ORIGINAL RESEARCH

Pharmacophore generation, atom-based 3D-QSAR, docking, and virtual screening studies of p38-α mitogen activated protein kinase inhibitors: pyridopyridazin-6-ones (part 2)

Sujit G Bhansali
Vithal M Kulkarni

Department of Pharmaceutical Chemistry, Poona College of Pharmacy, Bharati Vidyapeeth Deemed University, Pune, Maharashtra, India

Video abstract

Point your SmartPhone at the code above. If you have a QR code reader the video abstract will appear. Or use: http://dvpr.es/1bgsVnk

Correspondence: Vithal M Kulkarni
Department of Pharmaceutical Chemistry, Poona College of Pharmacy, Bharati Vidyapeeth Deemed University, Paud Road, Erandwane, Pune-411 038, Maharashtra, India
Tel +9198 9080 2623
Fax +9120 2543 9383
Email vmkulkarni60@gmail.com

Abstract: p38-α mitogen-activated protein kinase (MAPK) is a serine/threonine kinase activated by environmental stimuli, like stress, and by various proinflammatory cytokines, such as tumor necrosis factor-α and interleukin-1β. Excessive production of tumor necrosis factor-α and interleukin-1β may lead to various diseases, such as rheumatoid arthritis, psoriasis, and inflammatory bowel disease. Hence, inhibition of p38-α MAPK could be a novel approach for the development of new anti-inflammatory agents. In this study, a combination of pharmacophore generation, an atom-based three-dimensional quantitative structure-activity relationship (3D-QSAR), molecular docking, and virtual screening was performed for a series of pyridopyridazin-6-ones exhibiting p38-α MAPK inhibition activity. A five-point pharmacophore (AAAHR), ie, three hydrogen bond acceptors (AAA), one hydrophobic (H) group, and one aromatic ring (R) was obtained. A statistically significant 3D-QSAR model was obtained using this pharmacophore hypothesis with a good correlation coefficient (R^2=0.91) and a high Fisher ratio (F=90.3) for the training set of 47 compounds. The predictive power of the model generated was found to be significant, and was confirmed by the high value of the cross-validated correlation coefficient (q^2=0.80) and Pearson's R (0.90) for the test set of 16 compounds. Further, the docking study revealed the binding orientations of the active ligand on the amino acid residues Valine 30 (Val30), Glycine 31 (Gly31), Lysine 53 (Lys53), Leucine 75 (Leu75), Aspartic acid 88 (Asp88), Methionine 109 (Met109) of p38-α MAPK at the active site. The results of this ligand-based pharmacophore hypothesis and atom-based 3D-QSAR provide detailed structural insights and highlight the important binding features between pyridopyridazin-6-ones and p38-α MAPK. These findings may provide useful guidelines for rational design of compounds with better p38-α MAPK activity.

Keywords: pyridopyridazin-6-ones, p38-α mitogen-activated protein kinase, pharmacophore, three-dimensional quantitative structure activity relationship, docking, virtual screening

Introduction

p38-α mitogen-activated protein kinase (MAPK), also called cytokinin-specific binding protein, participates in a signaling cascade controlling cellular responses to cytokines and is activated by environmental stimuli such as stress, heat shock, lipopolysaccharides, ultraviolet light, growth factors, and inflammatory cytokines, including tumor necrosis factor-α and interleukin-1β.[1] Excessive production of these proinflammatory cytokines can lead to severe diseases, such as inflammatory bowel syndrome, Crohn's disease, psoriatic arthritis, and rheumatoid arthritis.[2–6] Four p38 MAPKs, ie, p38-α

(MAPK 14), p38-β (MAPK 11), p38-γ (MAPK 12), and p38-δ (MAPK 13), have been identified.[7,8] Several reports indicate that p38-α MAPK has an important role in rheumatoid arthritis and is involved in the expression of tumor necrosis factor-α and interleukin-1β at both the transcription and translation levels, while the roles of p38-β, p38-γ, and p38-δ MAPK are not clearly understood.

Although a number of structurally different p38-α MAPK inhibitors are reported,[9,10] with different degrees of selectivity, none have reached commercial status.[11] Several lead inhibitors, ie, SB203580, PH-797804, TAK-715, BIRB-796, and VX-745,[12–16] are undergoing clinical trials in inflammatory disorders.[17] Therefore, p38-α MAPK has become a target for development of novel anti-inflammatory agents. Efforts have continued to develop safe, potent, and active p38-α inhibitors, and one such class is the pyridopyridazin-6-ones, which are reported to be potent inhibitors of p38-α.[18,19]

A pharmacophore is an ensemble of steric and electronic features that is necessary to ensure optimal supramolecular interactions with a specific biological target and trigger (or block) its biological response.[20] A pharmacophore hypothesis collects common features distributed in three-dimensional (3D) space representing groups in a molecule that participate in important interactions between the drug and the active site.[21] To continue our research efforts in the development of a pharmacophore and a 3D quantitative structure-activity relationship (QSAR) for various therapeutic agents,[22] we report here studies on pharmacophore generation, atom-based 3D-QSAR model, docking, and virtual screening studies for a series of pyridopyridazin-6-ones using PHASE (pharmacophore alignment and scoring engine)[23] and ligand docking Glide (grid-based ligand docking with energetics)[24] incorporated in Schrodinger software (Portland, OR, USA).

The objective of the present work was to generate a ligand-based pharmacophore hypothesis and atom-based 3D-QSAR model to identify common features which may be responsible for the biological activity of pyridopyridazin-6-ones as potent p38-α MAPK inhibitors. Further, the binding of the active molecule with amino acid residues at the active site of p38-α MAPK was studied by docking. The ligand-based pharmacophore hypothesis and the cubes generated from the atom-based 3D-QSAR model highlight the important structural features required for p38-α MAPK inhibition which can be useful for design of potent p38-α inhibitors.

In addition, in silico screening of the ZINC "Clean druglike" database[25] was carried out by applying the Lipinski rule of five[26,27] and matching to the hypothesis, and the hits obtained were then subjected to virtual screening by docking

using three different docking parameters, ie, high throughput virtual screening, Glide SP, and Glide XP, to obtain the final hits potentially with potent p38-α MAPK inhibition activity.

Materials and methods
Dataset
In the present study, a set of 63 compounds was taken from the literature with their in vitro enzyme inhibitory data.[18,19] The dataset was divided randomly into a training set and a test set by considering 75% of the total molecules in the training set (47 compounds) and 25% in the test set (16 compounds). The training set was used to generate 3D-QSAR models and the test set was used to validate the quality of the model. All biological activities used in the present study were expressed as:

$$pIC_{50} = -\log_{10} IC_{50} \tag{1}$$

where IC_{50} is the nanomolar concentration of the inhibitor producing 50% inhibition. In all the models subsequently developed, pIC_{50} values were used as the dependent variable. Structures and related inhibitory activities (IC_{50} values) are reported in Table 1. Actual and predicted inhibitory activities (pIC_{50} values) and residual values are reported in Table 2.

Ligand preparation
All molecules were built in Maestro version 9.3 (Schrödinger, Portland, OR, USA) and prepared using LigPrep version 2.5 (Schrödinger)[28] to convert the two-dimensional structure to a 3D one, generate a stereoisomer, determine the most probable ionization state at user-defined pH, neutralize charged structures, add hydrogen, and generate the energy-minimized bioactive conformers using ConfGen (LigPrep version 2.5, Schrödinger) by applying optimized potentials for liquid simulations (OPLS)-2005 force field.[29] Conformational space was explored by combination of Monte-Carlo multiple minimum/low mode with the maximum number of conformers at 1,000 per structure and minimization steps of 10,000.[30,31] Each minimized conformer was filtered through a relative energy window of 50 kJ/mol and a redundancy check of 2 Å in the heavy atom positions.
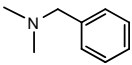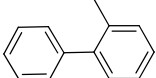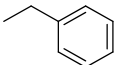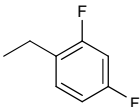
PHASE methodology
Pharmacophore modeling was carried out using PHASE implemented in Maestro 9.3.[32] PHASE also provides support for lead discovery, lead optimization, lead expansion, development of a structure-activity relationship, and generation of the 3D-QSAR model. The models generated can be used along with the hypothesis to mine a 3D database and obtain the hits that are most likely to have strong activity toward the target.

Table 1 Structures and inhibitory activity of p38-α mitogen-activated protein kinase of pyridopyridazin-6-one derivatives (1–63) used for training and test sets

Compound	X	p38α IC$_{50}$ (nm)	pIC$_{50}$
1		3.2	8.495
2*		720	6.143
3		4.7	8.328
4		10,000	5.000
5*		1,300	5.886
6*		160	6.796
7		10,000	5.000
8*		1,300	5.886
9		719	6.143
10		3.5	8.456
11		20	7.699

(*Continued*)

Table 1 (*Continued*)

Compound	X	p38α IC$_{50}$ (nm)	pIC$_{50}$
12		650	6.187
13		480	6.319
14		1,430	5.845
15		7.6	8.119
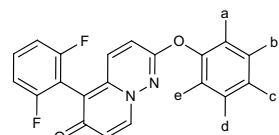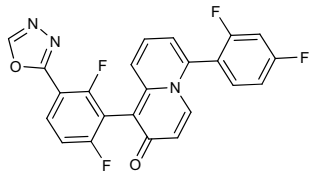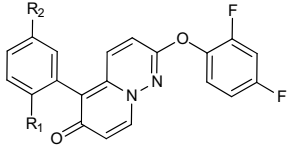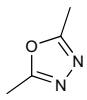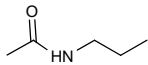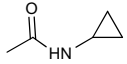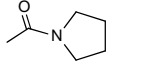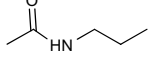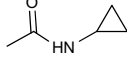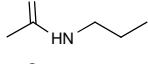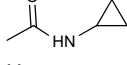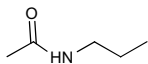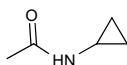
16–45

