-Journal of Applied Chemical Research, 7, 1, 64-74 (2013)

Journal of Applied Chemical Research
www.jacr.kiau.ac.ir

Application of 3D-QSAR on a Series of Potent P38-MAP Kinase Inhibitors

Reihaneh Safavi-Sohi, Jahan B. Ghasemi*

Chemistry Department, Faculty of Sciences, K. N. Toosi University of Technology, Tehran, Iran.

(Received 22 May 2012; Final version received 14 Nov. 2012)

Abstract

One of the most applied methods in drug industry for development of new drugs is 3D-QSAR methodology. As p38-mitogen-activated protein kinase (p38-MAPK) plays a crucial role in regulating the production of such proinflammatory cytokines as tumor necrosis factor-α (TNF-α) and interleukin-1, emerging as an attractive target for new anti-inflammatory agents, we used a 3D-QSAR based method of Comparative Molecular Field Analysis (CoMFA) on a series of 52 potent p38-MAP kinase inhibitors with IC_{50} ranging from 3.2 to 10,000 nM. An alignment rule for the compounds was defined using Distill in SYBYL 7.3. The best model was validated using a data set obtained by dividing the data set into a training set and test set using hierarchical clustering, based on the CoMFA fields and biological activities (pIC_{50}). The best predictions were obtained with a CoMFA region-focusing model (R^2_{ncv} = 0.952, q^2 = 0.678, R^2_{Pred} = 0.627). The statistical parameters from the model indicate that the data are well fitted and has high predictive ability. Moreover, the resulting 3D CoMFA contour maps provide useful guidance for designing highly active inhibitors.

***Keywords*:** *CoMFA, P38-MAP kinase inhibitors, Rigid alignment, Anti-inflammatory drugs, Hierarchical clustering.*

Introduction

The mitogen-activated protein kinases (MAPKs) are essential regulators for signal transduction pathways and play crucial roles in cellular processes such as transcription, apoptosis, and differentiation [1]. The p38-MAP kinase is highly expressed in severe invasive breast cancers and is involved in the regulation of cytokine biosynthesis (TNFα), which is associated with chronic inflammatory diseases such as rheumatoid arthritis, Crohn's disease, and inflammatory bowel syndrome [2-

***Corresponding author:** Dr. Jahan B Ghasemi, Prof., Chemistry Department, Faculty of Sciences, K. N. Toosi University of Technology, Tehran, Iran. Email address: jahan.ghasemi@gmail.com, Tel: +98 21 22850266; fax: +98 21 22853650.*

4]. The correlation of elevated levels of TNF-α with the pathophysiology of a number of inflammatory diseases has been studied for over two decades [5-6]. An alternative therapeutic approach would be suppressing the production of TNF-α with small molecule inhibitors designed to block the p38-MAP kinase signal transduction cascade. Although in vivo models indicate p38 is a viable therapeutic target to reduce symptoms of RA, no small molecule therapy has successfully reached the market [2, 7]. Therefore, there is already an urgent need to search for identification and development of novel inhibitors particularly those selective for all isoforms of P38-MAP kinase.

Quantitative structure–activity relationship (QSAR) methods have now become an essential part of modern drug design, since they represent an inexpensive and fast choice to the medium throughput in vitro assays. One would say that today no drug is developed without previous QSAR analyses [8]. The QSAR methodology is based on the concept that the differences observed in the biological activity, including affinity (e.g., K_i), efficacy (e.g., activation or stimulation of receptors, V_{max} of enzymes), pharmacokinetics (ADME), drug–drug interactions, or any biological properties of a set of compounds can be quantitatively correlated with differences in their structural or physicochemical properties [9].

Typically, a 3D-QSAR analysis allows the identification of the pharmacophoric arrangement of molecular features in space and provides guidelines for the design of next-generation compounds with enhanced bioactivity or selectivity. The first applicable 3D-QSAR method was proposed by Cramer et al. in 1988. His program, CoMFA, was a major breakthrough in the field of 3D-QSAR [10].

Herein, we present a 3D-QSAR study to investigate the correlation of a series of quinolizin-2-one and pyridopyridazin-6-one derivatives with the inhibition of p38-MAP kinase by employing comparative molecular field analysis (CoMFA).

