

# In Silico Exploration for Identifying Structure–Activity Relationship of MEK Inhibition and Oral Bioavailability for Isothiazole Derivatives

Georgia Melagraki<sup>1,\*</sup>, Antreas Afantitis<sup>2</sup>, Haralambos Sarimveis<sup>3</sup>, Olga Igglessi-Markopoulou<sup>3</sup>, Panayiotis A. Koutentis<sup>1</sup> and George Kollias<sup>4</sup>

<sup>1</sup>Department of Chemistry, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus

<sup>2</sup>Department of Chemoinformatics, NovaMechanics Ltd, Nicosia, Cyprus

<sup>3</sup>School of Chemical Engineering, National Technical University of Athens, Athens, Greece

<sup>4</sup>Biomedical Sciences Research Center 'Alexander Fleming', Athens, Greece

\*Corresponding author: Georgia Melagraki, georgiamelagraki@gmail.com

**In this study, quantitative structure–activity/property models are developed for modeling and predicting both MEK inhibitory activity and oral bioavailability of novel isothiazole-4-carboxamides. The models developed are thoroughly discussed to identify the key components that influence the inhibitory activity and oral bioavailability of the selected compounds. These selected descriptors serve as a first guideline for the design of novel and potent MEK inhibitors with desired ADME properties.**

**Key words:** ADME, *in silico* screening, isothiazole, MEK inhibitor, oral bioavailability, QSAR

Received 3 January 2010, revised 26 June 2010 and accepted for publication 22 August 2010

A goal of the modern cancer therapy is to identify molecules in signal transduction pathways that affect cell growth, and particularly those that cause a normal cell to become cancerous. Defects in the RAS/RAF/MEK/ERK signaling pathway are closely associated with the development of human tumors, such as melanoma, colon, lung, and thyroid cancers. MEK is a key component of the RAS/RAF/MEK/ERK signaling pathway that regulates cell proliferation and survival (1,2). This pathway has emerged as a significant focus for molecular-targeted cancer therapy, and MEK inhibitors have the potential for broad utility in the treatment of human cancers driven by activation of this pathway (3,4).

Big pharmaceutical companies have shown great interest in the area, and several MEK inhibitors have now entered clinical development. MEK inhibitors are the first highly selective inhibitors of the MAPK (Mitogen Activated Protein Kinase) pathway to enter the clinic, while data show promising hints that suppression of the MAPK pathway can be achieved within acceptable toxicity levels. The objective of many studies is the identification of small molecule, potent MEK inhibitors that are beneficially applied to the treatment of certain forms of cancer, with preference given to those molecules that possess good ADME (Absorption Distribution Metabolism Excretion) properties (5–7).

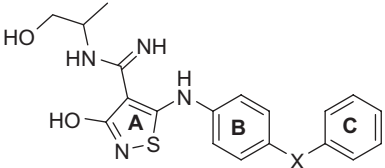
We have used both MEK inhibition and oral bioavailability data for 47 isothiazole-4-carboxamides (8). Isothiazole derivatives constitute a very important group of biologically active compounds that have recently gained attention as potent MEK inhibitors (9). Our aim is the identification of novel series of potent MEK inhibitors with favorable oral bioavailability. ADME properties and especially oral bioavailability are crucial for a molecule to be considered as an orally active drug. Oral bioavailability is one of the most difficult obstacles that a drug candidate faces. It is estimated that poor oral bioavailability is responsible for 40% of all drug failures. As such, it is very important to identify among active compounds those that are more likely to be orally active in humans (10).

We have developed quantitative structure–activity relationship (QSAR) models and searched for the most important descriptors that would help in the optimization of chemical structures as far as maximizing the activity and optimizing the ADME properties are concerned. The high oral bioavailability prediction will be used as an additional criterion for distinguishing among the most active compounds.

## Materials and Methods

### Data set

The chemical, biological, and bioavailability data that were used in this study have been presented by El Abdellaoui *et al.* (8) in their recent work (Tables 1–3). Different descriptors have been calculated to account for chemical, physicochemical, electronic, and quantum characteristics of compounds. All the descriptors were calculated using Chem3D<sup>a</sup>, MOPAC2007<sup>b</sup>, and Topix<sup>c</sup>. After removing useless descriptors by using the unsupervised attribute filter provided by Weka (11) in total 141 physicochemical constants, topological and structural descriptors (Table 4) were finally consid-

**Table 1:** Observed and Predicted values for the training and test set (compounds 1–7)


Compound	X	log(1/IC <sub>50</sub> )* observed	log(1/IC <sub>50</sub> ) predicted	log(1/AUC)* observed	log(1/AUC) predicted
1 <sup>a</sup>	O	-1.519	-1.599	-3.055	-3.060
2 <sup>b</sup>	NH	-2.380	-1.751	-3.092	-2.871
3	CH <sub>2</sub>	-1.763	-1.790	-2.342	-2.533
4 <sup>a</sup>	C=O	-2.079	-1.737	-1.778	-2.946
5	S	-1.653	-1.650	-2.423	-2.420
6	SO <sub>2</sub>	-2.176	-2.341	-3.628	-2.575
7 <sup>b</sup>	C(CH <sub>3</sub> ) <sub>2</sub>	-3.898	-3.897	-3.371	-2.150

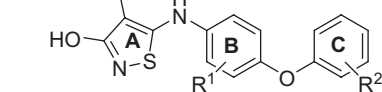
\*IC<sub>50</sub> (nM), oral-AUC (ng h/ml).<sup>a</sup>test set for inhibition activity.<sup>b</sup>test set for oral bioavailability.

ered as possible input candidates to the model. Before the calculation of the descriptors, the structures were fully optimized using PM6 method in MOPAC2007 suite that, as proposed in literature, offers a good balance between computational speed and accuracy (12,13).