Compound	a	b	c	d	e	p38α IC$_{50}$ (nm)	pIC$_{50}$
16	F	–	–	–	–	12	7.921
17	Cl	–	–	–	–	44	7.357
18	CH$_3$	–	–	–	–	92	7.036
19*	OCH$_3$	–	–	–	–	82	7.086
20	CF$_3$	–	–	–	–	2,400	5.620
21	–	F	–	–	–	420	6.377
22	–	Cl	–	–	–	240	6.620
23	–	CH$_3$	–	–	–	70	7.155
24	–	OCH$_3$	–	–	–	10,000	5.000
25*	–	CF$_3$	–	–	–	10,000	5.000
26	–	–	F	–	–	48	7.319
27	–	–	Cl	–	–	79	7.102
28*	–	–	CH$_3$	–	–	260	6.585
29	–	–	OCH$_3$	–	–	10,000	5.000
30	–	–	CF$_3$	–	–	10,000	5.000
31	F	F	–	–	–	32	7.495
32	Cl	Cl	–	–	–	32	7.495
33	F	–	Cl	–	–	11	7.959
34	Cl	–	F	–	–	14	7.854
35	Cl	–	Cl	–	–	30	7.523
36	F	–	–	F	–	130	6.886
37*	Cl	–	–	Cl	–	280	6.553
38	F	–	–	–	F	95	7.022
39*	Cl	–	–	–	Cl	90	7.046
40	F	–	F	–	F	200	6.699
41*	Cl	–	F	–	Cl	68	7.167
42*	F	F	F	F	F	4,300	5.367
43	–	F	F	–	–	270	6.569
44	–	Cl	F	–	–	140	6.854
45	–	Cl	Cl	–	–	260	6.585
46						14	7.854
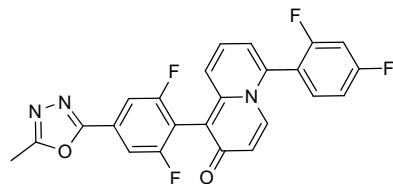
Table 1 (*Continued*)

Compound	a	b	c	d	e	p38α IC$_{50}$ (nm)	pIC$_{50}$
47						13	7.886

48–63

Compound	R$_1$	R$_2$	p38α IC$_{50}$ (nm)	pIC$_{50}$
48	F	H	17	7.770
49	F	COOCH$_3$	1.3	8.886
50*	F		2.3	8.638
51	F		32	7.495
52*	F		9	8.046
53*	F		99	7.004
54		COOCH$_3$	0.580	9.237
55*			3.3	8.481
56			2.3	8.638
57			5.8	8.237
58			9.9	8.004
59	OCH$_3$	H	8.8	8.056
60*	OCH$_3$	COOCH$_3$	0.580	9.237
61	OCH$_3$		1.6	8.796
62*	OCH$_3$		1.6	8.796
63	OCH$_3$		0.330	9.481

Note: *Test set compounds.
Abbreviation: IC$_{50}$, nanomolar concentration of the inhibitor producing 50% inhibition; pIC$_{50}$, the predicted concentration of the compound producing 50% inhibition.

Generation of common pharmacophore hypothesis

A common pharmacophore hypothesis is a spatial arrangement of chemical features common to two or more active ligands that is proposed to explain the key interactions involved in binding of a ligand with its receptor. PHASE provides a standard set of six pharmacophoric features, ie, a hydrogen bond acceptor (A), a hydrogen bond donor (D), a hydrophobic group (H), and a negatively ionizable (N), positively ionizable (P), and aromatic ring (R) to define the

Table 2 Actual activity, predicted activity, residual value, fitness score, and distribution of compounds in Pharm Set

Compound No	Actual activity (pIC$_{50}$)	Predicted activity	Residual value	Fitness	Pharm Set
1	8.495	8.820	−0.325	2.26	
2*	6.143	6.200	−0.057	2.40	
3	8.328	7.770	0.558	2.60	
4	5.000	5.320	−0.32	2.42	Inactive
5*	5.886	6.170	−0.284	2.41	
6*	6.796	6.790	0.006	2.59	
7	5.000	4.790	0.21	2.46	Inactive
8*	5.886	6.320	−0.434	2.15	
9	6.143	6.250	−0.107	2.12	
10	8.456	8.590	−0.134	2.27	
11	7.699	7.590	0.109	2.13	
12	6.187	6.300	−0.113	2.18	
13	6.319	6.420	−0.101	2.03	
14	5.845	6.620	−0.775	2.05	
15	8.119	7.610	0.509	2.01	
16	7.921	7.330	0.591	2.60	
17	7.357	7.190	0.167	2.60	
18	7.036	7.360	−0.324	2.58	
19*	7.086	6.790	0.296	2.56	
20	5.620	6.450	−0.83	2.56	
21	6.377	6.680	−0.303	2.58	
22	6.620	6.290	0.33	2.58	
23	7.155	6.770	0.385	2.58	
24	5.000	4.880	0.12	2.57	Inactive
25*	5.000	5.870	−0.87	2.52	Inactive
26	7.319	6.910	0.409	2.60	
27	7.102	6.940	0.162	2.60	
28*	6.585	6.040	0.545	2.37	
29	5.000	5.130	−0.13	2.59	Inactive
30	5.000	5.100	−0.1	2.33	Inactive
31	7.495	7.630	−0.135	2.01	
32	7.495	7.340	0.155	2.02	
33	7.959	7.470	0.489	2.61	
34	7.854	7.390	0.464	2.61	
35	7.523	7.290	0.233	2.61	
36	6.886	7.250	−0.364	2.60	
37*	6.553	7.180	−0.627	2.59	
38	7.022	6.980	0.042	2.37	
39*	7.046	7.480	−0.434	2.59	
40	6.699	7.340	−0.641	2.61	
41*	7.167	6.880	0.287	2.37	
42*	5.367	6.640	−1.273	2.36	Inactive
43	6.569	6.800	−0.231	2.59	
44	6.854	6.410	0.444	2.59	
45	6.585	6.420	0.165	2.59	
46	7.854	7.570	0.284	0.76	
47	7.886	7.680	0.206	0.72	
48	7.770	7.540	0.23	2.31	
49	8.886	9.140	−0.254	2.28	Active
50*	8.638	8.560	0.078	2.50	Active
51	7.495	7.510	−0.015	2.68	
52*	8.046	7.710	0.336	2.21	

(*Continued*)

Table 2 (*Continued*)

Compound No	Actual activity (pIC$_{50}$)	Predicted activity	Residual value	Fitness	Pharm Set
53*	7.004	7.420	−0.416	2.16	
54	9.237	9.460	−0.223	2.68	Active
55*	8.481	9.100	−0.619	2.63	
56	8.638	8.790	−0.152	2.23	Active
57	8.237	8.270	−0.033	2.70	
58	8.004	8.260	−0.256	2.26	
59	8.056	8.270	−0.214	2.90	
60*	9.237	8.790	0.447	3.00	Active
61	8.775	8.500	0.275	2.95	Active
62*	8.796	8.440	0.356	2.43	Active
63	9.481	9.480	0.001	2.78	Active

Note: *Test set compounds.
Abbreviation: IC$_{50}$, nanomolar concentration of the inhibitor producing 50% inhibition; pIC$_{50}$, the predicted concentration of the compound producing 50% inhibition; No, number.

chemical features of ligands. The Pharm Set column indicates whether a molecule is in the set of actives used to identify common pharmacophore hypotheses, in the set of inactives used to eliminate nondiscriminatory hypotheses, or in neither set. The pIC$_{50}$ ranged from 9.481 to 5.000, since only the active compounds are normally considered when developing a common pharmacophore hypothesis. A Pharm Set column was defined by setting a threshold for actives of pIC$_{50}$ >8.5 and a threshold for inactives of pIC$_{50}$ <6.5.

A five-point common pharmacophore hypothesis was identified from all the conformations of the active ligands having an identical set of features with very similar spatial arrangement and keeping a minimum intersite distance of 2.0 Å. Hypotheses were generated by systematic variation of number of sites and number of matching active compounds. The common pharmacophore hypothesis was considered, which indicated at least five sites common to all molecules.

Scoring pharmacophore hypothesis

The scoring procedure helps in ranking of the different hypotheses to yield the best alignment of the active ligands using an overall maximum root mean square deviation value of 1.2 Å with default options for distance tolerance. Thus, it helps in making a rational choice regarding which hypothesis is more appropriate for further investigation. The quality of alignment was measured by a survival score,[33,34] defined as:

$$S = W_{site}S_{site} + W_{vec}S_{vec} + W_{vol}S_{vol} + W_{sel}S_{sel}$$
$$+ W^m_{rew} - W_E\Delta E + W_{act}A \quad (2)$$

All the details regarding equation 2 and calculation of the survival score are provided in the supplementary data.

The hypotheses generated were scored and ranked to find out the best possible hypothesis. The best common pharmacophore hypothesis was selected depending on the adjusted survival score until one hypothesis was found and scored successfully.

Building QSAR models

A pharmacophore-based QSAR does not consider ligand features beyond the pharmacophore model, such as possible steric clashes with the receptor. This requires consideration of the entire molecular structure, so an atom-based QSAR model is more useful in explaining structure-activity relationships. In atom-based QSAR, a molecule is treated as a set of overlapping van der Waals spheres. Each feature is grouped according to a simple set of rules: hydrogens attached to polar atoms are classified as hydrogen bond donors (D); carbons, halogens, and C–H hydrogen are classified as hydrophobic/nonpolar (H); atoms with an explicit negative ionic charge are classified as negative ionic (N); atoms with an explicit positive ionic charge are classified as positive ionic (P); nonionic atoms are classified as electron-withdrawing (W); and all other types of atoms are classified as miscellaneous (X).

Validation of the pharmacophore model

The main purpose of developing the QSAR model was to predict biological activities of new compounds whereby the generated model would be statistically robust, both internally and externally. The dataset was divided into a training set and a test set. Atom-based 3D-QSAR models were generated for hypotheses using the 47 compounds in the training set. The best QSAR model was externally validated by predicting the activities of the 16 test set compounds.

The robustness of the developed pharmacophore hypotheses was internally validated by statistical parameters, including the squared correlation coefficient (R^2), q^2 (R^2 for test set), the standard deviation of regression, Pearson's correlation coefficient (Pearson's R), statistical significance (P), and variance ratio (F). The predicted pIC_{50} at the 5th partial least-squares (PLS) factor is shown in Table 3. A further increase in the number of PLS factors did not improve the statistics or predictive ability of the model.[33,34]

Docking

Glide[24] was used to perform the docking study. The crystal structure of p38-α was obtained from Protein Data Bank (pdb: 3FLY). The protein was refined using the protein preparation wizard. The missing residues of the side chain were added using the prime interface module incorporated in Maestro. All the water molecules were deleted. H atoms were added to the protein, including the protons necessary to define the correct ionization and tautomeric states of the amino acid residues. Each structure minimization was carried out with the impact refinement module, using the OPLS-2005 force field to alleviate steric clashes potentially existing in the structures. Minimization was terminated when the energy converged or the root mean square deviation reached a maximum cutoff of 0.30 Å. After preparation of the protein, a receptor grid was generated using the receptor grid generation panel. The ligand was selected to define the position and size of the active site.[35,36]

Virtual screening

The ZINC "Clean drug-like" database[25] containing 15,798,630 compounds with drug-like properties was used. In the ZINC database, several constraints were applied, including a maximum of 0.7 root mean square deviation, obeying 10 rotatable bond cutoffs, with a molecular weight range of 180–500 Da. Initially, the AAAHR hypothesis was used for screening of the ZINC database with drug-like properties to identify hits matching the pharmacophoric features of the hypothesis using the search for matches

Table 3 Summary of atom-based three-dimensional quantitative structure activity relationship results

