

Molecular Modeling on Pyrimidine-Urea Inhibitors of TNF- α Production: An Integrated Approach Using a Combination of Molecular Docking, Classification Techniques, and 3D-QSAR CoMSIA

Varnavas D. Mouchlis,[†] Georgia Melagraki,^{*,†} Thomas Mavromoustakos,[‡] George Kollias,[§] and Antreas Afantitis^{*,†,§}

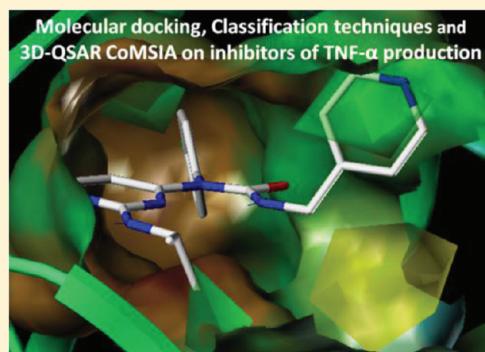
[†]Department of Chemoinformatics, NovaMechanics, Ltd., Nicosia, Cyprus

[‡]Laboratory of Organic Chemistry, Department of Chemistry, University of Athens, Athens 15771, Greece

[§]Institute of Immunology, Biomedical Sciences Research Center “Alexander Fleming”, Athens, Greece

Supporting Information

ABSTRACT: Molecular docking, classification techniques, and 3D-QSAR CoMSIA were combined in a multistep framework with the ultimate goal of identifying potent pyrimidine-urea inhibitors of TNF- α production. Using the crystal structure of p38 α , all the compounds were docked into the enzyme active site. The docking pose of each compound was subsequently used in a receptor-based alignment for the generation of the CoMSIA fields. “Active” and “inactive” compounds were used to build a Random Tree classification model using the docking score and the CoMSIA fields as input parameters. Domain of applicability indicated the compounds for which activity estimations can be accepted with confidence. For the active compounds, a 3D-QSAR CoMSIA model was subsequently built to accurately estimate the IC₅₀ values. This novel multistep framework gives insight into the structural characteristics that affect the binding and the inhibitory activity of these analogues on p38 α MAP kinase, and it can be extended to other classes of small-molecule inhibitors. In addition, the simplicity of the proposed approach provides expansion to its applicability such as in virtual screening procedures.



INTRODUCTION

Rheumatoid arthritis (RA) is a chronic autoimmune inflammatory disease that affects synovial joints by targeting the synovial membrane, articular cartilage, and bones.^{1,2} High levels of proinflammatory cytokines such as tumor necrosis factor alpha (TNF- α) and interleukin-1b (IL-1b) are associated in a variety of inflammatory diseases,^{3,4} such as RA, multiple sclerosis, inflammatory bowel disease, and Crohn’s disease (CD).^{5–7} The important role of TNF- α in the pathogenesis of RA was demonstrated both in experimental animal models and in RA patients.^{7–9} As a result, the blockade of TNF- α production may lead to the development of new anti-TNF- α therapies.^{10,11} They have been reported three drugs in use for treatment that block the activity of TNF- α : Infliximab (chimeric monoclonal antibody to human TNF), Adalimumab (human monoclonal antibody to TNF), and Etanercept (soluble TNF receptor construct). The pharmacological profile of these new drugs is described in a review article.¹² However, there are some drawbacks in the therapy of RA by these drugs including their high cost, inadequate clinical response, need of intravenous administration and several side effects such as increased risk of tuberculosis.¹²

One of the main signal transduction pathways implicated in RA involves p38 mitogen-activated protein (MAP) kinase.^{13,14} There are four recognized p38 enzymes with 40–60%

structural similarity which are referred as p38 α (also known as p38), p38 β , p38 γ , and p38 δ .^{15–18} It was demonstrated that p38 α regulates the expression of proinflammatory cytokines such as TNF- α and IL-1b, and thus plays a key role in RA.¹⁹ Therefore, the inhibition of p38 α MAP kinase is an attractive target for drug development,^{20,21} because it would prevent the expression of these proinflammatory cytokines and contribute to the development of therapeutic agents to treat inflammatory diseases such as RA.

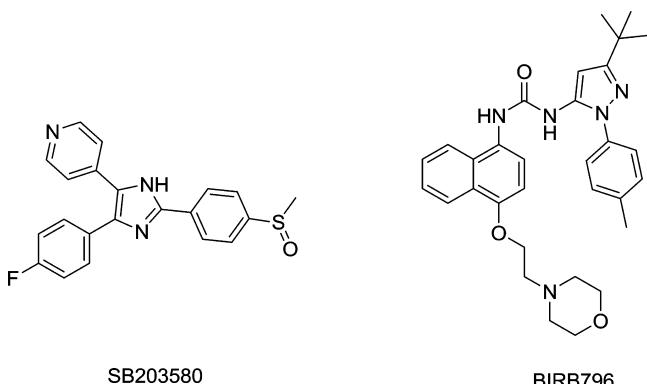
Because of their potential therapeutic applications, several small-molecules that serve as p38 inhibitors have been developed.²² It has been reported that potent p38 MAP kinase inhibitors reduce the expression of TNF- α in vitro and in vivo. The most representative p38 inhibitor is a 4-aryl-5-pyridylimidazole-based inhibitor known as SB203580 (Scheme 1).²³ This compound displayed oral efficacy in several models of cytokine inhibition and inflammatory diseases. The interesting in vivo pharmacological profile of SB203580 renders this compound a prototypical standard of other p38 inhibitors. However, into clinical trials SB203580 was characterized by several toxicological effects.^{24,25} Other structural classes reported to block cytokines activity include pyrroles, pyrimidines, pyridines,

Received: December 6, 2011

Published: February 23, 2012



Scheme 1. Chemical Structures of SB203580 and BIRB796



pyrimidones, indoles, heteroindoles, ureas, and various fused bicyclic heterocycles.²⁶ Several p38 inhibitors were progressed into clinical trials due to the tremendous interest of pharmaceutical companies in the development of agents against chronic inflammatory diseases.²⁷ The clinical candidate BIRB796 (Scheme 1) exhibited subnanomolar affinity against p38 MAP kinase and was entered phase II clinical trials.²⁸ Unfortunately, it was halted before development because of demonstrated liver toxicity.²⁹

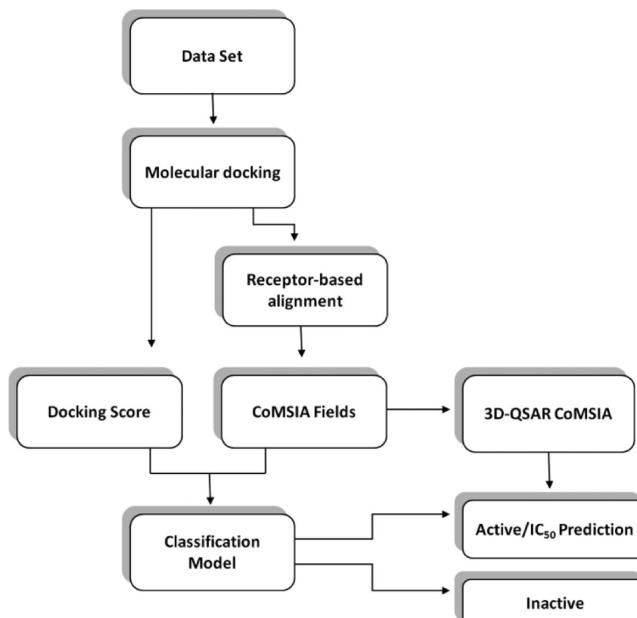
One of the major problems in rational drug discovery is the understanding of the mode of interaction of small-molecule inhibitors with their receptors. The elucidation of the three-dimensional structure of p38 MAP kinase through X-ray crystallography^{30,31} contributed to the understanding of key structural characteristics that could assist in the rational design of potent and selective p38 inhibitors. Moreover, computer-aided drug design (CADD) through in silico techniques has also been employed in rational drug discovery to understand the inhibitor-receptor interactions and predict the inhibitory activity of new compounds.^{32–37} Several in silico methods have emerged as a useful tool in the identification of novel compounds with improved characteristics.^{38–41} Different regression or classification methods have been employed for this purpose in an effort to minimize the time and cost associated with identifying new leads.^{42–48}

Previous CADD reports on p38 MAP kinase include molecular docking and 3D-QSAR studies.^{49–51} In this study, we describe a multistep framework by combining molecular docking, 3D-QSAR CoMSIA and classification techniques (Scheme 2) to understand the structural characteristics that affect the binding of 51 pyrimidine-urea analogues^{52,53} with p38 α MAP kinase receptor. These compounds were tested as inhibitors of TNF- α production, and two of them are cocrystallized with p38 α MAP kinase.^{52,53} All the compounds were docked into the enzyme active site and the docking pose of each compound was subsequently used in a receptor-based alignment for the generation of the CoMSIA fields. A classification model was then proposed with the primary purpose of discriminating compounds into two categories, active and inactive, by using the CoMSIA fields and the docking score (threshold $IC_{50} > 5 \mu M$). Finally, a 3D-QSAR CoMSIA model was developed that accurately estimates the IC_{50} values for the active compounds. This computational workflow gives insight into the structural characteristics that affect the binding and the inhibitory activity of these analogues on p38 α MAP kinase.

