International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

3D QSAR Analysis of a Set of *Pneumocystis carinii* DHFR Inhibitors through Pharmacophore Generation Approach

Ruchi Mishra*

Assistant professor, Department of Chemistry, Khandelwal College of Management Science and Technology, Bareilly, India

*Email: ruchimishra_0209[at]rediffmail.com

Abstract: ***A ligand-based pharmacophore model using Catalyst HypoGen algorithm was developed for set of pyrimidine and quinazoline analogues as pcDHFR inhibitors with an aim to obtain rational hypothetical image of the primary chemical features responsible for activity and this pharmacophore model was used as an in silico screening tool to retrieve novel and potential inhibitors against pcDHFR from various databases. The best pharmacophore model for selective pcDHFR inhibitors (Hypo-1) was obtained through a Cat-Scramble validation process. The best pharmacophore model (Hypo-1) for pcDHFR inhibitors consisting of one hydrogen bond acceptor lipid (HBAl), three hydrophobic (HY) and one ring aromatic (RA) features, with highest correlation coefficient (0.94), cost difference (45.1), low RMS (0.72), as well as it shows a high goodness of fit and predictive factor. Hydrophobic interactions are essential for ligand pharmacophore interaction. Pharmacophore models have been validated toward a test set containing 8 molecules. To further evaluate the model external test set comprising of known pcDHFR inhibitors were mapped on to developed pharmacophoric model which also showed five point mapping and estimated values in close range to actual values. The models were used for screening chemical data base. The validated pharmacophore model (Hypo-1) was used as a 3D query for virtual screening to retrieve potential inhibitors from the Maybridge and National Cancer Institute (NCI) databases. This resulted in identification of three druggable structurally diverse potent lead compounds. The results of our study will act as a valuable tool for retrieving potent compounds with desired biological activities and designing novel selective pcDHFR inhibitors.***

Keywords: Catalyst, HypoGen, pcDHFR, Pharmacophore, virtual screening

1. Introduction

Pneumocystis carinii pneumonia is one of the premier causes of morbidity and mortality in patients with acquired immunodeficiency syndrome (AIDS).[1] *Pneumocystis carinii* is a eukaryotic microorganism that is found worldwide. Its host range is wide and includes humans and other mammals such as rabbits, dogs, goats, swine, cats, chimpanzees, owl monkeys, and horses. It is generally understood that there are two ways by which *Pneumocystis* can begin infection after acquiring the organism: firstly by activation of latent organisms present in the host as a result of earlier acquisition, and secondly by reinfection. The metabolism of folate plays an important role in the biosynthesis of nucleic acid precursors. During the synthesis of purines and thymidylate, the cofactor tetrahydrofolate is oxidised to 7, 8-dihydrofolate and subsequently converted back to tetrahydrofolate by the enzyme dihydrofolate reductase (DHFR). The inhibition of DHFR causes the depletion of tetrahydrofolate and disrupts DNA synthesis, leading to cell death. For this reason, DHFR inhibitors such as methotrexate have been used as antitumor, anti-bacterial and antiprotozoan agents. Because MTX and other classical antifolates require an active transport mechanism for their uptake, they are not effective for the treatment of infections caused by *Pneumocystis carinii* that lack these mechanisms.[2, 3] *Pneumocystis carinii* infections are the principal cause of death in patients with AIDS, and also affect patients with other immune disorders.[4] Trimetrexate (TMQ) and piritrexim (PTX) are potent lipophilic inhibitors of pcDHFR and tgDHFR taken up by passive diffusion, but inhibit mammalian DHFR to a greater extent.[5, 6] This results in toxicity to mammalian tissue and requires that PTX or TMQ be co-administered with leucovorin, areduced folate which is taken up by active transport and protects the host tissue.[7] Treatment with leucovorin is costly and subject to serious side effects that may require interruption of treatment. As such, there is great interest in developing potent and selective inhibitors of *Pneumocystis carinii* DHFR.

Discovering and bringing one new drug to the public typically costs a pharmaceutical or biotechnology company nearly $900 million and takes an average of 10 to 12 years The application of computer-assisted drug design (CADD) methodologies to this problem has the potential to greatly decrease the time and effort required to discover new medicines or improve current ones in term of their efficacy. In view of this, the present work applies the development of pharmacophore which can help to visualize the potential interaction between ligands & target and can be used as a query in a 3D data base search to identify new structural classes of potential lead compounds. A pharmacophore represents the 3D arrangements of chemical features in a molecule (ligand) that may be essential for important binding interactions with a receptor. In the absence of any knowledge of the 3D structure of a receptor, pharmacophores may provide such important information in the drug design process. The pharmacophores may be used in several ways, for example, as a 3D query in searching 3D databases containing drug-like organic molecules to identify active and specific antagonists or in evaluating a new compound for mapping on a known pharmacophore.[8] Details of the pharmacophore development procedures have been described in the literature. The hypothesis generation methods of the

Catalyst software have been successfully used in drug discovery research and toxicology.[9-13] Our approach of pharmacophoric exploration via set of diverse 3D structures employing training lists should achieve two importantgoals: (i) The pharmacophore could predict and discriminate the activities of all the compounds and (ii) could explain the difference in activities of those highly active which were active in subnanomolar range over the others.