Compound No	Hypothesis	Survival score	Survival inactive score	Adjusted survival score	SD	R^2	F	P	q^2	Stability	RMSE	Pearson's R
1	AHRRR.1040	4.579	2.33	2.249	0.3057	0.9396	127.6	7.082E-024	0.6889	0.3550	0.6700	0.8562
2	AAHRR.1113	4.519	2.139	2.38	0.2927	0.9446	139.9	1.206E-024	0.5656	0.5186	0.7917	0.8408
3	AHRRR.1135	4.483	2.309	2.174	0.3123	0.937	122	1.685E-023	0.6378	0.3011	0.7229	0.8459
4	AAHRR.361	4.353	2.293	2.06	0.332	0.9288	107	2.043E-022	0.5218	0.3104	0.8306	0.8481
5	AAAHR.71	4.338	1.873	2.465	0.3591	0.9167	90.3	5.00E-021	0.8047	0.4082	0.5308	0.9080
6	AAHRR.2379	4.311	2.137	2.174	0.3878	0.9028	76.2	1.154E-019	0.6839	0.5013	0.6754	0.8554

Notes: R^2, correlation coefficient; F, variance ratio; P, significance level of variance ratio; q^2, for the predicted activities; Pearson's R, correlation between the predicted and observed activity for the test set.
Abbreviations: RMSE, root-mean-square error; SD, standard deviation of the regression; A, hydrogen bond acceptor; H, hydrophobic group; R, aromatic ring; No, number.

option in PHASE.[37,38] The virtual screening workflow in Maestro was used to dock and score the drug-like compounds. In the first step, the high-throughput virtual screening mode of Glide was used and the remaining 10% of the top-scoring ligands were further subjected to Glide SP docking. Again, 10% of the top-scoring leads from Glide SP were retained and all the ligands were subjected further to Glide XP docking. Only hits with a docking score less than −6.4 were retained. During the docking process, the docking score was used to select the best conformation for each ligand. The accuracy of the model was validated by the Güner-Henry scoring method using known actives and a decoy set.[39]

Güner-Henry scoring method
The Güner-Henry score (GH) has been used successfully to quantify the accuracy of hits and recall of actives from a dataset consisting of known actives and inactives. Güner-Henry analysis was done by computing the following variables: a) %A is the percentage of known active compounds retrieved from the database (precision); b) H_a is the number of actives in the hit list (true positives); c) A is the number of active compounds in the database; d) %Y, the percentage of known actives in the hit list (recall); e) H_t is the number of hits retrieved, f) D is the number of compounds in the database, g) E is the enrichment of the concentration of actives by the model relative to random screening without any pharmacophoric approach, and h) is the Güner-Henry score.[38]

$$\%A = H_a/A * 100 \tag{3}$$

$$\%Y = H_a/H_t * 100 \tag{4}$$

$$E = (H_a/H_t)/A/D \tag{5}$$

$$GH = H_a(3A + H_t)/4H_tA * (1 - H_t - H_a/D - A) \tag{6}$$

Applicability domain
The test of applicability domain was performed using SIMCA-P 12.0 demo version (UMETRICS, Umeå, Sweden)[40] to explore the activity predictions of the test set compounds. The QSAR applicability domain is the structural or biological space, physicochemical knowledge, or information based on the training set of the model and for which it is applicable to make predictions for new compounds. This can be attained by selecting different features and application of principal component analysis. Reliable prediction of a compound is unlikely when it is highly dissimilar to all compounds of the modeling set. To

avoid such unjustified exploration of activity predictions, we used the concept of applicability domain.[41]

Drug likeness analysis
The selected hits were further screened for absorption, distribution, metabolism, and excretion (ADME) using QikProp version 3.5 (Schrödinger),[42] which is studied to optimize the physiochemical properties in the early stages of drug design and to reduce the ADME problem later during the drug development process. The drug-likeness of the lead molecules was assessed for physicochemical properties by applying Lipinski's rule of five, and was found to be within the acceptable range for drug-like molecules.

Results and discussion
The PHASE module of the Schrodinger suite is used for pharmacophore modeling and QSAR studies. The main purpose of this study was to generate 3D pharmacophoric features of the pyridopyridazin-6-ones important for binding to the target and to predict the biological activity of p38-α MAPK inhibitors by generating an atom-based 3D-QSAR model. The model generated was further used for virtual screening to obtain hits which may have the p38-α MAPK inhibition activity predicted by the generated model. Four hits which all required features of the hypothesis were obtained by applying various constraints, including the Lipinski rule of five, the match hypothesis, and docking parameters, such as high-throughput virtual screening, Glide SP, and Glide XP docking score.

Pharmacophore generation and 3D-QSAR model analysis
Common pharmacophore hypotheses were generated from a set of eight active ligands in the Pharm Set because they contain important structural features crucial for binding at the receptor binding site. In find pharmacophore step we used four minimum sites and five maximum sites to generate an optimum combination of features common to the most active compounds. The three and four featured pharmacophore hypotheses were not selected since they have the low value of survival score because they were unable to define the complete binding space of the training set molecules used to develop the pharmacophore hypotheses. Five featured pharmacophore hypotheses were selected and subjected to stringent scoring function analysis.

In total, 115 common pharmacophore hypotheses were generated using different combinations of variants, ie, AAHRR, AHHRR, AHHHR, AHRRR, AAAHR, and

AAHHR, in which a few hypotheses were further considered for QSAR model generation and the results are given in Table 3. Among the various pharmacophores generated, the hypothesis showing the best alignment with the active compounds was identified by aligning hypothesis with active compounds and calculating the survival score. The survival score function helps in identifying the best hypothesis and for ranking all the hypotheses. The quality of each alignment is measured in three ways: the alignment score, which is the root-mean-squared deviation in the site-point positions; the vector score, which is the average cosine of the angles formed by corresponding pairs of vector features (acceptors, donors, and aromatic rings) in the aligned structures; and a volume score based on overlap of van der Waals models of nonhydrogen atoms in each pair of structures. The selected hypothesis should distinguish between the active and inactive molecules.

Further, to confirm that the pharmacophore hypotheses map well with more active than inactive features, they were aligned to inactive compounds and scored. If inactive ligands score well, the hypothesis may be considered to be poor and should be rejected since it does not distinguish between active and inactive ligands. Therefore, the adjusted survival score was calculated by subtracting the inactive score from the survival score of this hypothesis (Table 3). Finally, the hypothesis with the maximum adjusted survival score and lowest relative conformational energy was selected for generating the atom-based 3D-QSAR model. The top model with good predictive power was found to be associated with the five-point hypotheses which consist of three hydrogen bond acceptor (AAA), one hydrophobic (H) group, and one ring feature (R). This is denoted as AAAHR. The pharmacophore hypothesis showing the distance between pharmacophoric sites is depicted in Figure 1A. As shown in Figure 1B, among the three hydrogen bond acceptor groups (AAA), one feature is mapped to "N" of the pyridopyridazin-6-one ring, a second is observed at "O" attached to the pyridopyridazin-6-one ring, and a third is observed on C=O of the pyridopyridazin-6-one ring. One hydrophobic group (H) is mapped to the $-OCH_3$ group and one aromatic ring (R) is present on the benzene ring attached to the pyridopyridazin-6-one ring.

The alignment generated by the best pharmacophore hypothesis, AAAHR, is used for generation of the 3D-QSAR model. From Figure 1C and 1D it can be observed that active ligands have better alignment than inactive ligands.

An atom-based 3D-QSAR model was generated with the 5th partial least-squares factor having good statistical significance and predictive power. The partial least-squares factor was increased up to five since up to the 5th factor there is an incremental increase in predictive power and statistical value of the model.

A statistically significant 3D-QSAR model was obtained using this pharmacophore hypothesis with a good correlation coefficient (R^2=0.91) and a high Fisher ratio (F=90.3) for the training set of 47 compounds. Also, the predictive power of the generated model was found to be significant, which was confirmed by the high value of the cross-validated correlation coefficient (q^2=0.80) and Pearson's R (0.90) for the test set of 16 compounds. The large value of F (90.3) indicates a statistically significant regression model, which is also supported by the small value of the significance level of variance ratio (P), indicative of a high degree of confidence. A summary of the atom-based 3D-QSAR results is shown in Table 3. Graphs of observed versus predicted biological activity for the training and test set molecules are shown in Figures 2 and 3, respectively.

In order to visualize the generated 3D-QSAR model and to study its correlation with inhibitory activity, one or more ligands from the series having diverse inhibitory activity were taken into consideration. For better understanding, compound 63 is divided into three rings (A–C) as shown in Figure 4. Cubes were generated to represent the important structural features required for interaction of the ligand with the active site of the receptor, as shown in Figures 5 and 6D, respectively, for the most active and least active compounds in the series. In these representations, the blue cubes indicate favorable regions while red cubes indicate unfavorable regions for biological activity. Using the 3D-QSAR model, a comparison was done to study favorable and unfavorable interaction using the most active (compound 63) and least active (compound 29), as shown in Figures 5 and 6D, respectively.