MATERIALS AND METHODS

Computational Workflow. As aforementioned, a computational workflow was followed in order to study potent pyrimidine-urea inhibitors of TNF- α production, and to predict the inhibitory activity of newly designed analogues. The computational workflow is graphically depicted in Scheme 2. The

Scheme 2. Computational Workflow



data set of the pyrimidine-urea inhibitors was first docked into the p38 α active site for calculating the docking score. Then the docking pose of each inhibitor was used in a receptor-based alignment for generating the CoMSIA fields. The CoMSIA fields along with the docking score were then used in a classification framework to separate active from inactive compounds within the domain of applicability of the model. Finally, a 3D-QSAR CoMSIA model was developed using the active compounds to predict their inhibitory activity. After a thorough validation of all undertaken steps, the outcome for a given compound is to be either characterized as active or inactive with a prediction of its inhibitory activity.

Data Set. For the present molecular modeling study 51 pyrimidine-urea analogues (Tables 1 and S1–S5, see Supporting Information) were considered.^{52,53} These analogues were all tested for inhibition of TNF- α production in an LPS (LipoPolySaccharide) stimulated human monocytic (THP-1) whole cell-based assay, under the same experimental conditions.^{52,53} The structures of pyrimidine-urea analogues were built using the SYBYL 8.0 molecular sketcher. All the hydrogen atoms were added, and the structures were subsequently submitted to full energy minimization using the Conjugate Gradient energy minimization algorithm with gradient 0.01 kcal mol⁻¹ Å⁻¹. The Tripos force field⁵⁴ was used for the minimization procedure with nonbonded interactions cutoff of 8.0 Å, using a distance dependent dielectric function with dielectric constant of 1.

Preparation of the p38 α MAP Kinase File. Two high-resolution X-ray crystal structures of the holo form p38 α MAP kinase cocrystallized with the pyrimidine-urea analogues **15** (PDB ID 2GHL, 2.10 Å X-ray resolution) and **25** (PDB ID 2GHM, 2.35 Å X-ray resolution) were considered to examine

Table 1. IC₅₀ Values, the Class, the Predicted Class, the $-\log K_d$, the pIC₅₀, the Predicted pIC₅₀, and the Residuals for the Pyrimidine-Urea Analogue

ID ^a	compound	IC ₅₀ (μM)	class	predicted class	$-\log K_d$	pIC ₅₀	predicted pIC ₅₀	residuals
1	5a	0.122	active	active	7.380	6.914	6.747	0.167
2 ^{TS}	5b	0.176	active	active	8.200	6.754	6.773	-0.018
3	5c	63% ^b	active	active	6.850			
4	5d	3.100	active	active	7.510	5.509	5.556	-0.048
5 ^{TS}	5e	3.650	active	active	8.400	5.438	5.614	-0.177
6	5f	2.880	active	active	8.050	5.541	5.673	-0.133
7	5g	>5% ^b	inactive	inactive	6.690			
8	5h	0.843	active	active	6.250	6.074	6.374	-0.300
9	5i	>5% ^b	inactive	inactive	5.730			
10	5j	0.539	active	active	6.270	6.268	6.305	-0.036
11 ^{TS}	5k	3.610	active	active	7.360	5.442	5.965	-0.523
12	5l	1.420	active	active	7.930	5.848	5.787	0.060
13	5m	>5% ^b	inactive	inactive	4.920			
14 ^{TS}	5n	>5% ^b	inactive	inactive	4.700			
15	6a	0.092	active	active	8.540	7.036	7.048	-0.012
16	6b	0.135	active	active	9.190	6.870	6.842	0.028
17	6c	0.718	active	active	7.590	6.144	6.196	-0.052
18	6d	1.000	active	active	8.840	6.000	5.693	0.307
19 ^{TS}	6e	>5% ^b	inactive	inactive	5.480			
20	6f	>5% ^b	inactive	inactive	5.940			
21 ^{TS}	6g	6.16	inactive	active	6.800			
22	6h	>5% ^b	inactive	inactive	5.460			
23	6i	>5% ^b	inactive	inactive	5.800			
24 ^{TS}	6a	0.014	active	active	7.450	7.854	7.649	0.204
25	6b	0.009	active	active	8.280	8.046	7.695	0.351
26	6c	0.037	active	active	9.630	7.432	7.447	-0.015
27	6d	0.015	active	active	8.540	7.824	7.793	0.031
28 ^{TS}	6e	0.236	active	active	8.640	6.627	6.755	-0.128
29	6f	0.148	active	active	7.820	6.830	6.937	-0.107
30	6g	0.056	active	active	8.460	7.252	7.126	0.126
31	6h	0.611	active	active	8.910	6.214	6.154	0.060
32 ^{TS}	6i	2.070	active	active	9.450	5.684	5.921	-0.237
33	6j	0.016	active	active	6.830	7.796	7.847	-0.052
34	6k	0.016	active	active	6.810	7.796	8.067	-0.271
35	6l	35% ^b	inactive	inactive	2.690			
36 ^{TS}	6m	3% ^b	inactive	inactive	5.570			
37	6n	25% ^b	inactive	inactive	4.820			
38	7a	0.028	active	active	7.930	7.553	7.483	0.070
39	7b	0.324	active	active	8.960	6.489	6.654	-0.165
40 ^{TS}	7c	0.467	active	active	9.450	6.331	6.209	0.121
41	7d	0.819	active	active	8.920	6.087	6.114	-0.027
42	7e	0.276	active	active	8.600	6.559	6.581	-0.022
43	7f	0.086	active	active	7.560	7.066	7.311	-0.245
44 ^{TS}	7g	0.080	active	active	6.830	7.097	7.024	0.073
45	7h	0.074	active	active	6.780	7.131	6.966	0.165
46	10a	0.004	active	active	7.800	8.398	8.247	0.151
47 ^{TS}	10b	0.030	active	active	7.360	7.523	7.577	-0.054
48	10c	0.277	active	active	6.670	6.558	6.571	-0.014
49	10d	1.160	active	active	6.320	5.936	5.885	0.051
50 ^{TS}	10e	0.119	active	active	6.900	6.924	6.871	0.054
51	10f	0.016	active	active	6.400	7.796	7.865	-0.069

^aTS = test set. ^bPercentage inhibition at 10 μM .

which complex was suitable for the molecular docking procedure.^{52,53} It is well-known that proteins are flexible and can occupy different conformations upon the binding of small-molecule inhibitors because of induced fit effects. However, no significant conformational differences were observed on the two crystal structures. The crystal structure with the higher

X-ray resolution (PDB ID 2GHL) was chosen for the docking, and was prepared using the Protein Preparation Wizard implementation in Schrödinger suite 2009.⁵⁵ The bond orders were assigned, all the hydrogen atoms were added, the disulfide bonds were assigned and all the water molecules were deleted. The Epik 2.0 implementation was used to predict ionization

and tautomeric states of the ligand het groups.^{56,57} The hydrogen-bonding network was optimized by reorienting the hydroxyl groups, amide groups of Asn and Gln residues, and by selecting appropriate states and orientations of the imidazole ring in His residues. Finally, using the “impref utility” and the OPLS_2005 force field,⁵⁸ the hydrogen atom positions were optimized by keeping all the heavy atoms in place. The prepared structure was saved in PDB format and was used in the molecular docking. To avoid any inaccuracies depending on using different force fields, for the preparation of the protein (OPLS_2005) and for the molecular docking (Tripos), the prepared PDB structure was also examined using the Biopolymer in SYBYL 8.0 molecular modeling package.⁵⁹ The same combination of force fields was used successfully in a previous work.⁶⁰

Molecular Docking Procedure. The Surflex-Dock algorithm in SYBYL 8.0 molecular modeling package was used for the molecular docking procedure.⁶¹ Surflex-Dock supports a fully automated flexible docking procedure for the ligand and relies on the rigid-receptor approximation to treat ligand–receptor binding.⁶² The algorithm utilizes an idealized active site ligand called “protomol”, as a target to generate putative poses of molecules. These putative poses are scored using the Hammerhead scoring function. The scoring function contains the dominant hydrophobic and polar contact terms as well as a repulsive, an entropic, and a solvation terms. In this study, the ligand-based mode was adopted to generate the “protomol”, leaving the threshold and bloat parameters at their default values of 0.50 and 0 Å, respectively. The docking scores are expressed in -logK_d units to represent the calculated-binding affinities. The maximum number of poses per ligand was set to 20. No constraints were used for the molecular docking procedure. The docking complex assumed to represent the ligand–receptor interactions was selected based on the following three criteria: (i) the docking score of the pose possessed the highest docking score; (ii) its orientation of the aromatic rings of the ligand oriented into the active site in a similar with the cocrystallized ligands orientation; and (iii) the preservation of four key interactions, namely the two hydrogen bonds with Met109, the hydrogen bond with Lys53 as well as an intramolecular hydrogen bond.

