

A feature based pharmacophore for *Candida albicans* MyristoylCoA: protein *N*-myristoyltransferase inhibitors

Rajeshri G. Karki, Vithal M. Kulkarni*

Department of Chemical Technology, Pharmaceutical Division, University of Mumbai, Matunga, Mumbai 400019, India

Received 4 July 2000; revised 24 October 2000; accepted 15 November 2000

Abstract – A three-dimensional pharmacophore model has been generated for *Candida albicans* MyristoylCoA: protein *N*-myristoyltransferase (NMT) inhibitors, using the software program CATALYST. The in vitro NMT inhibitory activity of a series of peptidic inhibitors was used for pharmacophore generation. The effect of altering the control parameters and feature selection was studied to arrive at the pharmacophore model. The selection of the best hypothesis model was based on the total cost, predictive ability, difference in the cost from the null hypothesis and alignment of the training set compounds on to the hypothesis. The pharmacophore model selected has four features; one hydrophobic, two hydrogen bond acceptor and one positive ionisable function. Groups identified as necessary by scanning alanine mutagenesis studies of the peptidic substrate of *C. albicans* NMT, have been identified as pharmacophore features. Comparison of the ligand binding with the enzyme in the crystal structure of NMT and that proposed by the pharmacophore is consistent. The pharmacophore thus generated can be used as a template for designing non-peptidic inhibitors of NMT. © 2001 Éditions scientifiques et médicales Elsevier SAS

antifungals / catalyst hypothesis / feature mapping / NMT inhibitors / pharmacophore

1. Introduction

Myristoyl CoA: protein *N*-myristoyltransferase (NMT; EC 2.1.3.97) is a cytosolic monomeric enzyme which catalyses the transfer of cellular fatty acid myristate (C14:0), from myristoylCoA to the *N*-terminal glycine amine of a variety of eukaryotic proteins [1]. These protein substrates include kinases, phosphatases, α -subunits of many heterotrimeric G-proteins and endothelial cell nitric oxide synthase. Structural and nonstructural proteins encoded by many viruses, including HIV-1, are also *N*-myristoylated [2, 3]. Different *N*-myristoyl proteins use myristate for various purposes, such as promotion of protein–protein and protein–lipid interactions [3, 4]. Genetic studies have established that NMT is essential for growth and survival of *Saccharomyces cere-*

visiae, *Candida albicans* and *Cryptococcus neoformans* [5–7]. The latter two organisms are the leading causes of systemic fungal infections in immunocompromised humans. Hence, fungal NMTs may serve as a target for the design and development of a new class of antifungal agents. The differences in the substrate specificity of fungal and human NMTs have been exploited to develop species selective inhibitors of NMTs [8, 9]. A high-affinity octapeptide substrate GLYASKLS-NH₂ (figure 1) derived from ADP-ribosylation factor 2 (Arf2p), a yeast protein involved in intracellular protein and vesicular trafficking has been used as the starting point to identify elements critical for recognition by the acyl transferases peptide binding site. Scanning alanine mutagenesis and depeptisation of the octapeptide sequence has revealed three elements critical for molecular recognition. These are the *N*-terminal group of glycine 1, the hydroxyl group of serine 5 and the ϵ -amino group of lysine 6. Moreover, the substitution of alanine for glycine yielded the first competitive inhibitor of NMT, ALYASKLS-

* Corresponding author.

E-mail address: vithal@biogate.com (V.M. Kulkarni).

NH₂ (figure 1) [10]. Using this structure specificity data, several potent peptidic inhibitors of fungal NMT have been synthesised and tested. Replacement of the *N*-terminal tetrapeptide with an 11-aminoundecanoyl group results in a competitive inhibitor (11-aminoundecanoyl-SKLS-NH₂) that is ca. 40-fold more potent than the starting octapeptide. Removal of Leu-Ser from the *C*-terminus generates a competitive inhibitor (11-aminoundecanoyl-SK-NH₂), equivalent to that of the starting octapeptide. A dipeptide inhibitor containing a *C*-terminal *N*-cyclohexylethyl lysinamide moiety has the advantage of being more potent. Substituting a 2-methyl-imidazole for the *N*-terminal amine and structural rigidisation produces more potent inhibitors (IC₅₀ = 20–50 nM) that are up to 500-fold selective for the fungal compared to the human enzyme. Due to the peptidic nature of these

compounds poor in vitro antifungal activity was observed [8].

To derive the minimal structural requirements for NMT inhibition we made use of the three-dimensional Quantitative Structure Activity Relationship (3D-QSAR) software program CATALYST [11].

The CATALYST program models the drug–receptor interaction from the point of view of the receptor, using information derived only from the drug [12]. It automatically generates pharmacophore models that correlate the biological activity observed for a series of compounds to their chemical structure. Molecules are described as collection of chemical functions arranged in three-dimensional space. The conformational flexibility is modeled by creating multiple conformers, judiciously prepared to emphasise representative coverage over a specified energy range. A set of chemical features common to the input molecules

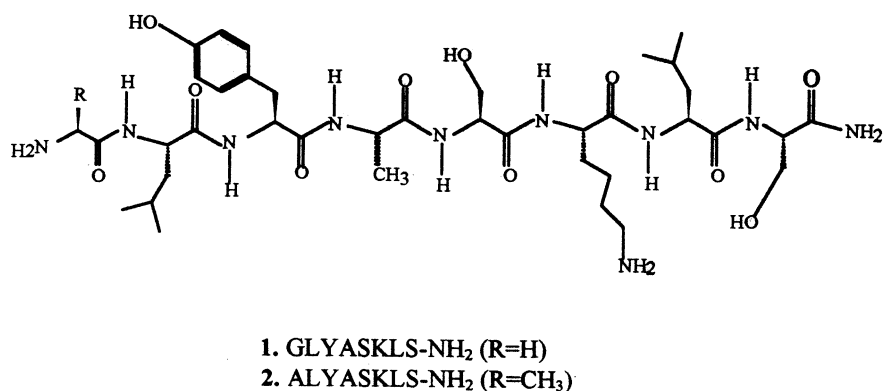


Figure 1. Octapeptide substrate (1) and inhibitor (2).

