Chem Biol Drug Des 2011; **78:** 391–407

Research Article

doi: 10.1111/j.1747-0285.2011.01160.x

Discovery of New Antifungal Leads via Pharmacophore Modeling and QSAR Analysis of Fungal *N*-Myristoyl Transferase Inhibitors Followed by *In Silico* Screening

Mutasem O. Taha[1,*], Amjad M. Qandil[2], Tariq Al-Haraznah[3], Reema Abu Khalaf[2], Hiba Zalloum[1] and Amal G. Al-Bakri[4]

[1]*Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan*
[2]*Department of Medicinal Chemistry and Pharmacognosy, Faculty of Pharmacy, Jordan University of Science and Technology, Irbid, Jordan*
[3]*Department of Pharmaceutical Sciences, Faculty of Pharmacy, Al Zaytoonah Private University of Jordan, Amman, Jordan*
[4]*Department of Pharmaceutics and Technology, Faculty of Pharmacy, University of Jordan, Amman, Jordan*
**Corresponding author: Mutasem O. Taha, mutasem@ju.edu.jo*

***N*-Myristoyl transferase is an essential enzyme for fungal growth and survival. The continuous interest in the development of new antifungal agents prompted recent interest in developing new potent inhibitors of fungal *N*-myristoyl transferase. In this context, we combined pharmacophore and QSAR modeling to explore the structural requirements for potent *N*-myristoyl transferase inhibitors employing 55 known *N*-myristoyl transferase ligands. Four binding pharmacophore models emerged in the optimal QSAR equations (r^2_{44} = 0.81–0.83, *F*-statistic = 47.89–58.83, r^2_{LOO} = 0.77–0.80, r^2_{PRESS} against 11 external test inhibitors = 0.61–0.71). The successful pharmacophores were complemented with exclusion spheres to optimize their receiver operating characteristic curve profiles. The QSAR equations and their associated pharmacophore models were validated by the identification and experimental evaluation of new promising antifungal leads retrieved from the NCI database and our *in-house*-built database of established drugs and agrochemicals.**

Key words: *in silico* screening, NMT, pharmacophore modeling, QSAR, ROC curves

Received 26 October 2009, revised 3 July 2010 and accepted for publication 19 May 2011

The frequency of systemic fungal infections, which are often life-threatening, has increased dramatically over the years owing to the increased number of immunocompromised patients (e.g., AIDS, cancer or organ transplantation patients) (1). On the other hand, the four classes of currently available antifungal drugs, namely, polyenes, azoles, flucytosine and candins, have restricted use because of their limited activity spectra, toxicity profiles, hazardous interactions, non-optimal pharmacokinetics and emergence of resistant fungal strains (1,2).

Lack of optimal antifungal agents prompted continuous search for new antifungal chemotherapeutic agents against novel fungal targets (2). One such target that has been identified recently for the development of antifungal agents is myristoyl-CoA: protein *N*-myristoyl transferase (NMT) (3,4).

NMT is a cytosolic monomeric enzyme present in eukaryotes such as protozoa, fungi and animals but absent in bacteria (3,4). NMT catalyzes the myristoylation of protein targets, which involve the transfer of myristate from myristoyl-CoA to the substrate protein. In fungi, myristoylation is involved in the regulation of interaction between certain signaling proteins and the cellular membrane, which is required in signal transduction cascades and vesicular and protein trafficking processes (5–8).

Knockout genetic studies have unequivocally established the necessity of NMT for the survival of *Candida albicans* and *Candida neoformans* (9,10). Moreover, *C. albicans* of defective NMT looses the ability to infect mice (11). The necessity of *N*-myristoylated proteins in various fungi makes NMT a suitable target for the development of antifungal agents. Furthermore, fungal and mammalian NMT differ in their peptide substrate specificities, which should minimize potential adverse effects related to cross-activity against mammalian NMT (12).

Several classes of NMT inhibitors were recently reported, including peptidomimetic inhibitors (3,13,14), myristic acid analogs (15,16), *p*-toluene sulfonamides (17), benzofurans (2,12,18,19), and benzothiazole inhibitors (20). Peptidomimetic inhibitors were designed based on the substrate peptide GLYASKLS-NH2 and its analogous inhibitor ALYASKLS-NH2 (14). However, peptide and peptidomimetic NMT inhibitors were generally devoid of *in vitro* antifungal activity, probably because of their inability to penetrate fungal cellular membranes. Still, ALYASKLS-NH2 was modified into a new class of non-peptide NMT inhibitors that illustrated limited fungistatic activities (13).

Random screening of Roche chemical library unveiled a new class of potent NMT inhibitors based on benzofuran and benzothiazole scaffolds. However, these compounds illustrated only marginal *in vitro* antifungal activities combined with significant β-adrenoreceptor-blocking properties, which prompted subsequent structure-based optimization *via* X-ray crystallographic analysis (18).

Clearly, the main focus of recent efforts toward the development of new NMT inhibitors concentrates on structure-based ligand design (21–25) with few ligand-based exceptions that focus mainly on 3D-QSAR methodologies, e.g., CoMFA and CoMSIA (17,26–31). To date, nine fungal NMT X-ray complexes are documented in the Protein Data Bank (PDB codes: 2P6G, 2P6F, 2P6E, 1IID, 2NMT, 1IIC, 1IYK, 1IYL, and 1NMT, resolution range: 2.20–3.20 Å) (22–25,32).

The continued interest in designing new antifungal agents combined with the drawbacks of structure-based design (33–36), e.g., limitations in dealing with the induced fit flexibility of NMT (18), and inappropriateness of CoMFA and CoMSIA methodologies to act as search queries to mine virtual three-dimensional (3D) databases for new hits (37), prompted us to explore the possibility of developing ligand-based 3D pharmacophore(s) integrated within self-consistent QSAR model. The pharmacophore model(s) can be used as 3D search query(ies) to mine 3D libraries for new NMT inhibitors, while the associated QSAR model can be used to predict the bioactivities of captured hits. We previously reported the successful use of this innovative approach to probe the induced fit flexibilities of activated factor X (fXa) (38) and toward the discovery of new inhibitory leads against glycogen synthase kinase 3β (GSK-3β) (39), hormone-sensitive lipase (40), bacterial MurF (41), protein tyrosine phosphatase 1B (PTP 1B) (42), and influenza neuraminidase (43).

Experimental

We employed HYPOGEN module of CATALYST package[a] to construct plausible binding hypotheses for NMT benzofuran inhibitors. Subsequently, genetic function algorithm (GFA) and multiple linear regression (MLR) analyses were employed to search for optimal QSAR model(s) that combine high-quality binding pharmacophores with other molecular descriptors and capable of explaining bioactivity variation across a large collection of diverse NMT inhibitors.

The QSAR-selected pharmacophores were validated by evaluating their ability to successfully classify a list of compounds as actives or inactives by assessing their receiver operating characteristic (ROC) curves. Optimal pharmacophores were complemented with exclusion spheres to enhance their ROC profiles and were used as 3D search queries to screen the National Cancer Institute (NCI) database and our *in-house*-built database of established drugs and agrochemicals (DAC) for new NMT inhibitors.

CATALYST models drug–receptor interaction using information derived only from the drug structure[a] (44–48). HYPOGEN identifies a 3D array of a maximum of five chemical features common to active training molecules, which provides a relative alignment for each input molecule consistent with their binding to a proposed common receptor site. The chemical features considered can be hydrogen bond donors and acceptors (HBDs and HBAs), aliphatic and aromatic hydrophobes, positive and negative charges, positive and negative ionizable groups and aromatic planes. CATALYST pharmacophores have been used as 3D queries for database searching and in 3D-QSAR studies (38–42,49).

Molecular modeling

Software and hardware

The following software packages were utilized: CATALYST (version 4.11; Accelrys Inc.; http://www.accelrys.com, USA), CERIUS2 (version 4.10; Accelrys Inc.; http://www.accelrys.com, USA), and CS CHEMDRAW ULTRA 7.01; Cambridge Soft Corp. (http://www.cambridgesoft.com), USA.

Pharmacophore modeling and QSAR analysis were performed using CATALYST (HYPOGEN module) and CERIUS2 software suites from Accelrys Inc. (http://www.accelrys.com) installed on a Silicon Graphics Octane2 desktop workstation equipped with a 2×600 MHz MIPS R14000 processor (1.0 GB RAM) running the IRIX 6.5 operating system.

Dataset

The structures of 55 antifungal NMT enzyme inhibitors were collected from published literature (2,18,19,50). The *in vitro* bioactivities of the collected inhibitors were expressed as the concentration of the test compound that inhibited the activity of NMT enzyme by 50% (IC). Figure 1 and Table S1 (Supporting Information) show the structures and IC_{50} values of the considered inhibitors. The logarithm of measured IC_{50} (μM) values was used in pharmacophore modeling and QSAR analysis, thus correlating the data linear to the free energy change.

The two-dimensional (2D) chemical structures of the inhibitors were sketched using CHEMDRAW ULTRA, installed on a PC, and saved in MDL-mol file format. Subsequently, they were imported into CATALYST, converted into corresponding standard 3D structures, and energy minimized to the closest local minimum using the molecular mechanics CHARMm force field implemented in CATALYST. The resulting 3D structures were utilized as starting conformers for conformational analysis.

Conformational analysis

The conformational space of each inhibitor (**1**–**55**, Figure 1 and Table S1, Supporting Information) was explored adopting the 'best conformer generation' option within CATALYST based on the generalized CHARMm force field implemented in the program. Default parameters were employed in the conformation generation procedure, i.e., conformational ensembles were generated with an energy threshold of 20 kcal/mol from the local minimized structure and a maximum limit of 250 conformers per molecule[a] (46,51).

Generation of pharmacophoric hypotheses

All 55 molecules with their associated conformational models were arranged into a spreadsheet. The biological data of the inhibitors

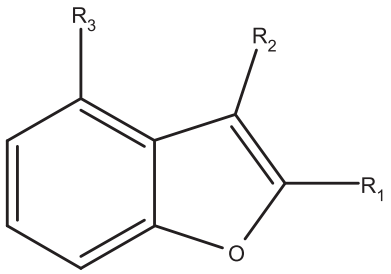

Figure 1: The chemical scaffold of training compounds, the detailed structures are as in Table S1, Supporting Information.

were reported with an 'Uncertainty' value of 3, which means that the actual bioactivity of a particular inhibitor is assumed to be situated somewhere in an interval ranging from one-third to three times the reported bioactivity value of that inhibitor (45,47,48). Subsequently, two subsets of inhibitors were selected and employed as training sets for pharmacophore modeling (Table 1). CATALYST requires informative training sets of at least 16 compounds of evenly spread bioactivities over at least four orders of magnitude (45,47,48,52).

Each training subset of compounds was utilized to conduct 12 modeling runs to explore the pharmacophoric space of NMT. Different binding hypotheses were generated by altering the interfeature spacing and the number of allowed features in the resulting pharmacophores (Table S2, Supporting Information). Pharmacophore modeling employing CATALYST proceeds through three successive phases: the constructive phase, subtractive phase, and optimization phase (see Supporting Information for information about CATALYST pharmacophore building algorithm) (45,47,48,52).

