Received Date : 12-Nov-2013

Revised Date : 31-May-2014

Accepted Date : 04-Jun-2014

Article type : Research Article

Targeting Dormant Tuberculosis Bacilli: Results for molecules with a novel Pyrimidone Scaffold

Rohit R. Joshi[a], Avinash Barchha[a], Vijay M. Khedkar[a], Raghuvir R. S. Pissurlenkar[a], Sampa Sarkar[b], Dhiman Sarkar[b], Rohini R. Joshi[c], Ramesh A. Joshi[c], Anamik K. Shah[d] and Evans C. Coutinho[a†]

[a]Department of Pharmaceutical Chemistry, Bombay College of Pharmacy, Kalina, Santacruz (E), Mumbai 400 098.

[b]Combi Chem Bio Resource Centre, National Chemical Laboratory, Pune 411 008.

[c]Organic Chemistry Division, National Chemical Laboratory, Pune 411 008.

[d]Department of Chemistry, Saurashtra University, Rajkot 360 005.

[†]**Corresponding Author**

Tel. No.: +91-22-26670905

Fax. No.: +91-22-26670816

Email: evans@bcpindia.org

Keywords

Antitubercular activity; dormant tuberculosis bacilli; pyrimidones; recursive partitioning

Abstract

Our inability to completely control TB has been due in part to the presence of dormant mycobacteria. This also renders drug regimens ineffective and is the prime cause of the appearance of drug resistant strains. In continuation of our efforts to develop novel antitubercular agents that especially target dormant mycobacteria, a set of fifty-five new compounds belonging to the pyrimidone class were designed on the basis of CoMFA and CoMSIA studies, these were synthesized and subsequently tested against both the dormant and virulent BCG strain of *M. tuberculosis.* Some novel compounds have been identified which selectively inhibit the dormant tuberculosis bacilli with significantly low IC_{50} values. This paper reports the second molecule after TMC-207, having the ability to inhibit tuberculosis bacilli exclusively in its dormant phase. The synthesis was accomplished by a modified multi-component Biginelli reaction. A classification model was generated using the binary QSAR approach - recursive partitioning (RP) to identify structural characteristics related to the activity. Physicochemical, structural, topological, connectivity indices and E-state key descriptors were used for generation of the decision tree. The decision tree could provide insights into structure-activity relationships that will guide the design of more potent inhibitors.

Introduction

Mycobacterium tuberculosis (MTB) infects about 32% of the world's population. M. tuberculosis has two characteristic features that render it the deadliest disease: its high virulence and its ability to enter into a dormant state with subsequent reactivation. Approximately ⅓ of the world population is infected with

latent TB caused by the dormant, non-replicating form of the mycobacterium(1). Although the physiology of latent MTB is still not clear, a hypoxic condition has been found to trigger the dormant state of mycobacterium. Normally, a small population of bacilli enters into the dormant stage in the granuloma formed by immune cells and fibroblasts. Dormant bacilli are adapted to anaerobic conditions and maintain their ability to resume growth and aggravate the disease by deterioration of the immune system. This unique property also renders tolerance to conventional antimycobacterial drugs. It has been now realized that the necessity of a minimum 6 months treatment for TB is due to the difficulty in eradicating non-replicating dormant mycobacterium tuberculosis. Therefore there is a pressing need for development of new drugs which are effective against MTB in both the active and dormant states.

The first molecule reported to possess selective activity against dormant MTB was metronidazole which undergoes reduction of the nitro group to a reactive species in a hypoxic environment, that causes DNA damage and subsequent cell death(2). Nitrofurans have also shown bactericidal activity against active as well as dormant mycobacterium. However, this activity was observed at very high concentrations of metronidazole or nitrofurans, rendering limited clinical value to the compounds. More than 350 products of natural origin have been reported as anti-mycobacterial agents from 2003 onwards(3). However, most of these efforts were focused on inhibiting the growth of the actively growing mycobacterium.

The current situation clearly shows that TB is far from under control and demonstrates the need for a re-evaluation of our approach to treating TB. Drug development for tuberculosis and other neglected diseases has been at a virtual standstill for decades, but increased awareness and advocacy in recent years have led to new initiatives in TB drug development which includes a search for new drug targets and new drugs. Recognizing these facts, we have initiated a program which includes design, synthesis and screening of some new compounds that target dormant mycobacteria. Efforts to optimize

the structure and biological activity have resulted in several classes of compounds with potential antitubercular activity (4-7). We had reported the discovery of ring-substituted quinolines as a promising new structural class of anti-tuberculosis agents, which emerged from a broad structure-based screening approach of new chemical entities against various pathogens including mycobacteria (8).

Currently we are focusing our efforts on developing an entirely new class of antitubercular agents possibly acting on completely novel targets with a mechanism of action different from those of the existing drugs. This effort has proven to be highly rewarding and has resulted in the identification of a new class of tetrahydropyrimidines with potent antimycobacterial activity against both drug-sensitive and dormant strains of mycobacteria with significantly low IC_{50} values. We report here the second molecule after TMC-207, which has the ability to inhibit dormant tuberculosis bacilli.

Pyridines(4), pyrimidines and pyrimidones(5) have been identified with antitubercular activity and bear a potential to be developed as novel structural classes of antitubercular agents (Figure 1). A class of pyridines known as N-alkyl-1,2-dihydro-2-thioxo-3-pyridine carbothioamides (Figure 1, I) have shown good antitubercular activity(9). N-pyridinylsalicylamides (Figure 1, II) which are related to the pyridine class of compounds have also shown activity against MTB. The ramifications of various substituents on the activity for this class of compounds were earlier studied by 3D-QSAR methods like CoMFA and CoMSIA(5). Within the pyrimidine group, 2,4-diaminopyrimidines (Figure 1, III) have been identified with IC_{50} of 5.8 nM and a safety index > 600(10). Thus, this structural class holds great hope for development of new and effective antitubercular agents. Various 1,4-dihydropyridines (Figure 1, IV) have been synthesized and evaluated as antitubercular drugs on the assumption that these compounds would act as prodrugs, which after penetration into the cell would be converted into the 3,5-dicarboxylate anion by enzymatic hydrolysis. The most effective analog has shown more than 97% inhibition on mycobacteria at a concentration of 2.50 μg/ml. This series was examined in detail by QSAR analysis (11, 12). In order to further explore this structural class, we have modified the 1,4-dihydropyridines as

tetrahydropyrimidines (Figure 1, V),which can also be viewed as aza analogs of pyridines, with the hope of developing a new class of antitubercular agents.