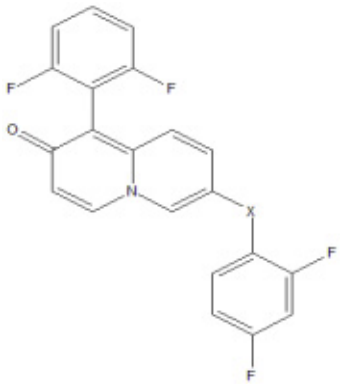
Experimental

Data set

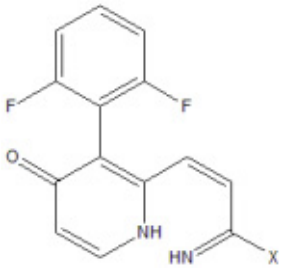
All P38-MAP Kinase inhibitors and their biological activities (IC_{50} values) were taken from Tynebor *et al.* [7]. They were divided into a training set of 35 compounds and a test set of 11 compounds from original data set by using hierarchical clustering based on the CoMFA fields and biological activities (pIC_{50}). The chemical structures and biological activity values of all of the compounds are presented in Table 1. The IC_{50} values in units of molarity (M) were transformed in pIC_{50} (–log IC_{50}) in order to provide numerically larger data values. The –log IC_{50} values of the training set span more than 3 orders of magnitude (5.00 – 8.49). The activity (IC_{50}) values of six molecules were qualitatively (i.e. they have not exact

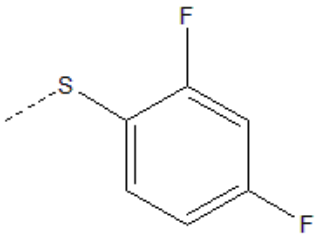
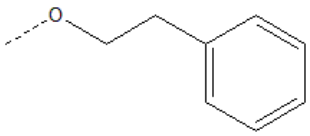
quantities for activities) higher than 10,000 nM, these molecules were removed from the data set. Molecules in the test set and those six molecules (which were removed at first) were used to evaluate the predictive quality of the QSAR model developed from the training set.

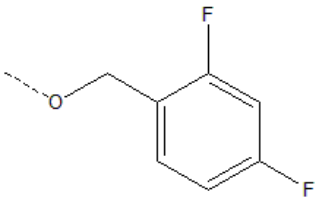
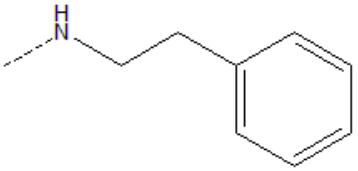
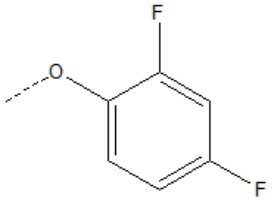
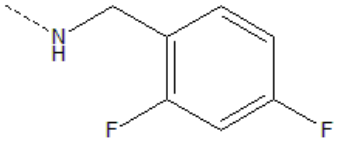
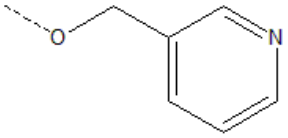
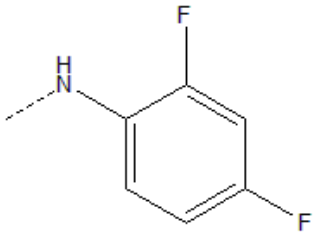
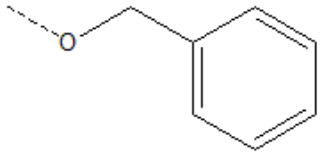
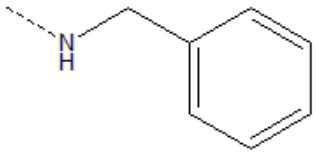
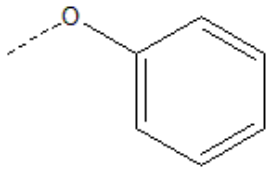
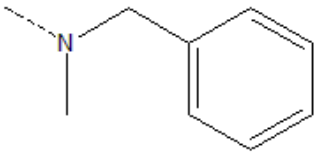
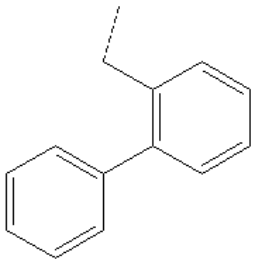
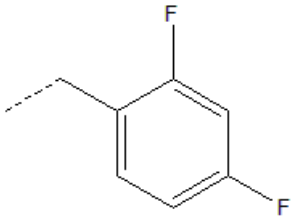
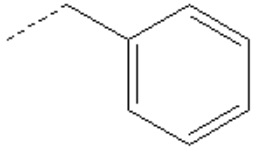
Table 1. Structures of training and test set compounds [a]