Assessment of the generated hypotheses

When generating hypotheses, CATALYST attempts to minimize a cost function consisting of three terms: weight cost, error cost and configuration cost[a] (44–48,52). Furthermore, CATALYST implements Fisher scrambling (Cat.Scramble algorithm) to validate that the generated binding hypotheses are not chance correlations (53,54) (see Supporting Information for information about CATALYST assessment of generated pharmacophore hypotheses).

Clustering the generated pharmacophoric hypotheses

Pharmacophore models originating from each training subset having Fisher confidence (Cat.Scramble) ≥85% were clustered utilizing the hierarchical average linkage method implemented in CATALYST in such a way that each five related models were categorized in one cluster, i.e., 206 successful models were clustered into 41 groups. Subsequently, the highest-ranking representatives, as judged from their significance *F*-values, were selected to represent their corresponding clusters in subsequent QSAR modeling.

QSAR modeling

A subset of 44 compounds was employed as the training set for QSAR modeling. The remaining 11 molecules (ca. 20% of the dataset) were employed as an external test subset for validating the QSAR models. The test molecules were selected as follows: the 55 collected inhibitors were ranked according to their IC_{50} values, and then every fifth compound was selected for the test set starting from the high-potency end. This selection considers the fact that the test molecules must represent a range of biological activities similar to that of the training set. The selected test inhibitors are **12**, **13**, **15**, **24**, **28**, **32**, **35**, **37**, **43**, **47** and **48** (numbers are as in Table S1 in supporting information and Figure 1).

The logarithm of measured $1/IC_{50}$ (μM) values was used in QSAR, thus correlating the data linear to the free energy change. The chemical structures of the inhibitors were imported into CERIUS2 as standard 3D single conformer representations (of the lowest energy within the conformational ensemble generated by CATALYST). Subsequently, different descriptor groups were calculated for each compound employing the C2.DESCRIPTOR module of CERIUS2. The calculated descriptors included various simple and valence connectivity indices, electro-topological state indices, and other molecular descriptors (e.g., logarithm of partition coefficient, polarizability, dipole moment, molecular volume, molecular weight, and molecular surface area).[b–e] Furthermore, the training compounds were fitted (using the best-fit option in CATALYST)[a] against successful pharmacophores (cluster centers of Cat.Scramble significance ≥85%), and the corresponding fit values were added as additional descriptors. The fit values were obtained automatically *via* eqn (D) (Supporting Information).[a]

GFA was employed to search for the best possible QSAR regression equation capable of correlating the variations in biological activities of the training compounds with variations in the generated descriptors employing Friedman's 'lack-of-fit' (LOF) fitness criteria.[b–e]

Preliminary diagnostic trials suggested the following optimal GFA parameters: explore linear, quadratic and spline equations at mating and mutation probabilities of 50%; population size = 500; number of genetic iterations = 30 000, and LOF smoothness parameter = 1.0. However, to determine the optimal number of explanatory terms (QSAR descriptors), it was decided to scan and evaluate all possible QSAR models resulting from 4 to 10 explanatory terms.

Table 1: The training subsets employed in exploring the pharmacophoric space of *N*-Myristoyl transferase inhibitors; numbers correspond to compounds in Table S1 (Supporting Information) and Figure 1

Training sets	Most active[a]	Intermediately active	Least active[a]
1	**42, 49, 46, 48, 41**	**52, 27, 31, 19, 44, 55, 38**	**53, 16, 36, 13, 32, 10, 14, 9, 7, 35, 18, 12**
2	**51, 46, 48, 41, 52**	**40, 31, 19, 44, 15, 17, 54, 37, 55**	**53, 36, 13, 11, 10, 35, 18**

[a]Potency categories as defined by eqns (A) and (B) under Supporting Information.

All QSAR models were validated employing leave-one-out cross-validation (r^2_{LOO}), bootstrapping (r^2_{BS})[b–e] (55), and predictive r^2 (r^2_{PRESS}) calculated from the test subsets. The predictive r^2_{PRESS} is defined as follows:

$$r^2_{press} = SD - PRESS/SD \quad (1)$$

where SD is the sum of the squared deviations between the biological activities of the test set and the mean activity of the training set molecules, PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set.[b–e]

Receiver operating characteristic (ROC) curve analysis

Selected pharmacophore models (i.e., Hypo1/2/1, Hypo9/2/1, Hypo2/9/2, Hypo1/7/1, and their sterically refined versions) were validated by assessing their abilities to selectively capture diverse NMT active compounds from a large test list of actives and decoys.

The testing list was prepared as described by Verdonk *et al.* (56,57). Briefly, decoy compounds were selected based on three basic one-dimensional (1D) properties that allow the assessment of distance (D) between two molecules (e.g., i and j): (i) the number of hydrogen bond donors (NumHBD); (ii) number of hydrogen bond acceptors (NumHBA), and (iii) count of non-polar atoms (NP, defined as the summation of Cl, F, Br, I, S, and C atoms in a particular molecule). For each active compound in the test set, the distance to the nearest other active compound is assessed by their Euclidean distance (eqn 2):

$$D(i,j) = \sqrt{(\mathrm{NumHBD}_i - \mathrm{NumHBD}_j)^2 + (\mathrm{NumHBA}_i - \mathrm{NumHBA}_j)^2 + (\mathrm{NP}_i - \mathrm{NP}_j)^2} \quad (2)$$

The minimum distances are then averaged over all active compounds (Dmin). Subsequently, for each active compound in the test set, around 27 decoys were randomly chosen from the ZINC database (58). The decoys were selected in such a way that they did not exceed Dmin distance from their corresponding active compound.

To diversify active members in the list, we excluded any active compound having zero distance (D (i, j)) from other active compound(s) in the test set. Active testing compounds were defined as those possessing NMT affinities ranging from 0.00039 to 4.1 μM. The test set included 22 active compounds and 590 ZINC decoys.

The test set (612 compounds) was screened by each pharmacophore employing the 'best flexible search' option implemented in CATALYST. Conformational ensembles were generated for the test compounds employing the 'fast conformation generation option' implemented in CATALYST. Compounds missing one or more features were discarded from the hit list. *In silico* hits were scored employing their fit values as calculated by eqn (D) (Supporting Information). The results were used to construct ROC curves for each pharmacophore model (59–61). The details of ROC analysis are shown in Supporting Information.

Addition of exclusion volumes

To account for the steric constraints of the binding pocket, and to optimize the ROC curves of the pharmacophores, it was decided to add exclusion volumes to Hypo9/2/1, Hypo2/9/2, and Hypo1/7/1 employing the HIPHOP-REFINE module of CATALYST. HIPHOP-REFINE uses inactive training compounds to add exclusion spheres to resemble the steric constraints of the binding pocket. It identifies spaces occupied by the conformations of inactive compounds and free from active ones. These regions are then filled with excluded volumes (38–42,49). In HIPHOP-REFINE, the user defines how many molecules must map the selected pharmacophore hypothesis completely or partially *via* the Principal and Maximum Omitted Features (MaxOmitFeat) parameters. Active compounds are normally assigned a MaxOmitFeat parameter of zero and principal value of 2 to instruct the software to consider all their chemical moieties to fit them against all the pharmacophoric features of the particular hypothesis. On the other hand, inactive compounds are allowed to miss one (or more) feature(s) by assigning them a MaxOmitFeat of 1 (or 2) and principal value of zero. Moderately active compounds are normally assigned a principal value of 1 and a MaxOmitFeat of 0 or 1 to encode their intermediate status.

A subset of training compounds was carefully selected for HIPHOP-REFINE modeling. It was decided to consider the IC_{50} value of 0.62 μM as an arbitrary activity/inactivity threshold. Accordingly, inhibitors of IC_{50} values $\leq$0.62 μM were regarded as 'actives' and were assigned principal and MaxOmitFeat values of 2 and 0, respectively, while inhibitors of $IC_{50} > 0.62$ μM were considered inactive and were assigned principal values of zero (38–42,49). However, each inactive compound was carefully evaluated to assess whether its low potency is attributable to missing one or more pharmacophoric features, i.e., compared with active compounds (MaxOmitFeat = 1 or 2), or related to possible steric clashes within the binding pocket (MaxOmitFeat = 0). HIPHOP-REFINE was configured to allow a maximum of 100 exclusion spheres to be added to the generated pharmacophoric hypotheses.

In silico *screening of databases for new NMT inhibitors*

Hypo1/2/1 and the sterically refined versions of Hypo9/2/1, Hypo2/9/2, and Hypo1/7/1 were employed as 3D search queries to screen two of our three-dimensional (3D) flexible structural databases, namely the National Cancer Institute (NCI) list of compounds (238 819 compounds) and our *in-house*-built database of established drugs and agrochemicals (DAC, 3002 structures). The screening was

performed employing the 'best flexible database search' option implemented within CATALYST.

NCI hits were subsequently filtered based on Lipinski's and Veber's rules (62,63). However, DAC hits were processed without subsequent postfiltering. Surviving hits were fitted against each corresponding pharmacophore model using the 'best-fit' option implemented within CATALYST. The fit values together with the relevant molecular descriptors of each hit were substituted in the optimal corresponding QSAR equation (i.e., the one bearing the particular pharmacophore that captured that hit). The highest-ranking hits were acquired and tested *in vitro*.

Microbiological assays

Materials

Nutrient, Sabouraud dextrose and Mueller–Hinton agar were obtained from Oxoid Ltd (Hampshire, UK) and were reconstituted according to the manufacturer's recommendations and sterilized by autoclaving at 121 °C for 30 min by means of RAYPA steam sterilizer (R. Espinar S.L., Athens, Greece). Dimethylsulfoxide (DMSO) was purchased from BDH Laboratory Supplies (Leicestershire, UK). Miconazole was used as reference antifungal agent and was purchased from Sigma-Aldrich (St. Louis, MO, USA). Hit compounds were kindly donated from the NCI (Bethesda, MD, USA) and were tested without further purification.

Microbial strain

The antifungal and antibacterial activities of hit molecules were assessed against standard reference *C. albicans* ATCC 10231 and *Escherichia coli* ATCC 8739 strains.

Broth microdilution method

The minimum inhibitory concentration (MIC) of each hit was determined according to the broth microdilution susceptibility assay described by the Clinical and Laboratory Standard Institute (CLSI) with some modifications (64). MIC test was performed in 96-well flat-bottom microtitre plates (TPP, Trasadingen, Switzerland). Each hit compound was dissolved in a minimum amount of DMSO and used as a stock solution. The stock solutions were stored at 4 °C in tightly capped amber-glassed bottles. From the stock solution, 100 μL was taken and was added to the first column of test wells, and then the volume was made up to 200 μL using double-strength Sabouraud dextrose broth for *C. albicans* and Mueller-Hinton broth for *E. coli* with a good mixing. A series of half-fold dilutions was then carried out across the plate using a micropipette, changing the tips at each dilution step. Then, 10 μL of preadjusted overnight microbial culture was used to inoculate each well in the microtitre plate to achieve a final inoculum size of ca. 6×10^5 CFU/well for *C. albicans* and ca. 2.44×10^7 CFU/well for *E. coli*. Then, the 96-well microtitre plate was incubated for 24 h using WTC binder static incubator (Tutlingen, Germany).

In all assays, positive growth controls (wells with overnight culture, DMSO, and microbial inoculum but without hit compound) and negative growth controls (wells with microbial inoculum without hit compound and DMSO) were included.

MICs were expressed as the average of two successive concentrations of the antimicrobial agent showing no growth and growth (turbidity), respectively. MIC determination was carried out in duplicate.