2. Material and Methods

Data set preparation and confirmation analysis

Catalyst 2.0 software (Discovery Studio, version 2.0; Accelrys Software Inc.: San Diego, CA 2001) was used to generate pharmacophore models. The biological activity data, (represented as IC50 in µM) were obtained from the studies reported in the literature Gangjee et al., 2003 & Gangjee et al., 2008.[14, 15] The present series was chosen in view of the established requirement of Catalyst pharmacophore generation module that is activity range of the compounds should span at least 3.5 orders of magnitude. The most active compound should be included so that they would provide critical information for pharmacophore requirements.

Several moderately active and inactive compounds were also included to spread the activity ranges as wide as possible. Any redundancy should be avoided in terms of structural features or activity range. The important aspect of this selection scheme is that each active compound should teach something new to the HypoGen module to help it uncover as much critical information as possible for predicting biological activity.[16] All structures were built using a 3D sketcher in catalyst discovery studio and were minimized to a local energy minimum using the CHARMm force field.[17] To generate 3D pharmacophore, each compound should have conformations to cover three-dimensional spaces. The single conformer 3D structures were used as starting point for conformational analysis and in the determination of various chemical features for QSAR-based pharmacophore modeling. The conformation is of great importance for the mode of drug action since it relies on the easy accessibility of the reactive groups. The conformational space of each inhibitor was extensively sampled utilizing the poling algorithm employed within Catalyst.[18] Poling promotes conformational variation via employing molecular mechanical force field algorithm that generates conformers that provide broad coverage of the accessible conformational space and achieve maximum diversity of a conformational model. Fast generation takes less time, but best generation provides more complete coverage of conformational space by optimizing the conformation in both torsional and cartesian space. Conformational models of training-set molecules for pcDHFR were generated using the best quality conformational search option in Catalyst using a constraint of 20 kcal mol-1 energy threshold above the global energy minimum and CHARMm force field parameters. A maximum of 250 conformations were generated to ensure maximum coverage in the conformational space. All other settings were kept as default. Instead of using just the lowest energy conformation of each compound, all conformational model for molecules in each training set were used in Catalyst for pharmacophore hypothesis generation. All the 28 compounds of the series and their different conformations were divided into training set of 20 compounds and test set of 8 compounds. The test was used to ascertain the predictive power of the model.

Feature mapping

A maximum of five features can be considered in the pharmacophore generation process using Catalyst hypogen algorithm.[19] Accordingly, from the 11 features available in the Catalyst features dictionary the features necessary to explain the variance in activity of the present pyrimidine and quinazoline analogues were identified by using feature mapping protocol. The feature mapping protocol generates all possible pharmacophore features for the given input ligands. In present work, the chemical features optimized for exploring the spatial pharmacophore map of series of pyrimidine and quinazoline analogues and their ester derivatives were hydrophobic, ring aromatic, hydrogen bond acceptor, hydrogen bond acceptor-lipid and hydrogen bond donor. Using these five features and varying the value of these parameters from min 0 to max 5, hypotheses were generated. The analysis revealed that hydrogen bond acceptor-lipid (0, 5), hydrophobic (0, 5) and ring aromatic (0, 5) features are the most important pharmacophoric feature for explaining the dependence of anti-pneumocystis activity of pyrimidine and quinazoline analogues.

Generation of pharmacophore hypotheses

Hypogen algorithm

In particular, here we report a study that applies a ligand based drug design (predictive pharmacophore generation) technique to rationalize the relationships between of pyrimidine and quinazoline analogues structures and their affinity data towards the pcDHFR that is to model drug receptor interactions using information derived from the drug structure. HypoGen identifies a 3D array of a maximum of five chemical features common to active training molecules, which provides a relative alignment for mapping of each molecule. The HypoGen run is divided into three phases. During the first one, known as the constructive phase, all the pharmacophore models, which are common to the most active compounds, are created. The second, called the subtractive phase, removes all the generated pharmacophore models in which the inactive compounds can fit. Finally, the third one, known as the optimization phase, makes some random modifications in the generated hypothesis's moving features, rotating vectors, or adding or removing features. Finally, correlation values are obtained by linear regression of the geometric fit index. The correlation coefficient is based on linear regression derived from the geometric fit index.[16, 20] In our pharmacophore exploration strategy, the pharmacophoric spaces of the selected training set were explored under reasonably imposed boundaries . The software was limited to search pharmacophoric models incorporating from zero to three features of any particular selected feature type instead of the default range of zero to five. With the default

setting (weight variation) of 0.302, the represented orders of magnitude were kept as close to 0.2 as possible. Drawing on the characteristic features of the training-set molecules, hydrogen bond acceptor-lipid (HBAl), hydrophobic (HY) and ring aromatic (RA) were selected from the feature dictionary of Catalyst to form the crucial basis for the hypothesis development. HypoGen process returned ten pharmacophore models with top ranking scores. These models were then evaluated and validated against the training and test sets. The quality of the generated pharmacophore models was evaluated using a cost function analysis and Fisher's randomization test.