### Variable selection – support vector machine regression

Before running the modeling methodology, the most significant attributes among the 141 available were preselected by using CfsSubset variable selection and BestFirst evaluator, which are included in Weka (11). Correlation-based feature subset selection (CfsSubset) algorithm evaluated the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that were highly correlated with the class while having low intercorrelation were preferred. The attribute selection mode was set to 10-fold cross-validation.

A great variety of the machine learning methods have been applied in QSAR studies (14–18), and the best approach for a specific

**Table 2:** Observed and Predicted values for the training and test set (compounds 8–32)


Compound	Ring-C R <sup>2</sup>	Ring-B R <sup>1</sup>	log(1/IC <sub>50</sub> )* observed	log(1/IC <sub>50</sub> ) predicted	log(1/AUC)* observed	log(1/AUC) predicted
8 <sup>a</sup>	6-F <sub>3</sub> C	H	-1.643	-1.809	-4.457	-3.929
9 <sup>b</sup>	5-F-6-F <sub>3</sub> C	H	-1.763	-1.764	-4.114	-4.171
10	5-Cl-6-F <sub>3</sub> C	H	-2.079	-1.805	-4.310	-4.313
11	5-Cl-8-F <sub>3</sub> C	H	-2.732	-2.017	-4.340	-4.362
12	6-F-8-F <sub>3</sub> C	H	-2.146	-1.808	-4.187	-4.147
13 <sup>a</sup>	6-Me <sub>2</sub> N	H	-2.204	-1.837	-2.954	-2.989
14	5-Cl-6-F <sub>3</sub> C	2-Cl	-1.826	-1.709	-3.757	-4.362
15	5-Cl-6-F <sub>3</sub> C	2-F	-1.716	-1.716	-3.653	-4.415
16 <sup>b</sup>	5-F-8-F <sub>3</sub> C	H	-1.778	-1.898	-4.495	-4.174
17	5,8-Cl <sub>2</sub>	H	-1.477	-1.622	-4.676	-3.785
18 <sup>a</sup>	5-Cl-8-F	H	-1.903	-1.637	-2.954	-4.044
19	5,8-Cl <sub>2</sub>	2-F	-1.415	-1.515	-4.347	-4.284
20 <sup>b</sup>	5,6-Cl <sub>2</sub>	H	-1.875	-1.575	-3.721	-3.818
21	5,6-F <sub>2</sub>	H	-1.301	-1.575	-3.974	-3.973
22	5,6-Cl <sub>2</sub>	2-Cl	-1.431	-1.458	-3.631	-3.862
23	5-F-6-Cl	H	-2.161	-1.579	-4.099	-4.094
24	6,8-Cl <sub>2</sub>	H	-1.301	-1.628	-4.068	-3.737
25	6,8-F <sub>2</sub>	H	-2.316	-1.628	-3.711	-3.885
26	5-Cl-7-F	H	-1.929	-1.645	-3.796	-4.049
27	5-Cl-7-F	2-Cl	-2.029	-1.552	-3.755	-4.109
28 <sup>b</sup>	5,7-Cl <sub>2</sub>	H	-1.806	-1.644	-4.419	-3.790
29	5,7-F <sub>2</sub>	H	-2.316	-1.644	-4.217	-3.949
30 <sup>a</sup>	5-Cl-6,8-F <sub>2</sub>	H	-1.588	-1.609	-4.308	-4.306
31 <sup>b</sup>	5-Cl-6,8-F <sub>2</sub>	2-F	-1.431	-1.455	-4.174	-4.522
32	5-Cl-7-F	1-Me	-1.756	-1.618	-4.104	-3.877

\*IC<sub>50</sub> (nM), oral-AUC (ng h/ml).<sup>a</sup>test set for inhibition activity.<sup>b</sup>test set for oral bioavailability.

**Table 3:** Observed and Predicted values for the training and test set (compounds 33–47)

Compound	X	log(1/IC <sub>50</sub> )* observed	log(1/IC <sub>50</sub> ) predicted	log(1/AUC)* observed	log(1/AUC) predicted
33 <sup>a</sup>		-1.505	-1.606	-3.459	-3.237
34		-1.602	-1.976	-2.790	-2.794
35		-2.167	-2.165	-3.836	-2.747
36		-1.531	-1.834	-3.296	-2.932
37		-2.137	-2.131	-2.418	-2.652
38 <sup>b</sup>		-2.474	-2.309	-3.384	-2.467
39		-1.643	-1.981	-1.940	-2.859
40		-1.663	-1.406	-4.163	-3.911
41 <sup>a</sup>		-1.663	-1.609	-4.096	-3.861
42		-1.544	-1.689	-3.708	-3.942
43		-1.531	-1.794	-3.305	-3.612
44		-0.903	-1.583	-4.379	-3.864
45 <sup>a</sup>		-2.872	-1.973	-4.369	-3.495
46 <sup>b</sup>		-1.602	-1.820	-4.163	-3.689
47		-1.531	-1.789	-2.628	-2.948

\*IC<sub>50</sub> (nM), oral-AUC (ng h/ml).