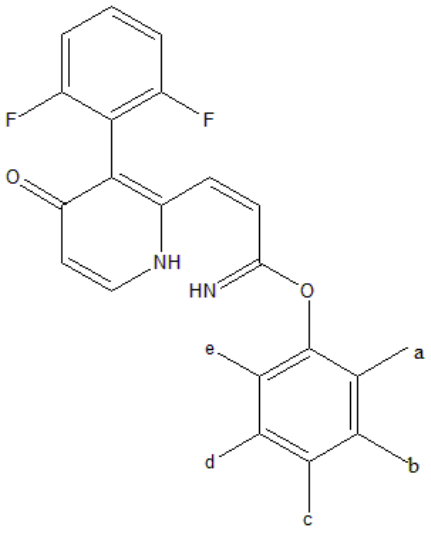


Comp.	X	pIC_{50}
1[b]	—	8.15
4	S	6.49
5	O	6.72
6[b]	CH_2	7.03
7	$NHCH_2$	6.06
8	C_2H_2	6.05
9	C_2H_4	6.31



Comp.	X	pIC_{50}	Comp.	X	pIC_{50}
17		8.49	**23**[b]		> 5

18[b]		6.14	**24**[b]		5.89
19		8.33	**25**[b]		6.14
20[b]		>5	**26**		8.46
21[b]		5.89	**27**		7.70
22[b]		6.80	**28**		6.19
29[b]		6.32	**31**		8.12
30[b]		5.84			



Comp.	a	B	C	d	e	pIC_{50}
32	F	—	—	H	H	7.92
33[b]	Cl	—	—	H	H	7.36
34	Me	—	—	H	H	7.04
35	OMe	—	—	H	H	7.09
36	CF_3	—	—	H	H	5.62
37	—	F	—	H	H	6.38
38	—	Cl	—	H	H	6.62
39	—	Me	—	H	H	7.15
40[b]	—	OMe	—	H	H	>5
41[b]	—	CF_3	—	H	H	>5
42	—	—	F	H	H	7.32
43	—	—	Cl	H	H	7.1
44	—	—	Me	H	H	6.59
45[b]	—	—	OMe	H	H	>5
46[b]	—	—	CF_3	H	H	>5
47	F	F	—	—	—	7.49
48[b]	Cl	Cl	—	—	—	7.49
49	F	—	Cl	—	—	7.96
50	Cl	—	F	—	—	7.85
51	Cl	Cl	—	—	—	7.52
52	F	—	—	F	—	6.89
53	Cl	—	—	Cl	—	6.55
54	F	—	—	—	F	7.02
55	Cl	—	—	—	Cl	7.05
56	F	—	F	—	F	6.70
57	Cl	—	F	—	Cl	7.17
58	F	F	F	F	F	5.37
59	—	F	F	—	—	6.57
60	—	Cl	F	—	—	6.85
61	—	Cl	Cl	—	—	6.59

[a] The structures of molecules are taken from Tynebor *et al*. [1]
[b] Prediction set

Computer hardware and software

All computation work was done on Ubuntu Linux 10.04 and windows 7 operating system. The molecular modeling software used were SYBYL 7.3 molecular modeling package (Tripose Inc., St. Louis, USA), and (ViewerLite 5.0, Accelrys, Inc., 2002).