Quality assessment of pharmacophore hypothesis

Cost function analysis

The HypoGen module in Catalyst performs two important theoretical cost calculations that determine the success of any pharmacophore hypothesis. One is known as the fixed cost, which represents the simplest model that fits all data perfectly, and the second one is known as 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. Because the null hypothesis is an empty hypothesis with no features, there is no contribution of the weight and configuration costs. All of these cost values are represented in bits, and Catalyst analyzes the pharmacophore models using the Occam's razor principle; that is, among equivalent possibilities (hypotheses), the simplest (less bits cost) is the best. Two other parameters that also determine the quality of any pharmacophore hypothesis with possible predictive values are the configuration cost, which is also known as the entropy cost and depends on the complexity of the pharmacophore hypothesis space, and the error cost, which is dependent on the RMS differences between the estimated and the actual activities of the training-set molecules.

The RMS deviations represent the quality of the correlation between the estimated and the actual activity data. The error cost is the most important part of the total cost and increases as the root mean-square (RMS) difference between the estimated and the actual affinity for the training set increases. The RMS value is related to the quality of prediction of the hypothesis. The weight cost is a value that increases in a gaussian form as the difference between the actual and ideal weights of the features deviates. According to the documentation, the ideal value of the weight is 2 because higher weight values tend to force unrealistic conformations of the compounds to fit such features. The configuration cost is a constant cost that depends on the complexity of the hypothesis space to be optimized. It describes the entropy of the hypothesis space and is related to the number of hypotheses that have been created in the constructive phase. The configuration is $\log_2 P$, where P is the number of initial hypotheses created in the constructive phase and that survived the subtractive phase. In standard HypoGen mode, the configuration should not be greater than 17.0. In each run, the resulting binding hypotheses were automatically ranked according to their corresponding total cost value, which is defined as the sum of error cost, weight cost, and configuration cost. Error cost provides the highest contribution to total cost and it is directly related to the capacity of the particular pharmacophore as 3D QSAR model, i.e., in correlating the molecular structures to the corresponding biological responses. Accordingly, the greater the difference from the null hypothesis cost, the more likely that the hypothesis does not reflect a chance correlation. Each feature of a hypothesis represents certain orders of magnitude of the compounds activity.

Validation of best pharmacophore model

Test set prediction

The ability of the models to predict the biological activity of compounds outside the model development procedure is a common method of validation. Test set was employed to assess statistical significance of the developed model.[21, 22] Here in this case, test set prediction was measured in terms of squared correlation coefficient.

Cat-scramble program

In order to assess the statistical significance of the generated pharmacophore hypotheses, a validation procedure based on Fischer's randomization test was performed. This was carried out by randomizing the activity data associated with the training set compounds using the *Cat-scramble* technique, available in the Catalyst/hypogen module. These randomized training sets were then used to generate pharmacophore hypotheses, employing the same features and parameters as used in the development of the original pharmacophore hypothesis. The number of spreadsheets generated depends on the level of statistical significance selected. Thus 19, 49, or 99 random spreadsheets have to be generated for attaining 95%, 98%, or 99% confidence level respectively.[23, 12] If the randomized data set results in generation of pharmacophoric models with similar or better cost values and correlation then the original hypothesis is considered to be generated by chance.

External validation set

To produce a strong validation of this model which can be utilized for virtual screening purposes, the external validations set comprising of twelve compounds with known pcDHFR inhibitory activity were taken from the reported literatures (Gangjee et al., 1995).[24] Various conformers of the compounds were generated and mapped on to the pharmacophoric features obtained from both ligand and structure based approach. The pcDHFR inhibitory activity was estimated and compared with the actual activity. This further supported the essential binding requirement for pcDHFR inhibitory activity of the pyrimidine and quinazoline analogues.

Database search for new hits

After generating a validated pharmacophore model containing pharmacophore features, shapes etc., National Cancer Institute (NCI) and MiniMaybridge 3D database search was performed to identify new molecules which

share its features and can thus exhibits the desired biological response. The screened compounds were subjected to drug-like property calculation by applying Lipinski's rule-of-five [25], which is a simple model to predict the absorption and intestinal permeability of a compound. According to the rule, compounds are well-absorbed when they possess Log P less than 5, molecular weight less than 500, number of H-bond donors less than 5, number H-bond acceptors less than 10 and number of rotable bonds less than 10.