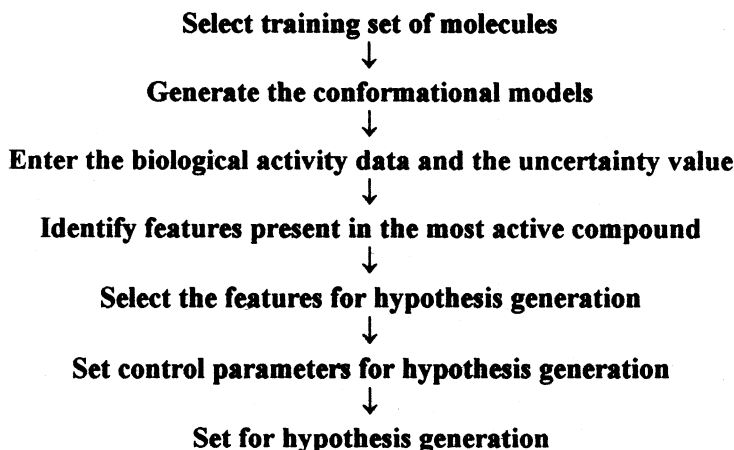
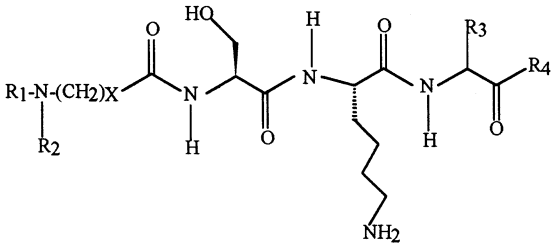
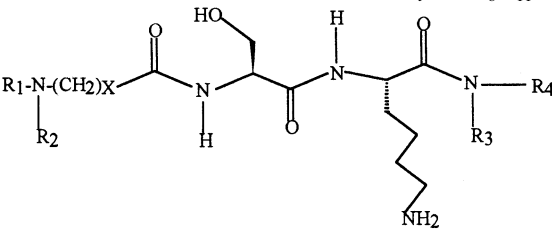
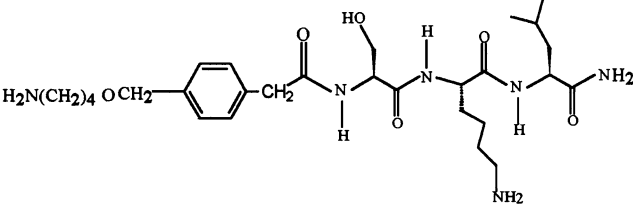
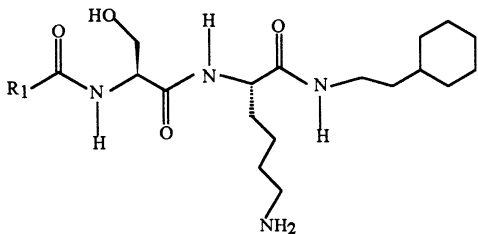
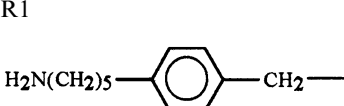
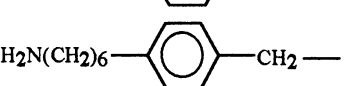


Figure 2. Steps involved in hypothesis generation.

Table I Structures and measured activities of the peptidic *C. albicans* NMT inhibitors

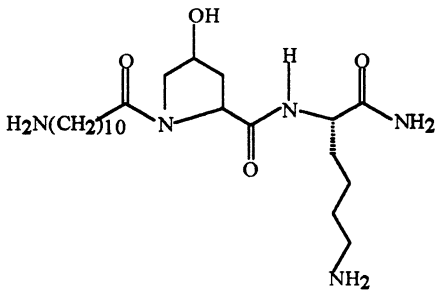
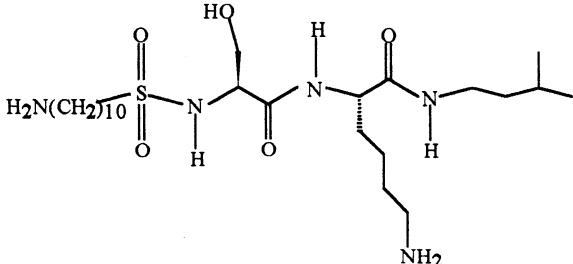
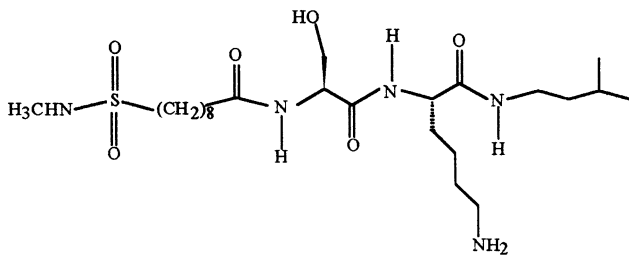
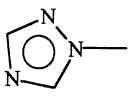
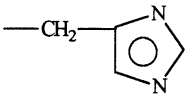
Compound	R ₁	R ₂	R ₃	R ₄	X	IC ₅₀ μM ^a
						
1	H	H	CH ₂ CH(CH ₃) ₂	OCH ₃	10	0.60 ± 0.03
2	H	H	CH ₂ CH(CH ₃) ₂	OCH ₃	11	0.27 ± 0.01
3	H	H	CH ₂ CH(CH ₃) ₂	OCH ₃	9	0.30 ± 0.01
4	H	CH ₃	CH ₂ CH(CH ₃) ₂	OCH ₃	10	0.35 ± 0.01
5	CH ₃	CH ₃	CH ₂ CH(CH ₃) ₂	OCH ₃	10	0.69 ± 0.06
6	H	H	CH ₂ - <i>cyclo</i> -C ₆ H ₁₁	NH ₂	10	0.36 ± 0.06
7	H	H	<i>cyclo</i> -C ₆ H ₁₁	NH ₂	10	0.76 ± 0.08
						
8	H	H	H	-(CH ₂) ₂ - <i>cyclo</i> -C ₆ H ₁₁	10	0.11 ± 0.03
9	H	H	H	(CH ₂) ₂ - <i>cyclo</i> -C ₈ H ₁₅	10	0.30 ± 0.03
10	H	H	H	-(CH ₂) ₂ - <i>cyclo</i> -C ₆ H ₁₁	9	0.20 ± 0.01
11	CH ₃	H	H	(CH ₂) ₂ - <i>cyclo</i> -C ₆ H ₁₁	10	0.19 ± 0.02
12	H	H	H	-(CH ₂) ₂ CH(CH ₃) ₂	10	0.78 ± 0.11
13						0.71 ± 0.16
						
14						0.19 ± 0.02
15						0.11 ± 0.04

Compound	R ₁	R ₂	R ₃	R ₄	X	IC ₅₀ μM ^a
16						0.75 ± 0.01
17						0.52 ± 0.12
18						0.17 ± 0.01
19						0.34 ± 0.06
20						0.75 ± 0.22
21						0.043 ± 0.006
22	R ₁	R ₂			X	0.42 ± 0.15
23	H	H			4	0.056 ± 0.01
24	CH ₃	H			4	0.02 ± 0.001
25	CH ₃	-(R)-CH ₃			4	0.31 ± 0.01
26		-(S)-CH ₃			4	
26						0.60 ± 0.16

Table I (Continued)

Compound	R ₁	R ₂	R ₃	R ₄	X	IC ₅₀ μM ^a
27						1.0 ± 2
28						0.26 ± 0.01
29						0.34 ± 0.05
30						80.0
31						64.0
32						65.0
33	R ₁	R ₂				170.0
34	CH ₂ CH ₂ OH	(CH ₂) ₄ NH ₂				43.0 ± 4
35	CH ₂ C ₆ H ₅	(CH ₂) ₄ NH ₂				490.0
36	CH ₂ F	(CH ₂) ₄ NH ₂				60.0 ± 2
37	CH ₂ OH	(S)-(CH ₂) ₃ CH ₃				98.0 ± 1
38	CH ₂ OH	(S)-CH ₂ -				21.0 ± 0.9

Table I (Continued)

Compound	R ₁	R ₂	R ₃	R ₄	X	IC ₅₀ μM ^a
39						1000.0
40						100.0
41						100.0
42		(CH ₃) ₂	-(CH ₂) ₄ NH ₂			20.5 ± 2
43	NH ₂	cyclo-C ₆ H ₁₁	-(CH ₂) ₃ CH ₃			74.0
44	NH ₂	cyclo-C ₆ H ₁₁				100.0

^a Potency against *C. albicans* NMT expressed as IC₅₀ μM.

