3D QSAR Based Study of Potent Growth Inhibitors of Terpenes as Antimycobacterial Agents

Neeraja Dwivedi[a,b,c], Sanjay Mishra[b,*], Bhartendu Nath Mishra[a], R. B. Singh[d] and Vishwa Mohan Katoch[c]

[a]*Department of Biotechnology, Institute of Engineering & Technology, Lucknow-226021, India*

[b]*Department of Biotechnology, College of Engineering & Technology, Moradabad-244001, India*

[c]*Department of Health Research (Govt. of India) & Indian Council of Medical Research, Ansari Nagar, New Delhi-110029, India*

[d]*Halberg Hospital and Research Center, Civil Lines, Moradabad 244 001, UP, India*

Abstract: The comparative molecular field analysis (CoMFA) based on three dimensional quantitative structure–activity relationship (3D-QSAR) studies were carried out employing, natural terpenes as potent antimycobacterial agents. The best prediction were obtained with a CoMFA standard model (q2 = 0.569, r2 = 0.999) using steric, electrostatic, hydrophobic and hydrogen bond donor fields. In the current study, a 3D QSAR model of natural product terpenes and their related derivative as antimycobacterial agents was developed. The resulted model exhibits wide-ranging *in vitro* potency towards *Mycobacterium tuberculosis*, with minimum inhibitory concentrations (MIC) from 0.25 µg/ml saringosterol through 200 µg/ml diaporthein A. In order to establish structure–activity relationships, 3D-QSAR studies were carried out using CoMFA for natural terpenes (secondary metabolite of plant origin products) as potent antitubercular agents. The *in vitro* Minimum Inhibitory Concentration (MIC) data against *M. tuberculosis* (Mtb) were used. The study was conducted using twenty four compounds. A QSAR model was developed using a training set of sixteen compounds and the predictive ability of the QSAR model was assessed employing a test set of eight compounds. The resulting contour maps produced by the best CoMFA models were used to identify the structural features relevant to the biological activity in this series of natural terpenes.

Keywords: CoMFA, *Mycobacterium tuberculosis*, terpene, MIC, QSAR.

INTRODUCTION

Natural products, or their direct derivatives, play crucial roles in the modern day chemotherapy of tuberculosis. Second-line natural product or related drugs include capreomycin and cycloserine is used in tuberculosis chemotherapy. While rifampicin and streptomycin are part of the front-line treatment regime [1]. Due to a number of factors however, tuberculosis still remains a leading cause of death in the world [2]. The combination of long treatment duration (6–9 months), increased incidence of (multi or extensive) drug resistance, co-morbidity with HIV-AIDS and lack of investment in anti-infectives drug discovery has led to a situation now where the discovery, development and introduction of new treatments for tuberculosis is critical [3].

There is currently a re-emerging interest in natural products as being able to provide novel structures for the drug discovery effort and being particularly effective as antibacterial leads [4]. An excellent minireview by Pauli *et al.* entitled "New perspectives on natural products in TB drug research", which provides an authoritative account of developments in both *in vitro* and *in vivo* antituberculosis bioassays and natural product isolation techniques [5]. While the global economic effects of eradication of tuberculosis have been reported, [6] the development of new drugs clearly requires considerable investment from the public and private sectors. One of the major worldwide facilitators of research and development of new antituberculosis drugs and treatment regimes is the Global Alliance for Tuberculosis Drug Development [7]. It is certainly exciting to see natural products being the focus of such development efforts.

To gain further insight into the relationship between the structure and biological activity, quantitative structure–activity relationships at the three-dimensional level (3D-QSAR), in general, is considered as a powerful approach for developing newer drug leads based on small ligand structures.

