In Silico Pharmacophore Modeling and Virtual Screening to Design Potential p38 MAP Kinase Inhibitors as New Leads

Neelakantan Suresh¹

¹GVK Biosciences Pvt. Ltd., Biocampus, 37, Sterling Road, Nungambakkam, Chennai 600034, Tamil Nadu, India.

ABSTRACT: p38 MAP kinase is one of the important targets in the treatment of osteoarthritis and inflammation. The best pharmacophore hypothesis (Hypo 1), consisting of four features, namely, one hydrogen-bond acceptor (HBA), one hydrophobic point (HY), and two ring aromatics (RA), showing a correlation coefficient of 0.954, a root mean square deviation (RMSD) of 0.898, and a cost difference of 61.786, suggesting that a highly predictive pharmacophore model was successfully obtained. A chemical feature based pharmacophore model has been generated from known p38 MAP kinase inhibitors (25 training set compounds) by HypoGen module implemented in CATALYST software. The top ranked hypothesis (Hypo1) contained four chemical feature types such as hydrogen-bond acceptor (HA), hydrophobic aromatic (HY), and two ring aromatic (RA) features. Hypo1 was further validated by 129 test set molecules giving a correlation coefficient of 0.923 between experimental and estimated activity. Thus, the Hypo1 was exploited for searching new lead compounds over chemical compounds in Medicem database and then the selected compounds were screened based on restriction estimated activity. Finally, we obtained 30 new lead candidates and the one best highly active compound structure was selected as a lead compound. The best highly active lead compound was docked into the active site of Human p38 MAP Kinase Inhibitor Complex 1IAN active site using Ligand fit of Cerius 2. The results demonstrate that hypothesis derived in this study could be considered to be a useful and reliable tool in identifying structurally diverse compounds with desired biological activity.

KEYWORDS: p38 MAP kinase; osteoarthritis; inflammation; catalyst; pharmacophore; ligand fit.

Introduction

p38 MAP kinase is a key regulator in stress, inflammation, development, and cell death. Osteoarthritis (OA) is a common rheumatic disease that is characterized by a progressive loss of articular cartilage.

Cartilage degeneration results from an imbalance between anabolic and catabolic processes due to the dedifferentiation and apoptosis of chondrocytes and increased synthesis of matrix degrading proteinases.

There is increasing evidence that inflammation plays an active role in pathophysiology of osteoarthritis. Proinflammatory cytokines are secreted from the inflamed synovium and from activated chondrocytes. Cytokines such as interleukin 1 beta (IL-1β) and tumor necrosis factor alpha (TNFa) upregulate numerous cytokines from chondrocytes and synoviocytes as well as prostaglandin E2 and proteinases such as the matrix metalloproteinases (MMPs) and aggrecanases. The aggrecanases and the matrix metalloproteinases are thought to mediate the structural degradation of cartilage in OA¹.
molecules of p38 MAP kinase with an aim to obtain Pharmacophore model that could provide a rational hypothetical picture of the primary chemical features responsible for activity and also docked the best highly active lead compound into the active site of Human p38 MAP Kinase Inhibitor Complex 1I1N obtained from Protein Data Bank receptor’s active site using Ligand fit and a best dock score was obtained for the active lead compound.

This study is expected to provide useful knowledge for developing new potentially active candidates targeting the p38 MAP kinase which can be useful for treatment of osteoarthritis. Pharmacophore modeling correlates activities with the spatial arrangement of various chemical features.

Materials and Methods

Selection of Molecules

We have selected a set of 154 compounds which are reported to be the inhibitors of p38 MAP kinase. The inhibitory activity of these compounds was expressed as IC_{50} (i.e., concentration of compound required to inhibit 50% of p38 MAP kinase was taken). The IC_{50} values span across a wide range from 0.002 μM to 1850 μM. Of these 154 compounds, 25 compounds were taken as training set and the rest of the 129 compounds as test set. The datasets were divided into the training set and test set. The training set was selected by considering both structural diversity and wide coverage of the activity range. These compounds were further distributed into most active, moderately active and least inactive compounds based on their IC_{50} values thereby to obtain critical information on pharmacophore requirements. The most important aspect of this selection was to ensure that each active compound would teach something new to the HypoGen module thus it can be able to uncover as much as critical information possible for predicting biological activity.