Results and Discussion

Data mining and conformational coverage

The literature was extensively surveyed to identify as many structurally diverse NMT inhibitors as possible to be used as training and test sets for modeling (**1**–**55**, see Table S1 in Supporting Information and Figure 1) (2,18,19,50). The fact that most reported fungal NMT inhibitors were evaluated employing distinct bioassay methodologies confined us to 55 inhibitors determined by a common bioassay technique. The conformational space of each inhibitor was extensively explored (47,48,51) to avoid inadequate conformational sampling of training compounds that might compromise pharmacophore generation and pharmacophore-based search procedures (65).

Pharmacophore modeling

CATALYST-HYPOGEN enables automatic pharmacophore construction by using a collection of at least 16 molecules with bioactivities spanning over 3.5 orders of magnitude[a] (44–48). As we have a diverse training set of 55 NMT inhibitors of evenly spread bioactivities over more than five orders of magnitude, we were encouraged to employ HYPOGEN to identify possible pharmacophoric binding modes assumed by these inhibitors within NMT. HYPOGEN evaluates large number of potential pharmacophoric models for a particular target (47,48). The number of assessed pharmacophores is reflected by the configuration (Config.) cost, which is generally recommended not to exceed 17 (corresponding to 2^{17} hypotheses to be assessed by CATALYST) to guarantee thorough analysis of all models (47,48).

The number of investigated pharmacophores is a function of training compounds, selected input chemical features and other CATALYST control parameters (47,48). Restricting the extent of explored pharmacophoric space should improve the efficiency of optimization by allowing effective evaluation of a limited number of pharmacophoric models. However, stern limitations on the number of explored pharmacophores might undermine the possibility of discovering optimal pharmacophoric hypotheses, particularly if they occur outside the 'boundaries' of the explored space. Therefore, we decided to explore the pharmacophoric space of NMT inhibitors under reasonable 'boundaries' through 24 HYPOGEN automatic runs and employing two carefully selected training subsets (i.e., from the collected compounds), subsets 1 and 2, in Table 1.

In line with our concept of restricting the evaluated 'pharmacophoric space', we instructed CATALYST-HYPOGEN to explore pharmacophoric models of zero to three features of any selected type (i.e., HBA, HBD, hydrophobic, positive ionizable, and RingArom) instead

of the default range of 0–5. Furthermore, we explored only 4- and 5-featured pharmacophores to represent the feature-rich nature of NMT binding pocket (Table S2, Supporting Information).

The resulting binding hypotheses from each run were automatically ranked according to their corresponding 'total cost' values (see section Assessment of the generated hypotheses)[a] (44–48,52,66). To avoid the possibility of having pharmacophore models based on chance correlation, HYPOGEN calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the null hypothesis cost (residual cost, Table S3, Supporting Information), the more likely that the hypothesis does not reflect a chance correlation[a] (44–48,52,66). Furthermore, pharmacophore models were validated using Fisher's randomization test (53) (see section Assessment of the generated hypotheses).

Eventually, 240 pharmacophore models emerged from 24 automatic HYPOGEN runs, out of which 206 models illustrated confidence levels ≥85% (based on Fisher's randomization test). The successful models were clustered, and their best representatives (41 models, details shown in Table S3, Supporting Information) were allowed to compete in subsequent QSAR modeling. The reason for our choice of 85% significance threshold, i.e., as framework to select pharmacophore models for subsequent QSAR manipulation, is because this level of statistical significance is quite reasonable, and it allows diverse pharmacophores to compete in the QSAR-based pharmacophore tournament. Higher significance levels tend to retain closely related pharmacophores (less diverse) and therefore tend to loose important pharmacophoric models that become more statistically significant upon complementation with 2D descriptors during QSAR modeling.

QSAR modeling

Despite that pharmacophore models provide excellent insights into ligand–macromolecule interactions and can act as 3D search queries, their predictive value as QSAR models is generally limited by steric and electronic effects (44). This point combined with the fact that our pharmacophore exploration furnished several good binding models of comparable statistical criteria (Table S3, Supporting Information) prompted us to employ classical QSAR analysis as a framework in which high-ranking pharmacophore models compete and their best winner, i.e., that can best explain bioactivity variations in combination with other 2D descriptors (characterizing electronic and steric molecular properties), is (are) selected for subsequent *in silico* screening. This task was performed *via* genetic function approximation-multiple linear regression QSAR (GFA-MLR-QSAR) analysis.

However, to avoid overwhelming GFA-MLR-QSAR with numerous inferior descriptor pharmacophores, which may allow the emergence of less-than-optimal regression models,[b–e] it was decided to cluster the generated pharmacophores (206 models) into 41 groups, such that models originating from a particular training subset were clustered separately. The best-performing cluster representatives were enrolled in GFA-MLR-QSAR analysis (41 models, Table S3, Supporting Information).

The molecular fit values obtained by mapping the selected representative pharmacophores (Table S3, Supporting Information) against NMT inhibitors were combined with other physicochemical descriptors and enrolled as independent variables in GFA-MLR-QSAR analysis[b–e] (67).

Finally, to avoid the possibility of missing a potentially significant pharmacophore model(s) owing to clustering, we decided to carry a GFA-MLR-QSAR scan against all successful pharmacophores (those of Fisher randomization confidence level ≥85%).

It remains to be mentioned that because it is essential to assess the predictive power of the resulting QSAR models on an external set of inhibitors, we randomly selected 11 molecules and employed them as external test molecules. The QSAR models were also cross-validated automatically using the leave-one-out cross-validation implemented in CERIUS2.[b–e]

Our GFA-MLR-QSAR procedure accessed several successful QSAR models, albeit mostly by scanning cluster representatives and other physicochemical descriptors. Still, in few instances, optimal QSAR models were achieved by scanning the whole list of successful pharmacophore models (e.g., Hypo1/7/1 in eqn 4).

Three QSAR models were rather self-consistent and predicative, warranting their use in subsequent *in silico* scanning. Equations (3), (4), and (5) represent the best QSAR models, while Figures 2, 3, and 4 shows the corresponding scatter plots of experimental versus estimated bioactivities for the training and testing inhibitors.

$$\text{Log}\,(1/\text{IC}_{50}) = -5.800 - 0.140(\text{SaaCH}) + 0.164(\text{Hypo9}/2/1) + 0.536({}^{3}\chi^{\text{P}}) + 0.156(\text{Hypo1}/2/1)$$

$$r^2_{44} = 0.831,\ F\text{- statistic} = 47.886,\ r^2_{\text{BS}} = 0.831,\ r^2_{\text{LOO}} = 0.795,\ r^2_{\text{PRESS}} = 0.711 \quad (3)$$

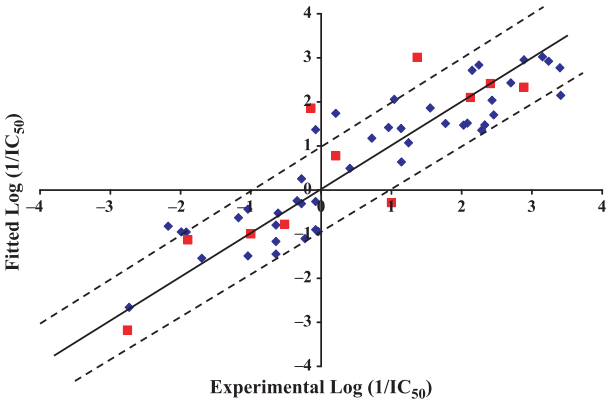

Figure 2: Experimental versus fitted (♦, 44 compounds, r^2_{LOO} = 0.795) and predicted (■, 11 compounds, r^2_{PRESS} = 0.711) bioactivities calculated from QSAR eqn 3. The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 1 log point error margins.

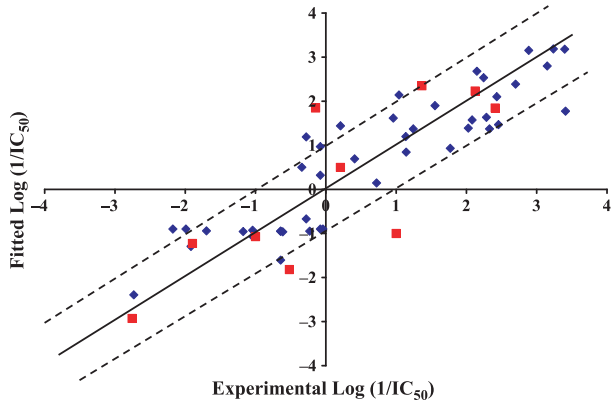

Figure 3: Experimental versus fitted (◆, 44 compounds, r^2_{LOO} = 0.815) and predicted (■, 11 compounds, r^2_{PRESS} = 0.621) bioactivities calculated from the QSAR eqn 4. The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 1 log point error margins.

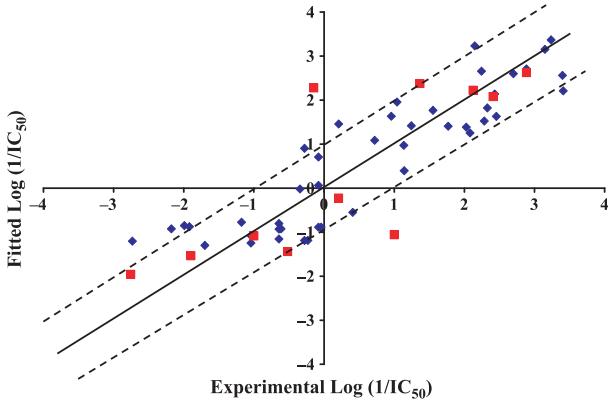

Figure 4: Experimental versus fitted (◆, 44 compounds, r^2_{LOO} = 0.772) and predicted (11 compounds, r^2_{PRESS} = 0.614) bioactivities calculated from QSAR eqn 5. The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 1 log point error margins.

$$\text{Log}\,(1/\text{IC}_{50}) = 0.154 - 0.197(\text{SaaCH}) + 0.446(\text{Hypo1/7/1}) + 1.903(\text{AtypeC43})$$

$$r^2_{44} = 0.815,\ F\text{- statistic} = 58.831,\ r^2_{BS} = 0.815,\ r^2_{LOO} = 0.787,\ r^2_{PRESS} = 0.621 \quad (4)$$

$$\text{Log}(1/\text{IC}_{50}) = -2.076 + 0.005(\text{JursWPSA3})^2 - 0.007(\text{SaaCH})^2 + 0.027(\text{Hypo2/9/2})^2$$

$$r^2_{44} = 0.808,\ F\text{- statistic} = 55.980,\ r^2_{BS} = 0.808,\ r^2_{LOO} = 0.772,\ r^2_{PRESS} = 0.614 \quad (5)$$

where r^2_{44} is the correlation coefficient against the training compounds, r^2_{LOO} is the leave-one-out correlation coefficient, r^2_{BS} is the bootstrapping regression coefficient, and r^2_{PRESS} is the predictive r^2 determined for the 11 test compounds.[b–e] Hypo9/2/1, Hypo1/2/1, Hypo1/7/1, and Hypo2/9/2 represent the fit values of the training compounds against these pharmacophores as calculated from eqn (D) under Supporting Information (Table S4 in Supporting Information shows the fit values of training compounds against the four QSAR-selected pharmacophores), $^3\chi^P$ is the third-order path connectivity index, and SaaCH is the electrotopological state index of aromatic CH atoms. AtypeC43 is atom-type-based descriptor encoding for the hydrophobic contributions of individual carbon atoms. JursWPSA3 is one of the surface-weighted charged partial surface area descriptors implemented in CERIUS2.[b–e]

Figures 5, 6, 7, and 8 show the pharmacophoric models and how they map the most potent training inhibitor **42** (IC_{50} = 0.39 nM) and the most active hit compound **57** (MIC = 187 μM), while Table 2 shows the pharmacophoric features, their tolerances, weights and *X*, *Y*, and *Z* coordinates of each binding pharmacophore. Incidentally, although our QSAR-selected pharmacophore models share HBA, Poslons, and Hbic features with previously reported CATALYST-based pharmacophore model developed for peptidic fungal NMT inhibitors (17), our pharmacophores exhibit additional vectored aromatic ring (RingArom) and hydrophobic features. Furthermore, our pharmacophores exhibit significantly different 3D distribution of features compared with the reported pharmacophore.