CoMSIA Interaction Energy. CoMSIA was originally developed by Gerhard Klebe to find the common features in a series of small-molecule inhibitors that are important in binding to the biologically relevant receptor.⁶³ In CoMSIA, five different similarity fields are calculated: steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor. These fields cover the major contributions to ligand binding⁶⁴ and the similarity indices are calculated at regularly spaced grid points (lattice) for the prealigned molecules using a probe atom. The similarity fields (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å. The similarity indices were calculated using the sp³ carbon as probe atom with radius 1 Å, charge +1, hydrophobicity +1, hydrogen bond donating +1, hydrogen bond accepting +1. The default value of 0.3 was set for attenuation factor α .⁶⁵

Classification Model. A classification model has been developed in order to separate active from inactive compounds and filter out the inactive ones. As input variables for the development of the classification model, the docking score and the CoMSIA similarity fields, steric, electrostatic, hydrophobic,

H-bond donor and H-bond acceptor were used. Random tree (RT) classification technique implemented in WEKA⁶⁶ program was used to discriminate between the different classes. Decision trees represent a supervised approach to classification with its simple structure consisting of root, nodes, branches and leaves. The first node is a root. A decision tree is usually drawn beginning from the root downward. Nonterminal nodes represent tests on attributes and each node corresponds with a certain characteristic. From each node two or more branches grow connecting the nodes and each branch corresponds with a range of values that give a partition of the set of values of the given characteristic. The terminal nodes are called leaves and reflect decision outcomes. A Random tree is a tree drawn at random from a set of possible trees using different features at each node. RT is a WEKA implementation of the random decision tree algorithm. It is a decision tree with no pruning and considering only log₂(N) of descriptors in each node (where N is a total amount of available descriptors). Random trees can be generated efficiently and the combination of large sets of Random trees generally leads to accurate models.⁶⁷

For the validation of the developed classification model various statistical parameters were calculated. In particular, the proposed classification model was fully validated using the following measurements:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (3)$$

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (4)$$

where TP = true positive, FP = false positive, TN = true negative, and FN = false negative

The confusion matrix is also given as shown below:

	positive predicted	negative predicted
positive observed (active)	TP	FN
negative observed	FP	TN
(inactive)		

Applicability Domain. In order for an in silico model to be used for screening new compounds, its domain of application^{37,68,69} must be defined and predictions for only those compounds that fall into this domain may be considered reliable. Similarity measurements were used to define the domain of applicability of the models based on the Euclidean distances among all the training set compounds.⁷⁰ The distance of a test set compound to its nearest neighbor in the training set was compared to the predefined applicability domain (APD) threshold. The prediction was considered unreliable when the distance was higher than APD. APD was calculated as follows:

$$\text{APD} = \langle d \rangle + Z\sigma \quad (5)$$

Calculation of $\langle d \rangle$ and σ was performed as follows: First, the average of Euclidean distances between all pairs of training compounds was calculated. Next, the set of distances that were lower than the average was formulated. $\langle d \rangle$ and σ were finally calculated as the average and standard deviation of all distances

included in this set. Z was an empirical cutoff value and for this work, it was chosen equal to 0.5.⁷⁰

3D-QSAR CoMSIA Model. The PLS analysis implemented in SYBYL 8.0^{71,72} was employed to obtain correlation between the CoMSIA similarity fields, which were used as independent explanatory variables, and the pIC_{50} values which were used as target dependent variables. PLS was performed in two stages. The first was the crossvalidation using the leave-one-out method (LOO) to determine the value of the crossvalidated q^2 , the crossvalidated standard error of predictions SPRESS and the optimal number of components (ONC). In the leave-one-out method, one molecule is omitted from the data set and the model is derived involving the rest of the molecules and using this model, the activity of the omitted molecule is predicted.⁷³ The crossvalidated q^2 is calculated by the following equation:

$$q^2 = 1 - \frac{\text{PRESS}}{\text{TSS}} \quad (6)$$

where PRESS is the predictive error sum of squares and TSS is the total sum of squares. The second stage was without cross-validation and using the ONC, which were determined by the first crossvalidation stage, to establish a CoMSIA model representing the data. The noncrossvalidated form is determined through the noncrossvalidated (conventional) r^2 , the standard error of estimate (SEE), the F value (F -ratio), and the probability of r^2 . To speed up the PLS analysis and reduce the noise, a column filtering value of 2.00 kcal mol⁻¹ was used.

To further assess the robustness and the statistical confidence of the derived model, bootstrap analysis for 100 runs was performed.⁷³ Bootstrap analysis relies on the generation of many new data sets from the original one which are obtained by randomly choosing samples from the original data set with repeated selection of the same sample being allowed. The statistical calculation is then performed on each of these bootstrap data sets. The difference between the parameters calculated from the original data set and the average of the parameters calculated from the many bootstrap data sets is a measure of the bias of the original calculations.

The CoMSIA model was assessed for its predictive ability using the molecules of the test set not included in the construction of the model. External validation is the most acceptable validation method for the predictive ability of a QSAR model, since the molecules of the test set are not included in the training set. The external r_{ext}^2 was calculated using the following equation:

$$r_{\text{ext}}^2 = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (y_i - \tilde{y}_i)^2}{\sum_{i=1}^{n_{\text{test}}} (y_i - \bar{y}_{\text{tr}})^2} \quad (7)$$

where n_{test} is the number of compounds that constitute the validation data set (test set), \bar{y}_{tr} is the averaged value for the dependent variable for the training set, y_i , \tilde{y}_i , $i = 1, \dots, n_{\text{test}}$ are the experimental values and the 3D-QSAR model predictions of the dependent variable over the available validation set.

γ -randomization test also ensures the robustness and the statistical significance of a 3D-QSAR model.⁷⁴ The dependent variable vector (pIC_{50}) is randomly shuffled and a new model is developed using the original independent variable matrix.^{75–78} The derived models after several repetitions are expected to have significant low q^2 and r^2 values than the ones of the original model. This method is usually performed to eliminate the possibility of chance correlation. If the opposite happens

then an acceptable 3D-QSAR model cannot be obtained for the specific modeling method and data.^{79,80}

RESULTS AND DISCUSSION

The previous mentioned computational workflow (Scheme 2) was implemented in order to study the data set of the 51 pyrimidine-urea inhibitors of TNF- α production. In the following sections, the computational workflow is described in detail.

Molecular Docking on Pyrimidine-Urea Inhibitors.

Molecular docking was conducted on all the 51 pyrimidine-urea analogues using the molecular docking algorithm Surflex-Dock. The docking scores ($-\log K_d$) are reported in Table 1. No correlation was found between the docking score and the IC_{50} values. The molecular docking results are described below.

Examination of p38 α Active Site. Enzymes in MAP kinase family are characterized by two domains which are separated by a conserved catalytic channel. The small N-terminal domain consists mainly of β -sheets, while the large C-terminal domain consists mostly of α -helices. The N-terminal domain creates a deep binding pocket for the adenine ring of ATP. The C-terminal domain contains the phosphorylation lip which is located at the opening of the channel. The phosphate groups of ATP bind near the phosphorylation lip, and the γ -phosphate group interacts with the residues of the C-terminal domain. Near the same area the substrate also binds to initiate the kinase reaction.^{30,81}

As aforementioned, two X-ray crystal structures are available for the analogues **15** (PDB ID 2GHL, 2.10 Å X-ray resolution) and **25** (PDB ID 2GHM, 2.35 Å X-ray resolution) with the mutated p38 α . The mutated p38 α is a double mutant (S180A, Y182F) of murine p38 α . The mutant enzyme cannot be phosphorylated and, therefore, it is not competent for activation.^{52,53} However, those residues are not included among the active site residues (SF1a, see Supporting Information) and, therefore, do not affect the molecular docking calculation. Since crystallography is a strong experimental evidence for the binding mode of those analogues, the cocrystallized with the studied analogues enzyme was used for the molecular docking. Proteins can exhibit induced fit effects during the binding of a ligand, in which the protein conformation changes significantly. The conformational differences caused by this effect can be observed when more than one cocrystallized complex is available. The two crystal structures were aligned in order to examine if they differ appreciably in conformation. For the alignment all the residues were used and the rmsd value was 1.240 Å. The low rmsd value designates that there are no significant conformational differences on the two crystal structures. The side chains of the active site residues (Val30, Val38, Ala51, Lys53, Leu75, Ile84, Leu86, Leu104, Thr106, Leu108, Met109, Ala157, and Leu167) do not appreciably differ in conformation (SF1a, see Supporting Information), and the two crystallographic ligands can fit into both active sites without significant steric clashes (SF1b, Supporting Information). Thus, the crystal structure with PDB ID 2GHL, which possesses the higher X-ray resolution, was chosen for the molecular docking.

Test of the Surflex-Dock Algorithm. According to the crystal structure of pyrimidine-urea **15** with the mutated p38 α (PDB ID 2GHL) the 2-aminopyrimidine functionality participates into two hydrogen bonds with the backbone of Met109 residue of the ATP-binding pocket. The N-4-MeO-Phenyl

group is accommodated in the well-defined hydrophobic pocket, which consists of residues, such as Leu75, Leu86, Leu104, and Thr106. The *N'*-2-chlorophenyl group is oriented toward the solvent exposed residues of the pocket. It is noteworthy that the conformation of the urea allows for the *N'*-urea to participate into an intramolecular hydrogen bonding with the nitrogen atom at the three position of the pyrimidine ring. This interaction creates a pseudobicyclic structure for these compounds and gives them rigidity.⁵² The second pyrimidine-urea **25** (PDB ID 2GHM) contains an *N'*-2-chlorobenzyl group instead of the *N'*-2-chlorophenyl group in the analogue **15**. The orientation and the interactions of the analogue **25** are identical to the ones of **15**, with the exception that the carbonyl moiety of urea interacts with the amino group of Lys53 unlike to the one of the analogue **15**.⁵³ The two crystallographic analogues **15** and **25** were docked into the active site of p38 α in order to examine if Surflex-Dock is able to reproduce the crystallographic binding mode (interactions, conformation, and orientation). Surflex-Dock successfully reproduces the crystallographic binding mode for both analogues (SF2 and SF3, for detailed discussion see Supporting Information).