Table II. Summary of the results of the hypothesis generated

Hypothesis no.	Training set ^a	Feature selection ^b	Features in the generated hypothesis	PRESS	Total cost	Cost of null hypothesis	Residual cost	Compound mappings
I	A	HBA, HBD HYD, POS	HBA, HBD HYD, POS	2.36	451.29	3912.42	3461.13	79.31% HBA 55.17% HBD 89.66% POS 100% HYD
II		IIHA, IIBD HYD, POS	IIBA, IIBD HYD, POS	3.14	421.4	3911.42	3490.02	89.66% HBA 34.00% HBD 93.00% POS 100% HYD
III	A	HBA, HBD HYD-A1 RA, POS	HBA, HBD 2 × HYD-A1 POS	4.00	495.7	3911.42	3415.72	58.60% HBA 44.80% HBD 79.30% POS 41.38% HYD-A1 86.21% HYD-A2
IV	A	HBA, HYD POS	2 × HBA HYD, POS	2.09	438.82	3911.42	3472.60	34.48% HBA 1 89.66% HBA 2 89.66% POS 89.66% HYD
V	B	HBA, HYD POS	2 × HBA, HYD, POS	1.55	212.89	3646.32	3433.43	82.61% HBA 1 39.13% HBA 2 100% POS 95.65% HYD
VI	C	HBA, HYD, POS	HBA, 2 × HYD POS	1.26	173.95	3543.58	3369.63	55.00% HBA 100% POS 55.00% HYD 1 95.00% HYD 2
VII	C	HBA, HYD POS	2 × HBA HYD, POS	0.92	168.4	3543.58	3375.18	90.00% HBA 1 15.00% HBA 2 100% POS 100% HYD
VIII	C	HBA, HYD POS	2 × HBA HYD, POS	0.77	168.19	3543.58	3375.39	85.00% HBA 1 30.00% HBA 2 95.00% POS 90.00% HYD
IX	C	HBA, HYD POS	2 × HBA HYD, POS	1.42	195.73	3543.58	3347.85	60.00% HBA 1 55.00% HBA 2 90.00% POS 100% HYD
X	D	HBA, HYD POS	2 × HBA HYD, POS	0.87	115.95	1827.12	1711.17	75.00% HBA 1 55.00% HBA 2 90.00% POS 90.00% HYD

^a A-compounds 1–29; B-compounds 8, 15, 17, 21, 23 and 26 excluded from A; C-compounds 3, 14 and 22 excluded from B; D-compounds 1–3, 5–7, 9–14, 18–20, 22, 24–25, 27–29

^b HBA-Hydrogen bond acceptor; HBD-Hydrogen bond donor; HYD-Hydrophobic; HYD-A-Hydrophobic aliphatic; RA-Ring aromatic; POS-Positive ionisable

that correlates best with the biological activity is determined. This 3D array of chemical features (called as hypothesis) provides a relative alignment for each input molecule consistent with their binding to a common receptor site. The chemical features considered can be hydrogen bond donors and acceptors, aliphatic and aromatic hydrophobes, positive and negative charges, positive and negative ionisable groups and aromatic planes. Successful examples involving the use of the CATALYST program have been reported, wherein the CATALYST derived pharmacophore has been used as a query for database searching and 3D-QSAR studies [13–19].

In this paper we present the feature based pharmacophore model for a series of peptidic *C. albicans*

NMT inhibitors. Such a pharmacophore model provides a hypothetical picture of the main chemical features responsible for activity [20]. This pharmacophore may also be utilised for database searching to derive non-peptidic structural leads.

2. Methods

2.1. Data set

The in vitro biological data of a series of 44 diverse peptidic and peptidomimetic inhibitors of *C. albicans* NMT used in the present study are reported by Sikorski et al [8]. The activity was expressed as IC₅₀ (concentration in micromoles required for 50% inhibition of enzyme activity) and was evaluated on recombinant *Candida* NMTs. The enzyme was expressed in *E. coli* and was purified to apparent homogeneity according to protocols described [21]. Standard solutions of inhibitors were prepared at 22 mM in 0.5% DMSO. The IC₅₀ determinations involved incubation of the compound at a series of known concentrations with *C. albicans* NMT in the presence of 0.11 nmol of [³H] myristoylCoA (1 μCi, 9.09 Ci/mmol) and 2.2 nmol of GNAASARR-NH₂ (peptide substrate) in a total volume of 60 μL. After a 10 min incubation at 24°C, the reaction was quenched by addition of ice-cold methanol. The products were separated by HPLC on a Vydac-C₄ column and quantitated by in-line scintillation counting [22].

2.2. Molecular modeling

The molecular modeling study was performed using CATALYST (version 3.1) running on a Silicon Graphics Indigo² computer. All structures were generated using 2D/3D-editor sketcher in CATALYST and were energy minimised to the closest local minimum

2.3. Conformational analysis

Conformational models were generated starting from the local minimised structure using the 'best conformer generation' option. A preliminary study was done with a small set of molecules, wherein conformations were generated by varying the energy threshold value in different runs. The default value of 255 for the maximum number of conformers to be generated per molecule was used. To check for reproducibility, different conformational ensembles were generated starting from different conformers of the same molecule with

Table III. The comparison between the measured and estimated activity of NMT inhibitors using hypothesis VIII

Compound	<i>C. albicans</i> NMT inhibitory activity IC ₅₀ μM		Residual
	Measured ^a	Estimated	
1	0.60	0.58	0.02
2	0.27	0.24	0.03
3	0.30	0.38	−0.08
4	0.35	0.37	−0.02
5	0.69	0.48	0.21
6	0.36	0.33	0.03
7	0.76	0.56	0.20
8	0.11	0.25	−0.14
9	0.30	0.23	0.07
10	0.20	0.21	−0.01
11	0.19	0.21	−0.02
12	0.78	0.33	0.45
13	0.71	0.39	0.32
14	0.19	0.21	−0.02
15	0.11	0.17	−0.06
16	0.75	0.75	0.00
17	0.52	0.27	0.25
18	0.17	0.27	−0.10
19	0.34	0.50	−0.16
20	0.75	0.59	0.16
21	0.043	0.300	−0.257
22	0.42	1.30	0.88
23	0.056	0.24	−0.184
24	0.02	0.019	0.001
25	0.31	0.36	−0.05
26	0.60	0.25	0.35
27	1.00	0.48	0.52
28	0.26	0.31	−0.05
29	0.34	0.53	−0.19

^a Potency against NMT as assessed by IC₅₀ μM at its apparent K_m and myristoylCoA at 1 μM.