3. Results and Discussion

A set of ten hypotheses were generated by the implemented algorithm on the basis of the activities and structures of the training compounds. The best pharmacophore hypothesis, named Hypo-1, was marked by the best correlation coefficient (0.94) and the highest cost difference (45.1) as shown in [Table 1]. The fixed cost of the best hypothesis was 176.9, and the cost of the null hypothesis is 242.3. The cost ranges between the best hypothesis (total cost: 197.2), the null hypothesis amount to 45.1. Because of the fact that the best hypothesis' total cost is much closer to the fixed cost than to the null cost and difference between total and fixed costs for the best hypothesis was only 11.58 bits, the correlation coefficient of 0.94 and the RMS value of 0.72 indicates a reliable ability of the generated pharmacophore model to predict training-set compounds activities and confirms that it does not show chance correlation. Moreover, despite of limitations on the extent of evaluated pharmacophore space, the weight costs of Hypo1 was high and exceeded the maximum limit of 2. Nevertheless, the reasonable cost and confidence criteria of pharmacophore model should overshadow any drawbacks related to the lessthan optimal weight costs. HypoGen generates hypotheses whose features contain a certain tolerance and weight that fit to the features of the training set and that correlate to the activity data. The Hypo1 model was utilized to predict the activities of all 21 training compounds. The estimated activities as predicted by Hypo1, the experimental activities are listed in [Table 2]. It is clear that the IC50 values were well predicted demonstrating the good predictive quality of Hypo1. Hypotheses included one hydrogen bond acceptor-lipid (HBAl), three hydrophobic (HY) and one ring aromatic (RA) features. On visual inspection of the hypotheses 'best' hypothesis was selected for the qualitative comparison was Hypo1. The features present in each hypothesis are shown in [Table 2].

Mapping of compounds 6, 08 and on the best

Pharmacophore In order to evaluate the predictive ability of Hypo1, the features of training-set compounds were mapped using Hypo1. The most active compound, 6, 08, mapped well to all the features of Hypo1. Most active compound of the series aligned to the developed pharmacophore is illustrated in [Figure 1]. One HY mapped to that ring of naphthalene which is connected with the quinazoline nucleus and one RA feature mapped to the terminal ring of naphthalene. second hydrophobic feature mapped to the ethyl group of side chain and third hydrophobic feature mapped at side chain bearing ring of quinazoline nucleus, one HBAl mapped to the 1' nitrogen of terminal ring of quinazoline nuclus, with a fit value of 9.20 (actual activity 2.5 nM and 1.65 nM) and is shown in [Table 2].

Mapping of compounds 6a, 03 and 13a, 03

Least active compound 6a, 03 of the series aligned with the developed pharmacophore is depicted in [Figure 2a]. One HY mapped to the phenyl ring of side chain and second hydrophobic feature mapped at side chain bearing ring of pyrimidine nucleus and one HBAl mapped to the 1' nitrogen of pyrimidine nucleus with a fit value of 6.24 (actual activity 6100 nM and estimated 1486.49 nM). Missing of one hydrophobic feature and one RA can be easily interpretated as no alkyl group substitution at nitrogen of side chain and absence of naphthalene ring in side chain in this compound [Table 2] thus fails to map this feature. To further investigate compounds with lower activity in the training set, the moderately active compound 13a, 03 [Figure 2b] was mapped on Hypo1. The mapping revealed that this compound missed one hydrophobic and one HBAl features. The compound with lower activity clearly missed one hydrophobic and one RA features of Hypo1 and compound with moderate activity missed one hydrophobic and one HBAl features, demonstrating the importance of these features in maintaining high potency.

Test set predictions

To evaluate the ability of Hypothesis 1 to identify pc DHFR inhibitors, a test set consisting of 8 ligands was submitted to phramcophore mapping analysis using the developed model. Objective of test set prediction is to verify whether generated pharmacophore models are capable of predicting the activities of compounds not included in training set and classifying them correctly as actives or inactives. Finally, the compounds were mapped onto the best hypothesis using the best fit and a conformational energy constraint of 10 kcal mol^{-1}. A correlation coefficient of 0.83 generated using the test set compounds indicates a good correlation between the actual and estimated activities, which means the Hypothesis 1 is convictive.

The estimated activities of test set were scored using hypothesis 1 as the pharmacophore and shown in [Table 2]. Three least active compounds were assigned least active by Hypothesis 1 [Table 2]. All highly active compounds (≤ 13 nM) were predicted to be highly active except compound 1b, 03 which was inaccurately predicted as moderately active. Among the moderately active compounds (42-84 nM) all were correctly predicted as moderately active.