Figures 5 and 6D were compared for the most significant favorable and unfavorable electron-withdrawing features at positions 2 and 4 of ring "A". In Figure 5, blue cubes are observed at positions 2 and 4, while in Figure 6D red cubes are observed at these positions. From this observations it is clear that the presence of electron-withdrawing groups at the 2 and 4 positions is favorable for activity. Thus, compounds having a fluoro group at these positions are more active than those lacking a fluoro group or compounds having only one fluoro substitution at these positions or having a fluoro substituent at other positions. These
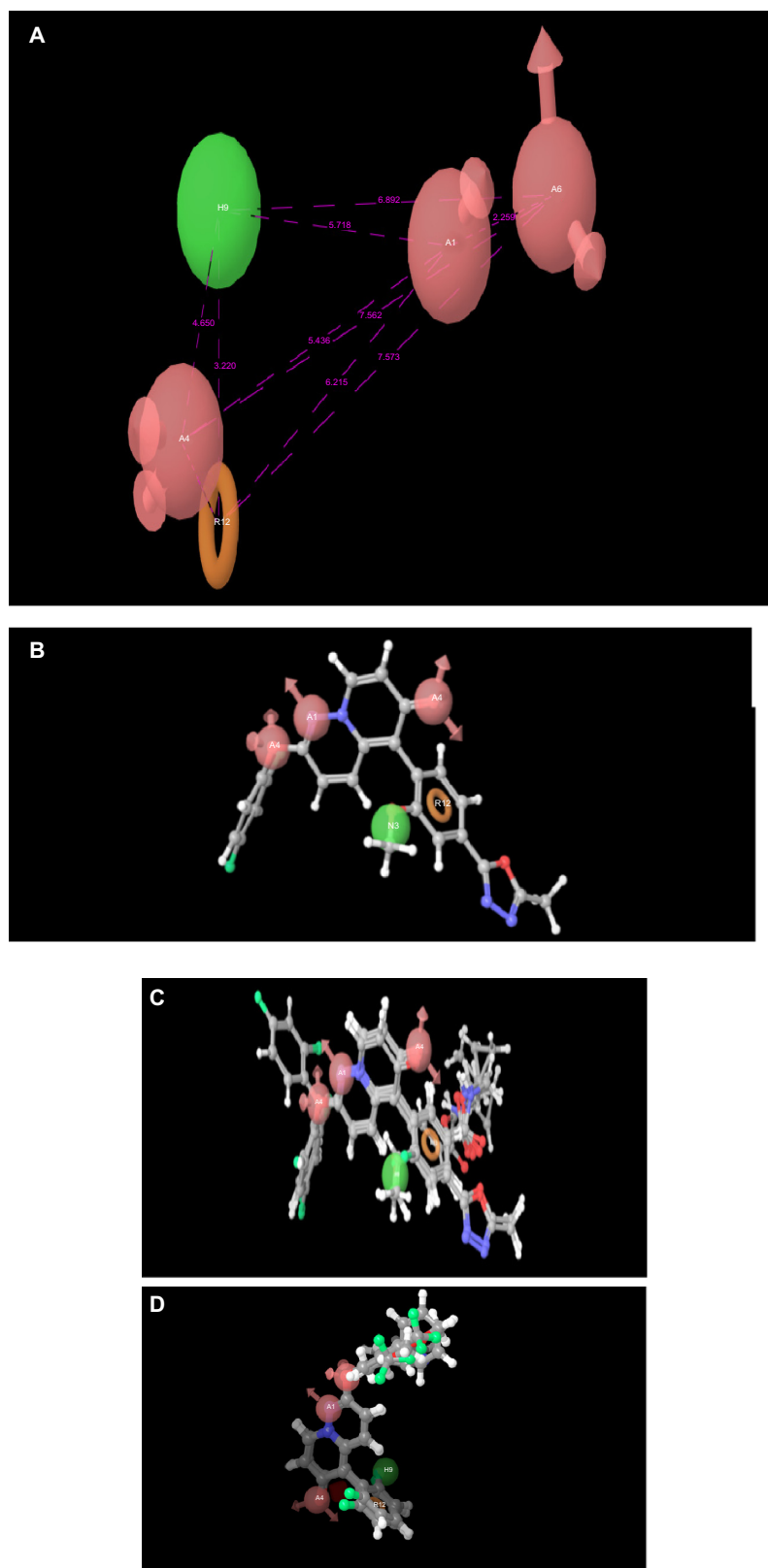
Figure 1 (**A**) AAAHR.71 pharmacophore hypothesis and distance between pharmacophoric sites. All distances are in Å units. (**B**) Common pharmacophore hypothesis aligned with most active ligand 63 (three hydrogen bond acceptor groups, one hydrophobic group, and one aromatic ring). (**C** and **D**) Common pharmacophore hypothesis aligned with all active ligands.
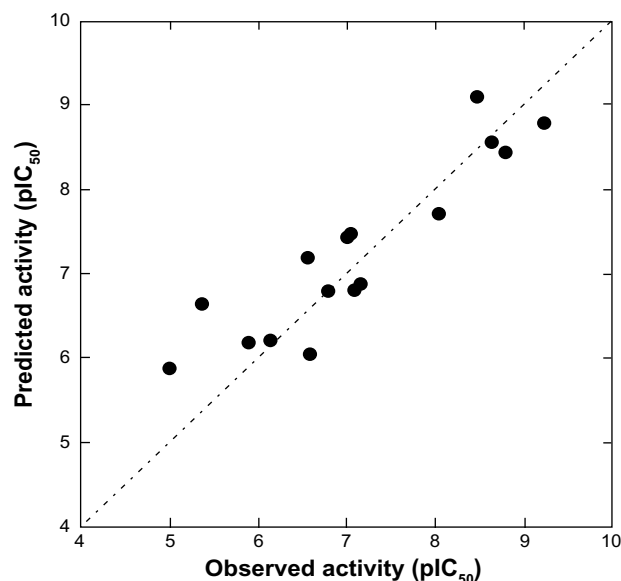**Notes:** The pharmacophoric feature **A**: acceptor appear as light red spheres centered on the atom with the lone pair, the arrows are pointing in the direction of the lone pairs, hydrophobic appear as a green sphere and the aromatic ring appears as an orange torus in the plane of the ring. The standard color scheme for the atoms in **B**, **C** and **D** is as follows: gray color for carbon atoms, blue color for nitrogen atoms, red color for oxygen atoms, white color for hydrogen atoms and light green color for fluoro atoms.
Abbreviations: A, hydrogen bond acceptor; H, hydrophobic group; R, aromatic ring.
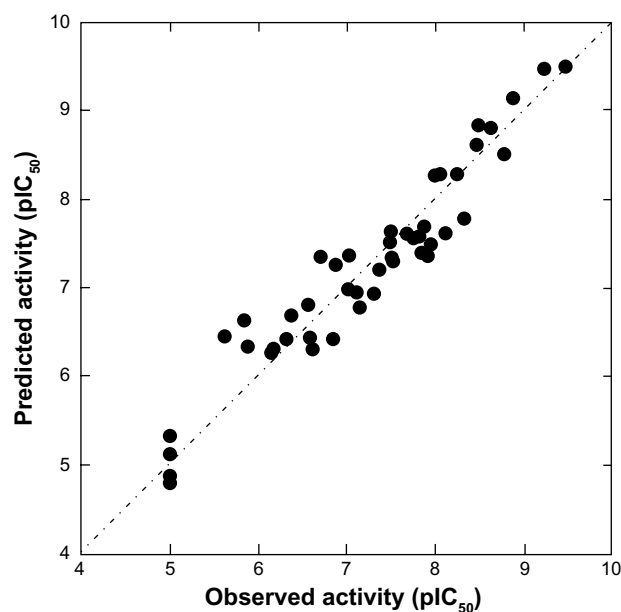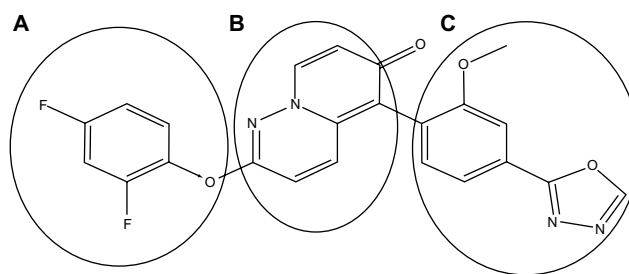
Figure 2 Graph of observed versus predicted biological activity of training set.
Abbreviation: pIC$_{50}$, the predicted concentration of the compound producing 50% inhibition.

results are supported by the higher activity of compounds having two electron-withdrawing groups at positions 2 and 4 (compounds 1, 3, 33–35, 48–63), moderate activity for compounds having one electron withdrawing group at any one of these positions (compounds 16–18, 26, 27, 31, 38), and the least active compounds which lack electron-withdrawing groups at these positions (compounds 4–5, 7–8, 14, 24–25, 29–30). A comparison of the atom-based 3D-QSAR model visualized in the context of compounds
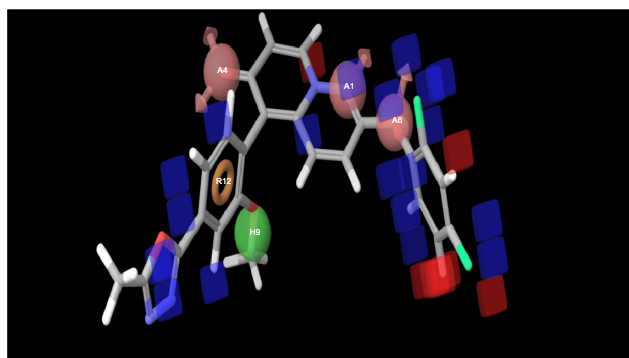

Figure 4 Compound 63 divided into (**A–C**) regions.

26, 27, 28, 29, and 30 is given in Figure 6A–E, respectively, which helps to distinguish between active and inactive compounds.
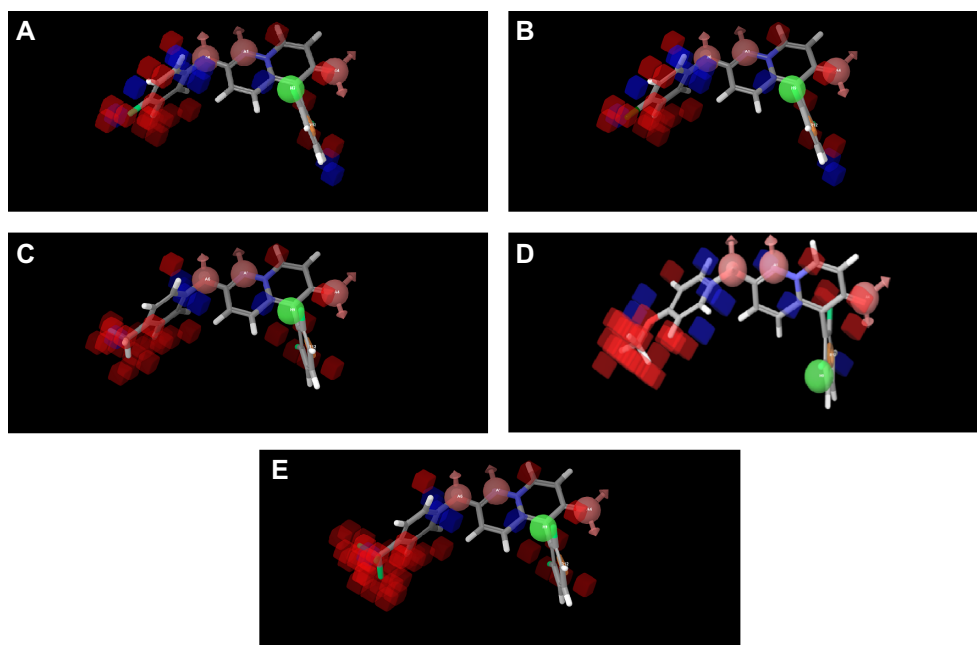
Figures 5 and 6D were again compared for the most significant favorable and unfavorable hydrogen bond acceptor group in the C ring. In Figure 5, it is observed that the blue cubes are mapped with the nitrogen of oxazole in the C ring, while as seen in Figure 6D it lacks this hydrogen bond acceptor group. Instead, it is replaced by a bulky and electron-withdrawing group in this region, which leads to a tremendous fall in activity; hence it is found to be the least potent compound. From the above information it is clear that the hydrogen bond acceptor group in C ring is essential for activity. This can be explained by analyzing the trend in the biological data for compounds having a hydrogen bond acceptor group in the C ring which have higher activity (compounds 48–63) and less potent compounds which lack the hydrogen acceptor group in the C ring (compounds 5, 7–8, 20, 24–25, 29–30).

Figure 3 Graph of observed versus predicted biological activity of test set.
Abbreviation: pIC$_{50}$, the predicted concentration of the compound producing 50% inhibition.

