer DA, Bodkin MJ (2007) 3D QSAR methods: phase and catalyst compared. *J Chem Inf Model* 47:1248–1257
 33. Carugo O, Pongor S (2001) A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci* 10(7):1470–3
 34. Carugo O (2007) Statistical validation of the root-mean-square-distance, a measure of protein structural proximity. *Protein Eng Des Sel* 20(1):33–7
 35. Hollup SM, Fuglebakk E, Taylor WR, Reuter N (2011) Exploring the factors determining the dynamics of different protein folds. *Protein Sci* 20(1):197–209
 36. Genheden S, Ryde U (2012) Will molecular dynamics simulations of proteins ever reach equilibrium? *Phys Chem Chem Phys* 14:8662–8677
 37. Honda M (2011) Successful treatment of methimazole-induced severe aplastic anemia by granulocyte colony-stimulating factor, methylprednisolone, and cyclosporin. *ISRN Endocrinol* 2011:732623
 38. Bahn RS, Burch HS, Cooper DS, Garber JR, Greenlee CM, Klein IL, Laurberg P, McDougall IR, Rivkees SA, Ross D, Sosa JA, Stan MN (2009) The role of propylthiouracil in the management of Graves' disease in adults: Report of a meeting jointly sponsored by the american thyroid association and the food and drug administration. *Thyroid* 19(7):673–4
 39. Gupta V (2012) Block and replace therapy using propylthiouracil and levothyroxine for Graves' disease during pregnancy. *Indian J Endocrinol Metab* 16(4):660–2