Molecular modeling and alignment

All the 3D structures of the compounds were sketched and processed using the CONCORD module [11, 12] of SYBYL 7.3. CONCORD quickly generates approximate structures for any compound that consists of H, C, N, O, F, Si, P, S, Cl, Br, I, Mg, B, and Se (Tripos Bookshelf 7.2, TRIPOS, Inc. St. Louis, Missouri). Information contained in a molecule's atomic connection table was converted into atomic coordinates. The structures were then subjected to energy minimization using tripos force field with a distance-dependent dielectric and the Powell conjugate gradient algorithm convergence criterion of 0.01 kcal/mol Å [13] and partial atomic charges were calculated by Gasteiger Hückel method. Compound **17** was used as a template (reference molecule) because of the highest activity and all other compounds were aligned on the basis of the common structure. Rigid body alignment of molecules in a Mol2 database was performed using maximum common substructures defined by Distill (without including bond types in rings). The common substructure between all compounds and aligned compounds on reference compound, **17**, are displayed in Figures 1A and 1B.

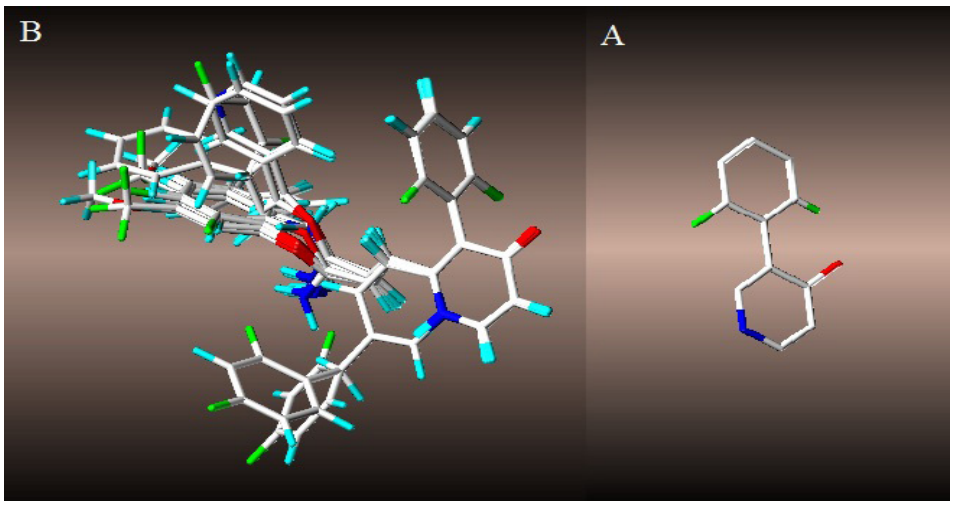


Figure 1. Common substructure between all compounds (A), Alignment of training and test set compounds on compound **17** (B).

Hierarchical cluster analysis

Crucial requisite in the development of a QSAR model is the division of the whole data set in training and test sets in order to maximize the diversity of the test set and to examine the predictive accuracy of the model

when extrapolating outside the training set [14, 15]. The chemometric technique of hierarchical cluster analysis (HCA) [16] was the approach used to remove outliers and then select the training and test sets based on structural similarities, CoMFA steric and electrostatic fields and biological activity data (expressed as pIC_{50}). As the name suggests, HCA attempts to find groupings within a set of data. This method was applied as implemented in SYBYL 7.3 (Tripos).

CoMFA study

The aligned sets of molecules were positioned inside a grid box with grid spacing value of 2.0 Å (default distance) in all Cartesian directions and CoMFA fields were calculated using the QSAR module of SYBYL. The interaction energies for each molecule were calculated at each grid point using a probe atom: an sp3 hybridized carbon atom with a +1.0 charge. The steric (vdW interaction) and electrostatic (Coulombic values with a 1/r distance-dependent dielectric function) fields were calculated at each intersection on the regularly spaced grid. Since CoMFA models are greatly sensitive to the different space orientations of the molecular collective with respect to the lattice, all-orientation search was also carried out on initial orientations of aligned structures by the rotation procedure written in SYBYL programming language [17].

In order to reduce noise and improve efficiency, several values for column filtering (minimum sigma) were examined from 1.0 to 2.0 kcal/mol, excluding from the analysis those columns (lattice points) whose energy variance was less than these values. Cutoff for both steric and electrostatic fields was set to 20 kcal/mol.