Compound 9a, 08 [Figure 3a] mapped all five features of the hypothesis quite well with a fit value of 7.89 (actual activity 6.9 nM and estimated 33.84 nM). Only highly active compounds mapped all the five features in the test set. The poorly active compound, 10a, 03 [Figure 3b] misses one of the HY and one HBAl completely. As all the highly active compounds with fit value 7.64–7.89 mapped five features, this suggested that presence of five features is minimum essential for the highly active compounds and three feature hypothesis missing HY is sufficient to account

for the least active compounds. The results of the activity prediction of these compounds are shown in [Table 2].

Fisher's cross-validation test

To further evaluate the statistical relevance of the model, the Fischer validation method at the confidence level of 99% was applied to the developed Hypogen. The experimental activities in the training set were scrambled randomly using CatScramble program, and the resulting training set was used for a HypoGen run. In this manner, all parameters were taken from the initial HypoGen calculation. The experimental activities of the molecules in the training set were scrambled 99 times to obtain spreadsheets with randomized activity data as illustrated in [Figure 4]. None of the outcome hypotheses had a lower cost score than the initial hypothesis. We have also performed rigorous cross validation analysis of the developed pharmacophore model using external test set prediction.

External validation

The results of the actual and estimated biological IC50 values of these twelve external test set compounds estimated by the Hypo-1 model are presented in [Table 3]. We took approximately 70% of compounds of the total dataset into our test set. These values do not match exactly in numerical terms because the evaluation of the test set compounds was carried out at a different location under different conditions compared with those used in the building of our model. Hence the validation of the Hypo-1 was based on the correlation of the predictive and observed values where the obtained predictive R^2 value [R^2 pred = 0.75] and the graph drawn between the actual and estimated activities for the external test set compounds using the best hypotheses Hypo-1 clearly show the strong validation of the model.

Database search

495 and 989 compounds were retrieved from the MiniMaybride and NCI data base respectively. The parameters included in Lipinski's rule of 5, were calculated for three compounds obtained from 3D data base search, which indicates that there is no violation to Lipinski rule and it is highly likely that these three compounds will have favorable pharmacokinetics profile. The screened lead compounds needs further evaluation in order to produce newer compounds for Pneumocystis crainii pneumoninia.

4. Conclusion

Our results suggest that pharmacophore modeling of pcDHFR inhibitors can be a useful tool for finding potential antipneumocystis agents. The exploration of the pharmacophoric features of pyrimidine and quinazoline analogues was done through the use of Catalyst-HypoGen to identify high-quality binding model. Pharmacophore models generated for pcDHFR inhibitors in this study highlight the structural requirements for antagonistic activity of pyrimidine and quinazoline analogues. All features in the model are essential for antagonistic activity. Compounds that can map all features in the pharmacophore model will be considered as potent pcDHFR inhibitors. External validation set of known pcDHFR inhibitors were also used to evaluate the derived pharmacophore features necessary for inhibitory activity of pyrimidine and quinazoline analogues.

The validated pharmacophore model was used for searching new lead compounds and we obtained three compounds with promising activity. The new lead candidate compounds were checked for druggable properties by applying Lipinski's rule. Thus, our pharmacophore model was able to retrieve few leads which had good estimated inhibitory activity with acceptable calculated drug-like properties and therefore they are subjected to further optimization.

References

[1] Barlett MS, Smith JW. Pneumocystis carinii, an opportunist in immunocompromized patients. Clin. Microbiol. Rev. 1991; 4: 137-49.
[2] Warren E, You J. Advances in the treatment and prophylaxis of Pneumocystis carinii pneumonia. Pharmacotherapy. 1997; 17: 900-16.
[3] Behbahani R, Moshfeghi M. Therapeutic approaches for AIDS-related toxoplasmosis. Ann. Pharmacother. 1995; 29: 760-68.
[4] Kovacs JA, Hiemenz JW. Pneumocystis carimi pneumonia: a comparison between patients with the the acquired immunodeficiency syn- drome. Ann. Intern. Med. 1984; 100: 663-71.
[5] Allegra CJ, Kovacs JA. Potent in vitro and in vivo antitoxoplasma activity of the lipid-soluble antifolate trimetrexate. J. Clin. Invest. 1987; 79: 478-82.
[6] Kovacs JA, Allegra CJ. Potent antipneumocystis and antitoxoplasma activities of piritrexim, a lipid-soluble antifolate Antimicrob. Agents Chemother. 1988; 32: 430-33.
[7] Kovacs JA, Allegra CJ. Characterization of de novo folate synthesis in Pneumocystis carinii and Toxoplasma gondii: potential for screening therapeutic agents. J. Infect. Dis. 1989; 160: 312-20.
[8] Hecker EV, Duraswami C, Andrea TA, Diller DJ. Use of catalyst pharmacophore models for screening of large combinatorial libraries. J Chem Inf Comput Sci. 2002; 42: 1204–11.
[9] Paliwal S, Yadav D, Yadav R, Paliwal S. In silico structure based drug design approach to develop novel pharmacophore model of human peroxisome proliferator activated receptor (PPARc) agonists. Med Chem Res. 2011a; 20: 656–59. doi:10.1007/s00044-010-9370-x.
[10] Sprague PW. Automated chemical hypothesis generation and database searching with catalyst. Perspect Drug Discov Des. 1995; 3: 1–20.
[11] Kurogi Y, Guner OF. Pharmacophore modeling and three dimensional database searching for drug design using catalyst. Curr Med Chem. 2001; 8: 1035–55.
[12] Langer T, Krovat EM. Chemical feature based pharmacophores and virtual library screening for