Molecular Modeling

The geometry of a compound is built with the Catalyst builder and optimized by the CHARMM like force field. All molecules were built using the builder module of Cerius2. All the structures were minimized using steepest descent algorithm with a convergence gradient value of 0.001 kcal/mol. Partial atomic charges were calculated using Gasteiger method. Further geometry optimization was carried out for each compound with the MOPAC 6 package using the semi-empirical AM1 Hamiltonian.

Pharmacophore Modeling

Multiple acceptable conformations were generated for all of 20.0 kcal/mol above the global energy minimum. Multiple acceptable conformations were generated for all ligands within the Catalyst ConFirm module using the ‘‘Poling’’ algorithm, instead of using lowest energy conformation of each compound. Maximum of 250 conformations were generated for each molecule within an energy threshold and all conformational models for each molecule in training set were used in Catalyst for pharmacophore hypothesis generation.

The training set molecules of 25 associated with their conformations were submitted to the Catalyst hypothesis generation (HypoGen). The four common chemical features that were identified and included were the features of hydrogen-bond acceptor (HBA), hydrophobic features (HY), hydrogen-bond donor (HBD) and ring aromatic (RA) features which were included for the pharmacophore generation on the basis of common features present in the study molecules. The statistical parameters like cost values determine the significance of the model. Ten pharmacophore models with significant statistical parameters were generated. The best model was selected on the basis of a high correlation coefficient (r), lowest total cost, and RMSD values. The final model was further validated by a test set of 129 molecules.

Generation of Pharmacophore Model

From the structures of the training set compounds and their experimentally determined inhibitory activities against p38 MAP kinase, 10 best pharmacophore (or hypotheses) were generated using HypoGen module implemented in Catalyst 4.11 software. On analysis, it was observed that four chemical feature types such as hydrogen-bond acceptor (HA), hydrophobic aromatic (HY), and two ring aromatic (RA) features could effectively map all critical chemical features of all molecules in the training and test sets. These features were further selected and used to build a series of hypotheses using the HypoGen module in Catalyst using default uncertainty value 3 (defined by Catalyst software as the measured value being within three times higher or three times lower of the true value). Catalyst thereby generates a chemical feature based model on the basis of the most active compounds.

The structure and activity correlations in the training set were examined for hypothesis generation. HypoGen identifies those features which will be common to the active compounds but excluded from the inactive compounds within conformationally allowable regions of space. This also further estimates the activity of each training set compound using the regression parameters. The greater is the geometric fit, greater would be the activity prediction of the compound. The fit function not only checks whether the feature is mapped or not but also checks whether it contains a distance term, which measures the distance that separates the feature on the molecule from the centroid of the hypothesis feature. Both these terms are used to calculate the geometric fit value. 


Pharmacophore Validation

The generated pharmacophore model should also be able to also predict the activity of the molecules accurately and also identify the active compound from the database. Therefore, the derived pharmacophore map was validated using (i) cost analysis, (ii) test set prediction.

Cost Analysis

The HypoGen module in Catalyst performs two important theoretical cost calculations determining the success of any pharmacophore hypothesis. One is the ‘fixed cost’ (termed as ideal cost), representing the simplest model that fits all data perfectly, and the second is the ‘null cost’ (termed as no correlation cost), representing the highest cost of a pharmacophore with no features and estimates activity to be the average of the activity data of the training set molecules.

A meaningful pharmacophore hypothesis may also result when the difference between null and fixed cost value is large; with values of 40-60 bits for a pharmacophore hypothesis may indicate that it has 75-90% probability of correlating the data (Catalyst 4.11 documentation).

Two other parameters determine the quality of any pharmacophore configuration cost or entropy cost depending on the complexity of the pharmacophore hypothesis space and should have a value <17, and the error cost, which is dependent on the root mean square differences between the estimated and the actual activities of the training set molecules. The RMSD represents the quality of the correlation between the estimated and the actual activity data. The best pharmacophore model has highest cost difference, lowest RMSD and best correlation coefficient.