Docking of the Pyrimidine-Urea Analogue. After successful reproduction of the crystallographic binding mode by Surflex-Dock all the pyrimidine-urea analogues were docked into the active site of p38 α . The aim was to examine if all the analogues are docked into the active site of p38 α in a similar with the crystallographic binding mode. The attention has been focused on the most characteristic receptor-ligand interactions of the most active analogue **46** (SF 4). The predicted by Surflex-Dock binding mode for the analogue **46** shows that the orientation and the conformation of this analogue is similar to the one of the crystallographic analogues **15** and **25**. The *N*-4-F-Phenyl group is accommodated in the hydrophobic pocket and interacts with the residues Leu75, Leu86, Leu104, and Thr106. The *N'*-methylpyridine group is oriented toward the solvent exposed area and interacts with the residues Tyr35 and Asp168. The 2-aminopyrimidine functionality participates into two hydrogen bonds with the backbone of Met109 residue of the ATP-binding pocket and the carbonyl moiety of urea interacts with the amino group of Lys53. The intramolecular hydrogen bond of the *N'*-urea NH with the nitrogen atom at the three position of the pyrimidine ring is also observed. Additionally, a hydrogen bond of the nitrogen atom of the pyridine ring with the backbone NH group of Ser32 is observed. It seems that this hydrogen bond contributes to the tight binding of the 4-substituted analogue **46**, because the corresponding 2-substituted analogue **47**, in which this hydrogen bond is not observed, is about 7-fold less active than the analogue **46** (Tables 1 and S5). By comparing the bioactivity of the analogues **46** and **25** with the analogue **15** indicates that a methylene group between the *N*'-NH and the pyridine or the 2-chlorophenyl group increases the inhibitory activity. This might be due to the extra flexibility that gives the methylene group to the aromatic rings so they are accommodated more favorably toward the solvent exposed area of the binding pocket. All the pyrimidine-urea analogues were docked into the active site of p38 α with a similar orientation and conformation. In most of the cases the three hydrogen bonds with Met109 and Lys53 were also observed.

Data Set Alignment and Atomic Charges. In the CoMSIA analysis, bioactive conformation and alignment rule selection are two important factors to construct a reliable

model. The docking pose of each analogue was used in a receptor-based alignment to calculate the CoMSIA fields. The most important parameter in 3D-QSAR is the alignment of the individual molecules since their fields are computed based on a 3D lattice. Properly aligned molecules have a comparable conformation and a similar orientation in Cartesian space. Molecular docking into the active site of the p38 α has the advantage of reducing the uncertainty about the bioactive conformation by adjusting the orientation and conformation of the molecules based on the interactions with the active site. The most active analogue **46** was used as a template molecule for the alignment using 11 atoms (SF 5a) that are common in all analogues. A rigid-body atom-by-atom superimposition of one molecule onto another was performed using the utility "Align database" available in SYBYL 8.0. The results of the alignment are depicted in SF 5b.

Atomic charges for the aligned molecules were calculated using the Gasteiger-Hückel method, which is a combination of two other charge computational methods: the Gasteiger-Marsili⁸² method to calculate the σ component of the atomic charge and the Hückel⁸³ method to calculate the π component of the atomic charge. The total charge is the sum of the charges calculated by the two methods. The aligned molecules were then used for the calculation of the CoMSIA fields.

Classification Model. For the development of the classification and 3D-QSAR CoMSIA models the available small molecules (51 pyrimidine-urea analogues) were separated into two independent sets, "actives" and "inactives". The cutoff value for the discrimination between "actives" and "inactives" was set to $IC_{50} > 5 \mu M$. The initial data set was split into a training and test set. The separation of the data set into training and test set was performed randomly.⁸⁴ The training set was used for the development of the models reported in this study. The test set was used to get an unbiased estimation of the predictive ability of the developed models (classification and 3D-QSAR CoMSIA models). A commonly used ratio of training to test set (70:30) was adopted.⁸⁵ The training set contains 37 compounds (29 "actives" and 8 "inactives") and the test set 14 compounds (10 "actives" and 4 "inactives") as shown in Table 1.

After the classification model was trained, prediction on the activity of test compounds was performed. The experimental values and the predictions for both training and test examples are presented in Table 1. The confusion matrix for the crossvalidation method and model predictions on the external test set, are presented in Tables 2 and 3.^{86,87} The performance

Table 2. Confusion Matrix (Training Set, 10-Fold Cross-Validation) Random Trees

	positive predicted	negative predicted
positive observed (active)	29	0
negative observed (inactive)	1	7

Table 3. Confusion Matrix (Test Set) Random Trees

	positive predicted	negative predicted
positive observed (active)	10	0
negative observed (inactive)	1	3

of the model was evaluated based on validation measurements described in the Materials and Methods section. The significance, accuracy and robustness of the model are illustrated by the corresponding statistics. In particular, the

application of the 10-fold crossvalidation method produced the following statistics: precision = 97%, sensitivity = 100%, specificity = 88%. By applying the model to the external test set, the following statistical results were obtained: precision = 91%, sensitivity = 100%, and specificity = 75%. The statistical results are summarized in Table 4.

Table 4. Specificity, Sensitivity, Precision, and Accuracy Statistics (Random Trees)

	specificity	sensitivity	precision	accuracy
cross-validation	0.88	1.00	0.97	0.97
external validation	0.75	1.00	0.91	0.93

The applicability domain was defined for the compounds that constitute the test set as described in the previous section. Since all validation compounds fell inside the domain of applicability, all model predictions for the external test set can be considered reliable (Table 5).

Table 5. Applicability Domain for the Test Set

ID	compound	distance (APD = 1.70)
2	SB	0.34
5	SE	0.44
11	SK	0.52
14	SN	0.88
19	6E	0.93
21	6G	0.73
24	P26A	0.40
28	P26E	0.64
32	P26I	0.72
36	P26M	0.98
40	P27C	0.73
44	P27G	0.74
47	P210B	0.54
50	P210E	0.65

Moreover, the classification technique gives insight into the model by visualizing the Random Tree. The visual inspection of the Random Tree helps in the analysis and interpretation of the data structure, the existence of cluster and outliers, the relationship between samples and the influence of each variable. The graphical representation of the Random Tree for the data set is presented in Figure 1.

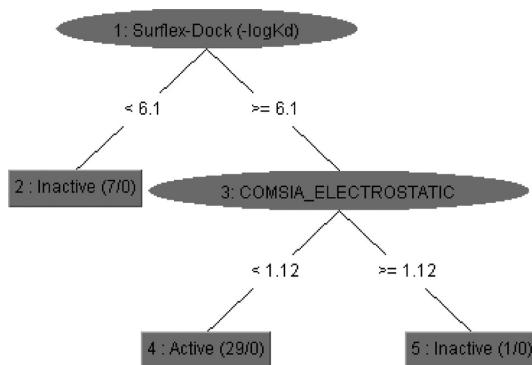


Figure 1. Graphical representation of the Random Tree for the data set.

The first classification on whether the compound is characterized as active or inactive was achieved by the

classification model. The next step was the development of the 3D-QSAR CoMSIA model using the compounds that were characterized as "actives" by the classification model. The aim was to establish a reliable CoMSIA model that could be used for predicting the inhibitory activity (IC_{50} values) of the compounds.

3D-QSAR CoMSIA Model. As aforementioned, a receptor-based alignment was used for the generation of the CoMSIA model based on the compounds which were characterized as "actives" by the classification model. The molecules possessing percentage inhibition of TNF- α production in the initial publication were excluded from the CoMSIA model. The separation of the data set into training and test set was performed randomly as described in previous section. The training set consists of 28 molecules and the test set consists of 10 molecules (Table 1). The Partial Least Squares (PLS) analysis was used for deriving the relationships among the CoMSIA fields and the inhibitory activity (IC_{50} values).^{71,72} The 3D-QSAR CoMSIA model was derived using the standard implementation in SYBYL 8.0 molecular modeling package.⁸⁸ The IC_{50} values were converted to the pIC_{50} scale ($-\log IC_{50}$).

CoMSIA Analysis. CoMSIA models were generated using the combination of the following similarity fields: steric, electrostatic, hydrophobic, H-bond donor and H-bond acceptor.⁶³ These fields illustrate the various properties into spatial locations where they play important roles in determining the inhibitory activity. The best CoMSIA model was derived using the steric, electrostatic, and H-bond acceptor fields (Table 6).

Table 6. Statistical Parameters of the CoMSIA Model

CoMSIA model	
ONC	4
q^2	0.813
S_{PRESS}	0.376
r^2	0.964
SEE	0.165
F ($n_1 = 4, n_2 = 23$)	154.8
prob. of $r^2 = 0$ ($n_1 = 4, n_2 = 23$)	0.000
field contribution	
Steric:Electrostatic:H-bond Acceptor	0.562:0.086:0.353
r_{bs}^2	0.971 \pm 0.010
SEE_{bs}	0.142 \pm 0.081
$q_{Y-randomization}^2$	-0.956 to -0.146
$r_{Y-randomization}^2$	0.215 to 0.347
r_{ext}^2	0.936

An ONC of four corresponds to the highest crossvalidated q^2 of 0.813 and the lowest S_{PRESS} of 0.376. A non-cross-validated r^2 of 0.964, a SEE of 0.165, and an F value of 154.8 with zero probability of r^2 show a good statistical correlation between the predicted and the experimental pIC_{50} values for the non-crossvalidated form of the CoMSIA model. Figure 2 represents the relationship between the predicted and the experimental pIC_{50} values. The CoMSIA analysis shows that the relative field contributions are 56.2%, 8.6%, and 35.3% for steric, electrostatic, and H-bond acceptor fields, respectively.