This paper describes the synthesis and biological evaluation of a series of substituted pyrimidone derivatives as potential antitubercular agents especially against dormant mycobacteria. Further, to classify these pyrimidone derivatives according to their activities, a recursive partitioning (RP) model was developed. The 'Recursive Partitioning' (RP) method is a specific case of the 'binary QSAR' approach that is capable of classifying class analog activity data by considering appropriate descriptors recursively.

Methods and Materials

Chemistry

The synthesis of N-phenyl-6-methyl-2-oxo-4-phenyl-1,2,3,4-tetrahydropyrimidine-5-carboxylates (Scheme 1) was accomplished with a multi-component Biginelli reaction. Acid catalyzed one pot cyclocondensation of various acetoacetanilides with the appropriate aromatic aldehyde and urea/thiourea afforded the substituted tetrahydropyrimidines. Various acetoacetanilides were synthesized from substituted anilines and ethyl/butyl acetoacetate. All molecules listed in **Table 1** contain at least one chiral centre in the pyrimidone ring with an equal preference for formation of both stereoisomers under the conditions of synthesis and were not separated.

Melting points were determined using a BUCHI-5300 melting point apparatus and are uncorrected. IR spectra were recorded on a PERKIN ELMER spectrophotometer (nujol mull method) and only characteristic peaks are reported in cm^{-1}. ^1H NMR spectra were recorded in DMSO-d_6, with TMS as internal standard using a Bruker AC-300F NMR Spectrometer (300 MHz). Chemical shifts (δ) are reported in parts per million (ppm) of the applied field. Reactions were monitored over silica gel-G (Merck) TLC plates and spots were visualized by iodine vapor or by ultraviolet light (254 nm). Spectral

data (IR and ^{1}H-NMR) of all the synthesized compounds were found to be in agreement with the proposed structures.

General procedure for the synthesis of N-phenyl-6-methyl-2-oxo-4-phenyl-1,2,3,4-tetrahydropyrimidine-5-carboxylates

An equimolar mixture (1:1) of substituted acetoacetanilide (**3**) (1 mol) and appropriate aldehyde (**4**) (1 mol) with urea or thiourea (1.5 mol) in absolute ethanol was heated under reflux for 4-8 hours in the presence of a catalytic amount of conc. HCl /pTSA (*para*-toluene sulfonic acid). The reaction was monitored with TLC. The precipitate (**5**) was filtered, dried and crystallized from the appropriate solvent. Details about the characterization of these derivatives are provided in the supplementary material.

Antitubercular activity

Determination of growth and NR activity

Growth of *M. bovis* BCG in microplate format of dormancy model was measured by reading the absorbance at 620nm as well as by determining the CFU/ml of the culture at different time intervals. The lowest concentration of the drug yielding a differential absorbance (A620) of approximately zero or less was defined as MIC. Nitro reductase (NR) activity of *M. bovis* BCG in microplate dormancy model was measured by following the Griess reaction method described earlier (13). Briefly, 1% solution of sulphanilic acid (in 20% HCl) and 0.1% of napthyl ethylene diamine hydrochloride solution (in distilled water) was added to a whole cell culture in 1:1:1 ratio and the absorbance at 540 nm was read after 15 min of incubation.

Computational Studies

All computational studies were carried out on a Linux workstation with the CentOS Enterprise Linux 4.9 operating platform. The computations were executed with the molecular modeling software – *Sybyl* v7.1 from Tripos Inc. St. Louis, MO, USA (14) and *Cerius2* v4.11 from Accelrys Inc, USA (15).

Recursive partitioning (RP) (16-20) is a simple, yet powerful statistical technique that seeks to decode the elusive relationships in complex data sets involving thresholds, interactions and nonlinearities. These factors impede an analysis that is based on assumptions of linearity such as multiple linear regression, principal component analysis or partial least squares. RP can be used for the division of a dataset into groups of higher and lower responses according to their appropriate descriptors. Further the decision tree derived from the RP analysis can be used to qualitatively predict activities or activity classes in structure–activity relationship analysis. It uses the best chemical features to split the large data set into smaller and more homogeneous subsets. At each non-terminal node in the decision tree, molecules are split into two groups by a particular descriptor. Each of the two subgroups is split further by another descriptor or variable. To identify the best split for a particular node, the algorithm considers all possible binary splits for each descriptor and selects the optimal one by some criterion. This recursive partitioning process is terminated when further splitting is impossible or a threshold value of the termination rule is reached. The tree display of the SAR, without considering complex statistical analysis, is clear, easy to interpret and often suggestive. It is inherently fast when compared with other grouping techniques, such as clustering and the 2D nature of this approach make it especially good for large datasets that are difficult to sieve into usable divisions of classification.

The *Cerius2* molecular modeling package was used to generate the 2D descriptors **(Table 2)**. RP analysis was carried out using the CSAR recursive partitioning method and the decision tree was generated from the results of RP. The antitubercular activity of compounds was described in terms of their IC_{50} values. It is necessary to classify the activity values into two classes, active (1) and less active

(0). Descriptors with zero variance as well as descriptors containing 95% of zero values were eliminated. The correlation matrices were built for the descriptors and some descriptors were eliminated on the basis of the correlation threshold of R = 0.5. The pruned set of descriptors was used as the independent variables (X) while the biological activity formed the dependent variable (Y) for all analyses. The CARTTM (classification and regression trees) method in *Cerius2* was applied to generate the decision tree model. The activity classes were weighted equally and the splits were scored using Twoing rule scoring function. The pruning factor was varied between 0-3. Various values were used for maximum tree depth (layers < 10) while the default values were set for maximum number of generic splits (30) and the number of knots per variable (20). The internal validation for the RP model was performed using cross-validation with the number of cross-validation groups set to 5.

Results and Discussion

Biological assay/anti-tubercular activity

As stated earlier all the synthesized compounds were obtained as a racemic mixture and were used as such for anti-tubercular testing. These compounds were evaluated for their antitubercular activity in accordance with protocols developed by Sarkar et al (21). They were also tested against dormant mycobacteria by the nitrate reductase model (21). IC_{50} values were determined from the dose-response curves and are provided in **Table 1**. Compound **14** (IC_{50} = 14.3 µg/ml) was found to be most active against normally growing mycobacteria. Interestingly, compounds **13** and **14** showed good (IC_{50} = 6.3 µg/ml) activity against dormant tuberculosis bacteria by the nitrate reductase model.