<sup>a</sup>test set for inhibition activity.

<sup>b</sup>test set for oral bioavailability.

**Table 4:** Physicochemical, topological, and structural descriptors

Chem3D	TOPIX	MOPAC	
MolWeight	NSglBnd	Wiener_1	Heat of formation
H Bond	NCCSgl	Dim	TotalE
Acceptors			
H Bond Donors	NDbIBnd	Bertz	ElectrE
SAS	Chi0 – Chi9	AtomCompMean	Core–core repulsion
MS	ChiInf0 ChiInf9	AtomCompTot	Dipole
SEV	Ki0 – Ki8	Zagreb1	No of filled levels
Ovality	KiInf0 – KiInf8	Zagreb2	Ionization potential
MR	ChiCl3	Quadr	HOMO
CLogP	ChiCl4	ScHultz	LUMO
Blndx	ChiCP30 – ChiCP39	Kappa1	
ClcC	ChiCP40 – ChiCP49	Kappa2	
MTI	KiCl3	Kappa3	
NRBo	KiCl4	WienerDistCode	
Polar Surface	KiCP30 – KiCP39	InfWiener	
ShA	KiCP40 – KiCP49	DistEqMean	
Wiener	ChiMod	DistEqTotal	
SDe	Xu1 – Xu3	InfMagnitDistTot	
SVDe	TopoJ	Polarity	
TCon	TopoRad	Gordon	
	TopoDia		
	NBranch		
	NRings		

problem needs to be explored. We considered Support Vector Machine regression (SVMreg) methodology as the machine learning method combined with the variable selection method previously described for performing regression to our available dataset. Support Vector Machine (SVM) was proposed in 1963 by Vapnik *et al.*, (19) and was shown as an effective tool for solving classification and regression problems.

For this work, we have used SVMreg methodology using the polynomial kernel with the complexity parameter *c* equal to 1. The training data were normalized, and the learning algorithm Reg-SMOimproved was chosen (20).

### Model validation

The models that performed best in predicting MEK inhibition and oral bioavailability were validated using external validation and cross-validation methods (21,22). External validation was applied, by randomly splitting the dataset into training and validation set in a proportion of 4:1. Thirty-eight compounds were used in the training set, and the remaining nine were used in the test set and were not involved by any means in the training procedure. Regarding cross-validation, both 10-fold cross-validation and leave-one-out (LOO) cross-validation methods were applied. The following statistical criteria were used to assess the robustness, reliability, and predictive activity of the model: the coefficient of determination between experimental values and model predictions ( $R^2$ ) and root mean square error.

### Applicability domain

Similarity measurements were used to define the domain of applicability of the two models based on the Euclidean distances among

all training compounds and the test compounds (23,24). The distance of a test 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:

$$APD = \langle d \rangle + Z\sigma$$

Calculation of  $\langle d \rangle$  and  $\sigma$  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  $\sigma$  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.

## Results and Discussion

The original dataset of 47 compounds was randomly partitioned into training and validation set consisting of 38 and 9 compounds, respectively. This random partitioning was performed separately for MEK inhibition activity and oral bioavailability, resulting in different training and test sets for the two end-points. The test compounds for MEK inhibition activity and oral bioavailability are clearly indicated in Tables 1–3.

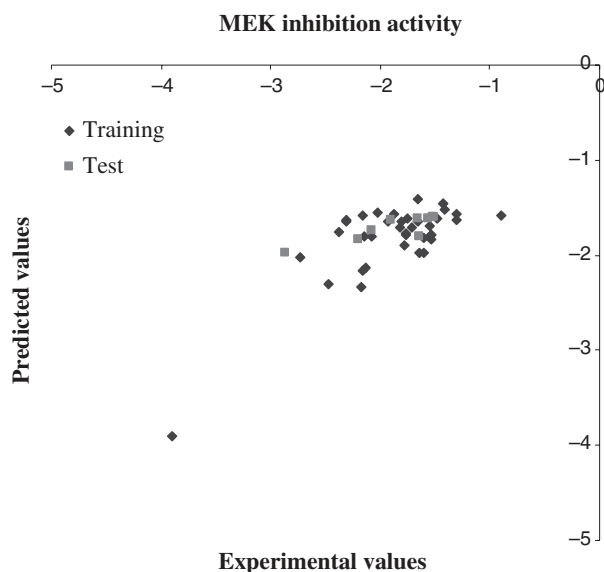
The CfsSubset variable selection with BestFirst evaluator method (which is included in WEKA platform) was then applied on the training data to select the most significant, among the 141 available descriptors. Six descriptors (namely NCCSgl, KiInf7, KiInf8, KiCP44, AtomCompMean, and AtomCompTot) were selected as most important to describe the MEK inhibition activity, and seven descrip-

**Table 5:** MEK inhibition activity model – Weights for normalized descriptors

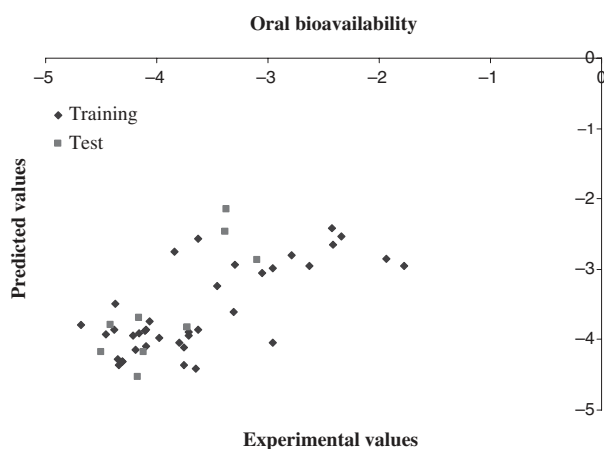
Weight	Descriptor
–0.111	NCCSgl
+0.2019	KiInf7
–0.0851	KiInf8
–0.5836	KiCP44
–0.0601	AtomCompMean
–0.2721	AtomCompTot
+0.8161	(constant)