The 3D-QSAR, comparative molecular field analysis (CoMFA) method, proposed by Cramer *et al.* in 1988, is extensively used, in the present practice of drug discovery [8]. As for as specificity of CoMFA is concerned it enables to predict biochemical activity of specific molecules by deriving a relationship between electrostatic ⁄ steric properties

*Address correspondence to this author at the Department of Biotechnology, College of Engineering & Technology, Moradabad 244 001, UP, India;
Tel: +91-591-2360818; Fax: +91-591-2360817;
E-mail: sanjaymishra66@gmail.com

1876-3960/11

and biochemical activities, which can be plotted on contour maps. Comparative molecular field analysis calculates steric fields using Lennard–Jones potential and electrostatic fields using a Coulombic potential. In particular, both of the potential functions are very steep near the van der Waals surface of the molecule, causing rapid changes in the surface descriptions and requiring the use of cut-off values so that calculations are not done within the molecular surface. In addition, a scalar factor is applied to the steric field, so that both fields can be used in the same partial least-square (PLS) analysis. Changes in the orientation of the superimposed molecular set, relative to the calculation grid, can cause significant change in the CoMFA results, again probably because of strict cut-off values. So, the alignment rules are one of the most sensitive input areas for CoMFA studies. Several improvements in the alignment methodology like addition of macroscopic descriptor(s) in the study table and a reverse method of CoMFA called adaptation of fields for molecular comparison and topomer CoMFA have been introduced [9-12].

This study is aimed at elucidating the structural features required for potent inhibitors to inhibit the infection of *Mycobacterium tuberculosis* and to obtain predictive 3D-QSAR models, which may guide rational synthesis of potent novel compounds. It would have been difficult to perform the QSAR study at two dimensional levels, as the series under investigation contain structurally diverse compounds with variations at different positions. Hence, 3D-QSAR study was performed to get insight into the structural requirements to be possessed by molecules possessing *Mtb* inhibitory activities. In this study, we report the development of 3D-QSAR models derived from the most widely used computational tools, CoMFA, for structurally diverse sets of natural terpenes as inhibitors of *Mycobacterium tuberculosis* infection from the literature considering their experimentally reported *in vitro* MIC values. The best developed model has been duly validated. Based on the developed model, some highly potent compounds could be designed and synthesized.

In this study, a 3D QSAR model of natural product Terpenes (Monoterpenes, Diterpenes, Sesquiterpenes, Triterpenes and their related derivatives) as antimycobacterial agents were developed. The resulted model exhibits wide-ranging *in vitro* potency towards *M. tuberculosis*, with minimum inhibitory concentrations (MIC) from **0.25** μg/ml saringosterol **through 200** μg/ml diaporthein A. In order to establish structure–activity relationships, 3D-QSAR studies were performed using the Comparative molecular field analysis (CoMFA) of natural terpenes (plant products) as potent antitubercular agents. The *in vitro* MIC data against *Mycobacterium tuberculosis* (Mtb) (already available) was used. The study was performed using Twenty four (24**)** compounds. A QSAR model was developed using a training set of sixteen (16) compounds and the predictive ability of the QSAR model was assessed using a test set of eight (8) compounds.

METHODOLOGY

Biological Activity Data

The antitubercular activity against *Mycobacterium tuberculosis f*or a group of natural terpenes (mono, di, tri and sesquiterpenes) containing 24 compounds as antitubercular agents were used for analysis. The general structure of the compounds, their source and references are indexed in **supplementary table** along with their names. The experimentally calculated MIC value of the compounds shown in Table **1**.

Table **2**, the observed and predicted biological activity in terms of pMIC= -log MIC where, MIC is the minimum inhibitory concentration, expressed in micro moles per milliliter (μM/ml)

Dataset for Analysis

Twenty four molecules selected for this study, were taken from the published work. (References are mentioned in supplementary table). The biological activity used in this study was expressed as pMIC= -log MIC where, MIC is expressed in micro moles per milliliter (μM/ml) (Table **2**) in *in vitro* assay to cause inhibition of *Mycobacterium tuberculosis*.

MOLECULAR MODELING

Molecular Structures and Optimization

For the present study the selected twenty four molecules are the natural terpenes taken from an earlier report [13]**.** The 3d structures of the compounds in the form of sdf files were taken from pubchem database and their biological activity data are taken from literature (related references are provided in Table # **1**). The MIC values were converted to the corresponding pMIC (-log MIC) and used as dependent variables in CoMFA analysis. The MIC values span a range of **0.25** μg/ml to **200** μg/ml providing a broad spectrum data set for 3D-QSAR study. The 3D QSAR models were generated using a training set of 16 molecules. Geometrical standardization followed by specific optimization was performed using MAXIMIN molecular mechanics and Tripos force field, Pullman charge supplied with Sybyl7.0 with the conversion criteria set at 0.05 kcal/(Å mol). Saringosterol is the most bioactive molecule, was used as template molecule for alignment. A common backbone structure was created on the basis of aligned molecules on template molecule. The structure of the template molecule Saningosterol is shown in Fig. (**1**).