discovery of new-leads. Curr Opin Drug Discov Dev. 2003; 6: 370–76.
[13] Clark DE, Westhead DR, Sykes RA, Murray CW. Active sitedirected 3D database searching: pharmacophore extraction and validation of hits. J Comput Aided Mol Des. 1996; 10: 397–416.
[14] Doweyko AM. Three-dimensional pharmacophores from binding data. J Med Chem. 1994; 37: 1769–78.
[15] Brooks BR, Bruccoleri RF, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy minimization and dynamics. J Comput Chem. 1983; 4: 187–217.
[16] Smellie A, Teig S, Towbin P. Poling: promoting conformational variation. J Comput Chem. 1995; 16: 171–87.
[17] Greene J, Kahn **S,** Savoy H, Sprague P, Teig S. J. Chem. Inf. Comput. Sci. 1994; 34: 1297.
[18] Mason JS, Good AC, Martin EJ. 3-D pharmacophores in drug discovery. Curr Pharm Des. 2001; 7: 567–97.
[19] Paliwal SK, Pal M, Siddiqui AA. Quantitative structure activity relationship analysis of angiotensinII AT1 receptor antagonist. Med Chem Res. 2010; 19: 475–89.
[20] Paliwal S, Seth D, Yadav D, Paliwal S, Yadav R. Development of a robust QSAR model to predict the affinity of pyrrolidine analogs for dipeptidyl peptidase IV (DPP-IV). J Enzym Inhib Med Chem. 2011b; 26: 129–40. doi:10.3109/14756361003777057.
[21] Funk OF, Kettmann V, Drimal J, Langer T. J. Med. Chem. 2004; 47: 2750–60.
[22] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Adv. Drug. Delivery Rev. 2001; 46: 3-26.

Table 1: Performance of different pharmacophoric hypotheses generated with training set of molecules using the HypoGen algorithm

Hypothesis no.	Total cost	Correlation	RMS	Weight	Configration	Maximum fit
1	197.23	0.89	0.94	1.23	14.13	8.60
2	197.66	0.88	1.01	1.31	14.13	6.49
3	200.53	0.85	1.11	1.28	14.13	6.48
4	201.47	0.85	1.11	1.76	14.13	6.69
5	201.86	0.83	1.18	1.19	14.13	6.43
6	202.41	0.82	1.21	0.94	14.13	5.76
7	202.41	0.83	1.19	2.10	14.13	6.81
8	203.90	0.79	1.30	1.14	14.13	6.56
9	204.70	0.78	1.31	0.84	14.13	6.96
10	205.49	0.78	1.32	0.92	14.13	6.20

Table 2: Experimental and predicted activity and feature mapping of training and test set compounds using the hypo1 model

Name	Actual IC50 (µM)	Predicted IC50 (µM)	Fit value	Mapped feature				
				HBAl HY HY HY RA				
1a, 03	86	42.958	7.787	-	1	1	1	1
1b, 03*	13	58.994	7.649	1	1	1	1	1
2a, 03	1500	926.83	6.453	1	-	1	1	1
2b, 03*	240	594.271	6.646	1	1	1	1	-
3a, 03*	84	127.855	7.313	1	1	1	1	-
4a, 03	2220	2237.8	6.07	1	-	1	1	-
5a, 03	520	834.919	6.498	-	-	1	1	1
6a, 03	6100	1486.49	6.248	1	-	1	1	-
7a, 03	430	835.388	6.498	1	-	1	1	1
8a, 03	440	1099.67	6.379	-	1	1	1	-
9a, 03	250	1123.41	6.369	-	-	1	1	1
10a, 03*	350	1151.55	6.359	-	-	1	1	1
11a, 03	70	204.569	7.109	1	-	1	1	1
12a, 03	2000	2169.39	6.084	1	-	1	1	-
13a, 03	100000	35701.5	4.867	-	-	1	1	1
14a, 03*	131	1316.33	6.301	-	1	1	1	-
15a, 03	196	563.935	6.669	1	1	1	1	-
1, 08	3800	628.921	6.621	-	1	1	1	-
3, 08	4600	1164.28	6.354	1	-	1	1	-
4, 08	87	49.91	7.722	1	1	1	1	-
5, 08	21	32.157	7.913	1	1	1	1	-
6, 08	2.5	1.652	9.202	1	1	1	1	1
8, 08	9.9	20.305	8.112	1	1	1	1	1
9, 08*	6.9	33.864	7.89	1	1	1	1	1
10, 08	38	24.62	8.029	-	1	1	1	1
11, 08	27	33.412	7.896	1	1	1	1	1
12, 08	25	25.962	8.006	-	1	1	1	1
13, 08*	42	70.649	7.571	1	1	1	1	-

*Compounds included in Test set.