Test Set Activity Prediction

In addition to the estimation of activity of the 25 training set molecules, the pharmacophore model should also be able to estimate the activity of new compounds. For external validation of the pharmacophore model, we have considered 129 compounds as test set, having wide range of activities (IC_{50} spanning from 0.002 uM to 1850 uM) and structural diversity. The best pharmacophore (Hypo1)
having high correlation coefficient (r), lowest total cost, and lower RMSD value was chosen to estimate the activity of test set. Test set compounds were classified on the basis of their activity as highly active ++++, < 1.5 μM (highly active); ++, 10-570 μM (moderately active); +, < 1850 μM (inactive).3-34

Results and Discussion
Pharmacophore models were generated by HypoGen present in Catalyst 4.113 and top 10 hypotheses were exported. Most hypotheses showed high correlation (>0.92). Interestingly, in the training set, all highly active compounds map all the features that is hydrophobic (HY), hydrogen-bond acceptor (HBA), and two ring aromatics (RA1 and RA2). With a few exceptions, in moderately active and inactive compounds one feature will be missing. All the compounds in the training set map HY and RA1 feature revealing that these two features should be mainly responsible for the high molecular bioactivity, thus, should be taken into account in discovering or designing novel p38 MAP kinase. The most active compound, 51, has a fitness score of 12.95 when mapped to Hypo 1 (Fig. 1) whereas the least active, 22, maps to a value of 8.76 as seen in Fig. 2B(1). On the basis of similar composition of the 10 hypotheses, hypothesis 1 (Hypo1), characterized by the best statistical parameters in terms of its predictive ability, as indicated by the highest correlation coefficient and lowest RMS deviations, has been chosen to represent ‘the pharmacophore model’. Remarkably, the highest active compound (compound 51) can be nicely mapped onto the Hypo1 model by the best fit values, which are shown in Fig. 2A(1) indicating that the Hypo1 model provides reasonable pharmacophoric characteristics of the p38 MAP kinase inhibitors for component of their activities.36

Cost Analysis
In addition to generating a hypothesis, Catalyst also provides two theoretical costs (represented in bit units) to help assess the validity of the hypothesis. The first is the cost of an ideal hypothesis (fixed cost), which represents the simplest model that fits all data perfectly. The second is the cost of the null hypothesis (null cost), which represents the highest cost of a pharmacophore with no features and which estimates activity to be the average of the activity data of the training set molecules. They represent the upper and lower bounds for the hypotheses that are generated. A generated hypothesis with a score that is substantially below that of the null hypothesis is likely to be statistically significant and bears visual inspection.

The greater the difference between the cost of the generated hypothesis and the cost of the null hypothesis, the less likely it is that the hypothesis reflects a chance correlation. A value of 40-60 bits between them for a pharmacophore hypothesis may indicate that it has 75-90% probability of correlating the data. The total fixed cost of the run is 123478, the cost of the null hypothesis 308.345, and the total cost of the Hypo1 is 140.975.

Then, the cost range between Hypo1 and the fixed cost is 24.945, while that between the null hypothesis and Hypo1 is 161.977, which shows that Hypo1 has more than 90% probability of correlating the data. Noticeably, the total cost of Hypo1 was much closer to the fixed cost than to the null cost. Furthermore, a high correlation coefficient of 0.989 was observed with RMS value of 1.567 and the configuration cost of 14.598, demonstrating that we have...
successfully developed a reliable pharmacophore model with high predictively.\(^3\)

**Score Hypothesis**

To verify Hypo1’s discriminability among p38 MAP kinase inhibitors with different order of magnitude activity, all training set compounds were classified by their activity as highly active +++ , \(<1.5 \text{ uM (highly active); } ++ , 10-570 \text{ uM (moderately active); } + , <1850 \text{ uM (inactive). The actual and estimated p38 MAP kinase inhibitory activities of the 25 compounds based on Hypo1 are listed in.**

The discrepancy between the actual and the estimated activity observed for the two compounds was only about one-order of magnitude, which might be an artifact of the program that uses different numbers of degrees of freedom for these compounds to mismatch the pharmacophore model. It shows that 20 molecules out of the 25 molecules in the training set have errors less than 10 which means that the activity prediction of these compounds falls between 10-fold greater and 1/10 of the actual activity.