Partial Least Squares (PLS) Analysis

PLS algorithm quantifies the structural parameters (CoMFA interaction energies) and biological activities relationship. PLS regression method is specifically advantageous in common cases where the number of descriptors (independent variables) is comparable to or greater than the number of compounds (data points) and/or there exist other factors leading to correlations between variables [14]. The cross-validation analysis was carried out using Leave-One- Out (LOO) method where one compound is removed from the dataset and its activity is predicted using the model derived from the rest of the dataset. The cross-validated q^2 value and the optimum number of components were obtained. A minimum column filtering value (σ) of 2.00 kcal/ mol was used for the cross-validation. It is also advantageous to speed up the analysis and reduce noise [15]. At last non-cross-validated analysis was performed for calculating non-cross-validated r^2 value by using the optimum number of previously identified components, employed to analyze the CoMFA result.

Table 1. Experimentally Calculated MIC Value of the Chemical Sample

S. No.	Chemical Sample	MIC µg/ml	-log (MIC)
1	CID_10483104_camaric acid	32	**0.3189**
2	CID_10007805_rehmanic acid	18	**0.3459**
3	CID_10018804_lecheronol A	4	**0.7213**
4	CID_10494_oleanolic	28.7	**0.2978**
5	CID_14136878_3epiursolic acid	8	**0.4809**
6	CID_14161394_saringosterol	0.25	**-0.7213**
7	CID_3003592_homopseudo	12.5	**0.3959**
8	CID_3008606_lecheronol B	128	**0.2061**
9	CID_460178_totarol	21.1	**0.3279**
10	CID_463811_litosterol	3.13	**0.8764**
11	CID_477494_Nephalsterol	12.5	**0.3959**
12	CID_485179_lantanolic acid	60	**0.2442**
13	CID_485707_3 epioleanolic acid	16	**0.3606**
14	CID_6475529_pseudopterazole	12.2	**0.3997**
15	CID_6480075_bornyl	25	**0.3106**
16	CID_64945_ursolic	41.9	**0.2677**
17	CID_64971_betulinic	62.1	**0.2422**
18	CID_72293_phorbol ester	25	**0.3106**
19	CID_72326_betulin	30	**0.2940**
20	CID_72943_heteronemin	6.25	**0.5456**
21	CID_9816893_ergog	12.5	**0.3959**
22	CID_10248341_diaporthein A	200	**0.1877**
23	CID_10473957_diaporthein B	3.1	**0.8838**
24	CID_44584761_lantinilic acid	73	**0.2330**

RESULTS AND DISCUSSION

Twenty four molecules were randomly divided into a training set of sixteen compounds and a test set of eight compounds having chemical and biological diversity in both the training set and the test set molecules. Regardless of the uncertainty of drug-receptor interactions in general, a statistically significant model was obtained from the CoMFA analysis. A "cross-validated q^2 value" can be defined by completely analogously to the definition of the conventional q^2, as

$$\text{Cross-validated } q2 = (\text{SD} - \text{press})/\text{SD}$$

Where press is the standard errors of the cross-validated predictions and SD is the sum of squared deviations of each biological property value from their mean and press, or predictive sum of squares, is the sum, over all compounds, of the squared differences between the actual and "predicted" biological property values [16].

Often for QSARs model developed by CoMFA a change in the $q2$ values is observed as the grid spacing is distorted. The control of the different grid spacing to CoMFA model is noticeable. The model with the grid spacing of 2.0 Å was selected as the best model by cross validating value ($q2$) after LOO cross-validation.

The statistical parameters of CoMFA analysis is compiled in Table **3**. A cross-validated value (q^2) of 0.569 of the best model was obtained through leave-one-out (LOO) cross validated PLS analysis, which suggest that the model is a helpful tool for predicting inhibitory activity of natural terpenes [17]. The correlation coefficient between the calculated and experimental activities, non-cross-validated value (r^2) of 0.999 with standard error estimate 0.024. The respective relative contributions of steric and electrostatic fields were 0.943 and 0.664, indicating that steric field is more predominant. Then the condition without electrostatic were studied and the new q^2 and r^2 values were found to be 0.739