A close appraisal of the molecules shows that R_1 groups which are electron withdrawing and positioned at C3 and C4 (**5-8**) decrease the activity of the compounds against aerobic bacteria (Table 1). Similarly, with reference to R_2, groups which are electron withdrawing and positioned at C3 and C4 (**9**

and **12**) also decrease the antitubercular activity. However, interestingly when the groups R_1 and R_2 are of electron withdrawing nature and simultaneously present at position 3 in their respective rings (**13** and **14**), the antimycobacterial activity increases and one of them *i.e*. compound **13** displays exclusive activity against bacteria under anaerobic conditions. One probable explanation for the differential activity could be that the intracellular environment of *M.tuberculosis* is different under aerobic and anaerobic condition. It is possible that compound **13** is metabolized faster compared to compound **14** under aerobic conditions consequently leading to early loss of activity. On the contrary compound **14** is relatively stable in this environment and is therefore able to retain the anti-tubercular activity. Such a metabolic degradation may not exist in the anaerobic environment and therefore no such variation in activity is observed.

Further, compounds with both R_1 and R_2 as electron withdrawing and with the former (R_1) stationed at positions 2 and 3 and the latter (R_2) resident at position 3 on their respective rings, exhibit activity against both aerobic as well as anaerobic bacteria. Compounds with R_2 as an aliphatic substituent for *e.g*. cyclohexyl group (**20** and **35**) also show medium activity but only when R_1 at positions 2, 3 and 4 is electron withdrawing by nature.

Recursive Partitioning (RP) Analysis

Two independent classification models were developed to classify the pyrimidone derivatives into their own activity class using the calculated descriptors (**Table 2**). For deriving the classification model, a dataset of 20 molecules was selected from the complete set of synthesized molecules on the basis of chemical and biological diversity using similarity search techniques viz. D-optimal design, Tanimoto similarity coefficient and the Euclidian distance matrix criteria defined in *Cerius2*. The selection was carried out such that the final set of molecules had structural diversity as well as a maximal distribution of biological activities.

The first classification model is based on the activity exhibited by the molecules against the dormant MTB while the second model is derived based on activity against virulent MTB. The RP models were developed by varying parameters discussed in the experimental section so as to improve the prediction for compounds correctly in their respective classes. This is defined by three terms- a percentage of the total number of compounds observed to be in each class *(Class % Observed Correct);* percent of the ratio of total number of compounds correctly classified to the number of compounds predicted in the class *(Overall % Predicted Correct)* and the *enrichment factor* which for a specific class is the ratio of the *(Overall % Predicted Correct)* to the original percentage of compounds belonging to that class.

RP model for the activity against dormant MTB

Recursive partitioning splits the dataset into smaller, homogeneous subsets by deriving a binary decision tree in which descriptor values are used as decision points. The dataset was divided into two classes, inactives (0) and actives (1) where class 1 contains 15 inactive compounds having IC_{50} values greater than 100µg/ml while the class 2 contains 5 active compounds with IC_{50} values less than 100µg/ml against dormant MTB. The statistical results of the RP model and activity predictions are summarized in **tables 3** and **4** respectively.

In class 1, all 15 'inactives' compounds were correctly classified as class 1. Also for the class 2 i.e. the class of actives, all five molecules active against dormant bacteria were correctly predicted as active, showing 100% class prediction. The number of true positives among the predictions in each activity class is listed as the term *'Overall % Predicted Correct'*. It is also noteworthy that for both classes ('actives' and 'inactives') 'Overall % Predicted Correct' was observed to be 100%. The enrichment factor (1.33 for class 1 and 4.00 for class 2 respectively) also suggest that the final RP model is statistically significant and can ably classify any new set of molecules.

A 6-leaf recursive partitioning decision tree was obtained for the pyrimidone analogs based on the 2D descriptors (**Figure 2**). The model contains 6 terminal and 5 non-terminal nodes where a true response to any given split follows the branch to the downside while a false response follow the branch to the upside in a decision tree.

The first primary split is observed on the radius of gyration which is a spatial descriptor related to the size of the molecule. The terminal node 1 contains two molecules (**16** and **39**) with radius of gyration greater than 4.36 which is shown in class 2 (active) while the remaining 18 molecules having the values above 4.36 formed the non-terminal node. These molecules were further split on the basis of molecular volume (Vm) which is a 3D spatial descriptor that defines the molecular volume inside the contact surface. It is calculated as a function of conformation and is related to the binding and transport properties of a molecule. The terminal node 2 which is the active node contains one molecule (**42**) with molecular volume greater than 339.13 and has been correctly assigned as active by the RP model.

The third split was observed on molecular surface area which is again 3D spatial descriptor that describes the van der Waals area of a molecule. It determines the extent to which a molecule exposes itself to the external environment. It is also related to the binding, solubility and transport properties of a molecule. The terminal node 3 contains ten molecules (**12, 20, 43, 44, 46, 48, 51, 54, 55** and **35)** with area greater than 394.54 which are correctly classified as in class 1 (inactive) while the remaining 7 compounds with area greater than 394.54 are placed on the non-terminal node. These were further split on the basis of density which is a 3D spatial descriptor calculated as the ratio of molecular weight to its volume. It reflects the types of atoms and how tightly they are packed in a molecule and therefore it can be related to transport and melt behavior of a molecule. The terminal node 6 which is the inactive node contains four molecules (**7, 22, 26** and **34**) with density less than 1.23 and have been correctly classified as class 1 i.e. inactives. Remaining 3 molecules in the dataset formed the non-terminal node which were

then split on the basis of Principal moment of inertia at a cut off value of 1383.28. Terminal node 4 which is the active node contains two correctly classified active molecules (**13** and **14**) with PMI greater than 1383.28 while the terminal node 5 was assigned molecule **49** with PMI less than the cut-off value and has been correctly classified as inactive by the classification model.

To avoid over fitting and to improve generalization of the classification model, this model was internally validated using the technique of 5-fold cross-validation. A set of 5 molecules are extracted each time from the whole dataset and a new model is generated using the reduced dataset which is then employed to predict the activity of the excluded molecules. The procedure is repeated iteratively until all molecules have been omitted and predicted once, so that the statistics can be derived from a comparison of the predicted data with the observed data. Approximately four-fifths of the compounds in the dataset formed the training set that was used to derive the models and the remaining compounds shaped the test set. The statistical results of cross-validation are summarized in **table 5**. An acceptable percent classification was achieved for the training set and the predictivity of the test set was also good. This reliability of the predictions for the molecules in the test set suggests that the model can be used with confidence for predictions of unknown but related molecules. The key descriptors determined in the trial set were consistent with those from the total set. The parameters used to build the RP model suggest that the model is well able to classify and predict the activity class of new candidates.