These results confirm that our hypothesis is a reliable model for describing the SAR in the training set. In this study, all but one highly active compound map the hydrogen-bond acceptor (HA) feature, and one least active inhibitor do not have this feature.

**Validation of the Constructed Pharmacophore Model**

A correlation coefficient of 150 generated using the test set compounds shows a good correlation of 0.959 between the actual and the estimated activities. Detailed, 8 out of 10 highly active, 36 of 55 moderately active, and 46 of 64 inactive compounds were predicted correctly.

Two highly active compounds were underestimated as moderately active; five moderately active compounds were underestimated as inactive and other seven moderately active compounds were overestimated as highly active; most of inactive compounds were overestimated as moderately active.

The most active compound 54 in the test set had a fitness score of 12.45 when mapped to the Hypo 1 as seen in Fig. 2A(2) and shows that all the features are being mapped accurately. The least active compound 4 in the test set had a fitness score of 8.76 when mapped to the Hypo 1 as seen in Fig. 2B(2) and shows that all the features are not being mapped accurately. In conclusion, most of the compounds in the test set were predicted correctly, which mean the hypothesis is suited for screening high active compounds from the database.

---

**Fig. 2B Pharmacophore mapping of the least active compound on the best hypothesis model Hypo1.**

(1) Compound 22 from the training set.

(2) Compound 4 from the test set.

---

**Model Validation and Knowledge based Screening**

The main purpose of this pharmacophore hypothesis generation was not only to predict the activity of the training set compounds accurately but also to verify whether the pharmacophore models would be capable of predicting the activities of compounds of the test set series and classifying them correctly as active or inactive. The best pharmacophore hypothesis was used initially to screen the p38 MAP kinase inhibitors. All these queries were performed using the Best Flexible search databases/Spreadsheet method. Hyporefine 1 was used to screen the known high, medium and low active inhibitors of the test set. Database mining was performed in Catalyst software using the BEST flexible searching technique. A number of
parameters such as hit list (Ht), number of active percent of yields (%Y), percent ratio of actives in the hit list (%A), enrichment factor of (E), False negatives, False positives and Goodness of hit score (GH) are calculated while carrying out the pharmacophore model and Virtual screening of test set molecules.

The number of molecules in the database is 304. Of these, 215 are highly active, 65 are moderately active and 32 are low active compounds. While the False positives and negatives, 16 and 12 respectively, are minimal, enrichment factor of 1.43 against a maximum value of 3.2 is a very good indication on the high efficiency of the screening. Of the 215 highly active molecules, 16 were predicted as moderately active and 3 were predicted as least active. In the 65 moderately active, 7 were predicted as low active and 2 as highly active.

The model also predicted 3 of the low active molecules as moderately active and 2 more molecules from the same set as highly active. The steric and other interaction effects might have a subtle, yet crucial role on the predicted activity.

Conclusions

The work presented in this study shows how chemical features of a selected set of compounds along with their activities ranging over several orders of magnitudes can be used to generate pharmacophore hypotheses that can successfully predict the activity. The models were capable of predicting the activities over a wide variety of scaffolds and exhibited distinct chemical features which may be responsible for the activity of the inhibitors. This knowledge can be used to identify and design inhibitors with greater selectivity.

Thus, the pharmacophores generated from the p38 MAP kinase can be used as the following:

1. To generate Pharmacophore models as powerful search tool to be used as a 3D query to identify lead molecules from chemical databases as potential p38 MAP kinase inhibitors.
2. Evaluation of how well any newly designed compound maps on the pharmacophore before undertaking any further study including the synthesis of compounds. These both applications may help in identifying or designing compounds for further biological evaluation and optimization.