**Table 6:** Oral bioavailability – Weights for normalized descriptors

Weight	Descriptor
–0.1976	ChiInf0
+0.0783	Ki8
+0.0932	ChiCP49
+0.3101	KiCP36
–0.1656	TopoJ
–0.6031	AtomCompMean
+0.0588	Dipole
+0.8	(constant)



**Figure 1:** Experimental versus Predicted values for MEK inhibition values.



**Figure 2:** Experimental versus Predicted values for oral bioavailability.

tors (dipole, TopoJ, ChiInf0, KiCP36, Ki8, ChiCP49, and AtomComp-Mean) were selected for oral bioavailability. The training data (selected descriptors and end-point values) were normalized in the range  $[-1,1]$ , before the final step of the modeling procedure, which was the application of the SVMreg methodology. The produced models predict the normalized end-point value of a specific compound (MEK inhibition activity or oral bioavailability) as a linear combination of the normalized values of the respective descriptors.

**Table 7:** Statistical measures

	$R^2_{\text{train}}$	$\text{RMS}_{\text{train}}$	$R^2_{\text{CV10fold}}$	$\text{RMS}_{\text{CV10fold}}$	$R^2_{\text{CVLOO}}$	$\text{RMS}_{\text{CVLOO}}$	$R^2_{\text{test}}$	$\text{RMS}_{\text{test}}$
MEK inhibition activity model	0.76	0.33	0.63	0.39	0.61	0.40	0.86	0.36
Oral bioavailability model	0.73	0.51	0.57	0.63	0.59	0.61	0.82	0.59

RMS, root mean square error.

The coefficients (weights) of the two linear models are shown in Tables 5 and 6.

The experimental values and the model predictions for both training and validation examples are shown in Tables 1–3. Figures 1 and 2 present plots of experimental versus predicted values of MEK inhibition activity and oral bioavailability, respectively. Validation of the models was performed using the techniques mentioned in the previous section. The corresponding statistics, presented in Table 7, illustrate the accuracy, significance, and robustness of the produced models.

The chemical meaning of the descriptors used in the development of each model is briefly discussed later.

Information indices (ChiInf0, KiInf7, KiInf8, AtomCompMean, and AtomCompTot) encode information on the adjacency and distance of atoms and the atomic composition in the molecular structure (25–29). Topological information indices (ChiInf0, KiInf7, KiInf8) are graph theoretical invariants that view the molecular graph as a source of different probability distributions to which the information theory is applied. These indices have several advantages such as unique representation of the compound and high discriminating power (isomer discrimination). Information connectivity (Chinf0, KiInf7, and KiInf8) indices are based on the partition of the edges in the graph according to the equivalence and the magnitudes of their edge connectivity values.

Total information content on atomic composition  $I_{\text{AC}}$  was calculated from the complete molecular formula, hydrogen included, using the following equation:

$$I_{\text{AC}} = A^h \cdot \log_2 A^h - \sum_g A_g \cdot \log_2 A_g$$

The mean information content on atomic composition  $\bar{I}_{\text{AC}}$  is the mean value of the total information content and can be calculated using the following formula:

$$\bar{I}_{\text{AC}} = - \sum_g \frac{A_g}{A^h} \cdot \log_2 \frac{A_g}{A^h} = - \sum_g p_g \cdot \log_2 p_g$$

where  $A^h$  is the total number of atoms (hydrogen included),  $A_g$  is the number of equal-type atoms in the  $g$ th equivalence class, and  $p_g$  is the probability of randomly selecting a  $g$ th type atom (30–34).

Dipole (DPL) is the electric dipole moment. The electric dipole is a vector quantity, which encodes displacement with respect to the center of gravity of positive and negative charges in a molecule. The DPL encodes information about the charge distribution in molecules and is important for modeling polar interactions. Large substituents decrease DPL value, which is not desirable.

**Table 8:** Test set applicability domain for MEK inhibition activity

Compound	Applicability domain (APD = 10.9564)
1	0.3900
4	2.0820
8	0.2458
13	1.2363
18	0.0807
30	0.1589
33	1.0119
41	0.2866
45	3.1695