CoMFA standard scaling applies the same weight to data from each lattice point in any given field. Region focusing is an iterative procedure which refines a model by increasing the weights for those lattice points which are most pertinent to the model. This enhances the resolution and predictive power (q^2; cross validated R^2) of a subsequent PLS analysis. Technically, this corresponds to rotating the model components through a high-order space [18]. PLS region focusing is intellectually analogous to the GOLPE approach and q^2-GRS [19, 20].

The leave-one-out cross-validation method using the SAMPLS method was employed. The optimal number of components for the final 3D-QSAR equation was defined as the number of components leading to the highest cross-validated R^2 (q^2) and the lowest standard error of prediction. The cross-validated coefficient q^2 (or r^2_{CV}) was evaluated as:

$$q^2 = 1 - \frac{\sum_{i=1}^{n}(Y_{Pridcited} - Y_{Observed})^2}{(Y_{Observed} - Y_{mean})^2}$$

where $Y_{Predicted}$, $Y_{Observed}$, Y_{mean} are the predicted, observed, and mean values of the pIC_{50},

respectively.

Results and discussion

The best model of all orientation search procedure was picked up, and then the effect of changing column filtering, cutoff value was investigated. The best results obtained at a column filtering of 1.8 kcal/mol and cutoff value 20 kcal/mol. Use of region focusing on the model yielded values R^2 = 0.952, F = 75.644, a standard error of estimation (SEE) of 0.191, q^2_{LOO} = 0.678 with 6 components, a standard error of prediction (SEP) of 0.541 and R^2_{Pred} = 0.627. The correlation between the predicted activities and the experimental activities are depicted in Figure 2. The experimental and predicted activities for training and test set compounds with CoMFA-RF are shown in Table 2. Interestingly the activities of 6 molecules which have not exact quantities were predicted properly by reported model (see Table 2 and Figure 2). As it can be seen, the model exhibited good predictive power confirmed by suitable values of cross validated correlation coefficient (q^2_{LOO}) and R^2_{Pred} .

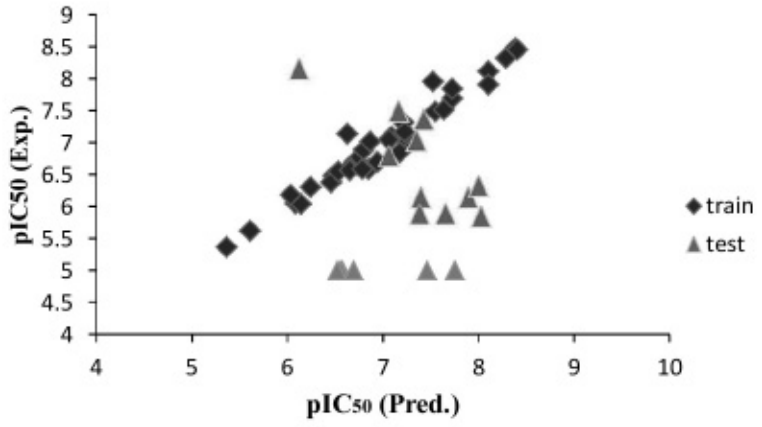


Figure 2. Experimental versus predicted activities for the training and test set compounds based on the CoMFA-RF model.

Table 2. The experimental pIC_{50} values, predicted pIC_{50} values (Pred.) of the training and test set compounds.

Comp. no	Exp.	Pred.	Comp. no	Exp.	Pred.
1*	8.15	6.12	**36**	5.62	5.61
4	6.49	6.48	**37**	6.38	6.45
5	6.72	6.71	**38**	6.62	6.65
6*	7.03	7.35	**39**	7.15	6.62
7	6.06	6.08	**40***	**>5**	**6.57**
8	6.05	6.14	**41***	**>5**	**6.56**
9	6.31	6.24	**42**	7.32	7.21
17	8.49	8.38	**43**	7.1	7.08
18*	6.14	7.39	**44**	6.59	6.84