Figure 5 Atom-based three-dimensional quantitative structure-activity relationship model visualized in the context of the most active compound, number 63.
Notes: Blue cubes indicate favorable regions while red cubes indicate unfavorable regions for the activity. The pharmacophoric feature: acceptor appear as light red spheres centered on the atom with the lone pair, the arrows are pointing in the direction of the lone pairs, hydrophobic appear as a green sphere and the aromatic ring appears as an orange torus in the plane of the ring. The standard color scheme for the atoms is as follows: gray color for carbon atoms, blue color for nitrogen atoms, red color for oxygen atoms, white color for hydrogen atoms and light green color for fluoro atoms.
Abbreviations: A, hydrogen bond acceptor; H, hydrophobic group; R, aromatic ring.

Figure 6 Comparison of atom-based three-dimensional quantitative structure-activity relationship model visualized in the context of the compounds. (**A**) Compound 26, (**B**) compound 27, (**C**) compound 28, (**D**) compound 29, and (**E**) compound 30.
**Notes:** Blue cubes indicate favorable regions while red cubes indicate unfavorable regions for the activity. The pharmacophoric feature: acceptor appear as light red spheres centered on the atom with the lone pair, the arrows are pointing in the direction of the lone pairs, hydrophobic appear as a green sphere and the aromatic ring appears as an orange torus in the plane of the ring. The standard color scheme for the atoms is as follows: gray color for carbon atoms, blue color for nitrogen atoms, red color for oxygen atoms, white color for hydrogen atoms and light green color for fluoro atoms.
Abbreviations: A, hydrogen bond acceptor; H, hydrophobic group; R, aromatic ring.

Docking studies

Structure-based docking studies were carried out to investigate the intermolecular interaction between the ligand and the targeted enzyme using Glide (Glide 5.8, Schrodinger, 2012). Docking was carried out to study the binding mode of the active compound 63 on p38-α MAPK and to obtain information for further structure optimization. Grid

generation for defining the binding site on the receptor was done using the receptor grid generation panel with the default settings. Glide XP docking was used for docking purposes. As seen in Figures 7 and 8, docking analysis of compound 63 at the active site of p38-α MAPK shows interactions and its docking score was −6.989. The C=O of the B ring and the oxygen present outside the B ring of compound 63 interact
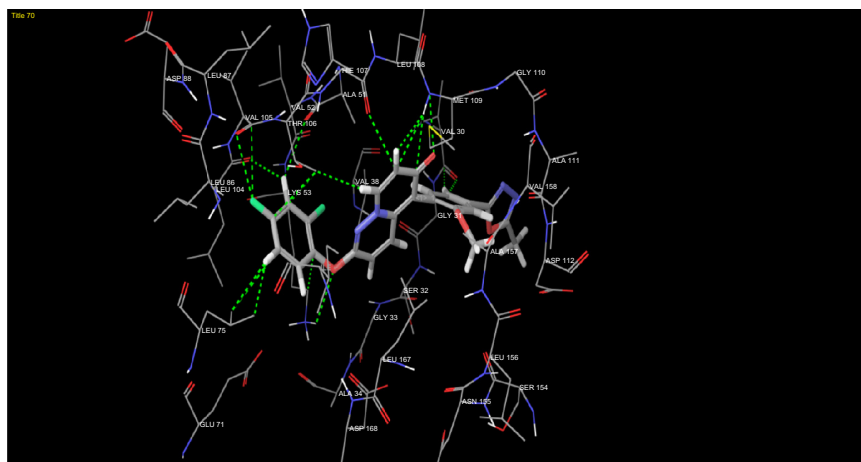
Figure 7 Docking of compound 63 in the active site of p38-α mitogen-activated protein kinase.
Notes: The standard color scheme for the atoms is as follows: gray color for carbon atoms, blue color for nitrogen atoms, red color for oxygen atoms, white color for hydrogen atoms and light green color for fluoro atoms. H-bonds are displayed as dotted green lines.
Abbreviations: VAL, valine; GLY, glycine; LEU, leucine; ASP, aspartate; SER, serine; ALA, alanine; LYS, lysine; ILE, isoleucine; MET, methionine; THR, threonine; GLU, glutamate; ASN, asparagine; HIE, histidine epsilon H.
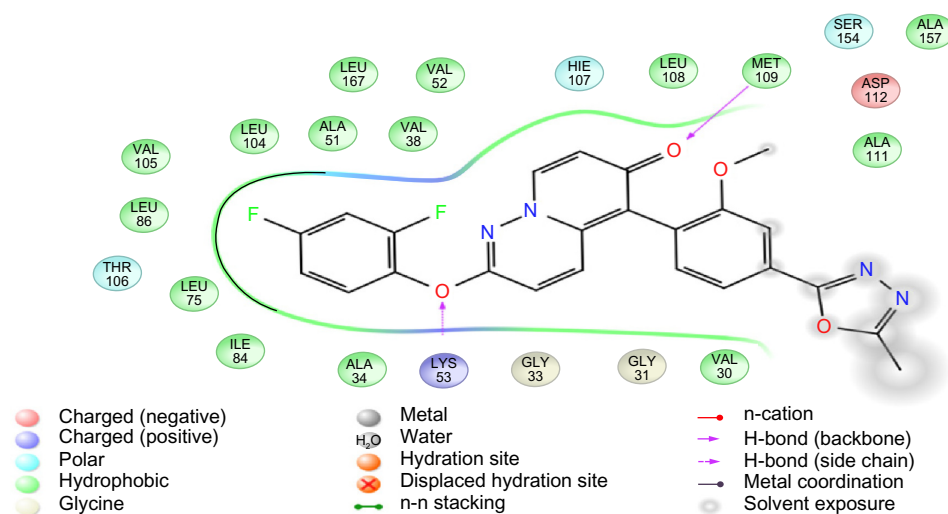
Figure 8 Two-dimensional view of the binding interaction of the most active compound, number 63, with active site of p38-α mitogen-activated protein kinase.
Abbreviations: VAL, valine; GLY, glycine; LEU, leucine; ASP, aspartate; SER, serine; ALA, alanine; LYS, lysine; ILE, isoleucine; HIE, histidine epsilon H; MET, methionine; THR, threonine.

via hydrogen bonding with amino acid residues Methionine 109 (Met109) and Lysine 53 (Lys53) within the active site. Both the A ring and B ring form good interactions with other amino acid residues present in both the phosphate-binding region and the hydrophobic region of Valine 30 (Val30), Glycine 31 (Gly31), Alanine 51 (Ala51), Leucine 75 (Leu75), Aspartic acid 88 (Asp88), Valine 105 (Val105), and Threonine

106 (Thr106), which play a crucial role for p38-α MAPK inhibition. Figure 9 clearly shows that the inactive compound 29 lacks the key interaction of binding with the amino acids Gly31, Val30, Leu75, and Asp88, and has a docking score of −5.412. Instead, compound 29 has H-bonding with Alanine 34 (Ala34), which leads to a decrease in activity, which supports the hypothesis of the generated model.
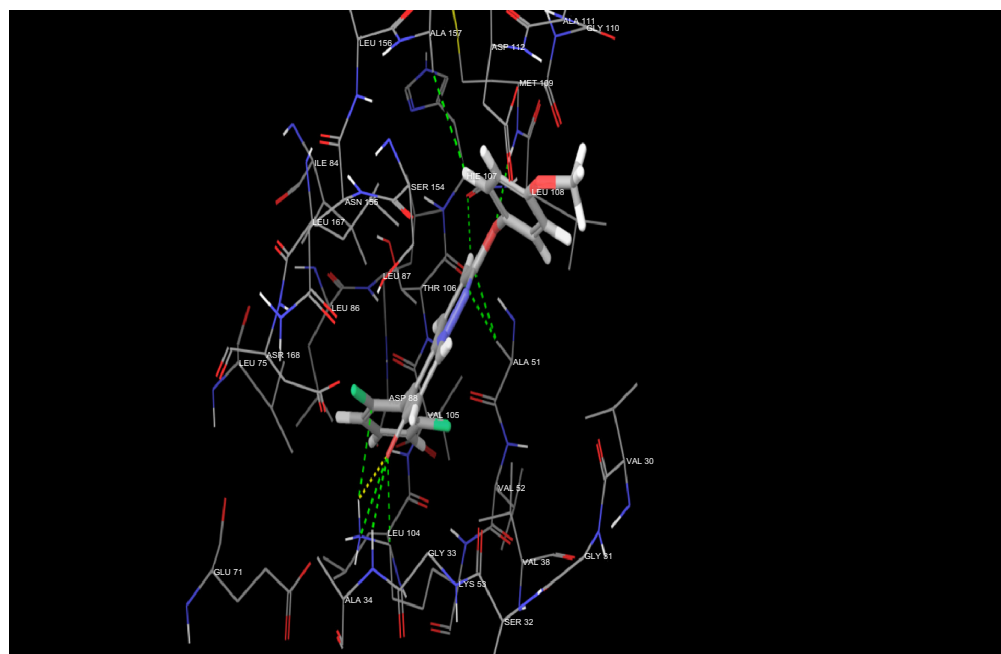
Figure 9 Docking of compound 29 in the active site of p38-α mitogen-activated protein kinase.
Notes: The standard color scheme for the atoms is as follows: gray color for carbon atoms, blue color for nitrogen atoms, red color for oxygen atoms, white color for hydrogen atoms and light green color for fluoro atoms. H-bonds are displayed as dotted green lines.
Abbreviations: VAL, valine; GLY, glycine; LEU, leucine; ASP, aspartate; SER, serine; ALA, alanine; LYS, lysine; ILE, isoleucine; HIE, histidine epsilon H; MET, methionine; GLU, glutamate; ASN, asparagine.
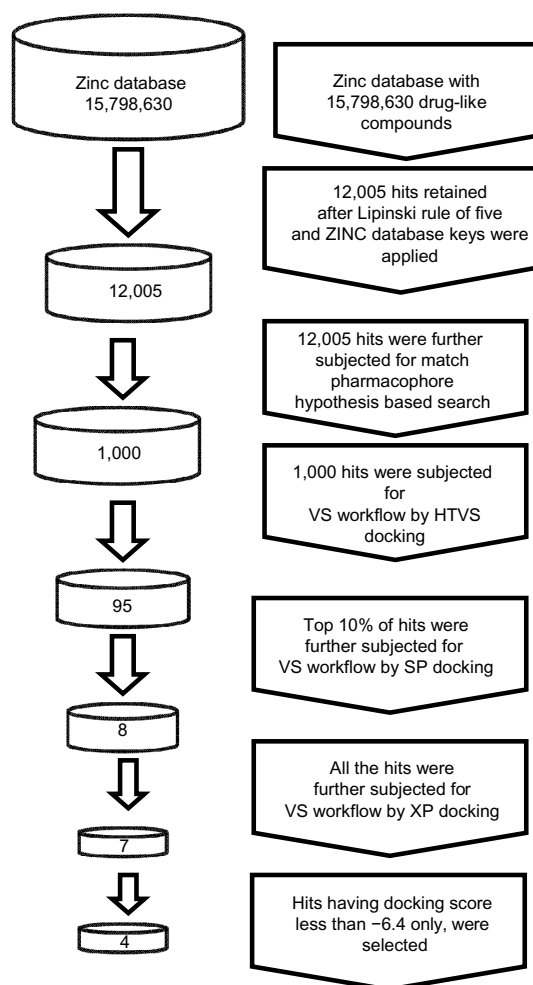
Figure 10 Workflow of the methodology used for virtual screening of ZINC database to get a novel p38-α mitogen-activated protein kinase inhibitor.
Abbreviations: VS, virtual screening; HTVS, high-throughput virtual screening mode; SP, standard-precision mode; XP, extra-precision mode.