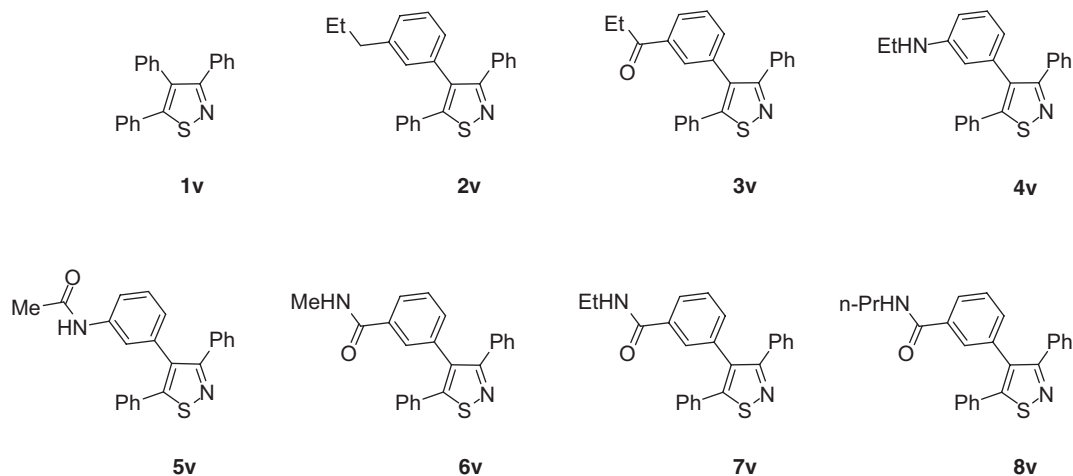
**Table 9:** Test set applicability domain for oral bioavailability

Compound	Applicability domain (APD = 0.8807)
2	0.1062
7	0.3793
9	0.1052
16	0.1084
20	0.0768
28	0.0760
31	0.2513
38	0.2832
46	0.1230

Topological indices such as Balaban index (TopoJ) and Ki8 are based on the two-dimensional representation of the molecule and give information about the atomic composition of a compound, the presence and character of chemical bonds, and the connectivity between atoms. Balaban index is one of the most discriminating molecular descriptors, and it is defined in terms of sums over each row of the distance matrix (35).

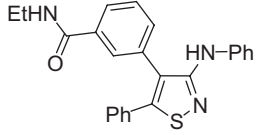
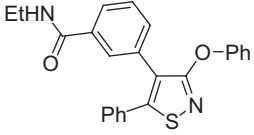
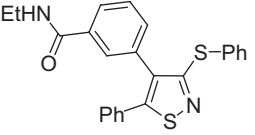
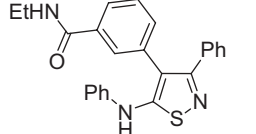
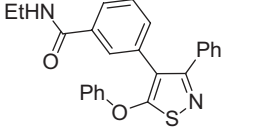
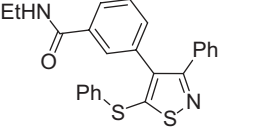
The number of single bonds (NCCSgl) accounts also for the presence of the saturated carbon-carbon bonds in the molecule and indicates that unsaturated bonds are desirable for an enhanced activity (27).

KiCP44, KiCP36, and ChiCP49 are cluster-path subgraph descriptors. Molecular subgraphs are subsets of atoms and related bonds that represent molecular fragments and functional groups. There are four commonly used subgraph types: path subgraph, cluster subgraph, path-cluster subgraph, and chain subgraph (or ring) emphasizing different aspects of atom connectivity within the molecule. They are defined according to the following rules: (i) if the subgraph contains a cycle, it is of type chain, (ii) if all vertex degrees in the subgraph are either greater than 2 or equal to 1, the subgraph is of type cluster, (iii) if all vertex degrees in the subgraph are either equal to 2 or 1, the subgraph is of type path, otherwise (iv) the subgraph is of type path-cluster. The order of a subgraph is the number of edges within it. An index with an order 3 for path refers to a path of length 3, and an index with an order 3 for cluster refers to a

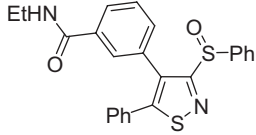
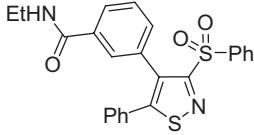
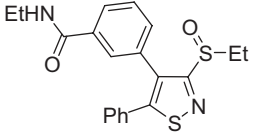
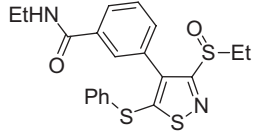
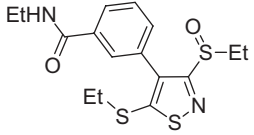
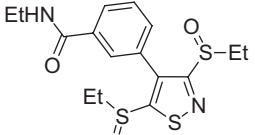
**Table 10:** Predicted MEK inhibition activity for virtual compounds 1v-8v

Compound	$\log(1/IC_{50})$ predicted	Applicability domain (APD = 10.9564)
1v	-1.277	43.4659
2v	-1.704	7.1970
3v	-1.674	9.5448
4v	-1.645	5.7452
5v	-1.689	10.6511
6v	-1.630	10.6465
7v	-1.779	8.3193
8v	-1.954	6.3574

**Table 11:** Predicted MEK inhibition activity for virtual compounds **9v–14v**

Compound	$\log(1/IC_{50})$ predicted	Applicability domain (APD = 10.9564)
	-1.773	4.1286
	-1.541	4.4486
	-1.616	4.4625
	-1.774	4.1306
	-1.542	4.4498
	-1.621	4.4718