19	8.33	8.28	**45***	**>5**	**6.69**
20*	**>5**	**7.75**	**46***	**>5**	**6.52**
21*	5.89	7.38	**47**	7.49	7.54
22*	6.8	7.06	**48***	7.49	7.16
23*	**>5**	**7.46**	**49**	7.96	7.52
24*	5.89	7.65	**50**	7.85	7.72
25*	6.14	7.89	**51**	7.52	7.63
26	8.397	8.46	**52**	6.89	6.79
27	7.719	7.7	**53**	6.55	6.53
28	6.034	6.19	**54**	7.02	6.86
29*	6.32	8.00	**55**	7.05	7.06
30*	5.84	8.02	**56**	6.7	6.94
31	8.12	8.1	**57**	7.17	7.22
32	7.92	8.1	**58**	5.37	5.36
33*	7.36	7.42	**59**	6.57	6.65
34	7.04	7.26	**60**	6.85	7.17
35	7.09	7.2	**61**	6.59	6.78

*Prediction set

The CoMFA-RF's steric and electrostatic field obtained from the best non-cross-validated analysis is shown as 3D colored contour maps in Figure 3. The field energies at each lattice point were calculated as the scalar results of the coefficient and the standard deviation associated with a particular column of the data table (SD*coeff), as always plotted as the percentages of the contribution of CoMFA equation. These maps show regions where differences in molecular fields are associated with differences in biological activity. The steric interactions are represented by green and yellow colored contours, while electrostatic interactions are represented by red and blue colored contours.

The contour maps of CoMFA show contribution for favorable and unfavorable interactions with the receptor in terms of steric (80% green, 20% yellow) and electrostatic (80% blue and 20% red). Greater values of bio-activity are correlated with more bulk near green, less bulk near yellow more positive charge near blue and more negative charge near red. The most potent compound **17** is overlaid on the map to help better visualization (Figure 3A). P38α crystal structure shows the active site is comprised of two hydrophobic pockets. The previous researches on quinolizin-2-one and pyridopyridazin-6-one derivatives as potent inhibitors pointed to that C ring and A ring (Figure 4) should be to occupy the hydrophobic pocket I and II respectively [2, 7].

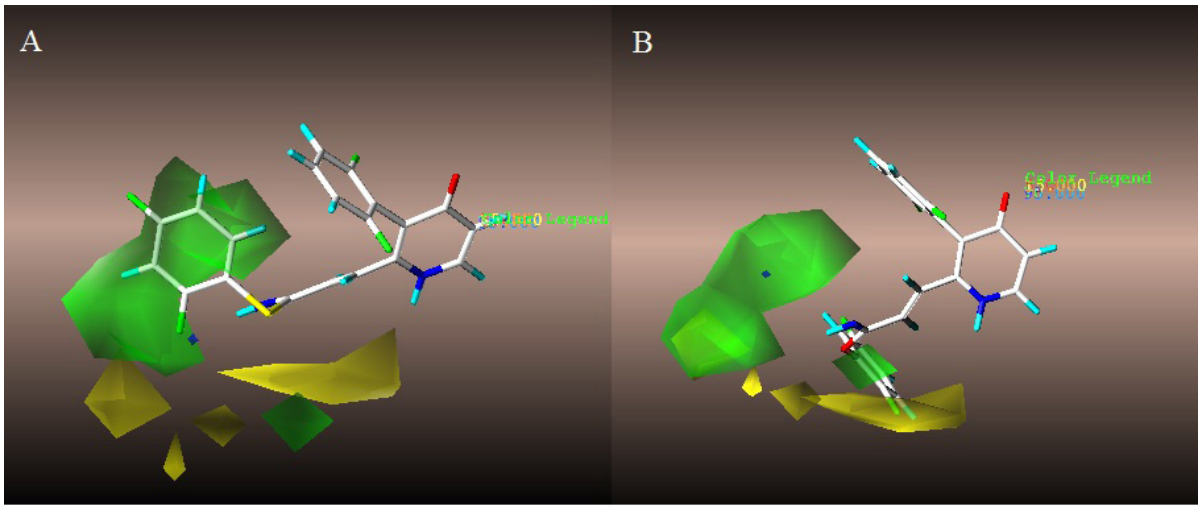


Figure 3. Steric and electrostatic contour maps of CoMFA-RF model in combination with compound **17** (the most active compound in data set) (A), and one of the inactive compounds **41** (B).