Table 3: Experimental and estimated IC_{50} values of the thirty external test set compounds based on the pharmacophore model Hypo-1.

Comp.	Actual activity	Estimated activity	Actual activity (-log value)	Estimated activity (-log value)
2a	46	88.903	-1.662	-1.948
2b	22.9	49.611	-1.359	-1.695
2c	316	145.982	-2.499	-2.164
2d	44	108.268	-1.643	-2.034
2e	76.7	471.098	-1.643	-2.034
3a	216	78.462	-1.643	-2.034
3b	130	23.4	-1.643	-2.034
3c	510	270.389	-2.707	-2.431
3d	320	159.488	-2.505	-2.202
3e	3100	259.835	-3.491	-2.414
5a	573	134.611	-2.758	-2.129
5b	41	76.241	-1.6127	-1.882

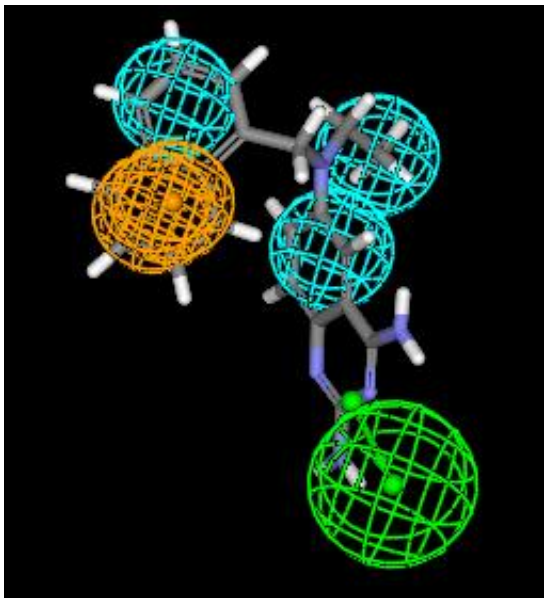

Figure 1: Best conformation of compound 6, 08 fit to the Catalyst generated pharmacophore model of pc DHFR inhibitor

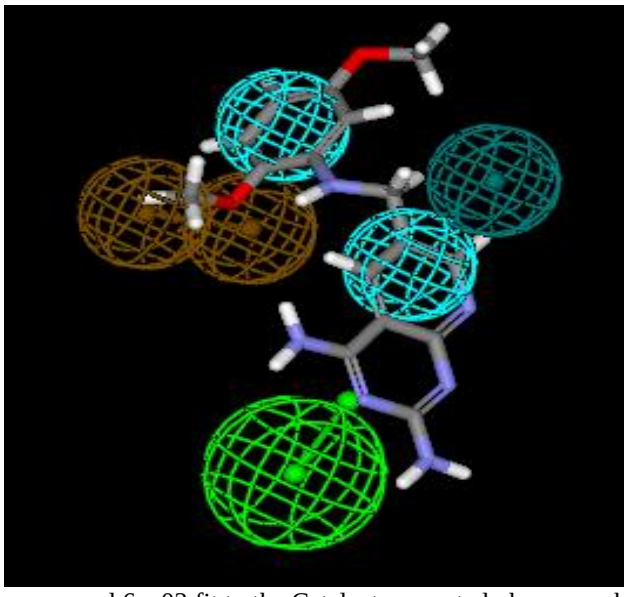

Figure 2a: Best conformation of compound 6a, 03 fit to the Catalyst generated pharmacophore model of pc DHFR inhibitors

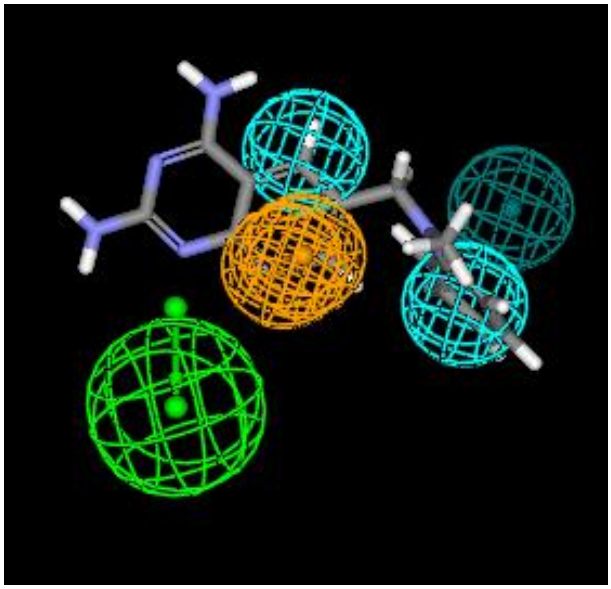

Figure 2b: Best conformation of compound 13a, 03 fit to the Catalyst generated pharmacophore model of pc DHFR inhibitors