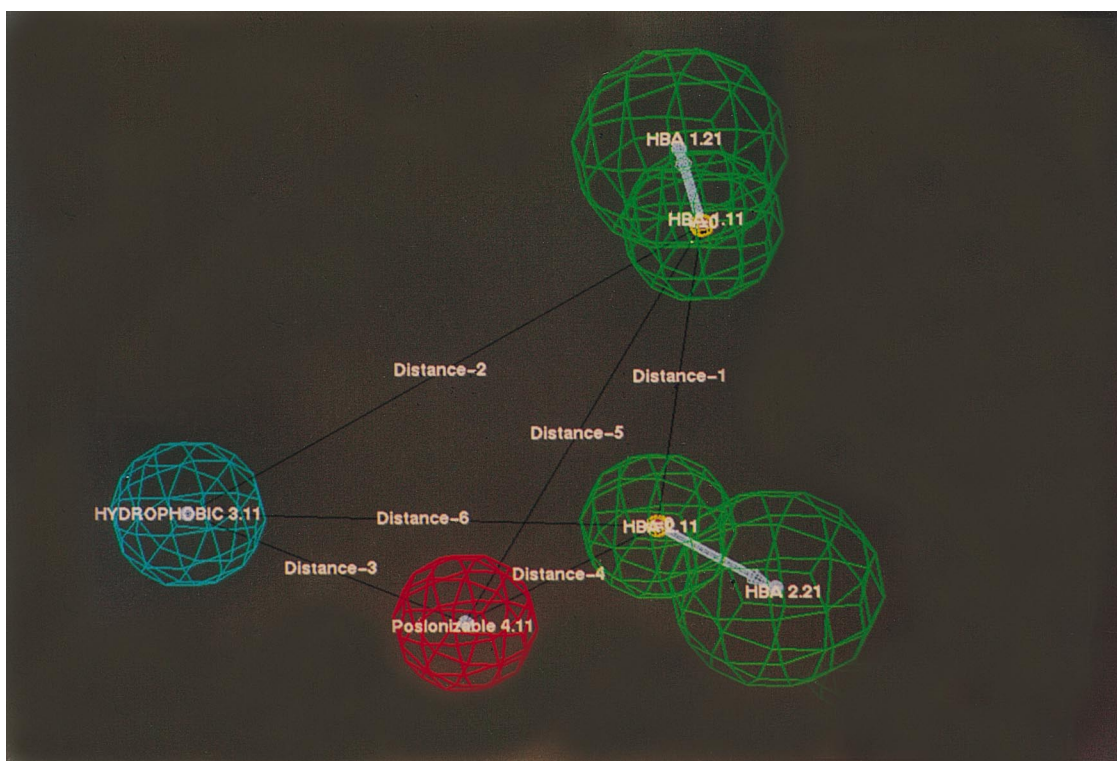


Figure 3. The CATALYST produced *C. albicans* NMT pharmacophore illustrating one hydrophobic feature (cyan), two H-bond acceptor features (green) and a positive ionisable feature (red). Distance-1 = 6.75 Å, Distance-2 = 12.23 Å, Distance-3 = 8.15 Å, Distance-4 = 6.58 Å, Distance-5 = 12.09 Å and Distance-6 = 9.73 Å.

optimised parameters. Following the analysis, conformational ensembles were generated with an energy threshold of 10 kcal/mol from the local minimised structure, for all the molecules.

2.4. Pharmacophore generation

All molecules with their associated conformational models were regrouped into a spreadsheet. The biological data of all the compounds were reported with an upper and a lower limit. To represent such information, the uncertainty value for each compound was calculated. The uncertainty value δ in the biological activity means the activity is situated somewhere in the interval from activity/ δ to activity $\times\delta$. The uncertainty value for each compound was calculated as follows:

$$\delta = \frac{\text{Upper range of biological activity for the compound}}{\text{Actual activity}}$$

It may also be calculated as:

$$\delta = \frac{\text{Actual activity}}{\text{Lower range of biological activity for the compound}}$$

The features present in the most active compound were identified using the 'Show function mapping' command. The compound showed the presence of hydrogen bond donor, hydrogen bond acceptor, hydrophobic and positive ionisable fluctuations. A schematic representation of the CATALYST hypothesis generation process is shown in figure 2. Different hypotheses were generated by altering the default values of the control parameters and altering the feature selection, so as to derive the best pharmacophore model. The control parameters used for hypothesis generation are [23]:

MinPoints: The default value of this parameter is 4, specifying a minimum of 4 individual feature components for a generated hypothesis.

MinSubset points: Only configurations of features in input molecules, with at least the number of points specified by this option, are considered when identifying a candidate hypothesis. The default value is 4.

Superposition error, Check superposition and Tolerance factor: The three control parameters together check the superposition of compounds for hypothesis genera-

tion. All three have a default value of 1. Reducing the value from its default, tightens the fit.

Feature misses: This specifies the number of compounds allowed to not map any particular feature in a generated hypothesis. The default value is 1.

Weight variation: In the catalyst generated pharmacophore model each feature represents certain orders of magnitude of activity of the compound. The extent to which this magnitude is explored by the hypothesis generator is controlled by the parameter weight variation. With the default value of 0.302, CATALYST keeps the value of function weights close to 2.

Mapping coefficient: This parameter controls the importance of having compounds with similar structure map to a hypothesis in a similar way. Increasing this

parameter will penalise the hypotheses that deviate from this behaviour. The default value is 0.

Default values of the parameters spacing (297 picometers), misses (1) and complete misses (0) were used.

2.5. Predictive sum of square (PRESS)

The predictive ability of each hypothesis was determined by calculating the PRESS value for the training set compounds. PRESS is the sum of the squared deviation between predicted and actual activity values for every molecule in the training set [24].

It is defined as:

$$\text{PRESS} = \sum_{i=1}^n (Y_1 - Y_2)^2$$

Table IV. Characteristics of the conformers mapped on hypothesis VIII

Compound	Number of conformers ^a	Conformer used	Fit value ^b	ΔE (kcal/mol) ^c
1	18	6	4.12	8.19
2	83	47	4.50	3.32
3	95	58	4.29	3.21
4	61	39	4.32	0.10
5	28	18	4.21	9.81
6	43	5	4.37	8.97
7	11	6	4.14	5.91
8	70	7	4.48	9.46
9	11	7	4.53	8.41
10	83	33	4.58	4.58
11	84	71	4.56	5.28
12	53	34	4.37	8.01
13	101	7	4.53	9.58
14	69	6	4.55	9.47
15	72	25	4.66	9.57
16	74	55	4.01	3.21
17	70	50	4.45	5.22
18	68	2	4.45	0.00
19	39	15	4.19	6.38
20	64	55	4.12	6.28
21	33	3	4.40	5.40
22	10	13	4.55	8.75
23	21	2	4.51	9.47
24	15	4	4.48	9.13
25	37	28	5.62	8.46
26	29	7	4.33	5.90
27	82	76	4.20	0.00
28	16	2	4.40	0.00
29	34	27	4.17	4.57

^a Number of conformers within the range of 10 kcal/mol from the local minimised structure.

^b Fit: $\sum \text{mapped hypofunction} \times W[1 - \sum \text{hypofunction}(\text{disp}/\text{tol})^2]$ where W = weight of the hypothesis function sphere, adaptively determined, disp = distance of the mapping function on the molecule to the hypothesis function sphere center, tol = radius of the hypothesis function sphere.

^c E (kcal/mol): Energy difference between the conformer used for mapping and the local minimum calculated by CATALYST.

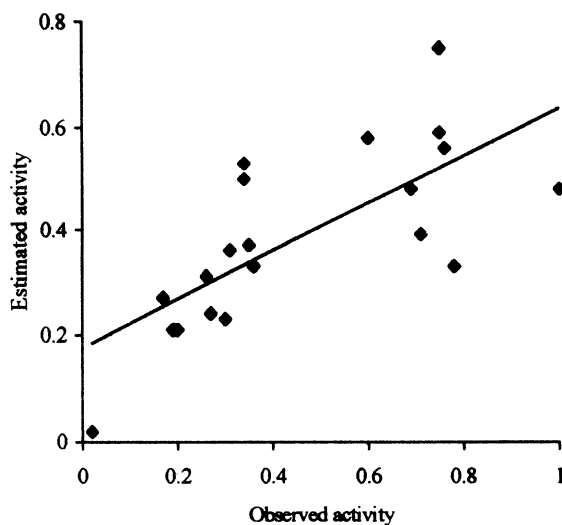


Figure 4. Plot of estimated versus experimental IC_{50} (μM) values for the set of 20 NMT inhibitors using hypothesis VIII.

where, Y_1 = actual biological activity, Y_2 = estimated biological activity and n = number of compounds.