A large green contour near C ring indicates a sterically bulky group is favored in this region, which contributes to the ligand stabilization at the binding site to fill hydrophobic pocket II. In compounds **20** and **33**, the most inactive compounds, oriented away from the green contour and do not occupy hydrophobic pocket II. A ring is a part of the scaffold, so this interaction would be present in all molecules and it does not appear in the CoMFA contours. The yellow contours are inversely related to the biological activities. Unfavorable effect might be correlated to bulky group, thus, introduction of bulky substituents in these regions are predicted to decrease activity. Overlay of all compounds on the CoMFA map shows that meta or para substituent of C ring in **40**, **41**, **45**, and **46** is oriented toward the yellow regions where the bulky groups are unfavored, leading to decrease the activity (see Figure 3B). On the other hand, substituents of **41** and other compounds are located relatively away from this region.

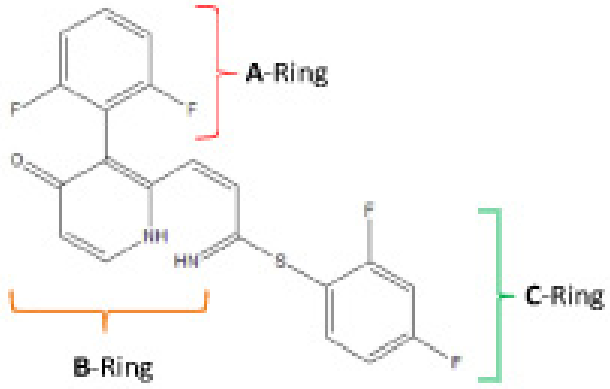


Figure 4. Three structure parts of the compounds in data set depicted on **17**.

Conclusion

In this study, by using the alignment scheme generated from Distill, a predictive CoMFA region focusing model was developed and was used to predict the pIC50 activity of a set of MAP-kinase inhibitors. The QSAR model gave good statistical results in terms of q^2 and R^2 values, and has been validated using a test set, obtained from the hierarchical clustering. From this study, it is possible to predict the ligand activities of newly designed MAP-kinase inhibitors, and design better inhibitors.

References

[1] G. Pearson, *Endocr. Rev*., 22, 153 (2001).

[2] R. M. Tynebor, *Bioorg. Med. Chem. Lett*., 20, 2765 (2010).

[3] M. A. Palladino, F. R. Bahjat, E. A. Theodorakis, L. L. Moldawer, *Nat. Rev. Drug. Discov.* 2, 736 (2003).

[4] *L. Chen, J.A. Mayer, T.I. Krisko, C. W. Speers, T. Wang, S. G. Hilsenbeck. Cancer Res*., 69, 8853 (2009).

[5] B. Wu, *J. Med. Chem*., 5,6398 (2010).

[6] C. A. Dinarello, *Cell*, 140, 935 (2010).

[7] R. M. Tynebor, *Bioorg. Med. Chem. Lett*., 21, 411 (2010).

[8] C. H. Andrade, K. F. M. Pasqualoto, E. I. Ferreira, A. J. Hopfinger, *Molecules*, 15, 3281 (2010).

[9] J. Shim, A. D. MacKerell *Med. Chem. Commun* 43, 122 (2011).

[10] R. D. Cramer, D. E. Patterson, J. D. Bunce, *J. Am. Chem. Soc*., 110, 5959 (1988).

[11] A. Rusinko, *J. Chem. Inf. Model*., 29, 251 (1989).

[12] E. M. Ricketts, *J. Chem. Inf. Model*, 33, 905 (1993).

[13] M. Clark, R. D. Cramer III, N. Van Opdenbosch, *J. Comp. Chem*., 10, 982 (1989).

[14] G. W. Milligan, M. C. Cooper, *Multivar. Behav. Res*., 21, 441 (1986).

[15] Y. Zhao, G. Karypis, U. Fayyad, *Data Min. Knowl. Disc*., 10, 141 (2005).

[16] HCA Manual, SYBYL 7.3, Tripos, St. Louis, MO (2006).

[17] R. Wang, Y. Gao, L. Liu, L. Lai, *J. Mol. Model*, 4, 276 (1998).

[18] QSAR ™ Manual, SYBYL 7.3, Tripos, St. Louis, MO, USA.

[19] S. J. Cho, A. Tropsha, *J. Med. Chem*., 38, 1060 (1995).

[20] M. Baroni, *Quant. Struct. Act. Relat*., 12, 9 (1993).