RP model for the activity against virulent MTB

The same dataset of 20 molecules identified in the previous study was again used to build the classification model to identify the features governing activity against virulent tuberculosis bacteria. The dataset was divided into two classes, inactives (0) and actives (1). Class 1 contains 8 inactive compounds with the IC_{50} values greater than 100μg/ml while class 2 contains 12 active compounds with IC_{50} values

less than 100μg/ml against the virulent tuberculosis bacteria. The statistical results of the RP model and activity predictions from it are summarized in **tables 6** and **7** respectively.

Analysis of the decision tree shows that in class 1, i.e. the class of 'inactives', 8 out of 8 compounds are correctly classified as belonging to class 1, while in class 2 which is the class of 'actives', 12 out of 12 compounds are correctly classified as well. It is noteworthy that for both classes ('actives' and 'inactives') the *'Overall % Predicted Correct'* is observed to be 100%. The enrichment factor (2.50 for class 1 and 1.67 for class 2 respectively) suggest that the final RP model is statistically significant and can be confidently used for classifying a new library of compounds.

Figure 3 displays the optimized 6-leaf recursive partitioning decision tree generated using the 2D descriptors. The decision tree consists of six terminal and five non-terminal nodes. The class 1 i.e. the class of inactives has been plotted in red while Class 2 which is the class of actives has been plotted with green color in the decision tree. The terminal nodes 2, 4 and 5 are associated with class 2 while the terminal nodes 1, 3 and 6 are linked to class 1. The descriptors- dipole moment, principle moment of inertia (PMI), AlogP98, radius of gyration and sum of atomic polarizability (Apol) form the decision points in the decision tree **(Table 7)**.

The first primary split is observed on the dipole moment which is an electronic descriptor reflecting the strength and orientation behavior of a molecule in an electrostatic field. Dipole moments have been found to be correlated to long range ligand–receptor recognition and receptor binding. The dataset of 20 molecules was split into two branches- those molecules with dipole moment less than 4.42 follows a branch to the downside while those with dipole moment above the cut-off follows a branch to the upside. Ten molecules (**48, 26, 22, 14, 16, 20, 46, 54, 34** and **35**) with dipole moment greater than 4.42 are further split on the basis of AlogP98 at a cut off mark of 4.05. LogP reflects the hydrophobic character of a molecule. In this atom-based approach, each atom of the molecule is assigned to a

particular class with additive contributions to the total value of logP. Terminal node 1 contains one molecule (**46**) with AlogP98 greater than 4.05 which is correctly classified as inactive by the RP model while the terminal node 2 which is the active node contains nine correctly classified molecules (**14, 16, 20, 22, 26, 34, 35, 48** and **54**) having logP less than 4.05.

Switching to the first primary split, ten molecules (**51, 49, 13, 43, 44, 42, 12, 39, 7** and **55**) with dipole moment values less than 4.42 follows a branch downside. These molecules were further split based on principal moment of inertia which is a spatial descriptor. The terminal node 6 contains 5 molecules with PMI less than 1519.07 (**49, 13, 44, 12 and 7**) correctly classified as inactives while the remaining five molecules (**51, 43, 42, 39** and **55**) with PMI above the cut off form the non-terminal node. These molecules are again split on the basis of radius of gyration, another spatial descriptor related to the size of the molecule. Molecules with radius of gyration less than 4.14 form the terminal node 5 (**42** and **55**) which are shown in class 2 while those having the value above 4.14 form the non-terminal node. These molecules are then split on the basis of sum of atomic polarizability (Apol) with a cut off value of 15390.03. The terminal node 3 contains two molecules (**43** and **51**) with Apol greater than 15390.03 correctly classified as inactives while molecule **39** with Apol less than 15390.03 forms the terminal node 4 which is the active node.

A 5-fold cross-validation was performed to gauge the stability and the statistical significance of the RP model. The statistical results for the same are summarized in **Table 8**. An acceptable classification percentage observed for the classes signifies the stability of this model for external predictability. Also the key descriptors identified in trial set are consistent with those observed in the whole set. The statistical parameters indicate that this RP model is valid for classifying and predicting the activity class of new candidate molecules that target tubercular bacteria.

A close appraisal of both the classification models reveal that features contributing towards antitubercular activity against dormant MTB are radius of gyration molecular volume, molecular surface area, density and principal moment of inertia while the activity against virulent MTB is shown to be governed by dipole moment, principle moment of inertia, AlogP98, radius of gyration and sum of atomic polarizability. This difference in contributing factors definitely provides some guidelines to channelize the activity against dormant or virulent strains of mycobacterium.

Conclusions

The primary goal of this endeavor was to explore pyrimidones as a novel structural class targeting dormant mycobacteria. The molecules were designed based on 3D-QSAR results and were synthesized by a modification of the Biginelli reaction. The compounds display promising antitubercular activity especially on dormant TB bacilli. Interestingly, some of the compounds show activity against both active as well as dormant bacilli while some compounds show activity only against dormant bacilli e.g. compound **13**. This is only the second compound reported so far to show activity specifically against dormant TB bacilli. Thus the hypothesis we initially proposed has been validated and the extensive QSAR analysis on this nucleus will surely lead to a more potent candidate for targeting dormant bacilli. RP helped to develop a discriminative model for analyzing the structure-activity relationships for these pyrimidone derivatives. The classification using 2D descriptors has considerable discriminative power in spite of high degree of structural similarity in the library. Also the two independent RP models derived for these molecules could identify features essential for activity against virulent and dormant tubercular bacteria. Cytotoxicity studies and elucidation of the mechanism of action of these molecules and determination of the site of action is a priority in our future efforts. In conclusion, considering the clinical importance of dormant tubercular bacilli, development of molecules active against dormant TB bacilli

might decrease the amount of time required to cure TB to almost half of the 6-24 months required currently and could improve therapy against TB thus limiting its spread across the world.

Acknowledgements

The computational facilities were jointly provided by the All India Council of Technical Education through grant (F. No. 8022/RID/NPROJ/RPS-5/2003-04), the Department of Science and Technology through their FIST program (SR/FST/LSI-163/2003), Department of Biotechnology (BT/PR11810/BRB/10/690/2009) and the Council of Scientific and Industrial Research (01/23/99/10-EMR-II).

REFERENCES

1. Wayne LG (1994) Dormancy of Mycobacterium tuberculosis and latency of disease. *Eur J Clin Microbiol Infect Dis;***13**: 908-14.

2. Papadopoulou MV, Bloomer WD, McNeil MR (2007) NLCQ-1 and NLCQ-2, two new agents with activity against dormant Mycobacterium tuberculosis. *Int J Antimicrob Agents;***29**: 724-7.