Validation of docking procedure

To validate our docking procedure, we eliminated the cocrystallized ligand from the active site and redocked within the inhibitor binding cavity of p38-α MAPK. In this study, the root mean square deviation value was found to be below 2 Å, showing that our docking method is valid for the inhibitors studied.

In silico screening and ADME studies

Initially, the ZINC "Clean drug-like" database of 15,798,630 compounds having drug-like properties was subjected to prefiltering by using ZINC database keys to eliminate the majority of molecules not matching the Lipinski rule of five and a default value for the feature matching tolerance. The applied filter gave 12,005 hits with the desired pharmacophore features. The compounds in the database were included by using the search criteria of log $P \leq 5.0$, molecular weight in the range of 150–500, H-bond donors ≤ 5, and H-bond acceptors ≤ 10. All the compounds in the database were optimized with Ligprep using OPLS 2005 force field along with generation of possible tautomers and ionization states at physiological pH. The 12,005 hits obtained in the previous step were further screened for a match to hypothesis (AAAHR) which gave 1,000 hits. It was confirmed from the results that very few inactive compounds from the decoy set were picked up by the hypothesis and thus the selected hypothesis was found to be suitable and have good predictive ability.

ZINC00646908

ZINC09110317

ZINC00645569

ZINC08383847

Figure 11 Structures of the final hits obtained after virtual screening.

Table 4 Pharmacophore model evaluation based on Güner-Henry scoring method

Model	%A	%Y	E	GH	D	A	H$_t$	H$_a$
AAAHR.71	75.18	82.4	6.843	0.788	1137	137	125	103

Notes: H$_t$ is number of hits retrieved; H$_a$ is number of actives in hit list; D is number of compounds in a database; %A is a ratio of actives retrieved in hit list; %Y is fraction of hits relative to size of database (hit rate or selectivity); E is enrichment of active bin by model relative to random screening; A is the number of active compounds in the database.
Abbreviation: GH, Güner-Henry score.

These 1,000 hits were next subjected to the virtual screening workflow to dock and score the drug-like compounds. In the first step, the high-throughput virtual screening mode of Glide was used and 10% of the top-scoring ligands (95 hits) were used for the next step,

Glide SP. Again, 10% of the top-scoring leads from Glide SP were retained (eight hits) and were further docked with Glide XP. Only hits with the best scoring state were retained (seven hits). Finally, four hits with a docking score <-6.4 were selected. All the virtual screening workflow is represented in Figure 10. Hits obtained after the virtual screening and docking studies have a good docking score and a good predictive and fitness value, and can be further modified using QSAR model to get a potent p38-α MAPK inhibitor. The diversity of the hits obtained demonstrates that the pharmacophore model was able to retrieve hits with features similar to the existing p38-α MAPK inhibitors as well as novel scaffolds (Figure 11). Further, the accuracy of the selected model was validated
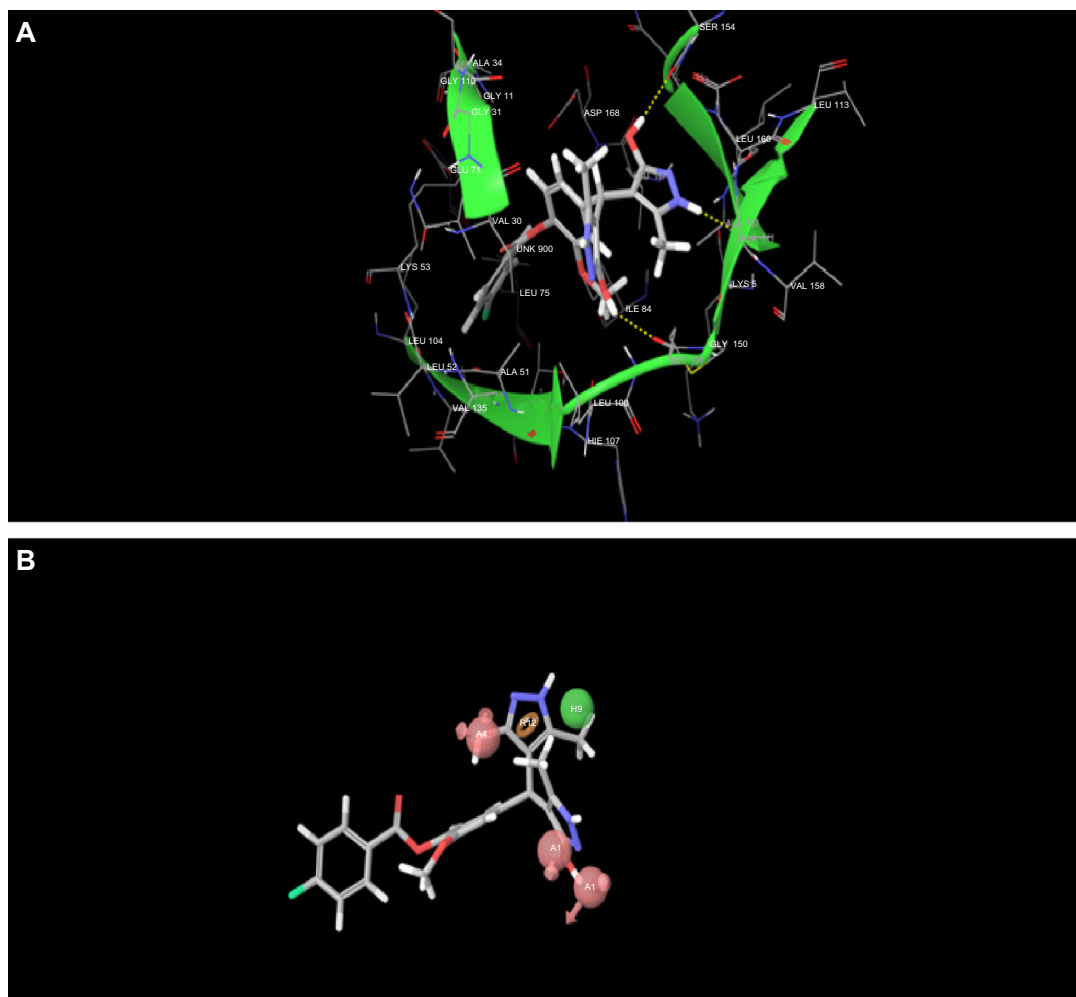
Figure 12 (**A**) Docking of ZINC08383847 in the active site of p38-α mitogen-activated protein kinase. (**B**) Common pharmacophore hypothesis aligned with ZINC08383847 (three hydrogen bond acceptor groups, one hydrophobic group, and one aromatic ring).
Notes: The pharmacophoric feature: acceptor appear as light red spheres centered on the atom with the lone pair, the arrows are pointing in the direction of the lone pairs, hydrophobic appear as a green sphere and the aromatic ring appears as an orange torus in the plane of the ring. The standard color scheme for the atoms is as follows: gray color for carbon atoms, blue color for nitrogen atoms, red color for oxygen atoms, white color for hydrogen atoms and light green color for fluoro atoms. H-bonds are displayed as dotted green lines.
Abbreviations: A, hydrogen bond acceptor; H, hydrophobic group; R, aromatic ring; VAL, valine; GLY, glycine; LEU, leucine; ASP, aspartate; SER, serine; ALA, alanine; LYS, lysine; ILE, isoleucine; HIE, histidine epsilon H; GLU, glutamate; UNK, field contains the number of amino acids in the chain for which the amino acid type is unknown.
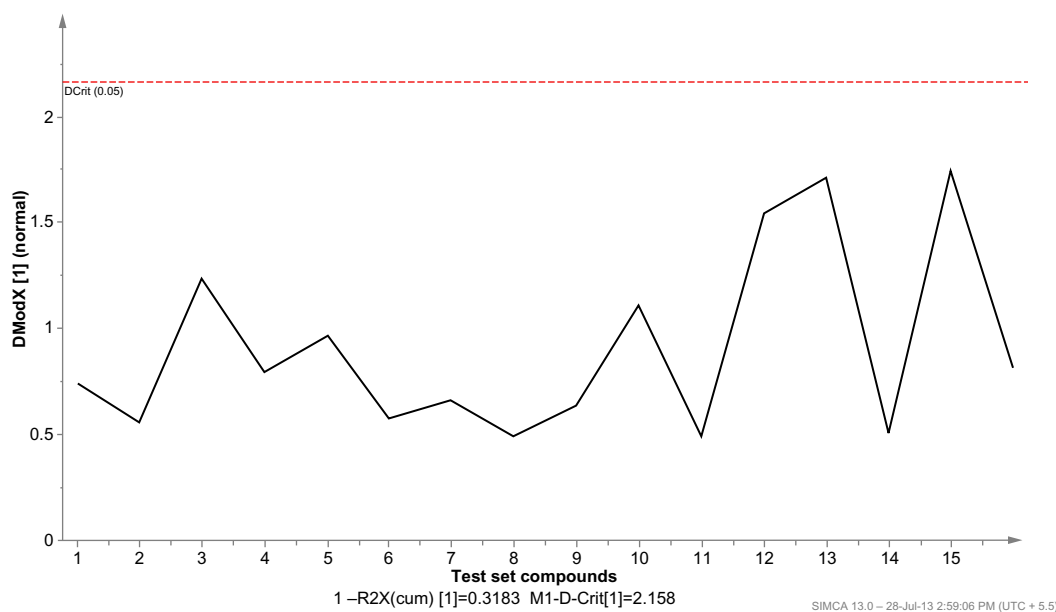
Figure 13 Residual standard deviation of X-residuals (DModX) of test set compounds for selected model.
Notes: In the plots the following standard deviations and targets are displayed: DModX = residual standard deviation; R2X[1] = fraction of sum of squares for the first component; Mxx-D-Crit[last component] = critical distance in the DModX plot; 1 − R2X(cum)[last component] = 1 − cumulative fraction of sum of squares up to the last component; DCrit = critical limit. The red DCrit line means points above this line have larger DModX than the critical limit which indicates that the observation is an outlier in the X space.

by assessing their ability to capture p38-α MAPK inhibitors selectively from a list of decoys (molecules which are presumably inactive against the examined target). The decoy database (1,000 compounds) of Schrodinger was mixed with the known 25 inhibitors of p38-α MAPK inhibitors from Stahl's data[39] set with their 133 conformers. The four hits obtained from virtual screening were mixed and Güner-Henry score was successfully applied to quantify the accuracy of the hits and recall of the actives from a dataset consisting of known actives and inactives.