Table 2. The Observed and Predicted Biological Activity in Terms of pMIC= -log MIC where, MIC is the Minimum Inhibitory Concentration, Expressed in Micro Moles Per Milliliter (μM/ml)

S. No.	Chemical Sample	pMIC (Observed)	pMIC (Predicted)	Residual
1	*CID_10483104_Camaric acid	**0.319**	**-0.410**	**0.729**
2	CID_10007805_rehmanic acid	0.346	0.342	0.004
3	*CID_10018804_lecheronol A	**0.721**	**0.710**	**0.011**
4	CID_10494_oleanolic	0.298	0.245	0.053
5	*CID_14136878_3epiursolic acid	**0.481**	**0.450**	**-0.031**
6	CID_14161394_saringosterol	-0.721	0.695	-0.026
7	*CID_3003592_homopseudo	**0.396**	**0.470**	**0.074**
8	*CID_3008606_lecheronol B	**0.206**	**-0.860**	**1.066**
9	CID_460178_totarol	0.328	0.096	0.232
10	CID_463811_litosterol	0.876	0.897	-0.021
11	CID_477494_Nephalsterol	0.396	0.325	0.071
12	CID_485179_lantanolic acid	0.244	0.221	0.023
13	CID_485707_3 epioleanolic acid	0.361	0.730	-0.369
14	CID_6475529_pseudopterazole	0.399	0.403	-0.004
15	CID_6480075_bornyl	0.311	0.317	-0.006
16	CID_64945_ursolic	0.268	0.295	-0.027
17	CID_64971_betulinic	0.242	0.238	0.004
18	CID_72293_phorbol ester	0.311	0.385	-0.074
19	CID_72326_betulin	0.294	0.814	-0.520
20	*CID_72943_heteronemin	**0.545**	**0.435**	**0.113**
21	*CID_9816893_ergog	**0.396**	**0.237**	**0.159**
22	*CID_10248341_diaporthein A	**0.188**	**-1.232**	**1.420**
23	CID_10473957_diaporthein B	0.884	0.878	0.006
24	CID_44584761_lantinilic acid	0.233	0.229	0.004

and 0.953 respectively. Practically the electrostatic contribution was taken to be negligible. The actual and predicted values of the best model of the training set are given in Table **2**. A test set of four compounds (in Table **2**) which were not included in the development of the model, were used for validation of 3D-QSAR model. This is further validated by the residual values of the test set (Table **2**) on the basis of the PLS statistics of CoMFA model.

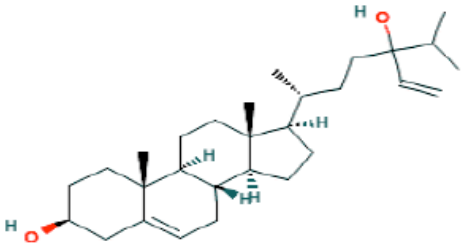

Fig. (1). Active structure of Saringosterol.

The contour plot representations of the CoMFA results for tuberculosis inhibitors is presented in Fig. (**2**) using compound **6** as reference structure. The green-colored regions indicate areas where steric bulk enhances tuberculosis inhibitory activity, while the yellow contours indicate regions where steric bulk is detrimental for the biological activity. Blue-colored regions show areas where electropositive charged groups enhance inhibitory activity, while red regions represent where electronegative charged groups improve the activity.

The electrostatic contour map displayed in Fig. (**3**) shows a region of red polyhedral space, indicating that the electron-rich groups are beneficial to the activity. Additionally, a blue polyhedron in the Fig. (**3**) indicates that electron-rich substituent will reduce the biological activity.

Table 3. PLS Statistics of CoMFA 3D-QSAR Model

PLS Statistics	CoMFA
*q*2 (leave-one out cross-validated predicted power of model)	0.569
*r*2 (correlation coefficient squared of PLS analysis)	0.999
N (optimum number of components obtained from cross-validated PLS analysis and the same used in final non cross-validated analysis)	3
Standard error of estimate (SEE)	0.024
F-test value (F-value) (n1= 6, n2= 9)	1690.169
R^2 prediction	0.507
Steric field contribution from CoMFA	0.943
Electrostatic field contribution from CoMFA	0.664