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

4. Manvar AT, Pissurlenkar RR, Virsodia VR, Upadhyay KD, Manvar DR, Mishra AK, et al. (2009) Synthesis, in vitro antitubercular activity and 3D-QSAR study of 1,4-dihydropyridines. *Mol Divers;*285-305.

5. Virsodia V, Pissurlenkar RR, Manvar D, Dholakia C, Adlakha P, Shah A, et al. (2008) Synthesis, screening for antitubercular activity and 3D-QSAR studies of substituted N-phenyl-6-methyl-2-oxo-4-phenyl-1,2,3,4-tetrahydro-pyrimidine-5-carboxami des. *Eur J Med Chem;***43**: 2103-15.

6. Virsodia V, Manvar A, Upadhyay K, Loriya R, Karia D, Jaggi M, et al. (2009) Synthesis of 1-(2,6-dichlorophenyl)-3-methylene-1,3-dihydro-indol-2-one derivatives and in vitro anticancer evaluation against SW620 colon cancer cell line. *Eur J Med Chem;***44**: 1355-62.

7. Manvar A, Malde A, Verma J, Virsodia V, Mishra A, Upadhyay K, et al. (2008) Synthesis, anti-tubercular activity and 3D-QSAR study of coumarin-4-acetic acid benzylidene hydrazides. *Eur J Med Chem;***43**: 2395-403.

8. Nayyar A, Malde A, Jain R, Coutinho E (2006) 3D-QSAR study of ring-substituted quinoline class of anti-tuberculosis agents. *Bioorg Med Chem;***14**: 847-56.

9. Ubiali D, Pagani G, Pregnolato M, Piersimoni C, Pedraz Munoz JL, Rodriguez Gascon A, et al. (2002) New N-alkyl-1,2-dihydro-2-thioxo-3-pyridinecarbothioamides as antituberculous agents with improved pharmacokinetics. *Bioorg Med Chem Lett;***12**: 2541-4.

10. Rosowsky A, Forsch RA, Queener SF (2002) Inhibition of Pneumocystis carinii, Toxoplasma gondii, and Mycobacterium avium dihydrofolate reductases by 2,4-diamino-5-[2-methoxy-5-(omega-carboxyalkyloxy)benzyl]pyrimidines: marked improvement in potency relative to trimethoprim and species selectivity relative to piritrexim. *J Med Chem;***45**: 233-41.

11. Desai B, Sureja D, Naliapara Y, Shah A, Saxena AK (2001) Synthesis and QSAR studies of 4-substituted phenyl-2,6-dimethyl-3, 5-bis-N-(substituted phenyl)carbamoyl-1,4-dihydropyridines as potential antitubercular agents. *Bioorg Med Chem;***9**: 1993-8.

12. Kharkar PS, Desai B, Gaveria H, Varu B, Loriya R, Naliapara Y, et al. (2002) Three-dimensional quantitative structure-activity relationship of 1,4-dihydropyridines as antitubercular agents. *J Med Chem;***45**: 4858-67.

13. Nicholas DJD, Nason A (1957) Determination of nitrate and nitrite. *Methods Enzymol;***3**: 981-4.

14. Sybyl. (2005) USA: Tripos Associates Inc.

15. Accelrys (1998) Cerius2. San Diego, CA, USA.

16. Strobl C, Malley J, Tutz G (2009) An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods;***14**: 323-48.

17. Choi SY, Shin JH, Ryu CK, Nam KY, No KT, Park Choo HY (2006) The development of 3D-QSAR study and recursive partitioning of heterocyclic quinone derivatives with antifungal activity. *Bioorg Med Chem;***14**: 1608-17.

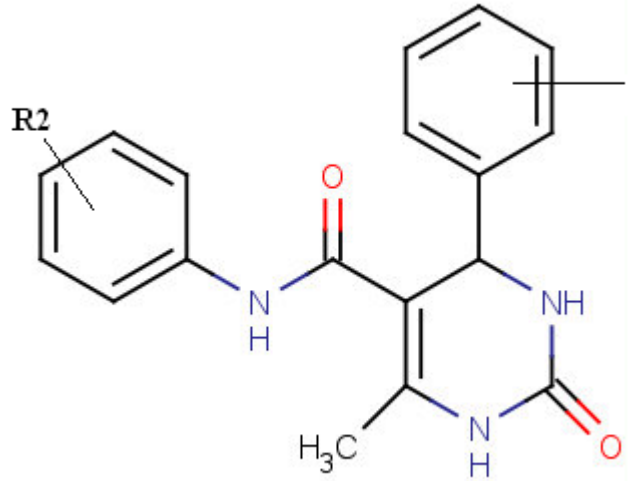
18. Young SS, Hawkins DM (2004) Using recursive partitioning analysis to evaluate compound selection methods. *Methods Mol Biol;***275**: 317-34.

19. Rusinko A, 3rd, Young SS, Drewry DH, Gerritz SW (2002) Optimization of focused chemical libraries using recursive partitioning. *Comb Chem High Throughput Screen;***5**: 125-33.

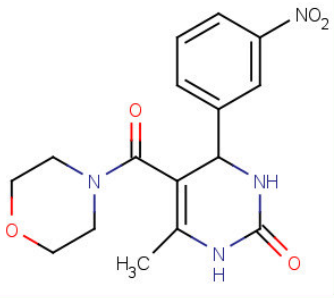
20. Blower P, Fligner M, Verducci J, Bjoraker J (2002) On combining recursive partitioning and simulated annealing to detect groups of biologically active compounds. *J Chem Inf Comput Sci;***42**: 393-404.

21. Khan A, Sarkar D (2008) A simple whole cell based high throughput screening protocol using Mycobacterium bovis BCG for inhibitors against dormant and active tubercle bacilli. *J Microbiol Methods;***73**: 62-8.

Table 1. Structures and activities of the molecules synthesized and tested against aerobic and dormant (NR) strains of *M. bovis*.