2.6. Activity prediction

The activity of the compounds outside the training set can be estimated from the generated hypothesis. Estimated activity = $10 (I - \text{Fit})$

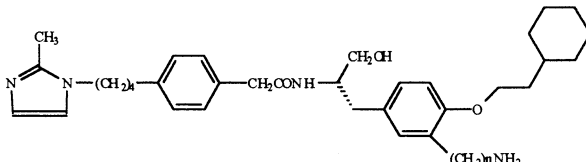
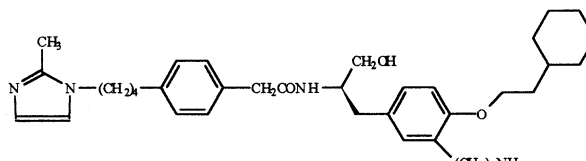
where, I = the intercept of the regression line obtained by plotting the log of the biological activity of the training set compounds against the Fit of the compounds.

Fit = Σ mapped hypo function

$$\times W[1 - \Sigma \text{hypofunction}(\text{disp}/\text{tol})^2]$$

where, W weight of the hypothesis function sphere, adaptively determined; disp = distance of the mapping function on the molecule to the hypothesis sphere center; tol = radius of the hypothesis function sphere.

Table V. Structure and enzyme potency of non-peptidic *C. albicans* NMT inhibitors

Compound	C. albicans NMT inhibitory activity IC ₅₀ μM		
	Measured ^a	Estimated ^b	
			
n			
T1	3	29.3 ± 0.7	19.0
T2	4	20.0 ± 0.03	18.0
T3		17.5	14.0
			

^a Potency against NMT as assessed by IC_{50} μM at its apparent K_m and myristoylCoA at $1 \mu M$.

^b Predicted activity using Hypothesis VIII.

2.7. Error

The error in estimating the activity of compounds is calculated by the formula:

(+) Error = $Y_{\text{estimated}}/Y_{\text{actual}}$ when actual activity
< estimated activity.

(-) Error = $Y_{\text{actual}}/Y_{\text{estimated}}$ when actual activity
> estimated activity

3. Results and discussion

CATALYST enables automatic pharmacophore construction by using a collection of molecules with activity ranging over a number of orders of magnitude. The pharmacophore (hypothesis) produced by CATALYST explains the variability of activity of the molecules with respect to the geometric localisation of the features present in the molecules used to build it. The pharmacophore model consists of a collection of features necessary for the biological activity of the

ligands arranged in 3D space, the common ones being hydrogen bond acceptor and donor, aliphatic and aromatic hydrophobes, positive and negative charges, positive and negative ionisable groups and aromatic planes.

Different hypotheses were generated for a series of peptidic inhibitors of *C. albicans* Myristoyl-CoA:protein *N*-myristoyltransferase (NMT). A total of 44 compounds were used for the study (table I). The compounds were divided into two groups based on their biological activity. The first group constituted 29 compounds with $IC_{50} < 1 \mu\text{M}$ and the second group constituted 15 compounds with $IC_{50} > 20 \mu\text{M}$. Both of the groups consisted of compounds with large structural variations and the chirality of the asymmetric centers was well defined. The biological activity in each group spanned two orders of magnitude.

3.1. Conformational analysis

For each of the training set compound, a conformational database was generated using the 'best conformer

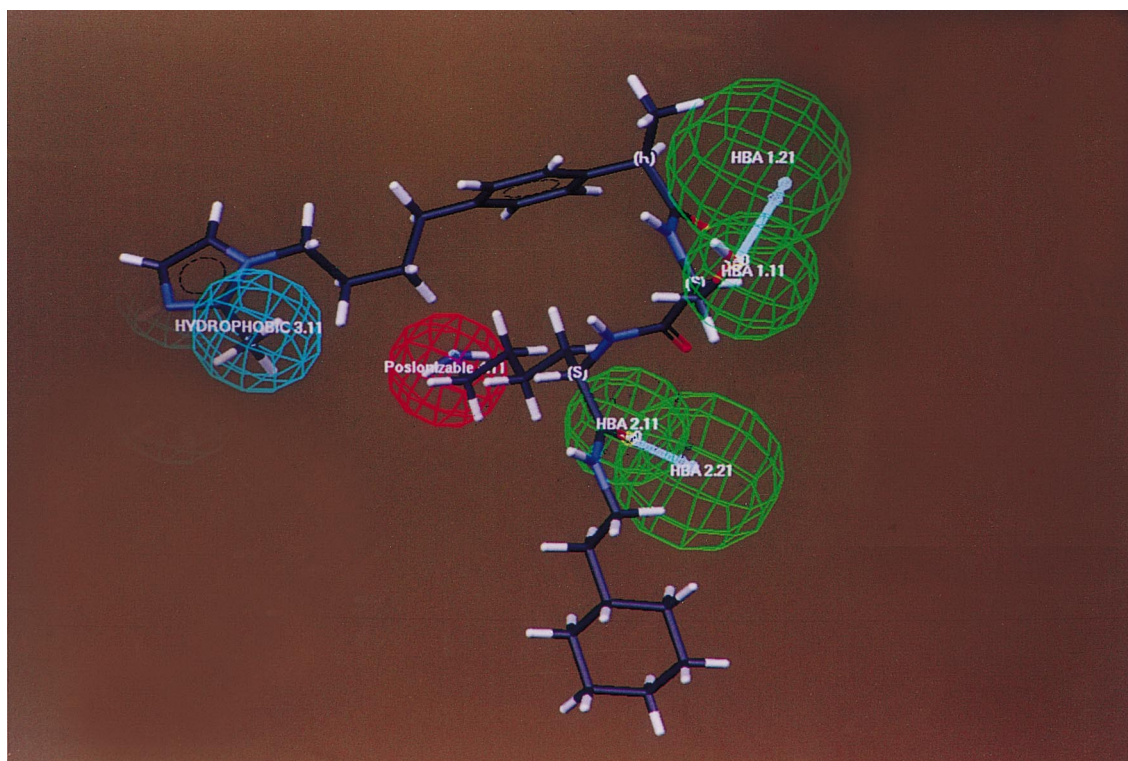


Figure 5. Mapping of compound **24** on hypothesis VIII (cyan: hydrophobic; green: H-bond acceptor, red: positive ionisable function).