The model retrieved around 75% of the active compounds including the final four hits. An enrichment factor of 6.843 and a Güner-Henry score of 0.788 indicates the quality of the model, as shown in Table 4.

The virtual screening procedure was further validated by subjecting the final four hits individually to Glide XP docking. The most active hit, ZINC08383847, had a docking score of −7.234. Its interaction with Met109, Ala111, and Ser154 on the active site of p38-α MAPK via H-bonding is shown in Figure 12A. Figure 12B shows how ZINC08383847 maps on the generated pharmacophore model. All the other hits had a docking score less than −6.4 and showed binding with the active site of p38-α MAPK. Further, the validation of results of applicability domain (AD) indicates that the predictions of the test set compound are quite reliable. The residual standard deviation (DModX) value of the 16 test set compounds are below the critical value of 2.8 (Figure 13).

Table 5 ADME properties of selected hits with the docking score

ZINC ID	QPlogPo/w	QPlogS	QPlogHERG	QPPCaco	QPlogBB	QPlogKhsa	Percent human oral absorption	Predicted activity	Docking score
ZINC08383847	3.739	−5.808	−5.749	89.686	−1.992	0.588	83.785	7.068	−7.23367
ZINC00646908	4.371	−6.579	−6.506	618.474	−0.891	0.332	100	7.633	−7.07034
ZINC09110317	4.565	−7.285	−5.743	430.717	−0.767	0.589	100	7.020	−6.51755
ZINC00645569	4.674	−6.537	−6.576	1151.681	−0.513	0.727	100	7.541	−6.40786

Notes: QPlogPo/w, predicted octanol/water partition coefficient (−2.0 to 6.5); QPlogS, predicted aqueous solubility in mol dm^{-3} (−6.5 to 0.5); QPlogHERG, pIC$_{50}$ value for blockade of HERG K+ channels (acceptable range below −6.0); QPPCaco, predicted apparent Caco-2 cell (gut-blood barrier model) permeability in nm/sec (<25 poor; >500 excellent); QPlogBB, predicted brain/blood partition coefficient (−3 to 1.2); QPlogKhsa, prediction of binding to human serum albumin (−1.5 to 0.5); percent human oral absorption (<25 poor; >80 high).
Abbreviation: ADME, absorption, distribution, metabolism, and excretion; pIC$_{50}$, the predicted concentration of the compound producing 50% inhibition.

Further, among the 44 pharmaceutically relevant descriptors available in QikProp, descriptors including predicted octanol/water partition coefficient (QPlogPo/w) (−2.0 to 6.5), predicted aqueous solubility(QPlogS) in mol dm^{-3}, predicted IC$_{50}$ value for blockade of heRgK+channels (QPlogHERG) (acceptable range below −6.0), predicted apparent Caco-2 cell (QPPCaco) (gut-blood barrier model) permeability in nm/sec (<25 poor; >500 excellent), predicted brain/blood partition coefficient (QPlogBB) (−3 to 1.2), prediction of binding to human serum albumin (QPlogKhsa) (−1.5 to 0.5), and percent human oral absorption <25 poor; >80 high (Table 5) were calculated for the final four hits. For all the hits, the partition coefficient (QPlogPo/w) and water solubility (QPlogS) critical for estimating the absorption and distribution of drugs within the body ranged between 4.674 to 3.739 and −7.285 to −5.808, respectively. The blood-brain barrier partition coefficient ranged from −1.982 to −0.513. QPlogHERG varied from −6.576 to −5.743. QPPCaco and QPlogKhsa varied from 1151.681 to 89.686 and 0.727 to 0.322, respectively. Further, the percentage human oral absorption varied from 100% to 95.568%. The pharmacokinetic parameters for all the four hits were found to be within the acceptable range defined for human use, revealing their potential drug-like properties.

Conclusion

A five-point pharmacophore was generated using a ligand-based approach for pyridopyridazin-6-one derivatives and highlights the importance of steric and electronic features of the ligand for binding with its target and how these are responsible for activity. A highly predictive atom-based 3D-QSAR model was generated using a training set of 47 molecules consisting of a five-point pharmacophore (AAAHR), ie, three hydrogen bond acceptor (A), one hydrophobic (H), and one aromatic ring feature (R). Further, the developed model was visualized with cubes to identify the important structural features. It was observed that a hydrogen bond acceptor group is necessary in the C ring and an electron-withdrawing or hydrophobic group is necessary in the A ring. Compounds lacking this group were found to be less potent. Finally, four hits were obtained after virtual screening, and docking studies that have a good docking score, and predictive and fitness value may be further modified using the QSAR model to obtain the most potent p38-α MAPK inhibitor. Hence, the results of an atom-based 3D-QSAR model and docking studies help in designing analogs with better activity prior to synthesis which may be potent p38-α MAPK inhibitors.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Lee JC, Laydon JT, McDonnell PC, et al. A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature*. 1994;372: 739–746.
2. Elliott MJ, Maini NM, Feldmann M, et al. Randomised double-blind comparison of chimeric monoclonal antibody to tumour necrosis factor α (cA2) versus placebo in rheumatoid arthritis. *Lancet*. 1994;344: 1105–1110.
3. Van Dullemen HM, Van Deventer SJ, Hommes DW, et al. Treatment of Crohn's disease with anti-tumor necrosis factor chimeric monoclonal antibody (cA2). *Gastroenterology*. 1995;109:129–135.
4. Feldmann M, Brennan FM, Maini RN. Role of cytokines in rheumatoid arthritis. *Annu Rev Immunol*. 1996;14:397–440.
5. Badger AM, Bradbeer JN, Votta B, Lee JC, Adams JL, Griswold DE. Pharmacological profile of SB 203580, a selective inhibitor of cytokine suppressive binding protein/p38 kinase, in animal models of arthritis, bone resorption, endotoxin shock and immune function. *J Pharmacol Exp Ther*. 1996;279:1453–1461.
6. Rutgeerts P, D'Haens G, Targan S, et al. Efficacy and safety of retreatment with anti-tumor necrosis factor antibody (infliximab) to maintain remission in Crohn's disease. *Gastroenterology*. 1999;117: 761–769.
7. Rachel BL, Steven H, Rebecca W, Hugh FP, Christopher JM. Nuclear export of the stress-activated protein kinase p38 mediated by its substrate MAPK AP kinase-2. *Curr Biol*. 1998;8:1049–1057.
8. Cowan JK, Storey BK. Mitogen-activated protein kinases: new signaling pathways functioning in cellular responses to environmental stress. *J Exp Biol*. 2003;206:1107–1115.
9. Dominguez C, Powers DA, Tamayo N. p38 MAPK inhibitors: many are made, but few are chosen. *Curr Opin Drug Discov Devel*. 2005;8:421–430.
10. Goldstein DM, Gabriel T. Pathway to the clinic: inhibition of P38 MAPK. A review of ten chemotypes selected for development. *Curr Top Med Chem*. 2005;10:1017–1029.
11. Hynes J, Leftheris K. Small molecule p38 inhibitors: novel structural features and advances from 2002–2005. *Curr Top Med Chem*. 2005;10:967–985.
12. Stelmach JE, Liu L, Patel SB, et al. Design and synthesis of potent, orally bioavailable dihydroquinazolinone inhibitors of p38 MAPK. *Bioorg Med Chem Lett*. 2003;13:277.
13. Hanson GH. Inhibitors of p38 kinase. *Expert Opinions on Therapeutic Patents*. 1997;7:729–733.
14. Salituro FG, Germann RA, Wilson KP, Bemis GW, Fox T, Su MS. Inhibitors of p38 MAPK: therapeutic intervention in cytokine-mediated diseases. *Curr Med Chem*. 1999;6:807–823.
15. Boehm JC, Adams JL. New inhibitors of p38 kinase. *Expert Opinions on Therapeutic Patents*. 2000;10:25–37.
16. Dumas J, Sibley R, Reidl B, et al. Discovery of a new class of p38 kinase inhibitors. *Bioorg Med Chem Lett*. 2000;10:2047–2050.
17. Mayer RJ, Callahan JF. p38 MAPK inhibitors: a future therapy for inflammatory diseases. *Drug Discov Today Ther Strateg*. 2006;3: 49–54.

18. Tynebor RM, Chen M, Natarajan SR, et al. Synthesis and biological activity of pyridopyridazin-6-one p38 MAPK inhibitors. Part 1. *Bioorg Med Chem Lett*. 2011;21:411–416.

19. Tynebor RM, Chen M, Natarajan SR, et al. Synthesis and biological activity of pyridopyridazin-6-one p38-α MAPK inhibitors. Part 2. *Bioorg Med Chem Lett*. 2012;22:5979–5983.

20. Van de Waterbeemd H, Carter RE, Grassy G, et al. Glossary of terms used in computational drug design. *Pure Appl Chem*. 1997;69: 1137–1152.

21. Marriott DP, Dougall IG, Meghani P, Liu YJ, Flower DR. Lead generation using pharmacophore mapping and three-dimensional database searching: application to muscarinic M(3) receptor antagonists. *J Med Chem*. 1999;42:3210–3216.

22. Shah UA, Deokar HS, Kadam SS, Kulkarni VM. Pharmacophore generation and atom-based 3D-QSAR of novel 2-(4-methylsulfonylphenyl) pyrimidines as COX-2 inhibitors. *Mol Div*. 2010;14:559–568.

23. PHASE version 3.4. New York, NY: Schrödinger LLC; 2012.

24. Glide version 5.8. New York, NY: Schrödinger LLC; 2012.

25. Irwin JJ, Shoichet BK. ZINC – a free database of commercially available compounds for virtual screening. *J Chem Inf Model*. 2005;45: 177–182.

26. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 1997;23:3–25.

27. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods*. 2000;44: 235–249.

28. LigPrep version 2.5. New York, NY: Schrödinger LLC; 2012.

29. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetic and properties of organic liquids. *J Am Chem Soc*. 1996;118: 11225–11236.

30. Chang G, Guida WC, Still WC. An internal-coordinate Monte Carlo method for searching conformational space. *J Am Chem Soc*. 1989;111: 4379–4386.

31. Kolossvary I, Guida WC. Low mode search. An efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides. *J Am Chem Soc*. 1996;118:5011–5019.

32. Maestro version 9.3. New York, NY: Schrödinger LLC; 2012.

33. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des*. 2006;20:647–671.

34. Dixon SL, Smondyrev AM, Rao S. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des*. 2006;67:370–372.

35. Friesner RA, Banks JL, Murphy RB, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 2004;47:1739–1749.

36. Halgren TA, Murphy RB, Friesner RA, et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem*. 2004;47:1750–1759.

37. Teli MK. Pharmacophore generation and atom-based 3D-QSAR of N-iso-propyl pyrrole-based derivatives as HMG-CoA reductase inhibitors. *Org Med Chem Lett*. 2012;2:25–34.