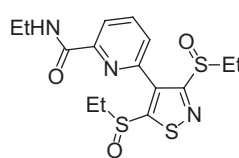
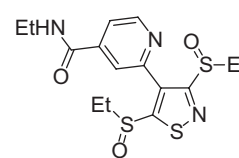
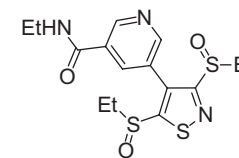
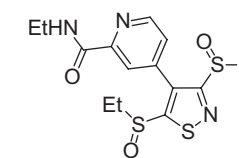
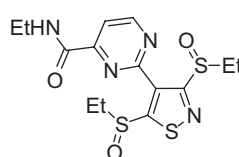
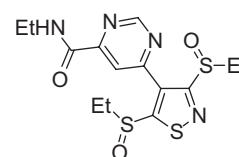
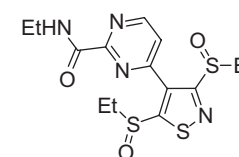
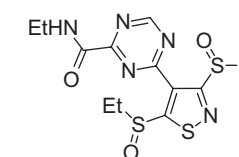
**Table 12:** Predicted MEK inhibition activity for virtual compounds **15v–20v**

Compound	$\log(1/IC_{50})$ predicted	Applicability domain (APD = 10.9564)
	-1.606	4.0814
	-2.778	4.2293
	-1.410	2.3859
	-1.329	1.1783
	-1.135	2.5108
	-1.150	1.4324

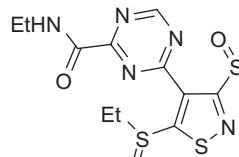
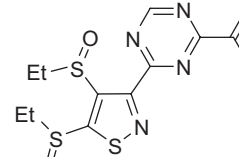
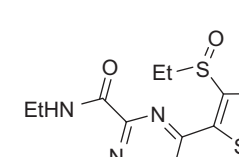
3-edge cluster (27). Based on the weights for cluster–path descriptors, the presence of a subgraph containing a cluster of order 4 combined with path of order 4 contributes negatively to the MEK inhibition activity, and as such this subgraph is undesirable. On the

contrary, the presence of a subgraph containing a cluster of order 4 combined with path of order 9 and the presence of a subgraph containing a cluster of order 3 combined with path of order 6 contribute positively to the oral bioavailability.

**Table 13:** Predicted MEK inhibition activity for virtual compounds **21v–28v**

Compound	$\log(1/IC_{50})$ predicted	Applicability domain (APD = 10.9564)
	-1.052	3.1626
	-1.041	3.1528
	-1.049	3.1369
	-1.051	3.1400
	-0.954	5.7089
	-0.993	5.7411
	-1.001	5.7639
	-0.912	3.1697

**Table 14:** Predicted MEK inhibition activity for virtual compounds **28v–30v**

Compound	Activity	Domain
	-0.912	3.1697
	-0.965	3.2568
	-0.955	3.2889

As shown from Tables 5 and 6, the aforementioned descriptors have different weights that influence the increase or decrease in MEK inhibition activity and oral bioavailability. Based on the previous discussion and the positive or negative influence of each descriptor, new derivatives with enhanced activity and desired oral bioavailability can be designed.

The APD was defined for all compounds that constituted the training sets for MEK inhibition and oral bioavailability models as described in the Materials and Methods section. As all validation compounds fell inside the domain of applicability for both models, all model predictions for the external test set were considered reliable. The results are presented in Tables 8 and 9.

With the model in hand, we screened an in-house library of 107 isothiazoles (36–39) in the hope that new isothiazole targets could be identified. The model gave a wide range of predicted activities (–5.377 to –0.041); however, all predictions were outside the domain of applicability (134.188–22.445, threshold of 10.968).

Of the 107 isothiazoles, we chose initially 3,4,5-triphenylisothiazole **1v** as a starting point for virtual modifications in the hope that theoretical structures of improved activity within the domain of applicability could be identified. 3,4,5-Triphenylisothiazole (**1v**) gave a good predicted activity (–1.277), and its domain of applicability (43.466) was deemed workable. Furthermore, this compound and



analogues can be readily prepared (38) and provide a large carbon periphery providing multiple sites for structural modification.

Initially substituents were introduced on the 4-phenyl substituent, and an immediate reduction in the domain of applicability was observed (Table 10). In particular, the introduction of alkylamino or carboxamide groups gave acceptable domains of applicability without affecting significantly the predicted activities.

The carboxamide **7v** was arbitrarily chosen for further structural modification. We note that isothiazoles **2v–6v** and **8v** could also have been taken as viable scaffolds. The introduction of heteroatoms to separate the phenyl substituents from the isothiazole at C3 and C5 could facilitate the synthesis, as these sites are known to be electrophilic and thus favor modification by nucleophilic substitution. Structures **9v–14v** were predicted to have comparable activities and were well within the domain of applicability (Table 11).

The model indicated that raising the oxidation level of the exocyclic sulfide to the sulfoxide was acceptable, while increasing to the sulfone was not favorable. Interestingly, switching from the phenyl-sulfoxide to the ethylsulfoxide showed a good increase in predicted activity within the domain of applicability. Repeating the sequence of structural modifications with the C5 phenyl gave the 3-(ethylsulfinyl)-5-(ethylthio)isothiazole **19v** and the 3,5-bis(ethylsulfinyl)isothiazole **20v** with predicted activities of  $-1.135$  and  $-1.150$ , respectively (Table 12).

Taking the isothiazole **20v** as a new starting point, the aryl ring at C4 was modified further by exchanging CH for N. The model tolerated this modification providing pyridyl, pyrimidyl, and 1,3,5-triazinyl C4 substituted isothiazoles with overall good predicted activities well within the required domain of applicability (Table 13). It was worth noting that the position of the nitrogen made little difference; however, increasing the number of nitrogens to give the 1,3,5-triazinyl analog **28v** gave the theoretical structure with the best predicted activity ( $-0.912$ ).