Validation of the Model. The robustness and the statistical confidence of the derived model were examined using the bootstrap validation method for 100 runs. This method produced an r_{bs}^2 of 0.971 ± 0.010 indicating the statistical validity of the CoMSIA model. The higher average value of the r_{bs}^2 is a demonstration of the robustness and the

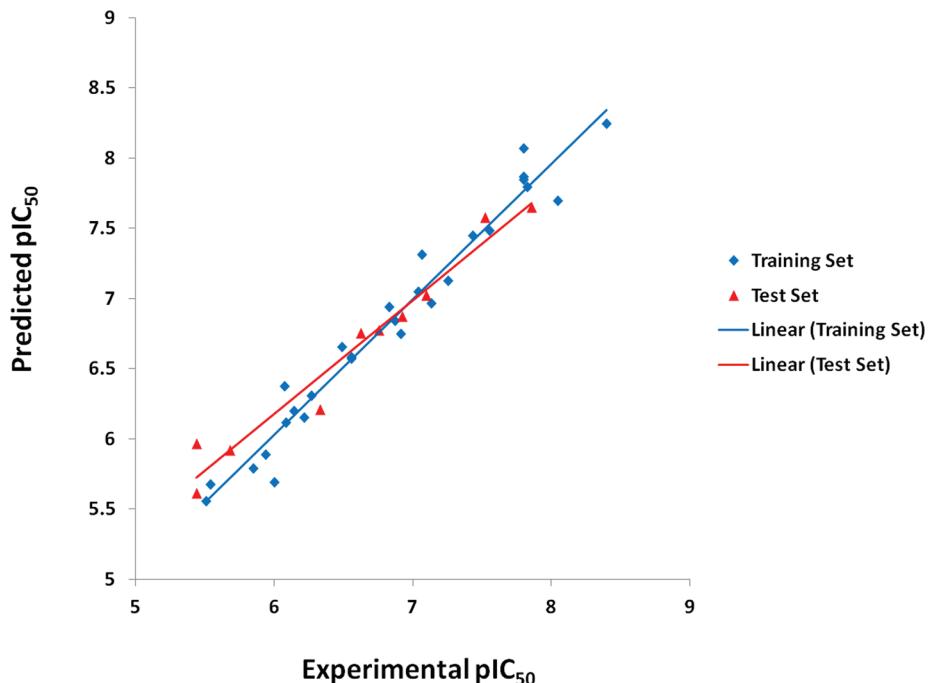


Figure 2. Plot of the predicted pIC_{50} values versus experimental ones for the CoMSIA model.

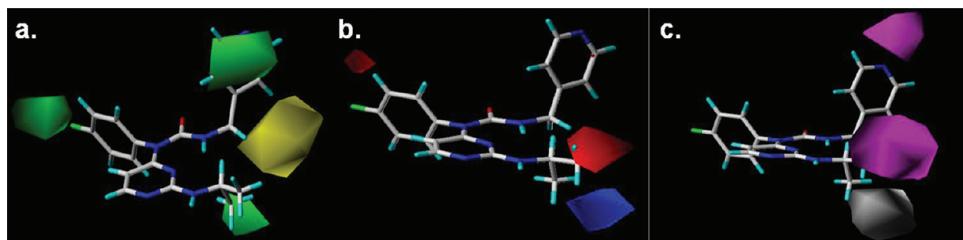


Figure 3. CoMSIA StDev*Coeff contour maps around the most active compound **46**: (a) for steric field (green: bulky groups are favored, yellow: bulky groups are disfavored); (b) for electrostatic field (red: electronegative groups are favored, blue: electronegative groups are disfavored); and (c) for H-bond acceptor field (magenta: groups with H-bond acceptor atoms are favored, white: groups with H-bond acceptor atoms are disfavored).

internal predictive ability of the CoMSIA model. Y-randomization is a method also for testing the robustness and the statistical significance of a QSAR model. This method was performed to eliminate the possibility of chance correlation. In particular, 10 random shuffles of the Y vector (pIC_{50} values) gave q^2 and r^2 values in the ranges of -0.956 to -0.146 and 0.215 to 0.347 , respectively. The negative values of q^2 and the low values of r^2 indicate that the results from the CoMSIA model were not due to chance correlation or structural dependency of the training set.

External validation was also performed to further assess the stability and the predictive ability of the CoMSIA model. This validation was performed using 10 molecules not included in the development of the model. The predicted pIC_{50} values are in correlation with the experimental ones within the tolerable error range (Figure 2). The external r_{ext}^2 has a value of 0.936 upholding the good correlation between predicted and experimental pIC_{50} values. The high value of the external r_{ext}^2 of the COMSIA model derived from the receptor-based alignment suggests that the receptor-based alignment can effectively take into consideration the ligand–receptor interactions and the CoMSIA model is reliable and could be used in the design of new inhibitors of TNF- α production within this structural motif of molecules.

An additional validation test has been carried out in order to further assess the predictability and the applicability of the model. The available data were randomly divided five times in a ratio of 70:30 for training and test set, respectively. The results are presented in Table S6 (see Supporting Information).

CoMSIA Contour Maps. The results of a CoMSIA analysis is a set of contour maps visualizing which areas around the molecule interact favorably or unfavorably with the receptor. The contour maps express the relationship found between variations in the magnitude of the similarity fields in a particular spatial position and variations in the inhibition of TNF- α production. The StDev*Coeff option was selected to control the type of CoMSIA fields to be viewed. It is based on the product of the “sigma” and “coefficients” fields and shows where variability in the fields around the molecule explains differences in the inhibitory activity. The default option “contribution” was selected to control the interpretation of the CoMSIA fields. The contour maps are essential because they represent 3D spatial position around a molecule where structural modifications correlate with changes in the inhibitory activity.

The CoMSIA fields around the most active compound **46** are displayed in Figure 3. The steric interactions are represented by green and yellow contour maps (Figure 3a),

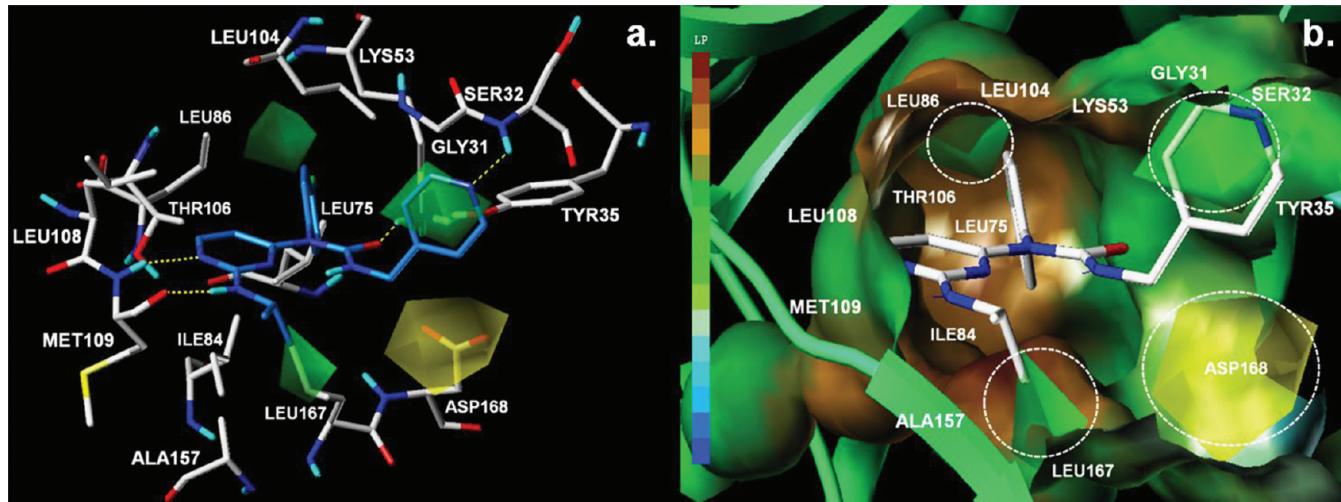


Figure 4. CoMSIA steric and hydrophobic fields around the most active compound **46** superimposed onto the active site of p38 α : (a) the match of the steric contour maps with the residues of the active site; (b) the MOLCAD lipophilic potential surface of the active site was created using the Connolly method, where brown color denotes the most lipophilic areas and blue color the most hydrophilic.

where bulky groups near the green regions increase the inhibitory activity but they cause the opposite effect near the yellow regions. The electrostatic interactions are displayed by red and blue contour maps (Figure 3b); where electronegative groups near the red regions increase the inhibitory activity whereas they decrease it near the blue regions. The H-bond acceptor interactions are displayed by magenta and white contour maps (Figure 3c), where groups with H-bond acceptor atoms near the magenta regions increase the inhibitory activity but they decrease it near the white regions.

Superimposition of the Contour Maps onto p38 α Active Site. As aforementioned, the 3D-QSAR CoMSIA model was derived using a receptor-based alignment and it was assumed that the CoMSIA contour maps would be in agreement with the active site of p38 α . Thus, the CoMSIA contour maps were superimposed onto the active site of the enzyme in an effort to examine if there is a relationship between the CoMSIA fields and the surrounding of the ligand residues of the active site.

The active site of p38 α is mostly hydrophobic and that is in agreement with the CoMSIA model, which gives 56.2% field contribution to the steric field. The superimposition of the CoMSIA steric fields around the most active compound **46** onto the active site of the enzyme is depicted in Figure 4. The CoMSIA steric contour maps showed a green contour that favors bulky substitutions near the N-4-F-Phenyl group (Figure 4a). On the basis of the docking results this group is accommodated inside the hydrophobic pocket and interacts with the residues Leu75, Leu86, Leu104, and Thr106. It seems that the pocket is deep enough (Figure 4b) and apparently bulkier groups at the para position of the phenyl ring would be able to optimize the interactions with the residues of the pocket. This can be confirmed by the fact that N-4-F-Phenyl **1** and N-4-MeO-Phenyl **15** analogues are 6-fold and 8-fold more potent, respectively, than the unsubstituted N-Phenyl analogue **17**. In addition, the analogue **15** is slightly more potent than the analogue **1** because the methoxy substitution is bulkier than the fluorine atom. Another green contour was observed near the N'-methylpyridine group (Figure 4a). Based on the lipophilic potential surface (Figure 4b) there is a small cavity near this group consisting of the residues Gly31, Ser32 and Tyr35 in

which the pyridine group is accommodated. This is in accordance with the experimental results where the N'-2-Chlorobenzyl analogues **24** and **25** are approximately 10-fold more potent than the corresponding N'-2-Chlorophenyl analogues **1** and **15**. Based on the CoMSIA contour maps the N'-2-Chlorophenyl group is directed toward the yellow contour that disfavors bulky groups and may clash with the around residues such as Asp168. On the other hand, the N'-2-Chlorobenzyl group is more flexible because of the methylene group and is accommodated near the previous mentioned small cavity. A green contour was also observed near the methyl group of pyrimidine substitution which is accommodated in a small hydrophobic cavity consisting of the residues Ala157 and Leu167 (Figure 4a and 4b) indicating that a bulkier substitution favors the inhibitory activity.