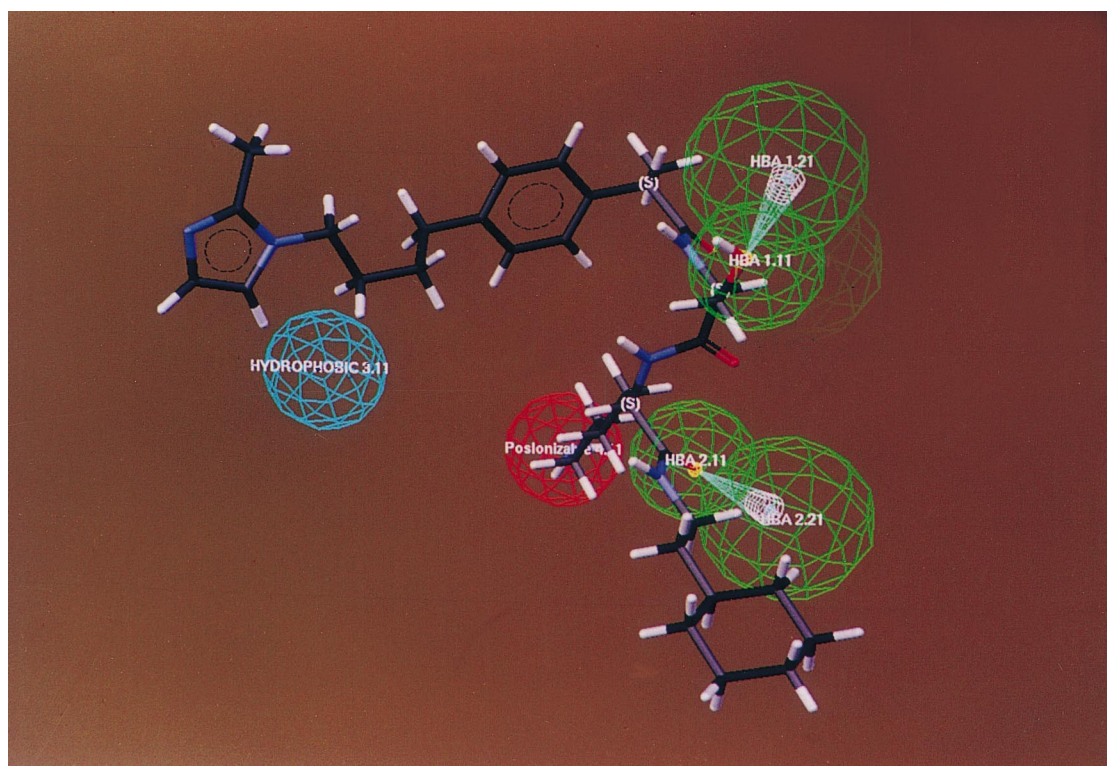


Figure 6. Mapping of compound **25** on hypothesis VIII (cyan: hydrophobic, green: H-bond acceptor, red: positive ionisable function).

generation' option. This option considers the arrangement in space of chemical features rather than simply the arrangement of atoms and uses the poling function. The algorithm intends to optimise the conformational coverage versus the size of the assembly [25, 26]. To explore the effect of control parameters of conformational search on the coverage of conformational space a series of runs was performed on a small set of compounds with varying energy threshold value. The conformational models generated were compared. It was observed that new conformers were generated when the energy threshold was 10 kcal/mol as compared to when the energy threshold was 5 kcal/mol. This indicated that an energy threshold of 5 kcal/mol could not adequately cover the conformational space of the molecules. So for all the molecules conformational ensembles were generated using molecular mechanics threshold of 10 kcal/mol. The default value of 255, for the maximum number of conformers to be generated per molecule was used. All the conformers generated had a *trans* peptide geometry and represented an adequate coverage of the conformational space. To check for reproducibility,

different conformer models were generated starting from different conformers of the same molecule with optimised parameters. The conformers generated were similar in each model thereby ensuring reproducibility. These databases were then used for generating the pharmacophore.

3.2. Pharmacophore generation

Initially to detect the different possible features that these molecules contain, feature mapping was performed on the most active compound **24**. This compound showed the presence of hydrogen bond donor (HBD),

Table VI. Comparison of hypothesis VIII and NMT active site residues

	Pharmacophore features	Active Site residues
1	Hydrophobic	Met 454, Thr 205, Asn 169
2	Hydrogen bond acceptor 1	Gly 418, His 221, Asp 417
3	Hydrogen bond acceptor 2	Asp 417, Gly 416
4	Positive ionisable	Asp 417, Asp 108, Asp 106

hydrogen bond acceptor (HBA), hydrophobic (HYD) and positive ionisable (POS) functions. These features were selected for initial hypothesis generation with the default value of the control parameters. Hypothesis generation was started with 29 compounds (Compounds 1–29, $IC_{50} < 1 \mu M$, *table I*) and the constraint that the hypothesis should have a minimum 1 and maximum 5 of each selected feature. Ten hypotheses were generated in each run. To generate the best hypothesis model, various CATALYST runs were performed, by varying the feature selections, control parameters and constraints. The best hypothesis from each run was selected based on the 'total cost' of the generated hypothesis and the predictive ability. When generating hypotheses, CATALYST tries to minimise a cost function consisting of three terms: Weight cost, Error cost and Configuration cost. Weight cost is a value that increases as the feature weight in a model deviates from an ideal value of 2. The deviation between the estimated activities of the training set and their experimentally determined values adds to the error cost. The third term penalises the complexity of the hypothesis, i.e. the configuration cost. This is a fixed cost, which is equal to the entropy of the hypothesis space. The more the number of features in a generated hypothesis, the higher is the entropy with subsequent increase in this cost. The overall cost of a hypothesis is calculated by summing over the three cost factors. CATALYST also 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. Hence the greater the difference in the cost of the null hypothesis, it is likely that the hypothesis does not reflect a chance correlation.

The results of the hypotheses generated in each run are summarised in *table II*. Hypothesis 1 was generated using 29 compounds. This hypothesis had four features, one each of HBA, HBD, HYD and POS functions. The total cost of the hypothesis was 451.29 and the PRESS value for the activity of the training set was 2.36. On this hypothesis, 79.31% of the compounds mapped the HBA function, 55.17% mapped the HBD function, 89.66% mapped the POS function and all the compounds mapped the HYD function. The percentage mapping of HBD function was low though all the compounds possess this feature.

To refine further the CATALYST hypothesis model I, two hypotheses were generated by altering the default value of the control parameter 'Min points' to 6 (hypothesis II) and altering the feature selection (hypothesis III). It was observed that there was no improvement in

the predictive ability or in the mapping of the compounds with the generated hypothesis.

After analysing the hypotheses generated so far, I–III, it was observed that only 44.66% (average of the three hypotheses) of the total compounds map the HBD function. Thus, the HBD function may not be contributing towards activity and hence may not be a part of the pharmacophore. To ascertain this fact, a hypothesis was generated with the exclusion of this feature (hypothesis IV). The value of the control parameters MinPoints and MinSubset Points was changed to 3, as HBD with a total of two individual feature components is not being considered in the feature selection. Hypothesis IV had reduced cost as compared to hypothesis I (438.82) and improved predictive ability (PRESS 2.09). The hypothesis had two HBA functions (HBA-1 and HBA-2), one HYD function and one POS function suggesting that HBD function may not be a part of the pharmacophore. This hypothesis without HBD function was found better in all respects, i.e. reduced cost, improved predictive ability and better mapping of the compounds on to the generated hypothesis. Hence all further studies were carried out using the feature selection HBA, HYD and POS functions and the values of MinPoints and MinSubsetpoints set to 3.

The hypotheses I–IV, showed an error of ± 3 in estimating the activity of the compounds **8**, **15**, **17**, **21**, **23** and **26**. The higher error could probably be because of the structural variations of the compounds **17**, **21** and **26** (*table I*) as compared to the other training set compounds. The compounds **8**, **15** and **23** though structurally quite similar to compounds **9**, **16** and **22** of the training set showed a large variation in the biological activity (*table I*), which the hypothesis is not able to correlate. In order to generate a better predictive model, these six compounds were kept aside from the training set and a hypothesis was generated with 23 compounds (hypothesis V). This model was able to predict the activity of the training set compounds with a PRESS value of 1.55.

To check the external predictivity of our CATALYST model we have partitioned our training set of 23 compounds into a training set of 20 compounds and a validation test set of 3 compounds. Compounds **3**, **14** and **22** were set aside as a test set as they have wide structural and activity diversity. Hypothesis VI was generated which could estimate the activity of the test set compounds with a very low PRESS value of 0.142, showing good external predictivity of the CATALYST generated pharmacophore model.

To improve upon the function mapping and predictive ability, hypotheses were generated by changing the control parameters. The parameter Weight Variation was changed to 0.2 and Check Superposition to 0.5. Out of the ten hypotheses generated, two models, hypothesis VII and VIII showed good feature mapping and predictive ability. The variation of the parameter 'Feature Misses' to 0, Tolerance Factor to 0.5 and Check Superposition to 0.5 (hypothesis IX) did not improve the feature mapping and predictive ability of the hypothesis model.