Finally, at this stage, we switched the C3, C4, and C5 substituents of isothiazole **28v** to see whether the actual substitution pattern made any significant difference in the predicted activity. As can be seen in Table 14, the model could not differentiate readily the isomers **28v–30v**.

In the manner demonstrated earlier, the model can be used to rapidly identify additional theoretical scaffolds. It is worth noting that modifications should be based on synthetically viable targets.

The synthesis and study of the active virtual compounds would be required to truly validate the model and as such is a worthy pursuit, but this is outside the scope of this present paper. It must therefore be noted that the virtual screening study acts only as an aid in proposing structural modifications to assist ongoing SAR (Structure Activity Relationship) studies. The high biological activities predicted are only indicative of which structures should be targeted for synthesis on the basis that they meet or approach the optimal values for the chosen descriptors for the given model.

## Conclusion

In this paper, the key components of the isothiazole scaffold that influence MEK inhibition and oral bioavailability have been identified. QSAR models that quantitatively describe and predict the relationship between structural characteristics, activity, and bioavailability have been developed and validated. Descriptors such as NCCSgl, Kilnf8, KiCP44, AtomCompMean, and AtomCompTot, which characterize the chemical entities, were found to be directly connected to MEK inhibition. Additionally, descriptors such as dipole, TopoJ, Chilnf0, KiCP36, Ki8, ChiCP49, and AtomCompMean were found to influence oral bioavailability of the compounds. The molecular descriptors used in QSAR encode information about the structure, branching, electronic effects, chains, and rings of the modules and thus implicitly account for cooperative effects between functional groups. Applicability domain was defined to identify the reliable predictions. A virtual screening study has been conducted to help researchers to design novel chemistry driven molecules with desired characteristics.

## Acknowledgments

The authors wish to thank the Cyprus Research Promotion Foundation (Grant ΔΙΑΚΤΩΡ/ΔΙΣΕΚ/0308/48) and the following organizations in Cyprus for generous donations of chemicals and glassware: the State General Laboratory, the Agricultural Research Institute, and the Ministry of Agriculture. Furthermore, we thank the A.G. Leventis Foundation for helping to establish the NMR facility in the University of Cyprus.

## References

- McCubrey J.A., Steelman L.S., Chappell W.H., Abrams S.L., Wong W.T.E., Chang F., Lehmann B. *et al.* (2007) Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance. *Biochim Biophys Acta*;1773:1263–1284.
- Thompson N., Lyons J. (2005) Recent progress in targeting the Raf/MEK/ERK pathway with inhibitors in cancer drug discovery. *Curr Opin Pharmacol*;5:350–356.
- Montagut C., Settleman J. (2009) Targeting the RAF- MEK-ERK pathway in cancer therapy. *Cancer Lett*;283:125–134.
- Ramnath N., Adjei A. (2007) Inhibitors of Raf kinase and MEK signaling. *Update on Cancer Therapeutics*;2:111–118.
- Wang D., Boerner S.A., Wonkler J.D., LoRusso P.M. (2007) Clinical experience of MEK inhibitors in cancer therapy. *Biochim Biophys Acta*;1773:1248–1255.
- Chen Z., Kim S.-H., Barbosa S.A., Huynh T., Tortolani D.R., Leavitt K.J., Wei D.D., Manne V., Ricca C.S., Gullo-Brown J., Poss M.A., Vaccaro W., Salvati M.E. (2006) Pyrrolopyridazine MEK inhibitors. *Bioorg Med Chem Lett*;16:628–632.
- Sebolt-Leopold J.S., Herrera R. (2004) Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nat Rev Cancer*;4:937–947.
- El Abdellaoui H., Varaprasad C.V.N.S., Barawkar D., Chakravarty S., Maderna A., Tam R., Chen H. *et al.* (2006) Identification