Interestingly, all our QSAR-selected models agreed on the significance of Poslons and RingArom features. This trend is probably due to the fact that the binding pocket of NMT is composed of aromatic floor and walls consisting of TYR107, TYR354, PHE117, and TYR335, in addition to the carboxylate of ASP110 that favors binding to cationic centers.

Despite the obscurity of connectivity and topological indices, emergence of $^3\chi^P$ and SaaCH in the optimal models is suggestive of certain role played by the ligands' topology in the binding process. Emergence of JursWPSA3 in one of the optimal QSAR models (eqn 5) is suggestive that in one of its binding modes (corresponding to Hypo2/9/2), NMT favors binding to more polar compounds.

Receiver operating characteristic (ROC) curve analysis

To further validate the resulting models (both QSAR and pharmacophores), we subjected our QSAR-selected pharmacophores to ROC curve analysis. In ROC analysis, the ability of a particular pharmacophore model to correctly classify a list of compounds as actives or inactives is indicated by the area under the curve (AUC) of the corresponding ROC curve as well as other parameters: overall accuracy, overall specificity, overall true-positive rate, and overall false-negative rate.

Figure 9 and Table S5 (Supporting Information) show the ROC results of our QSAR-selected pharmacophores. All QSAR-selected hypotheses illustrated excellent overall performance with an AUC

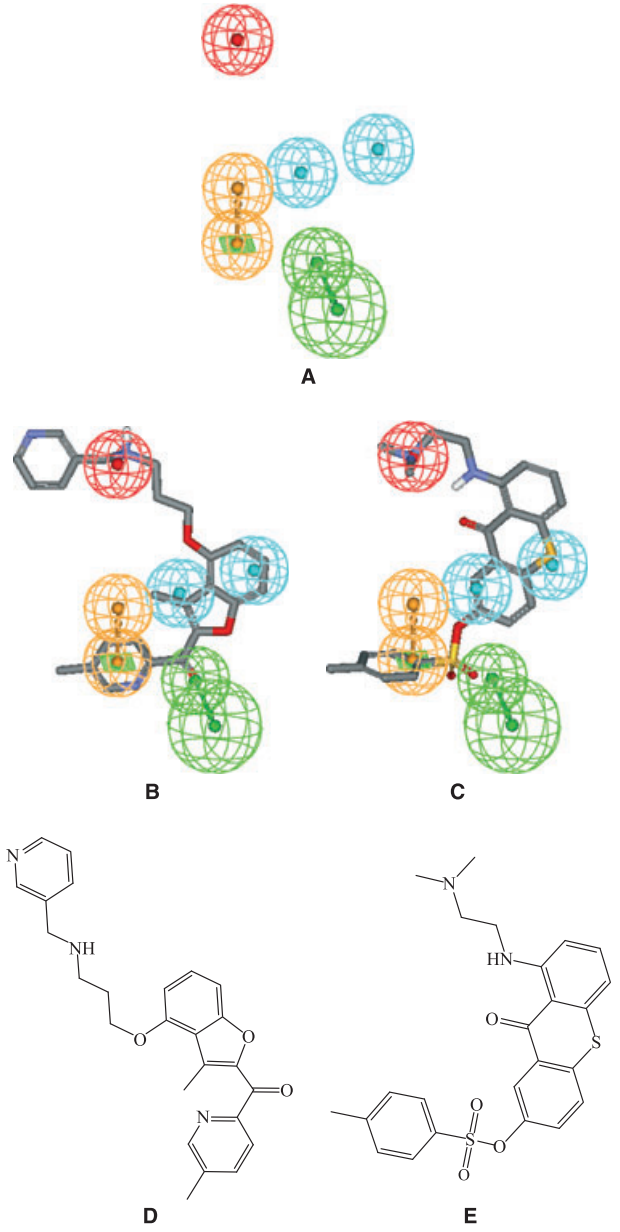

Figure 5: Hypo1/2/1: (A) The pharmacophoric features of the binding model: HBA as green vectored spheres, Hbic as blue spheres, Poslons as a red sphere, and RingArom as orange vectored spheres, (B) and (C) Hypo1/2/1 fitted against **42** (IC_{50} = 0.39 nM, fit value = 14.4644) and **57** (minimum inhibitory concentration = 187 μM, fit value = 8.596), respectively, (D) and (E) chemical structures **42** and **57**, respectively.

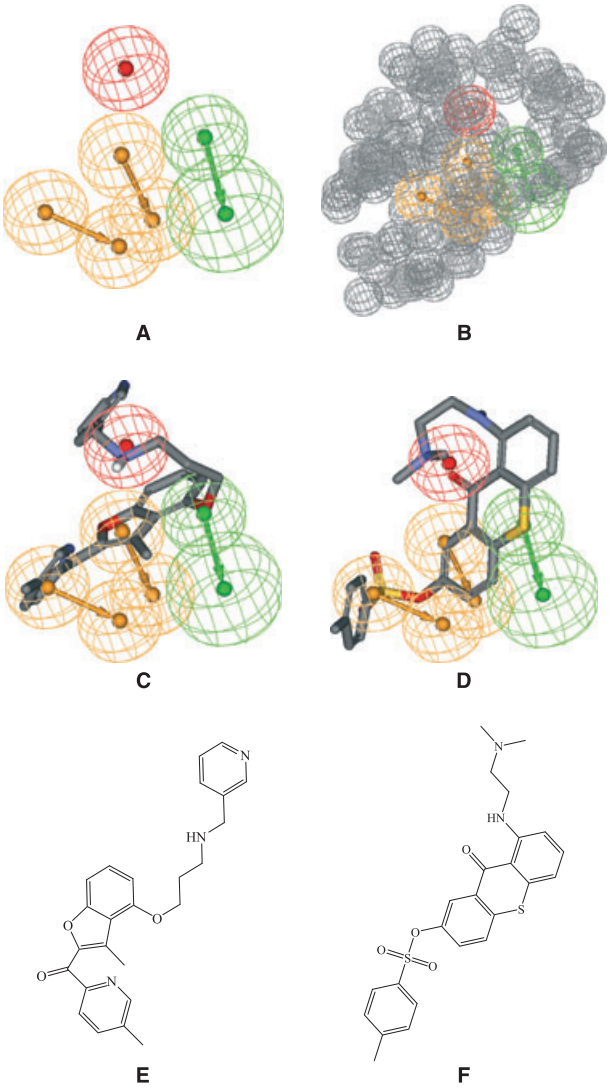

Figure 6: Hypo9/2/1: (A) The pharmacophoric features of the binding model: HBA as green vectored spheres, Hbic as blue spheres, Poslons as a red sphere and RingArom as orange vectored spheres, (B) Hypo9/2/1 with exclusion volumes (gray spheres), (C) and (D) Hypo9/2/1 fitted against **42** (IC_{50} = 0.39 nM, fit value = 13.9183) and **57** (minimum inhibitory concentration = 187 μM, fit value = 10.076), respectively, (E) and (F) chemical structures **42** and **57**, respectively.

ranging from 97.74% to 99.99%. However, Hypo1/2/1 illustrated the best ROC profile with AUC value of 99.99%. In order to enhance the ROC curves of the other three models, namely Hypo9/2/1, Hypo2/9/2, and Hypo1/7/1, we decided to complement them with exclusion spheres. Figure 9 and Table S5 (Supporting Information) show the ROC results of the sterically refined versions of these pharmacophores. Clearly, the performance of the sterically refined versions improved as reflected by their ROC-AUC.

Addition of exclusion volumes

Although ligand-based pharmacophores serve as excellent tools to probe ligand/macromolecule recognition and as useful 3D QSAR models and search queries, they lack steric constraints necessary to define the size of the binding pocket. This liability might render pharmacophoric models promiscuous. Therefore, we decided to

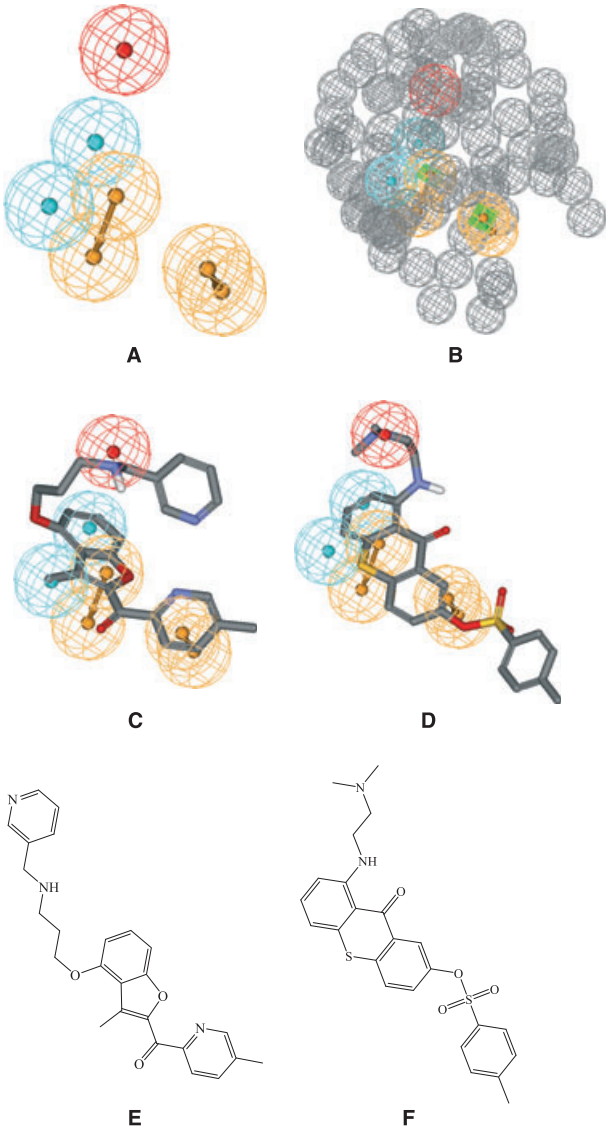

Figure 7: Hypo2/9/2: (A) The pharmacophoric features of the binding model: HBA as green vectored spheres, Hbic as blue spheres, Poslons as a red sphere and RingArom as orange vectored spheres, (B) Hypo2/9/2 with exclusion volumes (gray spheres), (C) and (D) Hypo2/9/2 fitted against **42** (IC_{50} = 0.39 nM, fit value = 11.9049) and **57** (minimum inhibitory concentration = 187 μM, fit value = 7.772), respectively, (E) and (F) chemical structures **42** and **57**, respectively.