Sr. No.	R_1	R_2	Activity (IC_{50} µg/ml)
5	H	H	> 100 µg/ml
6	H	3-Cl	> 100 µg/ml
7	H	4-NO_2	>100 µg/ml
8	H	2,3-diCl	> 100 µg/ml
9	2,4-diCl	H	> 100 µg/ml
10	2,4-diCl	3-Cl	>100 µg/ml
11	2,4-diCl	2,3-diCl	> 100 µg/ml
12	2,4-diCl	4-NO_2	>100 µg/ml
13	3-Cl	3-Cl	> 100 µg/ml (aerobic stage) 6.3 µg/ml (N.R. activity)

14	3-Cl	2,3-diCl	14.3 µg/ml (aerobic stage) 6.3 µg/ml (N.R. activity)
15	3-Cl	3-NO_2	> 100 µg/ml
16	3-Cl	4-NO_2	54.95 µg/ml (aerobic stage) 6.67 µg/ml (N.R. activity)
17	4-Cl	H	> 100 µg/ml
18	2,3-diOCH_3	3-Cl	> 100µg/ml
19	H	cyclohexyl	> 100 µg/ml
20	2,4-diCl	cyclohexyl	60.37 µg/ml (aerobic stage) > 100 µg/ml (N.R. activity)
21	3-NO_2	H	> 100 µg/ml
22	3-Cl	2-Cl	10.29 µg/ml (aerobic stage)
23	3-NO_2	2,3-diCl	> 100 µg/ml
24	3-Cl	4-OCH_3	> 100 µg/ml
25	3-NO_2	4-OCH_3	> 100 µg/ml
26	3-NO_2	3-Cl	10.52 µg/ml (aerobic stage)
27	3-Cl	2-CH_3	> 100 µg/ml
28	3-NO_2	morpholinyl	> 100 µg/ml
29	3-Cl	morpholinyl	> 100 µg/ml
30			> 100 µg/ml

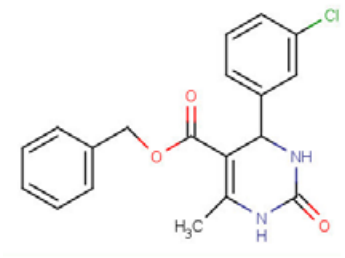
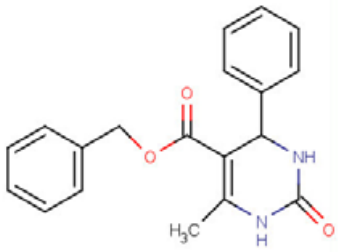
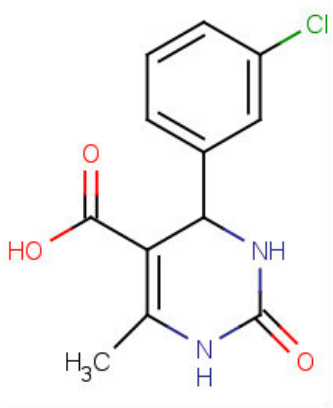
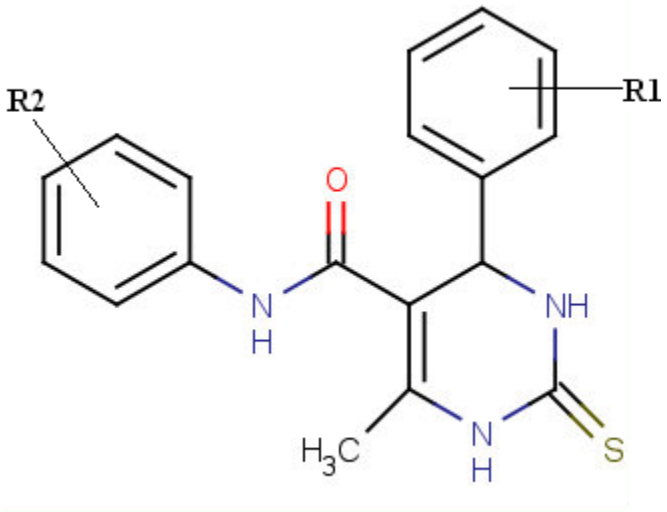
31			> 100 µg/ml
32			> 100 µg/ml
33			> 100 µg/ml
34	3-Cl	2-pyridinyl	12.5 µg/ml (aerobic stage)
35	3-Cl	cyclohexyl	27.17 µg/ml (aerobic stage)

Table 1. (contd.) Structures and activities of the molecules synthesized and tested against aerobic and dormant (NR) strains of *M. bovis*.



Sr. No.	R_1	R_2	Activity (IC_{50} µg/ml)
36	H	3-Cl	> 100
37	H	2,3-diCl	> 100
38	H	3-NO_2	>100
39	H	4-NO_2	80.69 (aerobic stage) 7.41 (N.R. activity)
40	2,4-diCl	3-Cl	>100
41	2,4-diCl	2,3-diCl	> 100
42	2,4-diCl	4-NO_2	25.1 (aerobic stage) 6.97 (N.R. activity)
43	3-Cl	3-Cl	> 100
44	3-Cl	2,3-diCl	> 100

45	H	cyclohexyl	> 100
46	2,4-diCl	cyclohexyl	>100
47	3-Cl	H	> 100
48	3-NO_2	2-Cl	18.96 (aerobic stage)
49	3-Cl	2-Cl	> 100
50	3-NO_2	2-CH_3	> 100
51	3-NO_2	2,3-diCl	> 100
52	3-Cl	4-OCH_3	> 100
53	3-NO_2	4-OCH_3	> 100
54	3-NO_2	3-Cl-4-F	3.45 (aerobic stage)
55	3-Cl	3-Cl-4-F	4.13 (aerobic stage)
56	3-Cl	2-CH_3	> 100
57	3-Cl	*morpholinyl*	> 100
58	3-Cl	*morpholinyl*	> 100
59	3-NO_2	*morpholinyl*	> 100
Rifampicin			0.04 (aerobic stage) 0.26 (N.R activity)
Isoniazid			0.04 (aerobic stage) 0.25 (N.R activity)
Streptomycin			0.17 (aerobic stage) 0.50 (N.R activity)
Ethambutol			0.16 (aerobic stage) 0.3 (N.R activity)
Pyrazinamide			0.5 (aerobic stage) 0.30 (N.R activity)

Table 2. Descriptors used to develop the classification model.

X-Descriptors	Description
Structural descriptor	the number of H-bond donor and acceptor
Kappa indices	the shape of molecule
E_state_keys	the electrotopological interaction for each atom
Electronic descriptors	Charge, Apol , HOMO, LUMO, Dipole moment.
Topological descriptors	Wiener index (*W*), Zagreb index (Zagreb), Hosoya index (*Z*), Kier & Hall molecular connectivity index (χ)
Spatial descriptors	Rad of Gyration, Jurs descriptors, Molecular volume, Area , Density, Principal moment of inertia

Table 3. Statistical results of recursive partitioning for the model built for activity against dormant mycobacteria.

Class	No. of molecules	Activity	% of molecules	Class % as Obs Correct	Overall % as Pred Correct	Enrichment Ratio
1	15	0	75	100	100	1.33
2	5	1	25	100	100	4.00

Table 4. Activity prediction in terms of binary codes by recursive partitioning for activity against dormant mycobacteria.