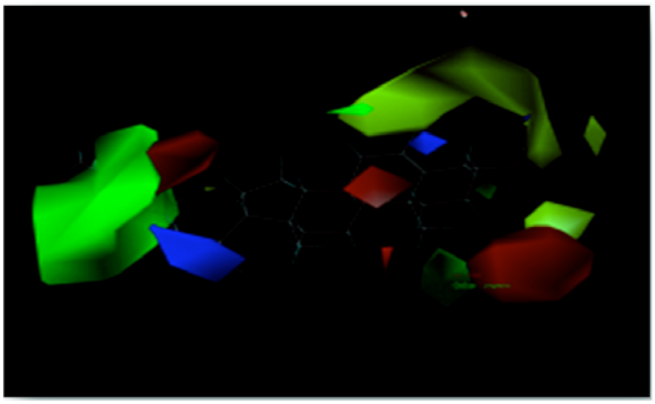

Fig. (2). CoMFA contour maps. In the steric contour map to the left, green contours indicate areas where steric bulk is predicted to increase antimycobacterial activity, while red contours indicate regions where steric bulk is predicted to decrease activity. The electrostatic contour map on the right displays yellow polyhedra where partial negative charge is correlated with antimycobacterial activity; the blue polyhedra indicating a relationship between partial positive charge and activity with bioactive compound *Saringosterol*.

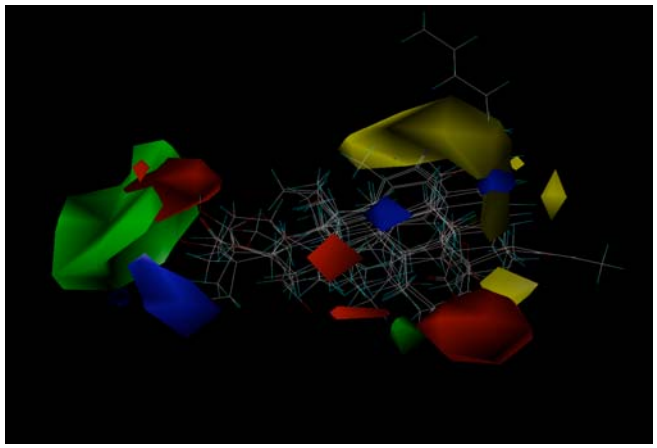

Fig. (3). CoMFA contour map of the steric field and electrostatic field with **a**lignment of the compounds used in the training set of 3D-QSAR analysis.

The steric contour map is displayed in Fig. (**3**) by big green polyhedrons. If a substituent, such as 3-pentyl group, is attached on molecule **6**, it occupies the green contour and enhances the biological activity. In compounds where the substituent group R is cyclopentyl and 1-methylcyclohex-1-yl, can fit into the green contour, and the activity of these compounds are higher. Additionally, the contour plot shows a yellow polyhedron in the top left corner of Fig. 2. The yellow contour will slow down the biological activity, if a bulky substituent exist, which represents the disfavored steric region. Consequently, these compounds almost have no biological activity.

CONCLUSION

In present study a 3D QSAR model was developed using CoMFA on a series of natural derivatives of terpenoid compounds having inhibitory activity against *Mycobacterium tuberculosis*. LOO cross-validation value (q^2) and non-cross-validated value (r^2) was obtained **0.569 and 0.999** respectively, shows the acceptable CoMFA model. The developed CoMFA model also possesses promising foretelling ability as validated by the testing on the external test set. It could be significant to elucidate the relationship between compound structures and biological activities to facilitate designing of more potent semi-synthetic terpene inhibitors. Our theoretical prediction may lead to the establishment of developing new semi-synthetic inhibitors for antimycobacterial activity using synthetic chemistry approaches.

ACKNOWLEDGEMENTS

The authors are grateful to Prof. R.M. Dubey (Managing Director) IFTM, Moradabad, U.P., INDIA and Prof. A. Srivastav (Director), College of Engineering and Technology, Moradabad, U.P., India, for providing research promotional grant and their generous help and encouragement during the course of experimental work and manuscript preparation.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers web site along with the published article.

REFERENCES

[1] Ormerod LP. Multidrug-resistant tuberculosis (MDR-TB): epidemiology, prevention and treatment. Br Med Bull 2005; 73/74: 17.

[2] The Global Plan to Stop TB, 2006–2015. Document available from www.stoptb.org .