- of isothiazole-4-carboxamidines derivatives as a novel class of allosteric MEK1 inhibitors. *Bioorg Med Chem Lett*;16:5561–5566.
9. Clerici F, Gelmi M.L., Pellegrino S., Pocar D. (2007) Chemistry of biologically active isothiazoles. *Top Heterocycl Chem*;9:179–264.
  10. Sheppard G.S., Bouska J.J. (2005) Why optimize cancer drugs for ADMET? *Drug Discov Today Ther Strateg*;2:343–349.
  11. Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.H. (2009) The WEKA Data Mining Software: an update. *SIGKDD Explor*;11:10–18.
  12. Stewart J.J.P. (2007) Optimization of parameters for semiempirical methodsV: modification of NDDOapproximations and application to 70 elements. *J Mol Model*;13:1173–1213.
  13. Stewart J.J.P. (2008) Application of the PM6 method to modeling the solid state. *J Mol Model*;14:499–535.
  14. Melagraki G., Afantitis A., Makridima K., Sarimveis H., Igglessi-Markopoulou O. (2006) Prediction of toxicity using a novel RBF neural network training methodology. *J Mol Model*;12:297–305.
  15. Melagraki G., Afantitis A., Sarimveis H., Koutentis P.A., Igglessi-Markopoulou O., Kollias G. (2009) Predictive QSAR workflow for the in silico identification and screening of novel HDAC inhibitors. *Mol Divers*;13:301–311.
  16. Toropov A.A., Toropova A.P., Benfenati E. (2009) Simplified molecular input line entry system-based optimal descriptors: quantitative structure-activity relationship modeling mutagenicity of nitrated polycyclic aromatic hydrocarbons. *Chem Biol Drug Des*;73:515–525.
  17. Toropov A.A., Toropova A.P., Benfenati E. (2009) QSAR modeling for mutagenic potency of heteroaromatic amines by optimal SMILES-based descriptors. *Chem Biol Drug Des*;73:301–312.
  18. Afantitis A., Melagraki G., Sarimveis H., Koutentis P.A., Igglessi-Markopoulou O., Kollias G. (2010) A combined LS-SVM & MLR QSAR workflow for predicting the inhibition of CXCR3 receptor by quinazolinone analogues. *Mol Divers*;14:225–235.
  19. Cortes C., Vapnik V. (1995) Support vector networks. *Mach Learn*;20:273–297.
  20. Shevade S.K., Keerthi S.S., Bhattacharyya C., Murthy K.R.K. (2000) Improvements to the SMO Algorithm for SVM Regression. *IEEE Trans Neural Netw*;11:1188–1193.
  21. Afantitis A., Melagraki G., Sarimveis H., Koutentis P.A., Markopoulos J., Igglessi-Markopoulou O. (2006) A novel QSAR model for evaluating and predicting the inhibition of dipeptidyl aspartyl fluoromethylketones. *QSAR Comb Sci*;25:928–935.
  22. Afantitis A., Melagraki G., Sarimveis H., Igglessi-Markopoulou O., Kollias G. (2009) A novel QSAR model for predicting the inhibition of CXCR3 receptor by 4-N-aryl-[1,4] diazepane ureas. *Eur J Med Chem*;44:877–884.
  23. Zhang S., Golbraikh A., Oloff S., Kohn H., Tropsha A. (2006) 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*;46:1984–1995.
  24. Melville J.L., Burke E.K., Hirst J.D. (2009) Machine learning in virtual screening. *Comb Chem High Throughput Screen*;12:332–343.
  25. Melagraki G., Afantitis A., Sarimveis H., Koutentis P.A., Markopoulos J., Igglessi-Markopoulou O. (2007) Optimization of Biaryl Piperidine and 4-Amino-2-biaryllurea MCH1 receptor antagonists using QSAR modeling, classification techniques and virtual screening. *J Comput Aided Mol Des*;20:83–95.
  26. Melagraki G., Afantitis A., Sarimveis H., Koutentis P.A., Markopoulos J., Igglessi-Markopoulou O. (2007) Identification of a series of novel derivatives as potent HCV inhibitors by a ligand – based virtual screening optimized procedure. *Bioorg Med Chem*;15:7237–7247.
  27. Todeschini R., Consonni V., Mannhold R. (2000) In: Kubinyi H., Timmerman H., Series editors. *Handbook of Molecular Descriptors*. Weinheim: Wiley-VCH.
  28. Singh J., Shaik B., Singh S., Agrawal V.K., Khadikar P.V., Deeb O., Supuran C.T. (2008) Comparative QSAR study on para-substituted aromatic sulphonamides as CAII inhibitors: information versus topological (distance-based and connectivity) indices. *Chem Biol Drug Des*;71:244–259.
  29. Konstantinova E.V. (2005) On some applications of information indices in chemical graph theory. *Electron Notes Discrete Math*;21:329–351.
  30. Shannon C.E. (1948) A mathematical theory of communication. *Bell Syst Tech J*;27:379–423.
  31. Shannon C.E., Weaver W. (1963) *The Mathematical Theory of Communication*. Urbana, Illinois: University of Illinois Press.
  32. Stahura F.L., Godden J.W., Xue L., Bajorath J. (2000) Distinguishing between natural products and synthetic molecules by descriptor Shannon entropy analysis and binary QSAR calculations. *J Chem Inf Comput Sci*;40:1245–1252.
  33. Bonchev D., Randic M. (2003) Shannon's entropy of proteomic 2D-gel maps. *Chem Phys Lett*;372:548–552.
  34. Godden J.W., Bajorath J. (2007) Analysis of chemical information content using Shannon entropy. *Rev Comput Chem*;23:263–289.
  35. Devillers J., Balaban A.T. (1999) *Topological Indices and Related Descriptors in QSAR and QSPR*. The Netherlands: Gordon and Breach Science Publishers.
  36. Christoforou I.C., Koutentis P.A., Rees C.W. (2003) Regiospecific Suzuki coupling of 3,5-dichloroisothiazole-4-carbonitrile. *Org Biomol Chem*;1:2900–2907.
  37. Christoforou I.C., Koutentis P.A. (2006) New regiospecific isothiazole C-C coupling chemistry. *Org Biomol Chem*;4:3681–3693.
  38. Christoforou I.C., Koutentis P.A. (2007) 3,4,5-triarylisothiazoles via C-C coupling chemistry. *Org Biomol Chem*;5:1381–1390.
  39. Ioannidou H.A., Koutentis P.A. (2009) The conversion of isothiazoles into pyrazoles using hydrazine. *Tetrahedron*;65:7023–7037.

## Notes

<sup>a</sup>CambridgeSoft Corporation, ChemOffice, <http://www.cambridgesoft.com>.

<sup>b</sup>MOPAC2007, Stewart J.J.P., Stewart Computational Chemistry, Version 7.295W, <http://www.openmopac.net>.

<sup>c</sup>Svozil D., Lohninger H, TOPIX, <http://lohninger.com/topix.html>.