Molecule ID	Biological activity in binary code	RP Predicted activity in binary code	Radius of Gyration	Molecular Volume (Vm)	Molecular Surface Area	Density	Principal moment of inertia

7	0	0	4.35	302.08	369.24	1.17	1459.97
12	0	0	4.06	329.24	413.14	1.28	1459.26
13	1	1	4.13	306.26	381.56	1.23	1495.92
14	1	1	4.13	319.37	393.77	1.29	1683.92
16	1	1	4.38	315.66	395.32	1.23	1709.37
20	0	0	4.04	324.25	419.31	1.18	1161.28
22	0	0	4.11	306.33	374.71	1.23	1276.94
26	0	0	4.21	315.32	393.55	1.23	1542.22
34	0	0	4.07	287.13	353.20	1.19	1083.97
35	0	0	4.08	310.00	407.93	1.12	1078.15
39	1	1	4.36	311.69	383.10	1.18	1584.96
42	1	1	4.08	339.34	429.73	1.29	1610.45
43	0	0	4.14	316.09	400.06	1.24	1594.63
44	0	0	3.88	329.76	417.22	1.29	1396.60
46	0	0	4.06	333.87	432.70	1.19	1264.83
48	0	0	4.20	325.00	407.36	1.24	1419.66
49	0	0	4.13	316.05	390.70	1.24	1369.96
51	0	0	4.23	338.91	420.44	1.29	1815.23
54	0	0	4.26	329.68	411.73	1.28	1882.99
55	0	0	4.14	320.67	404.45	1.28	1781.55

Table 5. Statistical results of recursive partitioning (cross validation)(dormant)

Class	No. of molecules	Activity	% of molecules	Class % as Obs Correct	Overall % as Pred Correct	Enrichment Ratio
1	15	0	75	66.67	71.43	1.95
2	5	1	25	50.00	66.67	1.97

Table 6. Statistical results of recursive partitioning for the model built for activity against virulent mycobacteria

Class	No. of molecules	Activity	% of molecules	Class % as Obs Correct	Overall % as Pred Correct	Enrichment Ratio
1	8	0	40	100	100	2.50
2	12	1	60	100	100	1.67

Table 7. Activity prediction in terms of binary codes by recursive partitioning for the model built using activity against virulent mycobacteria.

Molecule ID	Biological activity in binary code	RP Predicted activity in binary code	Apol	Dipole-mag	Radius of Gyration	PMI	AlogP98
7	0	0	13961.24	3.25	4.35	1459.97	1.89
12	0	0	17040.80	2.89	4.06	1459.26	3.22
13	0	0	15965.72	3.46	4.13	1495.92	3.32
14	1	1	17505.50	4.43	4.13	1683.92	3.99

16	1	1	15501.02	5.10	4.38	1709.37	2.55
20	1	1	15262.60	6.69	4.04	1161.28	3.60
22	1	1	15965.72	4.57	4.11	1276.94	3.32
26	1	1	15501.02	5.30	4.21	1542.22	2.55
34	1	1	14166.42	4.87	4.07	1083.97	2.05
35	1	1	13722.82	6.01	4.08	1078.15	2.94
39	1	1	15279.04	3.42	4.36	1584.96	2.79
42	1	1	18358.60	3.06	4.08	1610.45	4.12
43	0	0	17283.52	3.40	4.14	1594.63	4.22
44	0	0	18823.30	4.25	3.88	1396.60	4.89
46	0	0	16580.40	6.58	4.06	1264.83	4.50
48	1	1	16818.82	5.43	4.20	1419.66	3.45
49	0	0	17283.52	4.41	4.13	1369.96	4.22
51	0	0	18358.60	4.25	4.23	1815.23	4.12
54	1	1	16931.20	5.80	4.26	1882.99	3.66
55	1	1	17395.90	3.45	4.14	1781.55	4.43

Table 8. Statistical results of recursive partitioning (cross validation) (virulent)

Class	**No. of molecules**	**Activity**	**% of molecules**	**Class % as Obs Correct**	**Overall % as Pred Correct**	**Enrichment Ratio**
1	8	0	40	50.00	54.44	1.11
2	12	1	60	58.33	63.64	1.06