The CoMSIA H-bond acceptor contour showed a magenta contour around the nitrogen atom of the pyridine ring which favors the H-bond acceptor atoms. According to the docking results a hydrogen bond was observed between that nitrogen atom and Ser32 (Figure 5a). This is validated also by the experimental results where the 2-substituted compound **47** is about 7-fold less active than compound **46**. Additionally, compound **51** which has an H-bond donor atom at the para position is 4-fold less active than compound **46**. Another magenta contour was also observed near the methyl group of the pyrimidine substitution, which is directed toward the exterior of the active site, suggesting that H-bond acceptor atoms would be favorable for the inhibitory activity (Figure 5a). This is also in agreement with the electrostatic contour maps which showed a red contour near the same methyl group that favors electronegative groups (Figure 3b). A white contour that disfavors H-bond acceptor atoms is near the other methyl group (Figure 5a) which is placed in the small hydrophobic pocket consisting of the residues Ala157 and Leu167 (Figure 5b). On the basis of the electrostatic contour maps electronegative groups would also not be favorable for the inhibitory activity in that position (Figure 3b). In accordance with the experimental results, compounds **24** and **25** are approximately 16-fold more active than the corresponding compounds **28** and **29** which have an H-bond acceptor atom near the white contour. In addition, compound **33** which has an H-bond acceptor atom

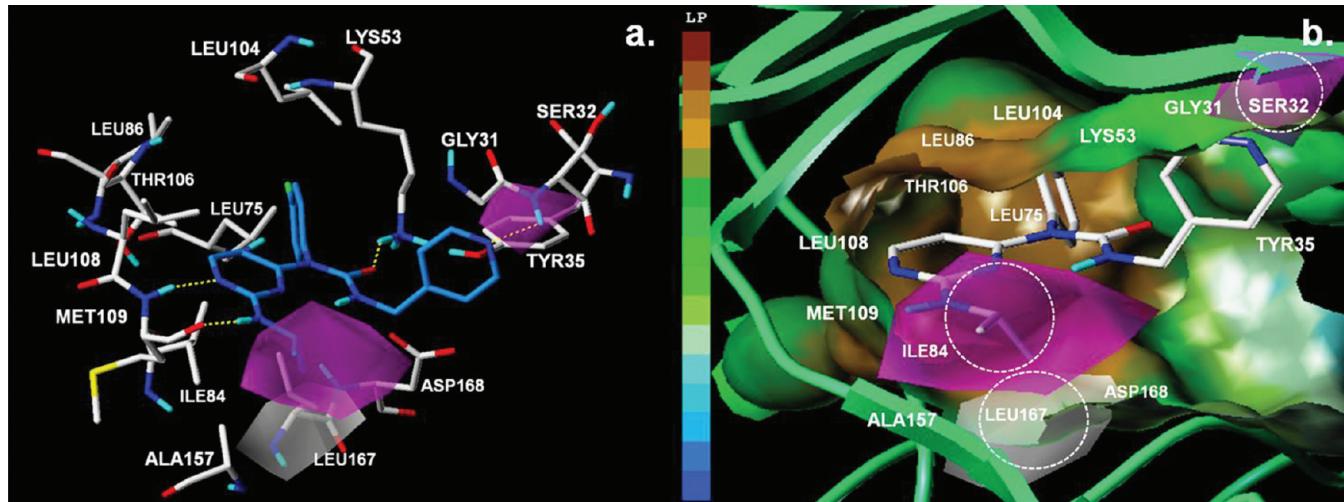


Figure 5. CoMSIA H-bond acceptor and hydrophobic fields around the most active compound **46** superimposed onto the active site of p38 α : (a) the match of the H-bond acceptor contour maps with the residues of the active site; (b) the MOLCAD lipophilic potential surface of the active site was created using the Connolly method, where brown color denotes the most lipophilic areas and blue color the most hydrophilic.

near the magenta contour is about 14-fold more active than the corresponding compound **28**.

CONCLUSIONS

The proposed computational workflow is initiated by molecular docking that produces the receptor-based alignment and the docking scores for the ligands using the crystal structure of p38 α MAP kinase. On the basis of the receptor-based alignment, five different similarity CoMSIA fields are calculated: steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor which cover the major contributions to ligand binding. A classification framework is then used as a filtering tool combining information from the docking scores and the CoMSIA fields to identify active compounds and filter out inactive compounds. Finally, the 3D-QSAR CoMSIA model identifies the common features in the series of small-molecule inhibitors that are important in binding to the receptor. Because of the high predictive ability³⁶ and simplicity, the proposed computational workflow could be a useful aid to the costly and time-consuming experiments for determining the TNF- α inhibition of pyrimidine-urea analogues. The two-stage approach that is proposed in this study increases the accuracy of the produced QSAR model, since it covers a narrower chemical space, compared to a model that uses all the available data.⁸⁵ A virtual screening procedure^{39,89} could be based on the proposed QSAR model. The design of new active molecules by the insertion, deletion, or modification of substituents on different sites of the molecule and at different positions could therefore be guided by the proposed CoMSIA model. The method³⁸ can also be used to screen existing databases or virtual combinations to identify derivatives with desired activity. In this scenario, the classification model will be used to screen out “inactive” compounds, while the applicability domain will serve as a valuable tool to filter out “dissimilar” combinations. The molecular descriptors used in QSAR workflow encode information about the stereoelectronic effects of the molecules and thus implicitly account for cooperative effects between functional groups. The proposed approach based on molecular docking, classification techniques and 3D-QSAR CoMSIA aims to help researchers to design novel chemistry driven molecules with desired biological activity.

ASSOCIATED CONTENT

S Supporting Information

Tables showing structures of urea analogs and classification and 3D-QSAR CoMSIA results and figures showing alignment of crystal structures, binding mode of crystal analogues, superimposition of crystallographic information, and receptor-based alignment. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: afantitis@novamechanics.com, melagraki@novamechanics.com.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by funding under the Seven Research Framework Programme of the European Union, Project Masterswitch (Grant agreement number HEALTH-F2-2008-223404)

ABBREVIATIONS

RA, rheumatoid arthritis; TNF- α , tumor necrosis factor alpha; IL-1 β , interleukin-1 β ; CD, Crohn’s disease; MAP, mitogen-activated protein; CADD, computer-aided drug design; 3D-QSAR, three-dimensional quantitative structure–activity relationship; CoMSIA, comparative molecular similarity indices analysis; LPS, lipopolysaccharide; PLS, partial least-squares; LOO, leave-one-out; S_{PRESS}, standard error of predictions; ONC, optimal number of components; PRESS, predictive error sum of squares; TSS, total sum of squares; rmsd, root-mean-square deviation

REFERENCES

- (1) McInnes, I. B.; Schett, G. Cytokines in the pathogenesis of rheumatoid arthritis. *Nat. Rev. Immunol.* **2007**, *7*, 429–42.
- (2) Chen, E.; Keystone, E. C.; Fish, E. N. Restricted cytokine expression in rheumatoid arthritis. *Arthritis Rheum.* **1993**, *36*, 901–10.