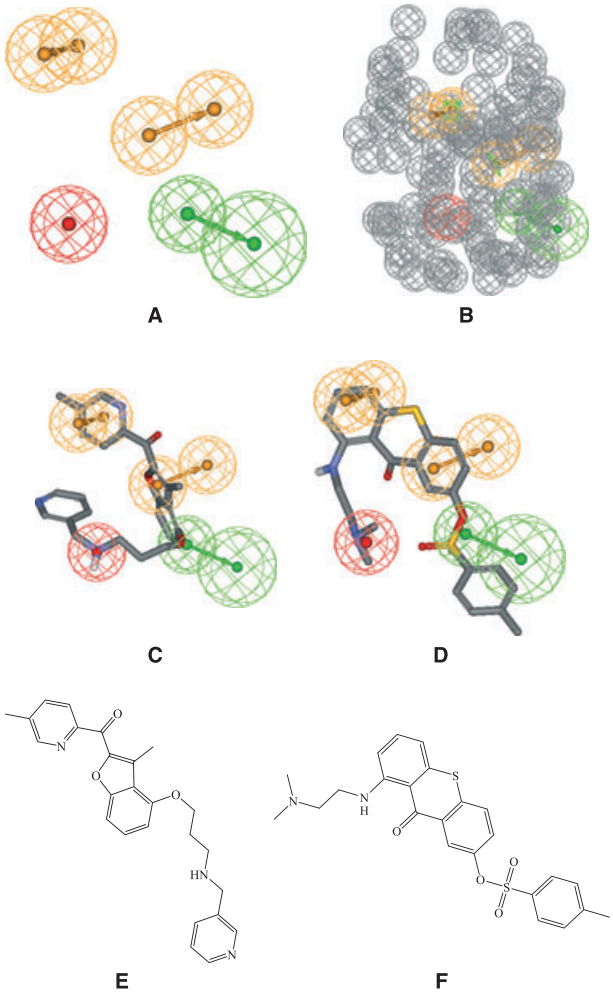

Figure 8: Hypo1/7/1: (A) The pharmacophoric features of the binding model: HBA as green vectored spheres, Hbic as blue spheres, Poslons as a red sphere and RingArom as orange vectored spheres, (B) Hypo1/7/1 with exclusion volumes (gray spheres), (C) and (D) Hypo1/7/1 fitted against **42** (IC_{50} = 0.39 nM, fit value = 11.805) and **57** (minimum inhibitory concentration = 187 μM, fit value = 9.08114), respectively, (E) and (F) chemical structures **42** and **57**, respectively.

complement pharmacophores Hypo9/2/1, Hypo2/9/2, and Hypo1/7/1 with exclusion spheres employing the HIPHOP-REFINE module implemented within CATALYST.[a] Excluded volumes resemble sterically inaccessible regions within the binding site[a] (68).

A subset of training compounds was carefully selected for HIPHOP-REFINE modeling, as in Table S6 (Supporting Information). The training compounds were selected in such away that the bioactivities of weakly active compounds are explainable by steric clashes with the binding pocket. Figures 6B, 7B, and 8B show sterically refined versions of Hypo9/2/1, Hypo2/9/2, and Hypo1/7/1, respectively.

In silico *screening of databases and subsequent experimental evaluation*

Hypo1/2/1 and the sterically refined versions of Hypo9/2/1, Hypo2/9/2, and Hypo1/7/1 were employed as 3D search queries against NCI and DAC. Hits are defined as those compounds that have their chemical groups spatially overlap with corresponding features within the particular pharmacophoric model. Table 3 shows the number of captured hits by each pharmacophore model. The

Table 2: Pharmacophoric features and corresponding weights, tolerances, and 3D coordinates of pharmacophore models in the optimal QSAR eqns (3), (4), and (5)

		Chemical features						
Model	Definitions	HBA		Hbic	Hbic	Poslons	RingArom	
Hypo1/2/1[a]	Weights	2.92678		2.92678	2.92678	2.92678	2.92678	
	Tolerances	1.6	2.20	1.60	1.6	1.6	1.6	1.6
	Coordinates							
	X	−0.34	−3.06	0.54	3.02	0.41	−2.62	−1.08
	Y	3.69	3.05	−1.66	−0.42	3.62	−4.70	−2.66
	Z	1.82	2.93	−1.62	−3.90	2.62	−0.88	0.69
		HBA		Poslons	RingArom		RingArom	
Hypo9/2/1[b]	Weights	3.49080		3.49080	3.49080		3.49080	
	Tolerances	1.6	2.20	1.60	1.6	1.6	1.6	1.6
	Coordinates							
	X	0.34	3.10	1.67	−2.86	−0.81	0.03	−0.02
	Y	−0.51	−3.02	1.35	−2.52	−3.91	0.76	0.02
	Z	−1.75	−3.38	1.24	−0.38	−2.07	−2.18	−5.09
		Hbic	Hbic	Poslons	RingArom		RingArom	
Hypo2/9/2[c]	Weights	2.41442	2.41442	2.41442	2.41442		2.41442	
	Tolerances	1.6	1.6	1.60	1.6	1.6	1.6	1.6
	Coordinates							
	X	3.96	0.83	1.25	−2.31	−0.08	1.33	2.70
	Y	0.63	2.14	2.85	−2.87	−4.78	−0.00	−1.18
	Z	−0.96	−3.40	2.65	−2.12	−2.76	−1.95	−4.34
		HBA		Poslons	RingArom		RingArom	
Hypo1/7/1[d]	Weights	2.07515		2.96450	2.96450		3.85385	
	Tolerances	1.6	2.20	1.60	1.6	1.6	1.6	1.6
	Coordinates							
	X	−4.68	−7.37	−1.91	−0.93	−1.56	3.63	3.53
	Y	0.77	1.95	−3.04	1.47	3.90	1.84	1.23
	Z	0.45	1.21	−0.94	1.14	2.79	−0.02	−2.95

[a]Hypo1/2/1 is hypothesis number 1 generated in run number 2 in set number 1.
[b]Hypo9/2/1 is hypothesis number 9 generated in run number 2 in set number 1.
[c]Hypo2/9/2 is hypothesis number 2 generated in run number 9 in set number 2.
[d]Hypo1/7/1 is hypothesis number 1 generated in run number 7 in set number 1.

anti-NMT bioactivities of captured hits were predicted by QSAR eqns (3), (4), and (5).

The highest-ranking available hits according to the QSAR models (a total of 51 compounds) were acquired for subsequent experimental validation against a laboratory strain of *C. albicans*. Twenty-five hits were found to possess fairly good antifungal activities. The fact that bacterial species lack NMT prompted us to evaluate our active hits against the Gram-negative bacteria *E. coli* to rule out the possibility of non-specific biocidal action of our active hits. Table 4 and Figure 10 show the active hits and their corresponding estimated NMT inhibitory bioactivities and experimental antifungal and antibacterial MIC values.

Interestingly, several active hit compounds illustrated selective antifungal activities, e.g., **57**, **59**, **60**, and **80**, thus strongly suggesting that their antifungal properties are related to NMT inhibition.

Although further work is required to establish the inhibitory activities of these hits against isolated NMT enzyme, our findings unveil a novel class of antifungal agents based on 1-amino-7-hydroxy-xanthenones and 1-amino-7-hydroxy-thioxanthenones, e.g., **59** and **57**, respectively. The fact that these classes of compounds were reported to possess some antitumor properties (69–71) supports our proposition regarding their anti-NMT activities because many tumor types overexpress human NMT, which is a close homolog to fungal NMT.

Intriguingly, a hit from our *in-house* list of DAC illustrated significant and selective antifungal properties, namely the beta-blocker carvedilol (**79**, Table 4 and Figure 10). Undoubtedly, carvedilol and xanthenone derivatives can serve as excellent leads for subsequent medicinal chemistry efforts toward potent NMT inhibitors.

Comparison of Hypo1/2/1 and the binding site of NMT

Despite the problems of crystallographic structures (see introduction) (32–36), pharmacophore features obtained by pharmacophore/QSAR modeling can be compared with the structure of NMT binding site to

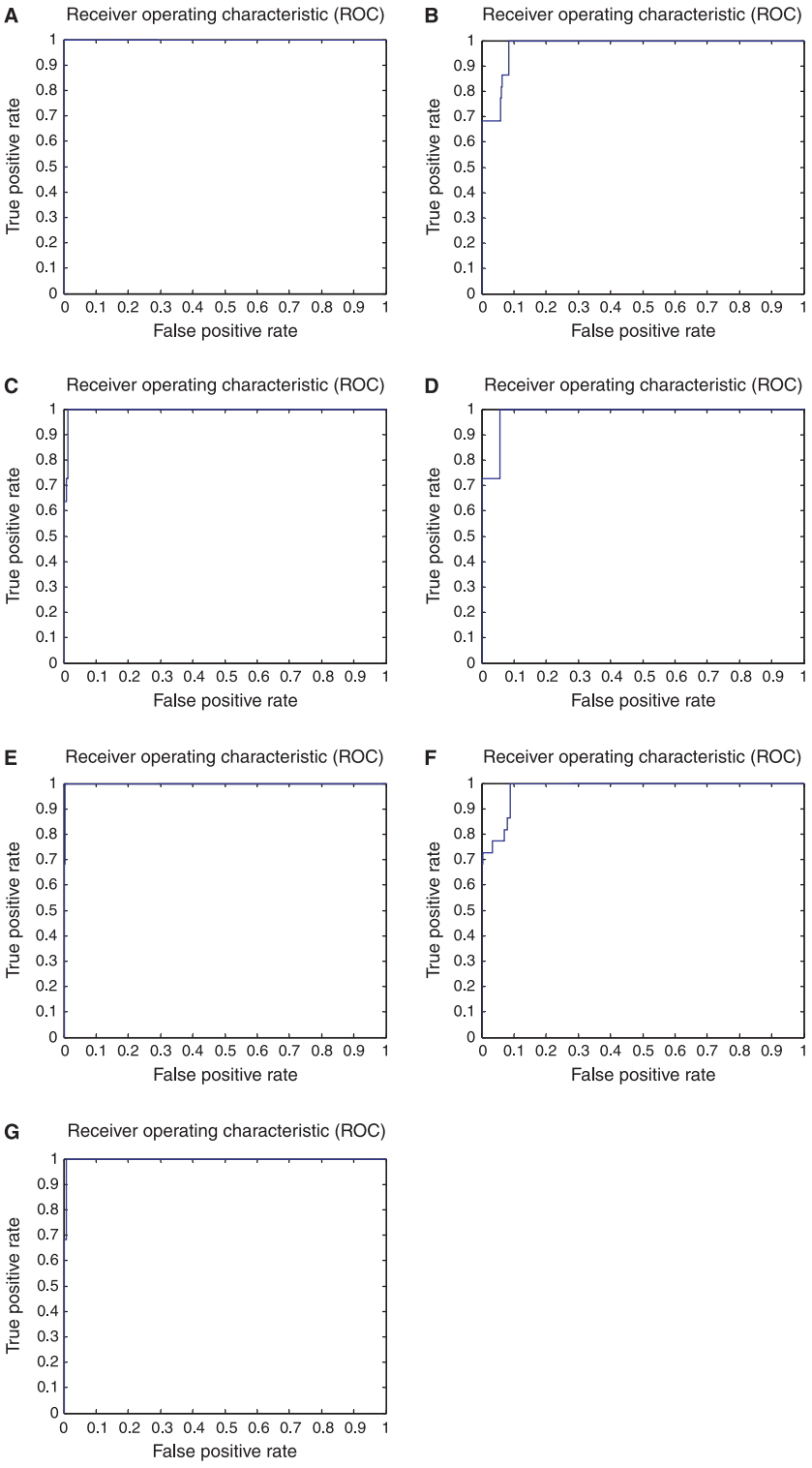

Figure 9: Receiver operating characteristic curves (ROCs) conducted for QSAR-selected models: (A) Hypo1/2/1, (B) Hypo9/2/1, (C) sterically refined Hypo9/2/1, (D) Hypo2/9/2, (E) sterically refined Hypo2/9/2, (F) Hypo1/7/1, and (G) sterically refined Hypo1/7/1.

identify probable residues important for inhibition. To illustrate this point, we compared the features of Hypo1/2/1, as an example, and the way it fits **42** (the most potent training compound) and **60** (one of the most potent tested hits with the highest fit value against Hypo1/2/1 and therefore should ideally behave in docking experiments) with the corresponding structures docked into the binding pocket of NMT (PDB code: 1IYL, resolution 3.2 Å) as in Figure 11.