[3] Newman DJ, Cragg GM, Snader KM. The influence of natural products upon drug discovery. Nat Prod Rep 2000; 17: 215.

[4] Economics of TB drug development. Report available from Global Alliance for TB Drug Development, www.tballiance.org.

[5] Pauli GF, Case RJ, Inui T, *et al*. New perspectives on natural products in TB drug research. Life Sci 2005; 78: 485-94.

[6] Economics of TB drug development. Report available from Global Alliance for TB Drug Development, www.tballiance.org

[7] Web source: www.tballiance.org.

[8] Cramer RD III, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA): Effect of shape on binding steroids to carrier proteins. J Am Chem Soc 1988; 110: 5959-67.

[9] Talale TT, Kulkarni SS, Kulkarni VM. Development of pharmacophore alignment models as input for comparative molecular field analysis of a diverse set of azole antifungal agents. J Chem Inf Comput Sci 1999; 39: 958-66.

[10] Murthy VS, Kulkarni VM. 3D-QSAR CoMFA and CoMSIA on protein tyrosine phosphatase 1B inhibitors. Bioorg Med Chem 2002; 10: 2267-82

[11] Gohlke H, Klebe G. DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. J Med Chem 2002; 45: 4153-70.

[12] Cramer RD. Topomer CoMFA: a design methodology for rapid lead optimization. J Med Chem 2003; 46: 374-88.

[13] Copp BR, Pearce AN. Natural product growth inhibitors of *Mycobacterium tuberculosis*. Nat Prod Rep 2007; 24: 278-97.

[14] Hokuldsson A. PLS regression methods. J Chemometrics 1988; 2: 211-28.

[15] Zhaoqi Y, Pinghua S. 3D-QSAR Study of Potent Inhibitors of Phosphodiesterase-4 Using a CoMFA Approach. Int J Mol Sci 2007; 8(7): 714-22.

[16] Burnouf C, Pruniaux MP. Recent advances in PDE4 inhibitors as immunoregulators and anti-inflammatory drugs. Curr Pharm Des 2002; 8: 1255-96.

[17] Xu M, Zhang AQ, Han SK, Wang LS. Studies of 3D-quantitative structure-activity relationships on a set of nitroaromatic compounds: CoMFA, advanced CoMFA and CoMSIA. Chemsphere 2002; 48: 707-15.

[18] Woldemichael GM, Gutierrez-Lugo M-T, Franzblau SG, Wang Y, Suarez E, Timmermann BN. Mycobacterium tuberculosis growth inhibition by constituents of Sapium haematospermum. J Nat Prod 2004; 67: 598.

[19] Gu J-Q, Wang Y, Franzblau SG, Montenegro G, Yang D, Timmermann BN. Antitubercular constituents of Valeriana laxiflora. Planta Med 2004; **70(6): 509-14.**

[20] Rodr´ıguez II, Rodr´ıguez AD. Homopseudopteroxazole, a new antimycobacterial diterpene alkaloid from Pseudopterogorgia elisabethae. J Nat Prod 2003; **66(6): 855-7**.

[21] Sayed KAE, Bartyzed P, Shen X, Zjawiony JK, Hamann MT. Marrine natural products as antituberculosis agents. Tetrahedron 2000; 56: 949-53.

[22] Rukachaisirikul T, Prabpai S, Kongsaeree P, Suksamrarn A. (+)-Bornyl piperate, a new monoterpene ester from Piper aff. pedicellatum roots. Chem Pharm Bull 2004; 52(6): 760-1.

[23] Chumkaew P, Karalai C, Ponglimanont C, Chantrapromma K. Antimycobacterial activity of phorbol esters from the fruits of *Sapium indicum*. J Nat Prod 2003; 66(4): 540-3.

[24] Guchi KI, Saitoh S, Yamata Y. Novel 19-oxygenated Saitoh S, Yamata Y. Novel 19-oxygenated sterols from the okinawan soft coral *Litophytom Viridis*. Chem Pharm Bull 1989; 37: 2553-4.

[25] Boruwa B, Kalita N, Barua JC, *et al.* Synthesis, Absolute Stereochemistry and Molecular Design of the New Antifungal and Antibacterial Antibiotic Produced by *Streptomyces sp.* Med Chem Lett 2004; 14: 3571.

Received: December 03, 2010 Revised: March 07, 2011 Accepted: March 10, 2011