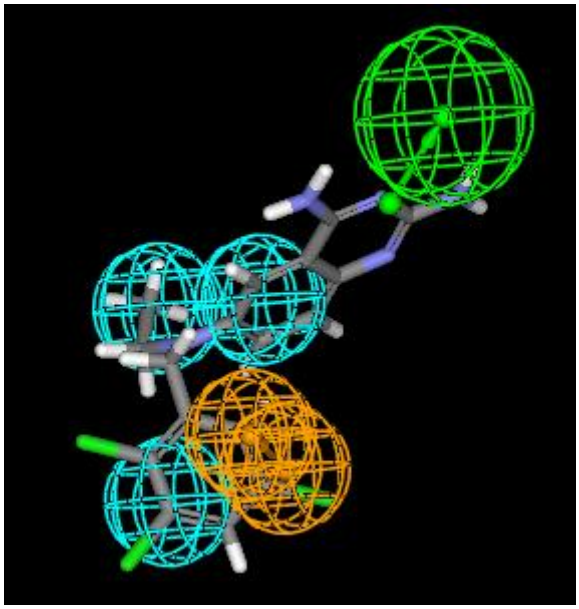

Figure 3a: Best conformation of compound 9a, 08 fit to the Catalyst generated pharmacophore model pc DHFR inhibitors

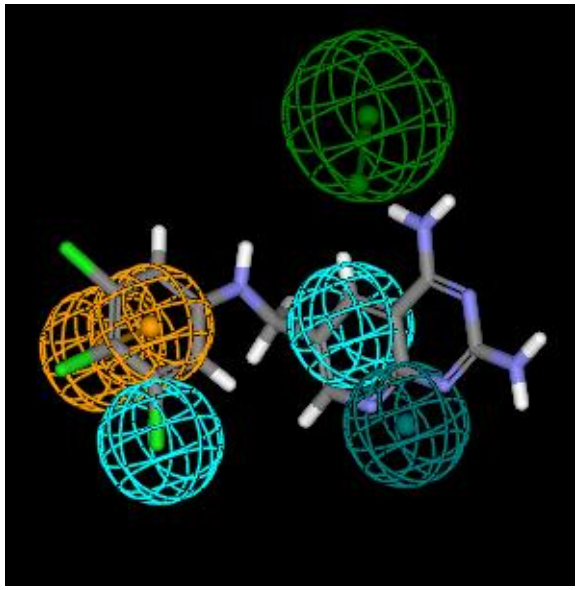

Figure 3b: Best conformation of compound10a, 03 fit to the Catalyst generated pharmacophore model of pc DHFR inhibitors

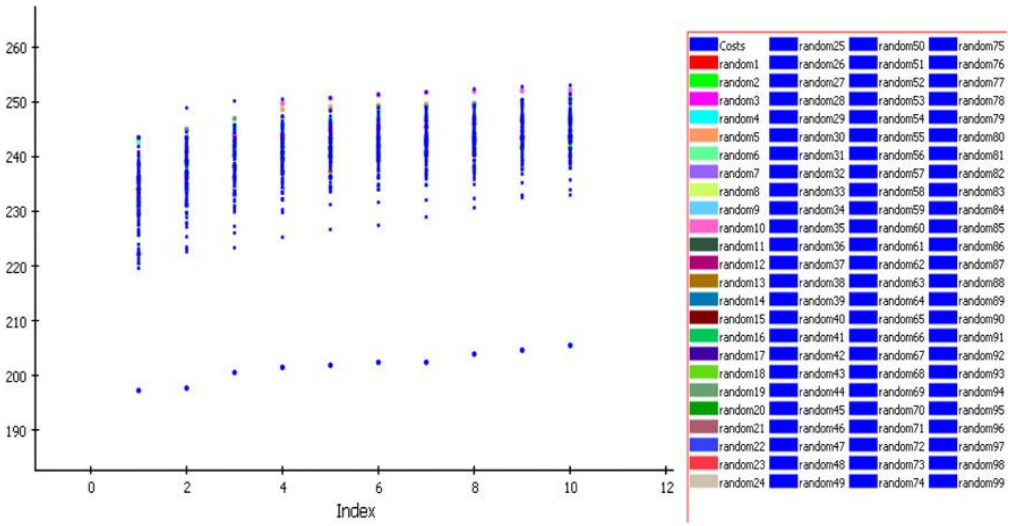

Figure 4: Graph of the catscrambled data generated from training set scrambled 99 times to obtain spreadsheets with randomized activity data.

None of the outcome hypotheses had a lower cost score than the initial hypothesis