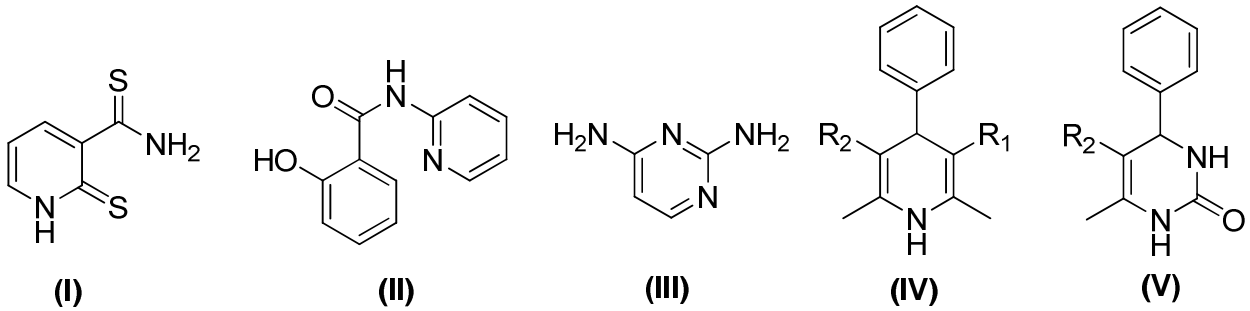


Figure 1. Chemical structures or cores known to have antitubercular activity

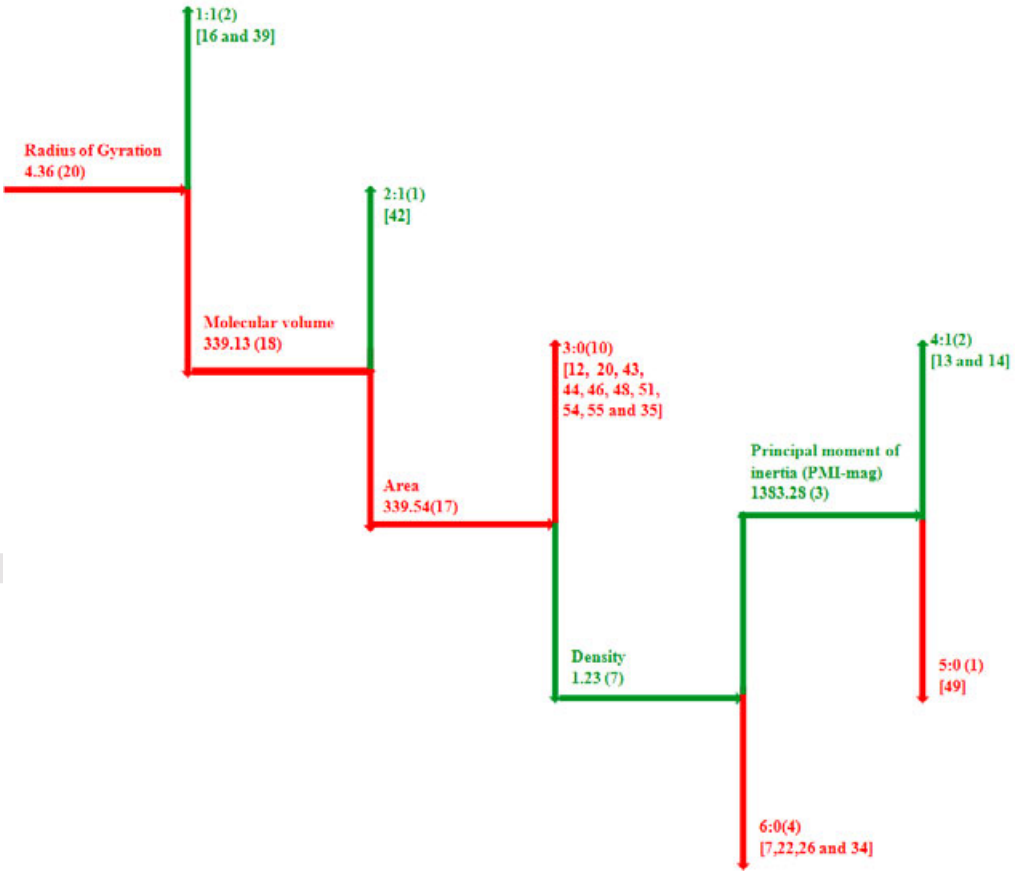


Figure 2 Recursive Partitioning tree generated with moderate pruning; score splits using Twoing Rule for prediction of antitubercular activity (dormant) classes. Those marked 1:1, 2:1, 3:0, 4:1, 5:0 and 6:0 correspond to terminal nodes 1-6 and each terminal node corresponds to the value of 1 (active) or 0 (inactive).

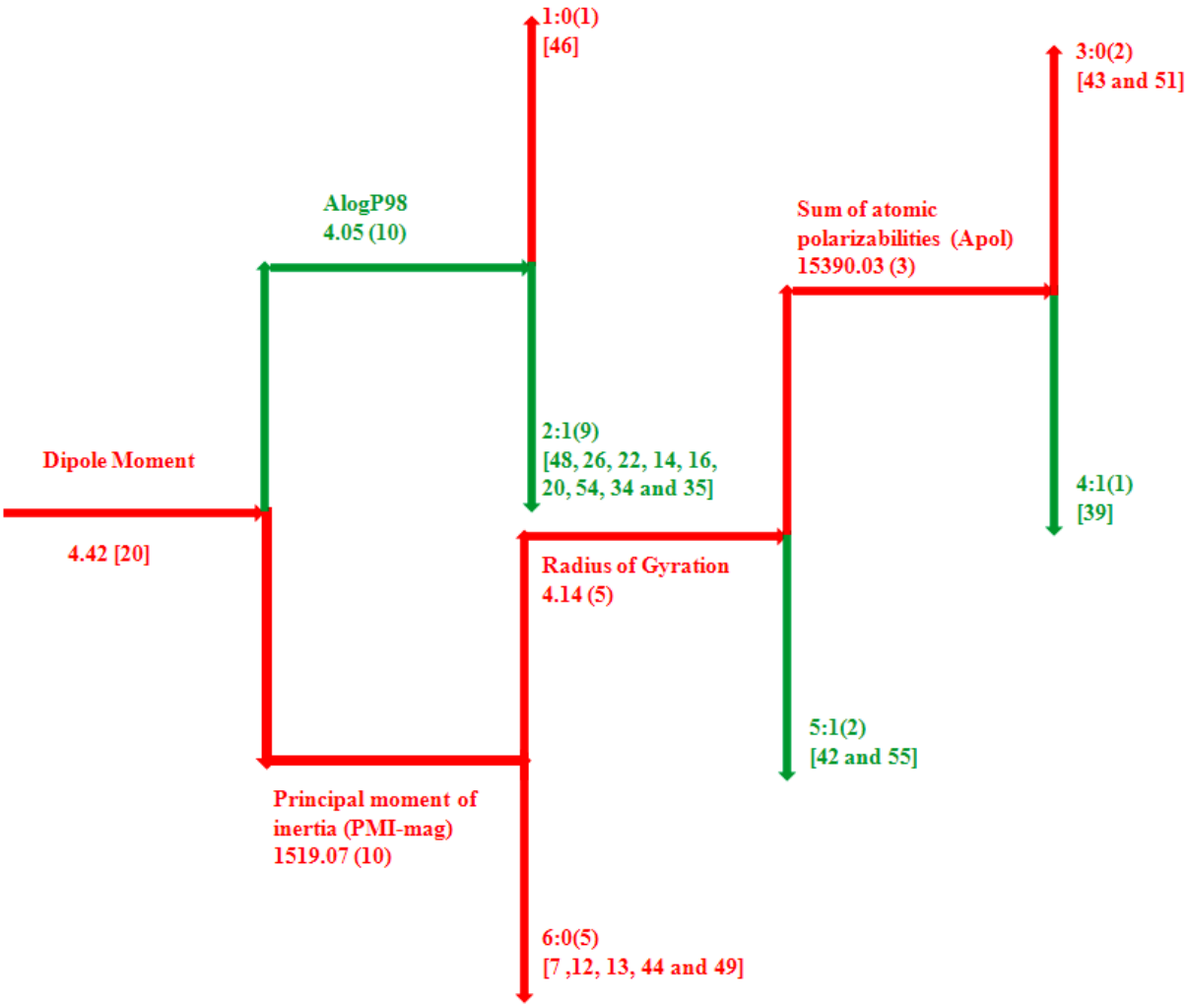


Figure 3 Recursive Partitioning tree generated with moderate pruning; score splits using Twoing Rule for prediction of antitubercular activity (virulent) classes. Those marked 1:0, 2:1, 3:0, 4:1, 5:1 and 6:0 correspond to terminal nodes 1-6 and each terminal node corresponds to the value of 1 (active) or 0 (inactive).

Scheme 1. The modified Biginelli reaction for synthesis of substituted tetrahydropyrimidone derivatives. (i) toluene; (ii) urea or thiourea in presence of conc. HCl / pTSA, absolute ethanol, 4-8 hours reflux.