The input training set of compounds may influence the features and the predictive ability of the hypothesis. To corroborate this fact different hypotheses were generated by shuffling the training and the test set compounds. The compounds **4**, **10** and **16** that exhibited structural variation and spanned a biological activity range from 0.2 μM to 0.75 μM were kept aside as the test set and compounds **3**, **14** and **22** were included in the training set. The hypothesis X generated had the same features as VII, VIII and IX, i.e. two HBA functions, one HYD and one POS function. This hypothesis had a total cost of 115.95 and the PRESS was 0.87.

The selection of the best model truly representing the pharmacophore was important. For this each of the best hypotheses (VI–X) were analysed with respect to the alignment of each of the compounds with compound **24**. After careful evaluation it was found that hypothesis VIII was the best as it had the lowest cost (168.19) and better predictive ability (PRESS 0.7697). The estimated activity from this hypothesis and the residuals are reported in *table III*. There was a larger difference in the total cost from the null hypothesis (3375.39), implying that the correlation is not by chance. Another interesting feature was the ability of the pharmacophore to distinguish enantiomers **24** and **25**. The molecular alignments of the compounds **24** and **25** are shown in *figures 5 and 6*, respectively. The close mapping of compound **24** with the hypothesis as against that of compound **25** explains the differences in its inhibitory potency.

Hypothesis VIII appeared to represent the most rational alignment when considering the structural diversity of all 20 compounds in the training set. The pharmacophore site points are shown in *figure 3* and the energies of each conformer selected for this alignment, relative to the lowest energy conformer found in the conformational search procedure are listed in *table IV*. A graph depicting the plot of predicted versus observed activities of the training set is shown in *figure 4*.

Having selected the best pharmacophore model, the activity of the six compounds **8**, **15**, **17**, **21**, **23** and **26** were estimated using the pharmacophore model VIII. The errors in estimation of the activity of these compounds when used for hypothesis generation I–IV were greater than ± 3 . The estimated activity of these compounds using hypothesis VIII is shown in *table III*. Except for compounds **21** and **23** the model is able to predict the activity of compounds **8**, **15**, **17** and **26** with precision (error $< \pm 2.5$). However, for compounds **21** and **23** error in estimation of the activity was +6.98 and +4.29, respectively. Compounds **21** and **23** thus demonstrate the limitations of model VIII. CATALYST attempts to produce a structure–activity correlation considering only the geometrical disposition of the pharmacophoric features, though a number of other physicochemical properties along with the 3D-geometrical disposition of pharmacophoric features possessed by the molecule may also affect the biological activity. The higher activity of compound **21** is attributed partly to the highly basic amino group at the *N*-terminus (pK_a 10) which it is assumed will most likely bind as protonated amine and thereby produce a potency advantage. As for compound **23**, it was not possible to ascertain the reason for the higher error in estimating its activity, as its binding with NMT is consistent with that proposed by the pharmacophore.

To further corroborate the predictiveness of the pharmacophore model, the activity of a set of three non-peptidic inhibitors (*table V*) reported by Devadas et al. [27] was predicted using pharmacophore model VIII. These are less potent compared to the peptidic analogues. The estimated activity of these compounds using pharmacophore model VIII is shown in *table V*. Thus it is able to predict the activity of even non-peptidic inhibitors with precision (error < -2). On mapping of the compounds **T1**–**T3** on model VIII it was observed that the molecules did not map all the pharmacophoric features though they did possess the three critical enzyme recognition elements. Further optimisation of the spatial orientation of the critical recognition elements may improve its mapping on the pharmacophore model and its potency against *C. albicans* NMT.

To evaluate the ability of the CATALYST hypothesis generator to discriminate a different set of input data and to determine the features responsible for inactivity, an additional hypothesis was generated using a new training set of inactive compounds (compounds **30**–**44**, $\text{IC}_{50} > 20 \mu\text{M}$, *table I*). The hypothesis generated XI had two HBA functions, two HYD functions and one POS

function. On this hypothesis, 66.67% of the compounds mapped the HBA-1 function, 73.33% mapped HBA-2 function, 80% mapped HYD-1 function, 66.67% mapped HYD-2 function and 40% mapped POS function. The differences in the pharmacophoric features in the two hypotheses VIII and XI indicate the ability of the hypothesis generator to distinguish two sets of data.

The pharmacophoric features of hypotheses VIII and XI were compared to explain the poor activity of compounds **30–44**. Compounds **30–44** did not possess either the ϵ -amino group of lysine or the serine hydroxyl group was modified. As a result, the POS and the HBA-1 features in the hypothesis were poorly mapped. These features are the major determinants of activity in the active model VIII and the poor activity of the compounds **30–44** may be attributed to the absence of these features. The hypothesis model XI contains an additional hydrophobic feature over hypothesis VIII. This may cause steric hindrance for binding with the enzyme thereby exhibiting higher IC₅₀ values.

3.3. Comparison of pharmacophore model with the active site of NMT

The structure of *C. albicans* apo-NMT has been reported recently while this manuscript was under preparation [28]. The structure of the ternary complex of *S. cerevisiae* Nmt1p with bound myristoylCoA and peptide inhibitor was determined to 2.9 Å resolution [29]. The bound analog is SC-58272 (compound **23** in our study). The overall structure of *S. cerevisiae* Nmt1p with the inhibitor structure is very similar to *C. albicans* apo-enzyme. The pharmacophore features can be compared with the enzyme active site to identify the residues important for activity [30]. However, one should not use these pharmacophores as receptor maps [31]. The alignment of the features in hypothesis models VI–X as well as the alignment of compound **24** as proposed by these hypotheses was compared with the crystal structure of the inhibitor bound with the enzyme. A marked similarity was observed, between the ligand binding features in the crystal structure and that proposed by the pharmacophore model VIII.

The methyl group of 2-methyl imidazole (compound **23**) is located in a cavity formed by amino acids Met 454, Thr 205, Asn 169 in NMT. This is a suitable site for a hydrophobic group because of the amino acids Thr 205 and Asp 169 which have uncharged side chains and the hydrophobic amino acid Met 454. The serine hydroxyl group is hydrogen bonded with the amide nitro-

gen of Gly 418 and the lysine carbonyl group is hydrogen bonded with the amide nitrogen of Asp 417. Both act as hydrogen bond acceptors. The charged ϵ -amino group of lysine is buried in a pocket formed by the side chains of Asp 106, Asp 108 and Asp 417 in the crystal structure. This pocket has a negative electrostatic potential because of the acidic amino acids and hence is a suitable site for binding of the positive ionisable group.