A total data set of test and training of 154 compounds of selective p38 MAP kinase inhibitors whose chemical features along with their respective activities ranging over a wide range of magnitude is used to generate pharmacophore hypotheses to successfully and accurately predict the activity. A highly predictive pharmacophore model was generated based on 25 training set molecules, which had hydrogen-bond acceptor, hydrophobic, hydrophobic bond donor and ring aromatic as chemical features which described their activities towards p38 MAP kinase. The validity of the model was based on 129 test set molecules, which finally showed that the model was able to accurately differentiate various classes of p38 MAP kinase inhibitors with a high correlation coefficient of 0.951 between experimental and predicted activity. This validated pharmacophore model, as such can be used as a query for identification of potential inhibitors of p38 MAP kinase while it can also be used to validate the potential of the compound to inhibit the enzyme prior to taking any step regarding the synthesis. p38 MAP kinase enzymes have proven to be exciting and promising novel targets for the treatment of osteoarthritis.

In-house build Medichem database was useful as a powerful resource to identify many p38 MAP kinase inhibitors with highly varied activities and chemotypes. These p38 MAP kinase inhibitors have been retrieved from the resource and some of them have been used to general a Pharmacophore model while other inhibitors have been used for virtual screening to validate the model.

The best quantitative Pharmacophore model in terms of predictive value consisted of four features like one hydrogen-bond acceptor (HA), one hydrophobic aromatic (HY), and two ring aromatic (RA) features, which is further validated by using a large set of 376 p38 MAP kinase inhibitors and gives a r value of 0.945. The most active molecule 51 (IC\textsubscript{50} = 0.005 uM) in the training set fits very well with this top scoring pharmacophore hypothesis.

Virtual screening produced some false positives and a few false negatives. This Pharmacophore model was further used to search the NCI database consisting of structurally diversified molecules, yielded 215 molecules as hits that satisfied the 3D query.

Thus, we hope that the model generated will be helpful to identify novel and potential lead molecules with improved activity against p38 MAP kinase.

Experimental

All molecular modeling works were performed on a Silicon Graphics Octane R12000 computer running Linux 6.5.12 (SGI, 1600 Amphitheatre Parkway, Mountain View, CA 94043) Catalyst 4.11 software was used to generate Pharmacophore models and Ligand Fit of Cerius 2 (Accelrys software) was used for docking of molecules in active site.
Acknowledgements

The author thanks Dr. J.A.R.P. Sarma, Senior Vice President and Dr. S. Vadivelan, Senior Scientist, GVK Biosciences Pvt. Ltd., Chennai for their valuable guidance, providing software facilities and a great chance to work there.

References


[7] Babu J. Mavunkel; Sarvajit Chakravarty; John J. Perumattam; Gregory R. Luedtke; Xi Liang; Don Lim; Yong-jin Xu; Maureen Laney; David Y. Liu; George F. Schreiner; John A. Lewicki; Sundeept Dugar, Indole-Based Heterocyclic Inhibitors of p38alpha MAP Kinase: Designing a Conformationally Restricted Analogue, Bioorg. Med. Chem. Lett., 2003, 13 (18), 3087-3090.


[12] Laszlo Revesz; Ernst Blum; Franco E. Di Padova; Thomas Buhl; Roland Feifel; Hermann Gram; Peter Hiestand; Ute Manning; Gerard Rucklin, Novel p38 inhibitors with potent oral efficacy in several models of rheumatoid arthritis, Bioorg. Med. Chem. Lett., 2004, 14 (13), 3595-3599.

[13] Laszlo Revesz; Ernst Blum; Franco E. Di Padova; Thomas Buhl; Roland Feifel; Hermann Gram; Peter Hiestand; Ute Manning; Gerard Rucklin, SAR of benzoylpyridines and benzophenones as p38alpha MAP kinase inhibitors with oral activity, Bioorg. Med. Chem. Lett., 2004, 14 (13), 3607-3610.