- (3) Dinarello, C. A. Inflammatory cytokines: Interleukin-1 and tumor necrosis factor as effector molecules in autoimmune diseases. *Curr. Opin. Immunol.* **1991**, *3*, 941–8.
- (4) Kontoyiannis, D.; Kotlyarov, A.; Carballo, E.; Alexopoulou, L.; Blackshear, P. J.; Gaestel, M.; Davis, R.; Flavell, R.; Kollias, G. Interleukin-10 targets p38 MAPK to modulate ARE-dependent TNF mRNA translation and limit intestinal pathology. *EMBO J.* **2001**, *20*, 3760–70.
- (5) Andreakos, E. T.; Foxwell, B. M.; Brennan, F. M.; Maini, R. N.; Feldmann, M. Cytokines and anti-cytokine biologicals in autoimmunity: present and future. *Cytokine Growth Factor Rev.* **2002**, *13*, 299–313.
- (6) Feldmann, M.; Brennan, F. M.; Maini, R. N. Role of cytokines in rheumatoid arthritis. *Annu. Rev. Immunol.* **1996**, *14*, 397–440.
- (7) Feldmann, M.; Maini, R. N. Anti-TNF alpha therapy of rheumatoid arthritis: what have we learned? *Annu. Rev. Immunol.* **2001**, *19*, 163–96.
- (8) Kollias, G.; Douni, E.; Kassiotis, G.; Kontoyiannis, D. On the role of tumor necrosis factor and receptors in models of multiorgan failure, rheumatoid arthritis, multiple sclerosis and inflammatory bowel disease. *Immunol. Rev.* **1999**, *169*, 175–94.
- (9) Kollias, G.; Douni, E.; Kassiotis, G.; Kontoyiannis, D. The function of tumour necrosis factor and receptors in models of multi-organ inflammation, rheumatoid arthritis, multiple sclerosis and inflammatory bowel disease. *Ann. Rheum. Dis.* **1999**, *58* (Suppl 1), I32–9.
- (10) Kollias, G.; Kontoyiannis, D. Role of TNF/TNFR in autoimmunity: specific TNF receptor blockade may be advantageous to anti-TNF treatments. *Cytokine Growth Factor Rev.* **2002**, *13*, 315–21.
- (11) Kollias, G.; Kontoyiannis, D.; Douni, E.; Kassiotis, G. The role of TNF/TNFR in organ-specific and systemic autoimmunity: implications for the design of optimized 'anti-TNF' therapies. *Curr. Dir. Autoimmun.* **2002**, *5*, 30–50.
- (12) Olsen, N. J.; Stein, C. M. New drugs for rheumatoid arthritis. *N. Engl. J. Med.* **2004**, *350*, 2167–79.
- (13) Westra, J.; Limburg, P. C. p38 mitogen-activated protein kinase (MAPK) in rheumatoid arthritis. *Mini Rev. Med. Chem.* **2006**, *6*, 867–74.
- (14) Johnson, G. L.; Lapadat, R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* **2002**, *298*, 1911–2.
- (15) Lee, J. C.; Laydon, J. T.; McDonnell, P. C.; Gallagher, T. F.; Kumar, S.; Green, D.; McNulty, D.; Blumenthal, M. J.; Heys, J. R.; Landvatter, S. W.; et al. A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature* **1994**, *372*, 739–46.
- (16) Jiang, Y.; Chen, C.; Li, Z.; Guo, W.; Gegner, J. A.; Lin, S.; Han, J. Characterization of the structure and function of a new mitogen-activated protein kinase (p38 β). *J. Biol. Chem.* **1996**, *271*, 17920–6.
- (17) Li, Z.; Jiang, Y.; Ulevitch, R. J.; Han, J. The primary structure of p38 gamma: a new member of p38 group of MAP kinases. *Biochem. Biophys. Res. Commun.* **1996**, *228*, 334–40.
- (18) Jiang, Y.; Gram, H.; Zhao, M.; New, L.; Gu, J.; Feng, L.; Di Padova, F.; Ulevitch, R. J.; Han, J. Characterization of the structure and function of the fourth member of p38 group mitogen-activated protein kinases, p38delta. *J. Biol. Chem.* **1997**, *272*, 30122–8.
- (19) Newton, R.; Holden, N. Inhibitors of p38 mitogen-activated protein kinase: Potential as anti-inflammatory agents in asthma? *BioDrugs* **2003**, *17*, 113–129.
- (20) Brennan, F. M.; Feldmann, M. Cytokines in autoimmunity. *Curr. Opin. Immunol.* **1996**, *8*, 872–7.
- (21) Camussi, G.; Lupia, E. The future role of anti-tumour necrosis factor (TNF) products in the treatment of rheumatoid arthritis. *Drugs* **1998**, *55*, 613–20.
- (22) Lee, J. C.; Kassis, S.; Kumar, S.; Badger, A.; Adams, J. L. p38 Mitogen-activated protein kinase inhibitors—Mechanisms and therapeutic potentials. *Pharmacol. Ther.* **1999**, *82*, 389–97.
- (23) Badger, A. M.; Bradbeer, J. N.; Votta, B.; Lee, J. C.; Adams, J. L.; Griswold, D. E. 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*, 1453–61.
- (24) Howard, M. O.; Schwartz, L. W.; Newton, J. F.; Qualls, C. W. Jr.; Yodis, L. A.; Ventre, J. R. Comparative biochemical and morphometric changes associated with induction of the hepatic mixed function oxidase system in the rat. *Toxicol. Pathol.* **1991**, *19*, 115–22.
- (25) Adams, J. L.; Boehm, J. C.; Kassis, S.; Gorycki, P. D.; Webb, E. F.; Hall, R.; Sorenson, M.; Lee, J. C.; Ayrton, A.; Griswold, D. E.; Gallagher, T. F. Pyrimidinylimidazole inhibitors of CSBP/p38 kinase demonstrating decreased inhibition of hepatic cytochrome P450 enzymes. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 3111–6.
- (26) Bolos, J. Structure-activity relationships of p38 mitogen-activated protein kinase inhibitors. *Mini-Rev. Med. Chem.* **2005**, *5*, 857–68.
- (27) Goldstein, D. M.; Gabriel, T. Pathway to the clinic: Inhibition of P38 MAP kinase. A review of ten chemotypes selected for development. *Curr. Top. Med. Chem.* **2005**, *5*, 1017–29.
- (28) Regan, J.; Breitfelder, S.; Cirillo, P.; Gilmore, T.; Graham, A. G.; Hickey, E.; Klaus, B.; Madwed, J.; Moriak, M.; Moss, N.; Pargellis, C.; Pav, S.; Proto, A.; Swinamer, A.; Tong, L.; Torcellini, C. Pyrazole urea-based inhibitors of p38 MAP kinase: from lead compound to clinical candidate. *J. Med. Chem.* **2002**, *45*, 2994–3008.
- (29) Lee, M. R.; Dominguez, C. MAP kinase p38 inhibitors: Clinical results and an intimate look at their interactions with p38alpha protein. *Curr. Med. Chem.* **2005**, *12*, 2979–94.
- (30) Wilson, K. P.; Fitzgibbon, M. J.; Caron, P. R.; Griffith, J. P.; Chen, W.; McCaffrey, P. G.; Chambers, S. P.; Su, M. S. Crystal structure of p38 mitogen-activated protein kinase. *J. Biol. Chem.* **1996**, *271*, 27696–700.
- (31) Wang, Z.; Harkins, P. C.; Ulevitch, R. J.; Han, J.; Cobb, M. H.; Goldsmith, E. J. The structure of mitogen-activated protein kinase p38 at 2.1-A resolution. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2327–32.
- (32) Mouchlis, V. D.; Barbayianni, E.; Mavromoustakos, T. M.; Kokotos, G. The application of rational design on phospholipase A₂ inhibitors. *Curr. Med. Chem.* **2011**, *18*, 2566–82.
- (33) Mouchlis, V. D.; Magrioti, V.; Barbayianni, E.; Cermak, N.; Oslund, R. C.; Mavromoustakos, T. M.; Gelb, M. H.; Kokotos, G. Inhibition of secreted phospholipases A₂ by 2-oxoamides based on α -amino acids: Synthesis, in vitro evaluation and molecular docking calculations. *Bioorg. Med. Chem.* **2011**, *19*, 735–43.
- (34) Mouchlis, V. D.; Mavromoustakos, T. M.; Kokotos, G. Molecular docking and 3D-QSAR CoMFA studies on indole inhibitors of GIIA secreted phospholipase A₂. *J. Chem. Inf. Model.* **2010**, *50*, 1589–601.
- (35) Mouchlis, V. D.; Mavromoustakos, T. M.; Kokotos, G. Design of new secreted phospholipase A₂ inhibitors based on docking calculations by modifying the pharmacophore segments of the FPL67047XX inhibitor. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 107–15.
- (36) Afantitis, A.; Melagraki, G.; Koutentis, P. A.; Sarimveis, H.; Kollias, G. Ligand-based virtual screening procedure for the prediction and the identification of novel β -amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial Neural Networks. *Eur. J. Med. Chem.* **2011**, *46*, 497–508.
- (37) Melagraki, G.; Afantitis, A.; Sarimveis, H.; Iglessi-Markopoulos, O.; Koutentis, P. A.; Kollias, G. In silico exploration for identifying structure-activity relationship of MEK inhibition and oral bioavailability for isothiazole derivatives. *Chem. Biol. Drug. Des.* **2010**, *76*, 397–406.
- (38) Melagraki, G.; Afantitis, A. Ligand and structure-based virtual screening strategies for hit-finding and optimization of hepatitis C virus (HCV) inhibitors. *Curr. Med. Chem.* **2011**, *18*, 2612–9.
- (39) Melagraki, G.; Afantitis, A.; Sarimveis, H.; Koutentis, P. A.; Kollias, G.; Iglessi-Markopoulos, O. Predictive QSAR workflow for the in silico identification and screening of novel HDAC inhibitors. *Mol. Diversity* **2009**, *13*, 301–11.
- (40) Roy, K.; Mitra, I. On various metrics used for validation of predictive QSAR models with applications in virtual screening and