The pharmacophore model VIII has one hydrophobic, two hydrogen bond acceptor features and one positive ionisable function. The mapping of the most active compound **24** with the pharmacophore is shown in figure 5. The methyl group overlaps the hydrophobic feature, the serine hydroxyl group maps the hydrogen bond acceptor feature, the lysine carbonyl maps the hydrogen bond acceptor feature and the lysine amino group maps the positive ionisable feature. Thus the 3D array of features is consistent with the binding of the compound in the crystal structure of NMT. Scanning alanine mutagenesis study of the octapeptide substrate GLYASKLS-NH₂ of *C. albicans* NMT has identified the presence of serine hydroxyl group and lysine amino group as necessary for enzyme inhibition [10]. The same is being proposed by the pharmacophore model. The superposition of CATALYST hypothesis on the active site of NMT shows the probable amino acid residues that can have interaction with the inhibitors (table VI). Putting together, it may be concluded, that for compounds to inhibit *C. albicans* NMT, it should possess a hydrophobic feature, able to interact with the side chain of Leu 450 (Met 454 in *S. cerevisiae*), a hydrogen bond acceptor feature which will form a hydrogen bond with the amide of Gly 413 (Gly 418 in *S. cerevisiae*), another hydrogen bond acceptor feature for hydrogen bond interaction with the amide proton of Asp 412 (Asp 417 in *S. cerevisiae*) and a group likely to be protonated at the physiological pH. This will bind the enzyme in the pocket formed by the side chain of amino acids Asp 110, Asp 112 and Asp 412 (Asp 106, Asp 108 and Asp 417 in *S. cerevisiae*).

4. Conclusions

A feature-based pharmacophore has been developed using the program CATALYST for *C. albicans* NMT inhibitors. Different hypotheses were generated by exploring the various possible input parameters. The selection of the best model truly representing the

pharmacophore was done based on its total cost, better predictive ability, larger difference in the total cost from the null hypothesis implying lower probability of chance correlation and alignment of the training set compounds on to the hypothesis. The pharmacophore model has two hydrogen bond acceptor, one hydrophobic and one positive ionisable function. This 3D-feature model can predict the activity of compounds, not included in model generation and can distinguish enantiomers. It can also predict the activity of nonpeptidic inhibitors with precision. However, the higher error in prediction of activity of compounds **21** and **23** demonstrates the limitations of the model. Comparison of the ligand binding with the enzyme in the crystal structure and that proposed by the pharmacophore is consistent. Further optimisation of the spatial orientation of the critical recognition elements in the non-peptidic inhibitors could lead to even more potent and selective NMT inhibitors with improved antifungal activity.

Acknowledgements

The authors thank University Grants Commission (UGC), New Delhi, for financial support under its department of special assistance (DSA) and COSIST programs and Department of Science and Technology (DST), Ministry of Science and Technology, Government of India for the research supports. R.G.K thanks Mr. Santosh, S. Kulkarni and Vijay M. Gokhale for helpful discussion.

References

- [1] Johnson D.R., Bhatnagar R.S., Knoll L.J., Gordon J.I., *Annu. Rev. Biochem.* 63 (1994) 869.
- [2] Boutin J.A., *Cell. Signalling* 9 (1997) 15.
- [3] Bhatnagar R.S., Gordon J.I., *Trends Cell. Biol.* 7 (1997) 14.
- [4] McLaughlin S., Aderem A., *Trends Biochem. Sci.* 20 (1995) 272.
- [5] Duronio R.J., Towler D.A., Heuckeroth R.O., Gordon J.I., *Science* 243 (1989) 796.
- [6] Weinberg R.A., McWherter C.A., Freeman S.K., Wood D.C., Gordon J.I., Lee S.C., *Mol. Microbiol.* 16 (1995) 241.
- [7] Lodge J.K., Jackson-Machelski E., Toffaletti D.L., Perfect J.R., Gordon J.I., *Proc. Natl. Acad. Sci. U.S.A.* 91 (1994) 12008.
- [8] Sikorski J.A., Devadas B., Zupiec M.E., Freeman S.K., Brown D.L., Lu H., Nagarajan S., Mehta P.P., Wade A.C., Kishore N.S., Bryant M.L., Getman D.P., McWherter C.A., Gordon J.I., *Biopolymers* 43 (1997) 43.
- [9] White T.C., Marr K.A., Bowden P.A., *Clin. Microbiol. Rev.* 11 (1998) 382.
- [10] McWherter C.A., Rocque W.J., Zupiec M.E., Freeman S.K., Brown D.L., Devadas B., Getman D.P., Sikorski J.A., Gordon J.I., *J. Biol. Chem.* 272 (1997) 11874.
- [11] Computational results obtained using software program CATALYST 3.1, Molecular Simulations Inc., San Diego, CA, 1996.
- [12] Sprague P.W., *Drug Discovery Des.* 3 (1995) 1.
- [13] Kaminski J.J., Rane D.F., Snow M.E., Weber L., Rothofsky M.L., Anderson S.D., Lin S.L., *J. Med. Chem.* 40 (1997) 4103.
- [14] Greenidge P.A., Carlson B., Bladh L., Gillner M., *J. Med. Chem.* 41 (1998) 2503.
- [15] Langer T., Hoffmann R.D., *J. Chem. Inf. Comput. Sci.* 38 (1998) 325.
- [16] Grigorov M., Weber J., Tronchet J.M.J., Jefford C.W., Milhouse W.K., Maric D., *J. Chem. Inf. Comput. Sci.* 37 (1997) 124.
- [17] Ekins S., Bravi G., Ring B.J., Gillespie T.A., Gillespie J.S., Vandenbranden M., Wrighton S.A., Wikel J.H., *J. Pharmacol. Exp. Ther.* 288 (1998) 21.
- [18] Daveu C., Bureau R., Baglin I., Prunier H., Lancelot J., Rault S., *J. Chem. Inf. Comput. Sci.* 39 (1999) 362.
- [19] Keller P.A., Bowman M., Dang K.H., Garner J., Leach S.P., Smith R., McCluskey A., *J. Med. Chem.* 42 (1999) 2351.
- [20] Hariprasad V., Kulkarni V.M., *J. Comput.-Aided Mol. Des.* 10 (1996) 284.
- [21] Rudnick D.A., Durino R.J., Gordon J.I., in: Hooper N.M., Turner A.J. (Eds.), *Lipid Modifications of Proteins: A Practical Approach*, IRL, Oxford, 1992, pp. 37–61.
- [22] Devadas B., Freeman S.K., Zupiec M.E., Lu H., Nagarajan S.R., Kishore N.S., Lodge J.K., Kuneman D.W., McWherter C.A., Vinjamoori D.V., Getman D.P., Gordon J.I., Sikorski J.A., *J. Med. Chem.* 40 (1997) 2609.
- [23] CATALYST Hip Hop Users Reference 3.1, Molecular Simulations Inc., San Diego, CA, 1996.
- [24] Gokhale V.M., Kulkarni V.M., *J. Med. Chem.* 42 (1999) 5348.
- [25] Smellie A., Kahn S.D., Teig S.L., *J. Chem. Inf. Comput. Sci.* 35 (1995) 285.
- [26] Smellie A., Kahn S.D., Teig S.L., *J. Chem. Inf. Comput. Sci.* 35 (1995) 295.
- [27] Devadas B., Freeman S.K., McWherter C.A., Kishore N.S., Lodge J.K., Jackson-Machelski E., Gordon J.I., Sikorski J.A., *J. Med. Chem.* 41 (1998) 996.
- [28] Weston S.A., Camble R., Colls J., Rosenbrock G., Taylor I., Egerton M., Tucker A.D., Tunnicliffe A., Mistry A., Mancía F., Fortelle E., Irwin J., Bricogne G., Paupit R.A., *Nature Struct. Biol.* 5 (1998) 213.
- [29] Bhatnagar R.S., Futterer K., Farazi T.A., Korolev S., Murray C.L., Jackson-Machelski E., Gokel G.W., Gordon J.I., Waksman G., *Nature Struct. Biol.* 5 (1998) 1091.
- [30] Kulkarni S.S., Kulkarni V.M., *J. Med. Chem.* 42 (1999) 373.
- [31] Cramer R.D., III, Patterson D.E., Bunce J.D., *J. Am. Chem. Soc.* 110 (1988) 5959.