Table 3: The number of captured compounds by each pharmacophore model

3D Database[a]	Postscreening filtering[b]	Pharmacophore models[c] Hypo1/2/1	Sterically refined Hypo9/2/1	Sterically refined Hypo2/9/2	Sterically refined Hypo1/7/1
NCI	Before	236	3605	734	1407
	After	77	2663	477	917
DAC[d]		4	76	35	39

DAC, drugs and agrochemicals.
[a]NCI: National Cancer Institute list of available compounds (238 819 structures), DAC: the list of established drugs and agrochemicals (3002 structures).
[b]Postscreening filtering employing Lipinski's and Veber's rules. Two Lipinski's violations were tolerated.
[c]The number of captured hits by the sterically refined versions of the pharmacophore models.
[d]This list of compounds was *in silico* scanned without postscreening filtering.

Table 4: The hit molecules captured by Hypo1/2/1, Hypo9/2/1, Hypo2/9/2, Hypo1/7/1, and their corresponding QSAR estimates from eqns (3, 4, and 5) and their *in vitro* bioactivities

Tested hits[a]	NCI codes	Fit values against[b] Hypo1/2/1	Hypo9/2/1	Hypo2/9/2	Hypo1/7/1	QSAR-based estimates IC_{50} (μM) Equation 10	Equation 12	Equation 11	MIC[c] (μM) against *Candida albicans*	MIC[c] (μM) against *Escherichia coli*
56	15792	6.5	11.2	–	5.3	1.7	–	0.7	157	157
57	375390	8.6	10.1	7.8	–	0.1	6.8	–	187	NA[d]
58	121428	5.6	10.1	–	–	0.8	–	–	238	951
59	354676	7.2	10.0	–	–	0.1	–	–	276	NA
60	314594	10.5	7.9	7.1	–	0.4	2.3	–	289	NA
61	14071	5.0	6.2	9.2	9.7	9.8	0.1	0.0	361	361
62	32929	8.6	4.8	9.1	9.9	5.5	0.5	0.0	363	NA
63	261408	6.8	3.9	10.0	–	1.4	0.0	–	380	NA
64	330722	–	–	8.9	8.6	–	0.1	3.6	541	135
65	62323	–	–	–	8.4	–	–	0.0	460	115
66	303772	–	–	9.5	–	–	0.2	–	383	NA
67	370864	5.5	11.1	–	9.6	0.3	–	0.0	572	572
68	116491	6.6	10.8	–	–	1.2	–	–	601	1202
69	342912	5.3	8.6	–	–	0.6	–	–	728	NA
70	656354	–	–	8.9	9.2	–	0.2	0.0	545	NA
71	130815	6.3	10.5	9.7	9.8	0.3	0.0	0.1	495	124
72	23677	5.0	9.4	8.7	9.7	2.0	3.0	0.0	1141	143
73	153423	6.6	10.9	–	7.1	0.8	–	4.5	1061	NA
74	102844	8.8	12.2	–	9.9	0.9	–	2.5	1189	NA
75	109836	3.4	7.3	8.6	–	10.2	0.2	–	1215	608
76	222808	6.1	9.9	–	9.4	6.1	–	0.1	1166	NA
77	33112	0.0	8.4	8.5	8.8	41.8	0.7	0.1	1748	NA
78	12503	0.0	10.0	7.9	7.9	91.7	10.5	0.2	2338	146
79	Carvedilol	0.0	11.3	–	–	16.9	–	–	577	NA
80	Thioridazine HCl	–	–	8.4	9.8	–	27.2	0.0	288	288
	Miconazol[e]	–	–	–	–	–	–	–	27.2	NA

MIC, minimum inhibitory concentration.
[a]Compound numbers as in Figure 10.
[b]Best-fit values against each binding hypothesis calculated by eqn (D), under Supporting Information.
[c]Minimal inhibitory concentration determined for each hit compound against *C. albicans* and *E. coli*.
[d]Not active (at concentrations ≤2 mg/mL).
[e]Reference antifungal agent.

The docking experiments were performed employing LigandFit docking engine, CFF as force field, and through default docking parameters and consensus scoring based on LigScore1, LigScore2, PLP1, PLP2, JAIN, and PMF scoring functions[b] (72,73). However, we confined the docking simulation to rigid docking of pharmacophore-fitted conformers to avoid unnecessary exploration of irrelevant docked conformers and therefore to focus on binding residues corresponding to features within our QSAR-selected pharmacophores.

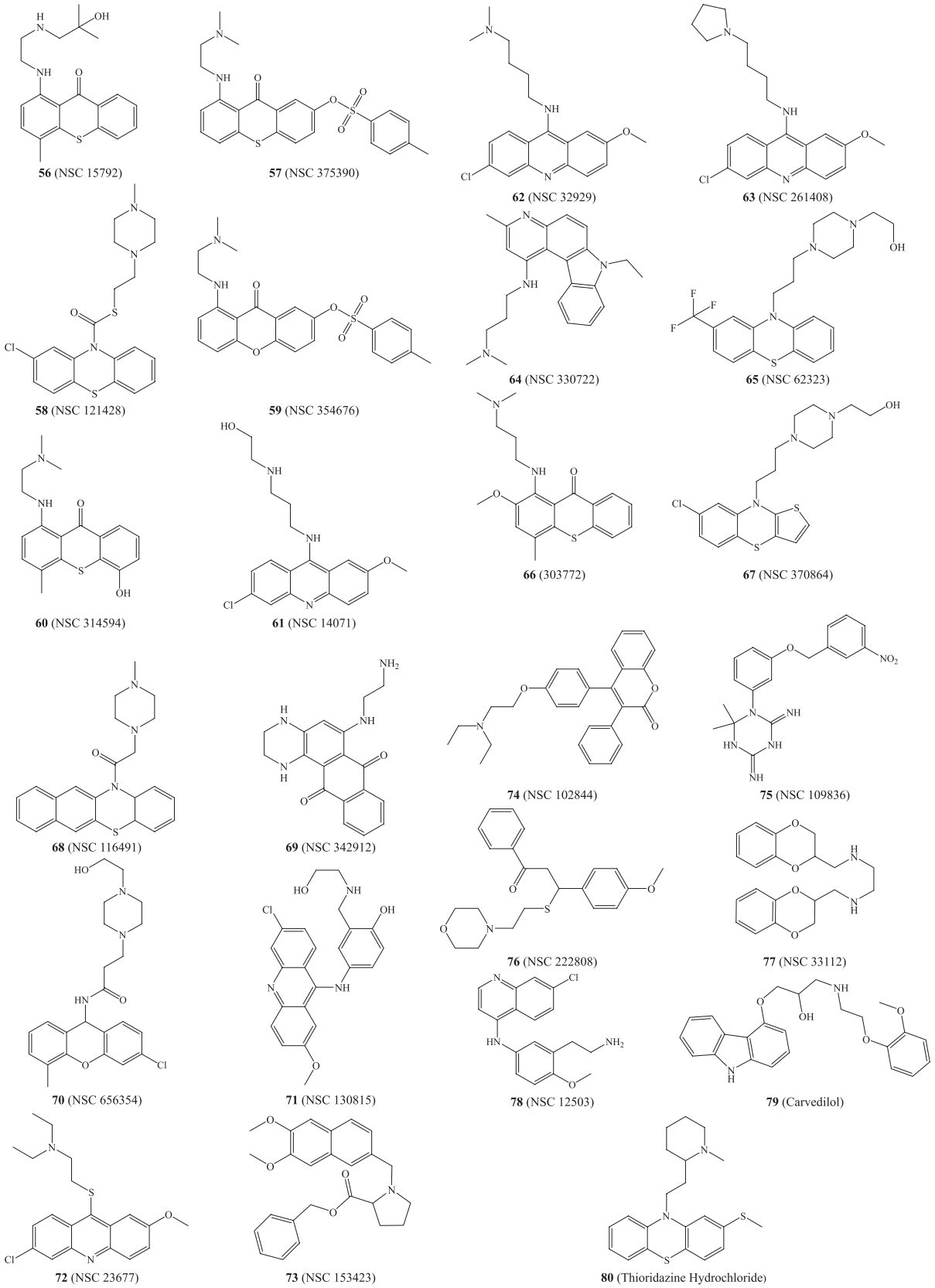

Figure 10: The chemical structures of the tested highest-ranking hits predicted by the best QSAR models and their associated pharmacophores.

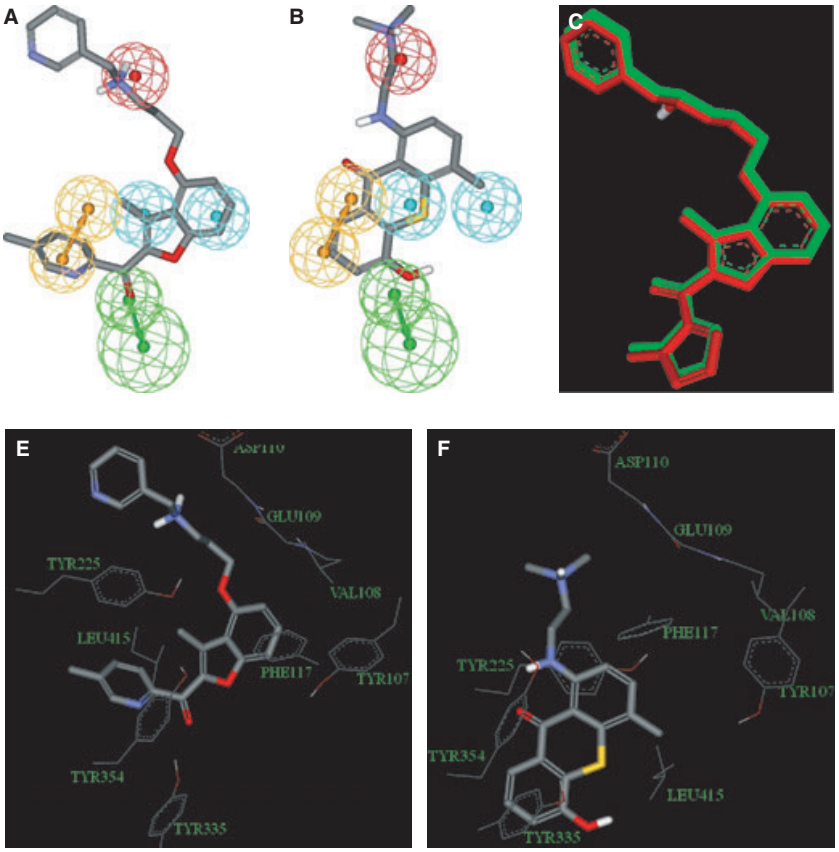

Figure 11: (A) Hypo1/2/1 mapped against **42**, (B) Hypo1/2/1 mapped against **60**, (C) the docked conformer/pose of R64452 as suggested by our docking-scoring conditions (red) compared with the crystallographic conformer/pose of this inhibitor bound within *N*-Myristoyl transferase (green, PDB code 1IYL), (D) docked pose of **42**, (E) docked pose of **60**.