- focused library design. *Comb. Chem. High Throughput Screen* **2011**, *14*, 450–74.
- (41) Barreca, M. L.; Iraci, N.; Manfroni, G.; Cecchetti, V. Allosteric inhibition of the hepatitis C virus NS5B polymerase: in silico strategies for drug discovery and development. *Future Med. Chem.* **2011**, *3*, 1027–55.
- (42) Toropova, A. P.; Toropov, A. A.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. CORAL: Quantitative structure-activity relationship models for estimating toxicity of organic compounds in rats. *J. Comput. Chem.* **2011**, *32*, 2727–33.
- (43) Toropova, A. P.; Toropov, A. A.; Benfenati, E.; Gini, G. Simplified molecular input-line entry system and International Chemical Identifier in the QSAR analysis of styrylquinoline derivatives as HIV-1 integrase inhibitors. *Chem. Biol. Drug Des.* **2011**, *77*, 343–60.
- (44) Yan, A.; Chong, Y.; Wang, L.; Hu, X.; Wang, K. Prediction of biological activity of Aurora-A kinase inhibitors by multilinear regression analysis and support vector machine. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 2238–43.
- (45) Wen, J. H.; Zhong, K. J.; Tang, L. J.; Jiang, J. H.; Wu, H. L.; Shen, G. L.; Yu, R. Q. Adaptive variable-weighted support vector machine as optimized by particle swarm optimization algorithm with application of QSAR studies. *Talanta* **2011**, *84*, 13–8.
- (46) Fernandez, M.; Caballero, J.; Fernandez, L.; Sarai, A. Genetic algorithm optimization in drug design QSAR: Bayesian-regularized genetic neural networks (BRGNN) and genetic algorithm-optimized support vectors machines (GA-SVM). *Mol. Divers.* **2011**, *15*, 269–89.
- (47) Sinha, N.; Sen, S. Predicting hERG activities of compounds from their 3D structures: Development and evaluation of a global descriptors based QSAR model. *Eur. J. Med. Chem.* **2011**, *46*, 618–30.
- (48) Nicolotti, O.; Giangreco, I.; Introcaso, A.; Leonetti, F.; Stefanachi, A.; Carotti, A. Strategies of multi-objective optimization in drug discovery and development. *Expert Opin. Drug Discovery* **2011**, *6*, 871–884.
- (49) Sperandio da Silva, G. M.; Sant'Anna, C. M.; Barreiro, E. J. A novel 3D-QSAR comparative molecular field analysis (CoMFA) model of imidazole and quinazolinone functionalized p38 MAP kinase inhibitors. *Bioorg. Med. Chem.* **2004**, *12*, 3159–66.
- (50) Xiao, Z.; Varma, S.; Xiao, Y. D.; Tropsha, A. Modeling of p38 mitogen-activated protein kinase inhibitors using the Catalyst HypoGen and k-nearest neighbor QSAR methods. *J. Mol. Graph. Model.* **2004**, *23*, 129–38.
- (51) Ravindra, G. K.; Achaiah, G.; Sastry, G. N. Molecular modeling studies of phenoxypyrimidinyl imidazoles as p38 kinase inhibitors using QSAR and docking. *Eur. J. Med. Chem.* **2008**, *43*, 830–8.
- (52) Brugel, T. A.; Maier, J. A.; Clark, M. P.; Sabat, M.; Golebiowski, A.; Bookland, R. G.; Laufersweiler, M. J.; Laughlin, S. K.; Vanrens, J. C.; De, B.; Hsieh, L. C.; Mekel, M. J.; Janusz, M. J. Development of N-2,4-pyrimidine-N-phenyl-N'-phenyl ureas as inhibitors of tumor necrosis factor alpha (TNF-alpha) synthesis. Part 1. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3510–3.
- (53) Maier, J. A.; Brugel, T. A.; Clark, M. P.; Sabat, M.; Golebiowski, A.; Bookland, R. G.; Laufersweiler, M. J.; Laughlin, S. K.; Vanrens, J. C.; De, B.; Hsieh, L. C.; Brown, K. K.; Juergens, K.; Walter, R. L.; Janusz, M. J. Development of N-2,4-pyrimidine-N-phenyl-N'-alkyl ureas as orally active inhibitors of tumor necrosis factor alpha (TNF-alpha) synthesis. Part 2. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3514–8.
- (54) Clark, M.; Crammer, D. R. III; Van Opdenbosch, N. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (55) Schrödinger Suite 2009 Protein Preparation Wizard; Epik version 2.0, Schrödinger, LLC, New York, NY, 2009; Impact version 5.5, Schrödinger, LLC, New York, NY, 2009; Prime version 2.1, Schrödinger, LLC, New York, NY, 2009.
- (56) Epik, version 2.0, Schrödinger, LLC, New York, NY, 2009.
- (57) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A software program for pK(a) prediction and protonation state generation for drug-like molecules. *J. Comput.-Aided. Mol. Des.* **2007**, *21*, 681–91.
- (58) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (59) SYBYL/Biopolymer, molecular modeling software packages, version 8.0, 2007, Tripos Inc., 1699 South Hanley Rd., St. Louis, MO 63144-2917.
- (60) Mouchlis, V. D.; Michopoulos, V.; Constantinou-Kokotou, V.; Mavromoustakos, T.; Dennis, E. A.; Kokotos, G. Binding conformation of 2-oxoamide inhibitors to group IVA cytosolic phospholipase A₂ determined by molecular docking combined with molecular dynamics. *J. Chem. Inf. Model.* **2011**, *52*, 243–54.
- (61) SYBYL/Surflex-Dock, molecular modeling software packages, version 8.0, 2007, Tripos Inc., 1699 South Hanley Rd., St. Louis, MO 63144–2917.
- (62) Jain, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46*, 499–511.
- (63) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–46.
- (64) Klebe, G.; Abraham, U. Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 1–10.
- (65) Bohm, M.; St rzebecher, J.; Klebe, G. Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J. Med. Chem.* **1999**, *42*, 458–77.
- (66) Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.* **2009**, *11*, 10–18.
- (67) Zhao, Y.; Zhang, Y. Comparison of decision tree methods for finding active objects. *Adv. Space Res.* **2008**, *41*, 1955–1959.
- (68) Papa, E.; Kovarich, S.; Gramatica, P. Development, Validation and Inspection of the Applicability Domain of QSPR Models for Physicochemical Properties of Polybrominated Diphenyl Ethers. *QSAR Comb. Sci.* **2009**, *28*, 790–796.
- (69) Liu, H.; Yao, X.; Gramatica, P. The applications of machine learning algorithms in the modeling of estrogen-like chemicals. *Comb. Chem. High Throughput Screen.* **2009**, *12*, 490–6.
- (70) Zhang, S.; Golbraikh, A.; Oloff, S.; Kohn, H.; Tropsha, A. A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. *J. Chem. Inf. Model.* **2006**, *46*, 1984–95.
- (71) Frank, I.; Feikema, J.; Constantine, N.; Kowalski, B. Prediction of product quality from spectral data using the partial least-squares method. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 20–24.
- (72) Rännar, S.; Lindgren, F.; Geladi, P.; Wold, S. A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm. *J. Chemometrics* **1994**, *8*, 111–125.
- (73) Cramer, R. D. III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.–Act. Relat.* **1988**, *7*, 18–25.
- (74) Baumann, K. Cross-validation as the objective function for variable-selection techniques. *Trends Anal. Chem.* **2003**, *22*, 395–406.
- (75) Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Igglekiss-Markopoulou, O. A novel QSAR model for predicting induction of apoptosis by 4-aryl-4H-chromenes. *Bioorg. Med. Chem.* **2006**, *14*, 6686–94.
- (76) Nair, P. C.; Sobhia, M. E. Quantitative structure activity relationship studies on thiourea analogues as influenza virus neuraminidase inhibitors. *Eur. J. Med. Chem.* **2008**, *43*, 293–9.
- (77) Jalali-Heravi, M.; Asadollahi-Baboli, M.; Shahbazikhah, P. QSAR study of heparanase inhibitors activity using artificial neural networks

and Levenberg–Marquardt algorithm. *Eur. J. Med. Chem.* **2008**, *43*, 548–56.

(78) Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Iglessi-Markopoulou, O. A novel simple QSAR model for the prediction of anti-HIV activity using multiple linear regression analysis. *Mol. Diversity* **2006**, *10*, 405–14.

(79) Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Markopoulos, J.; Iglessi-Markopoulou, O. Development and evaluation of a QSPR model for the prediction of diamagnetic susceptibility. *QSAR Comb. Sci.* **2008**, *27*, 432–436.

(80) Melagraki, G.; Afantitis, A.; Sarimveis, H.; Koutentis, P.; Markopoulos, J.; Iglessi-Markopoulou, O. A novel QSPR model for predicting θ (lower critical solution temperature) in polymer solutions using molecular descriptors. *J. Mol. Model.* **2007**, *13*, 55–64.

(81) Tong, L.; Pav, S.; White, D. M.; Rogers, S.; Crane, K. M.; Cywin, C. L.; Brown, M. L.; Pargellis, C. A highly specific inhibitor of human p38 MAP kinase binds in the ATP pocket. *Nat. Struct. Biol.* **1997**, *4*, 311–6.

(82) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3228.

(83) Purcell, W. P.; Singer, J. A. A brief review and table of semiempirical parameters used in the Hueckel molecular orbital method. *J. Chem. Eng. Data* **1967**, *12*, 235–246.

(84) Novotarskyi, S.; Sushko, I.; Korner, R.; Pandey, A. K.; Tetko, I. V. A comparison of different QSAR approaches to modeling CYP450 1A2 inhibition. *J. Chem. Inf. Model.* **2011**, *51*, 1271–80.

(85) Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P. A.; Iglessi-Markopoulou, O.; Kollias, G. A combined LS-SVM & MLR QSAR workflow for predicting the inhibition of CXCR3 receptor by quinazolinone analogs. *Mol. Diversity* **2010**, *14*, 225–35.

(86) Perez-Garrido, A.; Helguera, A. M.; Borges, F.; Cordeiro, M. N.; Rivero, V.; Escudero, A. G. Two new parameters based on distances in a receiver operating characteristic chart for the selection of classification models. *J. Chem. Inf. Model.* **2011**, *51*, 2746–59.

(87) Du-Cuny, L.; Chen, L.; Zhang, S. A critical assessment of combined ligand- and structure-based approaches to HERG channel blocker modeling. *J. Chem. Inf. Model.* **2011**, *51*, 2948–60.

(88) SYBYL/QSAR and COMFA, molecular modeling software packages, version 8.0, 2007, Tripos Inc., 1699 South Hanley Rd., St. Louis, MO 63144-2917.

(89) Asadollahi, T.; Dadfarnia, S.; Shabani, A. M.; Ghasemi, J. B.; Sarkhosh, M. QSAR models for CXCR2 receptor antagonists based on the genetic algorithm for data preprocessing prior to application of the PLS linear regression method and design of the new compounds using in silico virtual screening. *Molecules* **2011**, *16*, 1928–55.