[16] Dumas J; Sibley R; Riedl B; Monahan MK; Lee W; Lowinger TB; Redman AM; Johnson JS; Kingery-Wood J; Scott WJ; Smith RA; Bobko M; Schoenleber R; Ranges GE; Housey TJ; Bhargava A; Wilhelm SM; Shrikhande A, Discovery of a new class of p38 kinase inhibitors, Bioorg. Med. Chem. Lett., 2000, 10 (18), 2047-2050.

[17] Dumas J; Hatoum-Mokdad H; Sibley R; Riedl B; Scott WJ; Monahan MK; Lowinger TB; Brennan C; Natero R; Turner T; Johnson JS; Schoenleber R; Bhargava A; Wilhelm SM; Housey TJ; Ranges GE; Shrikhande A; 1-Phenyl-5-pyrazolyl ureas: potent and selective p38 kinase inhibitors, Bioorg. Med. Chem. Lett., 2000, 10 (18), 2051-2054.


[19] Dumas J; Hatoum-Mokdad H; Sibley RN; Smith RA; Scott WJ; Khire U; Lee W; Wood J; Wolanin D; Cooley J; Bankston D; Redman AM; Schoenleber R; Caringal Y; Gunn D; Romero R; Osterhout M; Paulsen H; Housey TJ; Wilhelm SM; Pirro J; Chien DS; Ranges GE, Synthesis and pharmacological characterization of a potent, orally active p38 kinase inhibitor, Bioorg. Med. Chem. Lett., 2002, 12 (12), 1559-1562.

[20] Dumas J; Sibley R; Riedl B; Monahan MK; Lee W; Lowinger TB; Redman AM; Johnson JS; Kingery-Wood J; Scott WJ; Smith RA; Bobko M; Schoenleber R; Ranges GE; Housey TJ; Bhargava A; Wilhelm SM; Shrikhande A, Discovery of a new class of p38 kinase inhibitors, Bioorg. Med. Chem. Lett., 2000, 10 (18), 2047-2050.

[21] Dumas J; Hatoum-Mokdad H; Sibley R; Riedl B; Scott WJ; Monahan MK; Lowinger TB; Brennan C; Natero R; Turner T; Johnson JS; Schoenleber R; Bhargava A; Wilhelm SM; Housey TJ; Bhargava A; Ranges GE; Shrikhande A; 1-Phenyl-5-pyrazolyl ureas: potent and selective p38 kinase inhibitors, Bioorg. Med. Chem. Lett., 2000, 10 (18), 2051-2054.

[22] Redman AM; Johnson JS; Dally R; Swartz S; Wild H; Paulsen H; Caringal Y; Gunn D; Renick J; Osterhout M; Kingery-Wood J; Smith RA; Lee W; Dumas J; Wilhelm SM; Housey TJ; Bhargava A; Ranges GE; Shrikhande A; Young D; Bombara M and Scott WJ, p38 Kinase Inhibitors for the Treatment of Arthritis and Osteoporosis: Thieryl, Furyl, and Pyrrolyl Ureas., Bioorg. Med. Chem. Lett., 2001, 11 (1), 9-12.


[26] Laszlo Revesz; Franco E. Di Padova; Thomas Buhl; Roland Feifel; Hermann Gram; Peter Hiested; Ute Manning and Alfred G. Zimmerlin, SAR of 4-Hydroxyperipérimeride and Hydroxysalyl substituted Heterocycles as Novel p38 MAP Kinase Inhibitors, Bioorg. Med. Chem. Lett., 2000, 10 (11), 1261-1264.

[27] Stephen E. de Laszlo; Denise Visco; Lily Agarwal; Linda Chang; Jayne Chin; Gift Croft; Amy Forsyth; Daniel Fletcher; Betsy Frantz; Candice Hacker; William Hanlon; Coral Harper; Matthew Kostura; Bing Li; Sylvie Luell; Malcolm MacCoss; Nathan Mantlo, PYRROLES AND OTHER HETEROCYCLES AS INHIBITORS OF P38 KINASE, Bioorg. Med. Chem. Lett., 1998, 8 (19), 2689-2694.


[31] Mark A. Dombroski; Michael A. Letavic; Kim F. McClure; John T. Barberia; Thomas J. Carty; Santo R.