We validated the docking configuration by comparing the docked pose of R64452 with the crystallographic structure of this inhibitor bound within NMT (PDB code: 1IYL), as in Figure 11C.

The best docked pose of **42** scored 1.33, 1.83, −60.31, −63.2, 0.18, and −122.16 based on LigScore1, LigScore2, PLP1, PLP2, JAIN, and PMF docking scoring functions, respectively, while the best docked pose of **60** scored 2.0, 4.81, −83.23, −83.85, −0.38, and −121.18 according to the same scoring functions.

A marked similarity was observed between the features proposed by the pharmacophore models and the ligand-binding features in the docked structures. The features in Hypo1/2/1 as well as the alignment of training compound **42** (IC_{50} = 0.39 nM) and hit molecule **60** (antifungal MIC = 289 μM) as proposed by Hypo1/2/1 were compared with the way these compounds dock into the binding pocket of NMT, as shown in Figure 11. In the highest-ranking docked poses of **42** and **60** (Figures 11D,E), the side chain ammonium fragment is electrostatically bonded to the carboxylic acid moieties of ASP110. This interaction corresponds to a PosIon feature mapping the same amine in Hypo1/2/1 (Figures 11A,B). Moreover, TYR107, VAL108, PHE117, TYR225, and LEU415 constitute a hydrophobic pocket surrounding the benzofuran and thiochromone systems in **42** and **60**, respectively, and therefore nicely corresponds to hydrophobic features mapping these regions in Hypo1/2/1. Moreover, mapping the carbonyl oxygen of **42** and hydroxyl of **60** against a HBA feature in Hypo1/2/1 correlates nicely with hydrogen bonding interactions tying the two groups with the phenolic hydroxyl of TYR335. Finally, mapping the aromatic rings of both compounds against a RingArom feature in the Hypo1/2/1 seems to correspond to π-stacking against the aromatic ring of TYR354.

Conclusions

Our results suggest that this pharmacophore modeling, QSAR analysis, and *in silico* screening of NMT can be a useful tool for finding potential antifungal agents based on NMT inhibition. The pharmacophoric space of a diverse set of NMT inhibitors was explored using CATALYST-HYPOGEN. Subsequently, QSAR analysis was performed to select an optimal model capable of complementing other 2D and physicochemical descriptors in a self-consistent and predictive QSAR models capable of explaining bioactivity variation (*F*-statistic = 47.886, r^2_{LOO} = 0.795, r^2_{PRESS} against 11 external test inhibitors = 0.711 for eqn (3); *F*-statistic = 58.831, r^2_{LOO} = 0.787, r^2_{PRESS} against 11 external test inhibitors = 0.621 for eqn (4) and *F*-statistic = 55.980, r^2_{LOO} = 0.772, r^2_{PRESS} against 11 external test inhibitors = 0.614 for eqn 5).

The validation of the QSAR equations and the associated pharmacophoric models was experimentally established by the identification of several moderately potent NMT inhibitors retrieved from NCI structural database. Although CATALYST-based pharmacophore models were previously reported for NMT inhibitors (17), we believe that our currently presented pharmacophores are superior as they were selected on competitive bases (*via* QSAR analysis) from a large pool of pharmacophores (206 hypotheses clustered into 41 diverse pharmacophores); furthermore, they were complemented with appropriate exclusion spheres and QSAR-based descriptors, which improved their predictive values significantly. Moreover, experimental *in vitro* validation and comparison with docked poses further enhanced the value of our pharmacophores compared with previously reported counterparts.

Acknowledgments

The authors thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan and the Deanship of Research at Jordan University of Science and Technology for their generous funds.

References

1. Fridkin S., Jarvis W. (1996) Epidemiology of nosocomial fungal infections. Clin Microbiol Rev;9:499–511.
2. Kawasaki K., Masubuchi M., Ebiike H., Morikami K., Hayase M., Shindoh H., Sogabe S., Fujii T., Sakata K., Shiratori Y., Aoki Y., Ohtusuka T., Shimma N., Aoyama T., Niizuma S. (2003) Design and synthesis of novel benzofurans as a new class of antifungal agents targeting fungal N-myristoyltransferase. Part 3. Bioorg Med Chem Lett;13:87–91.
3. Devadas B., Freeman S.K., Zupec M.E., Lu H., Nagarajan S.R., Kishore N., Lodge J., Kuneman D., McWherter C., Vinjamoori D., Getman D., Gordon J., Sikorski J. (1997) Design and synthesis of novel imidazole-substituted dipeptide amides as potent and selective inhibitors of *Candida albicans* myristoylCoA:protein N-myristoyltransferase and identification of a related tripeptide inhibitor with mechanism-based antifungal activity. J Med Chem;40:2609–2625.
4. Mcwherter C., Rocque W., Wood D., Gordon J. (1993) A comparative analysis of the kinetic mechanism and peptide substrate specificity of human and *Saccharomyces cerevisiae* myristoyl-CoA:protein N-myristoyltransferase. J Biol Chem;268:9964–9971.
5. Farazi T., Waksman G., Gordon J. (2001) The biology and enzymology of protein N-myristoylation. J Biol Chem;276:39501–39504.
6. Gordon J., Duronio R., Rudnick D., Adams S., Gokel G. (1991) Protein N-myristoylation. J Biol Chem;266:8647–8650.
7. Peitzsch M., McLaughlin S. (1993) Binding of acylated peptides and fatty acids to phospholipid vesicles: pertinence to myristoylated proteins. Biochemistry;32:10436–10443.
8. Wilcox C., Hu J., Olson E. (1987) Acylation of proteins with myristic acid occurs cotranslationally. Science;238:1275–1278.
9. Freeman S., McWherter C., Gordon J., Wood D., Weinberg R., Lee S. (1995) Genetic studies reveal that myristoylCoA:protein N-myristoyltransferase is an essential enzyme in *Candida albicans*. Mol Microbiol;16:241–250.
10. Lodge J., Jackson-Machelski E., Toffaletti D., Perfect J., Gordon J. (1994) Targeted gene replacement demonstrates that myristoyl-CoA:protein N-myristoyltransferase is essential for viability of *Cryptococcus neoformans*. Proc Natl Acad Sci U S A;91:12008–12012.
11. Nakayama H., Mio T., Nagahashi S., Kokado M., Arisawa M., Aoki Y. (2000) Tetracycline-regulatable system to tightly control gene expression in the pathogenic fungus *Candida albicans*. Infect Immun;68:6712–6719.
12. Masubuchi M., Sogabe S., Sakata K., Fukami T., Morikami K., Shiratori Y., Ebiike H., Kawasaki K., Aoki Y., Shimma N., Allan D., Winkler F., Banner D., Tatsuo O. (2002) Crystal structures of *Candida albicans* N-myristoyltransferase with two distinct inhibitors. Chem Biol;9:1119–1128.
13. Devadas B., Freeman S., McWherter C., Kishore N., Lodge J., Jackson-Machelski E., Gordon J., Sikorski J. (1998) Novel biologically active nonpeptidic inhibitors of myristoylCoA:protein N-myristoyltransferase. J Med Chem;41:996–1000.
14. Devadas B., Zupec M., Freeman S., Brown D., Nagarajan S., Sikorski J., McWherter C., Getman D., Gordon J. (1995) Design and syntheses of potent and selective dipeptide inhibitors of *Candida albicans* myristoyl-CoA:protein N-myristoyltransferase. J Med Chem;38:1837–1840.
15. Paige L., Zheng G., DeFrees S., Cassady J., Geahlen R. (1990) Metabolic activation of 2-substituted derivatives of myristic acid to form potent inhibitors of myristoyl CoA:protein N-myristoyltransferase. Biochemistry;29:10566–10573.
16. Parang K., Knaus E., Wiebe L., Sardari S., Daneshtalab M., Csizmadia F. (1996) Synthesis and antifungal activities of myristic acid analogs. Arch Pharm;329:475–482.
17. Karki G., Kulkarni V. (2001) A feature based pharmacophore for *Candida albicans* MyristoylCoA: protein N-myristoyltransferase inhibitors. Eur J Med Chem;36:147–163.
18. Masubuchi M., Kawasaki K., Ebiike H., Ikeda Y., Tsujii S., Sogabe S., Fujii T., Sakata K., Shiratori Y., Aoki Y., Ohtsuka T., Shimma N. (2001) Design and synthesis of novel benzofurans as a new class of antifungal agents targeting fungal N-myristoyltransferase. Part 1. Bioorg Med Chem Lett;11:1833–1837.
19. Masubuchi M., Ebiike H., Kawasaki K., Sogabe S., Morikami K., Shiratori Y., Tsujii S., Fujii T., Sakata K., Hayase M., Shindoh H., Aoki Y., Ohtsuka T., Shimma N. (2003) Synthesis and biological activities of benzofuran antifungal agents targeting fungal N-myristoyltransferase. Bioorg Med Chem;11:4463–4478.
20. Ebara S., Naito H., Nakazawa K., Ishii F., Nakamura M. (2005) FTR1335 is a novel synthetic inhibitor of *Candida albicans* N-myristoyltransferase with fungicidal activity. Biol Pharm Bull;28:591–595.
21. Telvekar V.N., Kundaikar H.S., Patel K.N., Chaudhari H.K. (2008) 3-D QSAR and molecular docking studies on aryl benzofuran-2-yl ketoxime derivatives as *Candida albicans* N-myristoyl transferase inhibitors. QSAR Comb Sci;27:1193–1203.
22. Wu J., Tao Y., Zhang M., Howard M.H., Gutteridge S., Ding J. (2007) Crystal structures of *Saccharomyces cerevisiae* N-myristoyltransferase with bound myristoyl-CoA and inhibitors reveal the functional roles of the N-terminal region. J Biol Chem;282:22185–22194.
23. Farazi T.A., Waksman G., Gordon J.I. (2001) Structures of *Saccharomyces cerevisiae* N-myristoyltransferase with bound myri-