38. Vyas VK, Ghate M, Goel A. Pharmacophore modeling, virtual screening, docking and in silico ADMET analysis of protein kinase B (PKB β) inhibitors. *J Mol Graph Model*. 2013;42:17–25.

39. Cheminformatics programs and QSAR datasets [webpage on the Internet]. USA: Stahl data set. Available from: http://cheminformatics.org/datasets/stahl/index.html. Accessed October 21, 2013.

40. UMETRICS SIMCA-P 12.0. Umea, Sweden, 2002. Available from: http://www.umetrics.com. Accessed September 13, 2013.

41. Zhang L, Zhu H, Oprea T, Golbraikh A, Tropsha A. QSAR modeling of the blood-brain barrier permeability for diverse organic compounds. *Pharm Res*. 2008;25:1902–1914.

42. QikProp version 3.5. New York, NY: Schrödinger LLC; 2012.

Supplementary data
Scoring pharmacophore hypothesis

The quality of alignment was measured by a survival score,[33,34] defined as:

$$S = W_{site}S_{site} + W_{vec}S_{vec} + W_{vol}S_{vol} + W_{sel}S_{sel}$$
$$+ W^m_{rew} - W_E\Delta E + W_{act}A \qquad (2)$$

where Ws are weights and Ss are scores; S_{site} represents alignment score, the root mean square deviation in the site point position; S_{vec} represents vector score, and averages the cosine of the angles formed by corresponding pairs of vector features in aligned structures; S_{vol} represents volume score based on overlap of van der Waals models of nonhydrogen atoms in each pair of structures; and S_{sel} represents selectivity score, and accounts for what fraction of molecules are likely to match the hypothesis regardless of their activity toward the receptor. $W_{site}, W_{vec}, W_{vol}, W_{rew}$ have default values of 1.0, while W_{sel} has a default value of 0.0. In hypothesis generation, default values have been used. W_{sel} represents reward weights defined by $m - 1$, where m is the number of actives that match the hypothesis. Hypotheses for which the reference ligand has a high energy relative to the lowest-energy conformer for that ligand are less likely to be good models of binding, because of the energetic cost hence a penalty for high-energy structures can be included by subtracting a multiple of the relative energy from the final score, $W_E\Delta E$. Similarly, for the hypothesis in which the reference ligand activity has relatively lower energy than the highest activity can be penalized by adding a multiple of the reference ligand activity to the score, $W_{act}A$ where A is the activity.[33,34]

The fitness score is a linear combination of the site and vector alignment scores and the volume score, and is related to the default survival score. The fitness score is defined by:

$$S = W_{site}(1 - S_{align}/C_{align}) + W_{vec}S_{vec} + W_{vol}S_{vol}$$

Where S_{align} is alignment score and C_{align} is alignment cutoff. Also, the terms used in calculating the survival score are described in Table S1.

Significance: Fitness is a score that measures how well the matching pharmacophore site points align to those of the hypothesis, how well the matching vector features (acceptors, donors, aromatic rings) overlay those of the hypothesis, and how well the matching conformation superimposes, in an overall sense, with the reference ligand conformation. The refer-

Table S1 Terms used for calculating the survival score

Parameter	Description
S_{align}	Alignment score: RMS deviation between the site point positions in the matching conformation and the site point positions in the hypothesis
C_{align}	Alignment cutoff: user-adjustable parameter; default is 1.2
W_{site}	Weight of site score: user-adjustable parameter; default is 1.0
S_{vec}	Vector score: average cosine between vector features in the matching conformation and the vector features in the reference conformation
W_{vec}	Weight of vector score: user-adjustable parameter; default is 1.0
S_{Vol}	Volume score: ratio of the common volume occupied by the matching conformer and the reference conformer, to the total volume (the volume occupied by both) Volumes are computed using van der Waals models of all nonhydrogen atoms
W_{Vol}	Weight of volume score: user-adjustable parameter; default is 1.0

Abbreviations: S_{align}, alignment score; C_{align}, alignment cutoff; W_{site}, weight of site score; S_{vec}, vector score; W_{vec}, weight of vector score; S_{vol}, volume score; W_{vol}, weight of volume score.

ence ligand, which matches exactly, has a perfect fitness score. Hits are then fetched in order of decreasing fitness score.

Validation of pharmacophore model

The PLS factor was increased up to 5, since up to the 5th factor there is an incremental increase in predictive power and statistical value of the model. A further increase in the number of PLS factors did not improve the statistics or predictive ability (q^2) of the model. A detailed explanation on why the PLS factor was increased up to 5 is given in Table S2.

The regression is done by constructing a series of models with increasing number of PLS factors. The accuracy of the model increases with increasing number of the PLS factor until overfitting starts to occur. There is no limit on the maximum number of PLS factors, but as a general rule, adding factors should be stopped when the standard deviation of regression is approximately equal to the experimental error. This point occurs beyond the 5th factor in our case. Also, the statistical parameters like R^2 and q^2 were high (0.9167 and 0.8047, respectively) at the 5th PLS factor with minimum standard deviation of regression. Therefore, we selected the 5th standard deviation of regression factor for generation of our atom-based three-dimensional quantitative structure-activity relationship model.

The details of the atom-based three-dimensional quantitative structure-activity relationship results using pharmacophore AAAHR.71 up to the 6th PLS factor indicate that there

Table S2 Details of statistical parameters of hypothesis AAAHR.71 up to 6 PLS factors

Pharmacophore ID	PLS factors	SD	R^2	F	P	Stability	RMSE	q^2	Pearson-R
AAAHR.71	1	0.8893	0.4394	35.3	3.84e-007	0.8917	0.7933	0.5638	0.8055
	2	0.7869	0.5708	29.3	8.27e-009	0.8351	0.722	0.6387	0.8607
	3	0.5449	0.7989	56.9	5.06e-015	0.4814	0.5893	0.7593	0.8964
	4	0.4062	0.8908	85.7	1.24e-019	0.3574	0.5786	0.768	0.8876
	5	0.3591	0.9167	90.3	5e-021	0.4082	0.5308	0.8047	0.908
	6	0.3171	0.9366	98.5	2.22e-022	0.4194	0.5669	0.7772	0.887

Abbreviations: PLS, partial least-squares; SD, standard deviation; RMSE, root mean square of error.

is an increase in statistical value up to the 5th PLS factor, as shown in Table S2.

The PHASE quantitative structure-activity relationship model uses distinct training and test sets for validation techniques. An explanation regarding how R^2 and q^2 are calculated and how they are correlated with each other is given below.

Training set and model

R^2 is generally calculated for training set compounds. Leave-n-out techniques are useful for assessing the stability of the model to changes in the training set. In PHASE quantitative structure-activity relationship models, leave-n-out models are built, and the R^2 value is computed between the leave-n-out predictions and the predictions from the model built on the full training set. This value is reported as the stability value, and has a maximum value of 1. If the stability value is high, the model built from the full training set is fairly insensitive to changes in that training set, ie, the predicted values do not change much.

R^2 can never be negative, because the regression coefficients are optimized to minimize *sse* (sum of squared errors). The values go in negatives when the independent variables have no statistical relationship with activity. In our case, the R^2 value at 5th PLS factor was found to be 0.9167, which is quite stable.

Different statistical terms used for describing the training set and the quantitative structure-activity relationship model are defined below:

m	Number of PLS factors in the model
n	Number of molecules in the training set
$df_1 = m + 1$	Degrees of freedom in model
$df_2 = n - m - 2$	Degrees of freedom in data
y_i	Observed activity for training set molecule i
y'_i	Predicted activity for training set molecule i
$\bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$	Mean observed activity

$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad \text{Variance in observed activities}$$

$$sse = \sum_{i=1}^{n}(y'_i - y_i)^2 \qquad \text{Sum of squared errors}$$

$$\sigma_{err}^2 = \frac{sse}{n} \qquad \text{Variance in errors}$$

$$ssy = \sum_{i=1}^{n}(y'_i - \bar{y})^2 \qquad \text{Variance in model}$$

$$SD = \sqrt{sse/df_2} \qquad \text{Standard deviation of regression}$$

$$F = \frac{ssy/df_1}{sse/df_2} \qquad \text{F statistic; overall significance of model}$$

$$R^2 = 1 - \frac{\sigma_{err}^2}{\sigma_y^2} \qquad \text{R^2; coefficient of determination.}$$

Test set predictions

q^2 is generally calculated for test set compounds for external cross-validation of model.

Statistical quantities describing the test set predictions are listed below:

T	The test set of molecules
n_T	Number of molecules in T
y_j	Observed activity for molecule $j \in T$
$\hat{y}_j$	Predicted activity for molecule $j \in T$

$$RMSE = \sqrt{\frac{1}{n_T}\sum_{i \in T}(\hat{y}_j - y_j)^2} \qquad \text{Root-mean-squared error}$$

$$r = \frac{\sum_{i \in T}(y_j - \bar{y}_T)(y_j - \dot{\bar{y}}_T)}{\sqrt{\sum_{i \in T}(y_j - \bar{y}_T)^2(y_j - \dot{\bar{y}}_T)^2}} \qquad \text{Pearson's r value, Pearson's correlation coefficient}$$

$$Q^2 = R^2(T) \qquad \text{Q-squared}$$

The formulae for R^2 and q^2 are equivalent, with the only difference being that q^2 is computed using the observed and predicted activities for the test set. However, q^2 can take on negative values. This happens whenever the variance in the test set error is larger than the variance in the observed test set activities. Often, the test set does not have as large a range

of activity values as the training set (so the variance in y is smaller), and the errors for the test set tend to be larger than those for the training set (so the variance in the errors is larger), therefore it might be possible that q^2 values sometimes have a negative value. Pearson's R value is calculated for the correlation between the predicted and observed activity for the test set compounds. A high q^2 and a Pearson's R value of 0.8047 and 0.9080, respectively, were obtained at the 5th PLS factor. This explains the correlation between R^2 and q^2 and also the basis for selecting the PLS factor as 5.

In silico screening and absorption, distribution, metabolism, and excretion studies

We subjected 15.8 million compounds in the ZINC "Clean drug-like" database to prefiltering by using the ZINC database keys to eliminate the majority of molecules which did not match the Lipinski rule of five and default value of the feature matching tolerance. The default zinc database keys used for search of the 15.8 million compounds of "ZINC Clean drug-like" database which reduce the compounds to 12,005 compounds were as follows:

- molecular weight: 150 to 500 g/mol
- xlogP: −4 to 5
- net charge: −5 to 5
- rotatable bonds: 0 to 8
- polar surface area (A^2): 0 to 150
- hydrogen donors: 0 to 10
- hydrogen acceptors: 0 to 10
- polar desolvation (kcal/mol): −400 to 1
- apolar desolvation (kcal/mol): −100 to 40.