stoylCoA and peptide provide insights about substrate recognition and catalysis. Biochemistry;40:6335–6343.
24. Weston S.A., Camble R., Colls J., Rosenbrock G., Taylor I., Egerton M., Tucker A.D., Tunnicliffe A., Mistry A., Mancia F., de la Fortelle E., Irwin J., Bricogne G., Pauptit R.A. (1998) Crystal structure of the anti-fungal target N-myristoyl transferase. Nat Struct Biol;5:213–221.
25. Bhatnagar R.S., Futterer K., Farazi T.A., Korolev S., Murray C.L., Jackson-Machelski E., Gokel G.W., Gordon J.I., Waksman G. (1998) Structure of N-myristoyltransferase with bound myristoyl-CoA and peptide substrate analogs. Nat Struct Biol;5:1091–1097.
26. Bitencourt M., Freitas M.P. (2009) Bi- and multilinear PLS coupled to MIA-QSAR in the prediction of antifungal activities of some benzothiazole derivatives. J Med Chem;5:79–86.
27. Deokar H.S., Puranik P., Kulkarni V.M. (2009) QSAR analysis of N-myristoyltransferase inhibitors: antifungal activity of benzofurans. Med Chem Res;18:206–220.
28. Li W., Zhao X., Liu F., Hou B. (2008) Investigation on inhibition behavior of S-triazole-triazole derivatives in acidic solution. Corros Sci;50:3261–3266.
29. Zhu J., Sheng C.Q., Zhang M., Song Y.L., Chen J., Yu J.X., Yao J.Z., Miao Z.Y., Zhang W.N. (2006) 3D-QSAR study of a series of novel benzofuran NMT inhibitors. Gaodeng Xuexiao Huaxue Xuebao;27:287–291.
30. Hasegawa K., Morikami K., Shiratori Y., Ohtsuka T., Aoki Y., Shimma N. (2003) 3D-QSAR study of antifungal N-myristoyltransferase inhibitors by comparative molecular surface analysis. Chemom Intell Lab Syst;69:51–59.
31. Purushottamachar P., Kulkarni V.M. (2003) 3D-QSAR of N-myristoyltransferase inhibiting antifungal agents by CoMFA and CoMSIA methods. Bioorg Med Chem;11:3487–3497.
32. Beeley N.R.A., Sage C. (2003) GPCRs: an update on structural approaches to drug discovery. Targets;2:19–25.
33. Klebe G. (2006) Virtual ligand screening: strategies, perspectives and limitations. Drug Discov Today;11:580–594.
34. Steuber H., Zentgraf M., Gerlach C., Sotriffer C.A., Heine A., Klebe G.J. (2006) Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions. Mol Biol;363:174–187.
35. Stubbs M.T., Reyda S., Dullweber F., Moller M., Klebe G., Dorsch D., Mederski W., Wurziger H. (2002) pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design. ChemBioChem;3:246–249.
36. DePristo M.A., de Bakker P.I.W., Blundell T.L. (2004) Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. Structure;12:831–838.
37. Akamatsu M. (2002) Current state and perspectives of 3D-QSAR. Curr Top Med Chem;12:1381–1394.
38. Taha M., Qandil A., Zaki D., AlDamen M. (2005) Ligand-based assessment of factor Xa binding site flexibility via elaborate pharmacophore exploration and genetic algorithm-based QSAR modeling. Eur J Med Chem;40:701–727.
39. Taha M.O., Bustanji Y., Al-Ghussein M., Mohammad M., Zalloum H., Al-Masri I.M., Atallah N. (2008) Pharmacophore modeling, quantitative structure–activity relationship analysis, and in silico screening reveal potent glycogen synthase kinase-3β inhibitory activities for cimetidine, hydroxychloroquine, and gemifloxacin. J Med Chem;51:2062–2077.
40. Taha M.O., Dahabiyeh L.A., Bustanji Y., Zalloum H., Saleh S. (2008) Combining ligand-based pharmacophore modeling, quantitative structure–activity relationship analysis and in silico screening for the discovery of new potent hormone sensitive lipase inhibitors. J Med Chem;51:6478–6494.
41. Taha M.O., Atallah N., Al-Bakri A.G., Paradis-Bleau C., Zalloum H., Younis K., Levesque R.C. (2008) Discovery of new MurF inhibitors via pharmacophore modeling and QSAR analysis followed by in-silico screening. Bioorg Med Chem;16:1218–1235.
42. Taha M.O., Bustanji Y., Al-Bakri A.G., Yousef M., Zalloum W.A., Al-Masri I.M., Atallah N. (2007) Discovery of new potent human protein tyrosine phosphatase inhibitors via pharmacophore and QSAR analysis followed by in silico screening. J Mol Graphics Models;25:870–884.
43. Abu Hammad A.M., Taha M.O. (2009) Pharmacophore modeling, quantitative structure–activity relationship analysis, and shape-complemented in silico screening allow access to novel influenza neuraminidase inhibitors. J Chem Inf Model;49:978–996.
44. Bersuker I.B., Bahçeci S., Boggs J.E. (2000) In: Güner O.F., editor. Pharmacophore Perception, Development, and Use in Drug Design. California: International University Line; p. 457–473.
45. Kurogi Y., Güner O. (2001) Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. Curr Med Chem;8:1035–1055.
46. Accelrys Inc. (2006) Proceedings of the 9th European CATALYST User Group Meeting Advanced Seminars in CATALYST. Frankfurt, Germany. San Diego, CA: Accelrys Inc.
47. Sutter J., Güner O., Hoffmann R., Li H., Waldman M. (2000) In: Güner O.F., editor. Pharmacophore Perception, Development, and Use in Drug Design. California: International University Line; p. 501–511.
48. Sutter H., Li J., Hoffmann R., Güner O.F. (2000) In: Güner O.F., editor. Pharmacophore Perception, Development, and Use in Drug Design. California: International University Line; p. 73–189.
49. Taha M., Al-Bakri A., Zalloum W. (2006) Discovery of potent inhibitors of pseudomonal quorum sensing via pharmacophore modeling and in silico screening. Bioorg Med Chem Lett;16:5902–5906.
50. Ebiike H., Masubuchi M., Liu P., Kawasaki K., Morikami K., Sogabe S., Hayase M., Fujii T., Sakata K., Shindoh H., Shiratori Y., Aoki Y., Ohtsukaa T., Shimmaa N. (2002) Design and synthesis of novel benzofurans as a new class of antifungal agents targeting fungal N-myristoyltransferase. Part 2. Bioorg Med Chem Lett;12:607–610.
51. Smellie A., Kahn S., Teig S. (1995) Analysis of conformational coverage. 1. Validation and estimation of coverage. J Chem Inf Comput Sci;35:285–294.
52. Poptodorov K., Luu T., Hoffmann R. (2006) Methods and principles in medicinal chemistry. In: Langer T., Hoffmann R., editors. Pharmacophores and Pharmacophores Searches. Weinheim: WILEY-VCH; p. 17–47.
53. Fisher R. (1966) The Principle of Experimentation Illustrated by a Psycho-Physical Exp., 8th edn, Chapter II. New York: Hafner Publishing.
54. Krovat E.M., Langer T. (2003) Non-peptide angiotensin II receptor antagonists: chemical feature based pharmacophore identification. J Med Chem;46:716–726.

55. Tropsha A., Gramatica P., Gombar V. (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci;22:69–77.
56. Kirchmair J., Markt P., Distinto S., Wolber G., Langer T. (2008) Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection-what can we learn from earlier mistakes? J Comput Aided Mol Des;22:213–228.
57. Verdonk M.L., Marcel L., Berdini V., Hartshorn M.J., Mooij W.T.M., Murray C.W., Taylor R.D., Watson P. (2004) Virtual screening using protein ligand docking: avoiding artificial enrichment. J Chem Inf Comput Sci;44:793–806.
58. Irwin J.J., Shoichet B.K. (2005) ZINC – a free database of commercially available compounds for virtual screening. J Chem Inf Comput Sci;45:177–182.
59. Triballeau N., Acher F., Brabet I., Pin J.-P., Bertrand H.-O. (2005) Virtual screening workflow development guided by the ``Receiver Operating Characteristic'' curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. J Med Chem;48:2534–2547.
60. Jacobsson M., Liden P., Stjernschantz E., Bostroem H., Norinder U. (2003) Improving structure-based virtual screening by multivariate analysis of scoring data. J Med Chem;46:5781–5789.
61. Gao H., Williams C., Labute P., Bajorath J. (1999) Binary quantitative structure–activity relationship (QSAR) analysis of estrogen receptor ligands. J Chem Inf Comput Sci;39:164–168.
62. Lipinski C., Lombardo F., Dominy B., Feeney P. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev;46:3–26.
63. Veber D., Johnson S., Cheng H., Smith B., Ward K., Kopple K. (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem;45:2615–2623.
64. National Committee for Clinical Laboratory Standards (1993) Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria that Grow Aerobically. Approved Standard M7-A3. Villanova, PA: National Committee for Clinical Laboratory Standards.
65. Sheridan R., Kearsley S. (2002) Why do we need so many chemical similarity search methods? Drug Discov Today;7:903–911.
66. Li H., Sutter J., Hoffmann R. (2000) In: Güner O.F., editor. Pharmacophore Perception, Development, and Use in Drug Design. California: International University Line; p. 173–189.
67. Ramsey L., Schafer W. (1997) The Statistical Sleuth, 1st edn. USA: Wadesworth Publishing Company.
68. Clement O.O., Mehl A.T. (2000) In: Güner O.F., editor. Pharmacophore Perception, Development, and Use in Drug Design. California: International University Line; p. 71–84.
69. Ostrov D.A., Law B.K., Corsino P. (2007) Cyclin-dependent kinase inhibitors, methods for their identification, and their use for the treatment of proliferative disorders. PCT Int Appl WO; 2007050673:172.
70. Archer S., Zayed A.H., Rej R., Rugino T.A. (1983) Analogs of hycanthone and lucanthone as antitumor agents. J Med Chem;26:1240–1246.
71. Brown S., Sandhu G.S. (1997) Lyophilized thioxanthenone antitumor agents. PCT Int Appl WO;9710809:60.
72. Venkatachalam C.M., Jiang X., Oldfield T., Waldman M. (2003) LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J Mol Graphics Models;21:289–307.
73. Gehlhaar D.K., Verkhivker G.M., Rejto P.A., Sherman C.J., Fogel D.B., Fogel L.J., Freer S.T. (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem Biol;2:317–324.

Notes

[a]CATALYST 4.11 User Guide, 2005, Accelrys Software Inc., San Diego, CA.
[b]CERIUS2 4.10 LIGANDFIT User Manual, 2005, Accelrys Inc., p. 3–48.
[c]CERIUS2 4.10 QSAR Users' Manual, 2005, Accelrys Inc., p. 43–88.
[d]CERIUS2 4.10 QSAR Users' Manual, 2005, Accelrys Inc., p. 221–235.
[e]CERIUS2 4.10 QSAR Users' Manual, 2005, Accelrys Inc., p. 237–250.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1: The structures of NMT inhibitors utilized in modeling. The corresponding scaffold is in Figure 1

Table S2: Training sets and CATALYST run parameters employed for exploring NMT pharmacophoric space

Table S3: Performance of the best representatives of clustered pharmacophore hypotheses generated for NMT inhibitors

Table S4: Fit values of training compounds against QSAR-selected pharmacophores as calculated using eqn (D) in S1-1

Table S5: Performance of QSAR-selected pharmacophores and their sterically refined versions as 3D search queries

Table S6: The training compounds used for adding excluded spheres for Hypo9/2/1, Hypo1/7/1 and Hypo2/9/2 using HIPHOP-REFINE module of CATALYST

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.