ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
http://web.icf.ro/rrch/

Rev. Roum. Chim.,
2015, *60*(2-3), 167-173

Dedicated to Professor Zeno Simon
on the occasion of his 80th anniversary

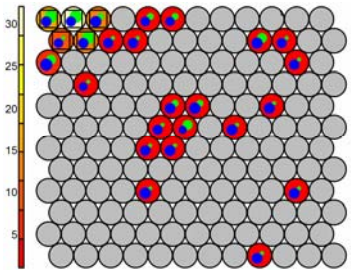
SELF-ORGANIZING MAP CLASSIFICATION MODEL FOR THE PREDICTION OF MEK1 INHIBITORS**

Sorin AVRAM,[a,*] Stefana AVRAM,[b,*] Luminita CRISAN,[a] Liliana PACUREANU,[a]
Ludovic KURUNCZI[a,b] and Alina BORA[a,*]

[a] Timişoara Institute of Chemistry of Roumanian Academy, 24 Mihai Viteazul, 300223, Timişoara, Roumania
[b] Victor Babeş University Of Medicine and Pharmacy, 2 E. Murgu, 300041 Timişoara, Roumania

Received August 25, 2014

In this study, we aim to model inhibitors of the dual specificity mitogen-activated protein kinase kinase 1 (MEK 1), which acts as a key component of the mitogen-activated protein kinase signal transduction pathway modulating cell growth in various types of tumors. We applied supervised self-organizing map (SOM) modeling using topological autocorrelation descriptors to identify inhibitors of MEK1 in virtual screening conditions. We found a 12x12 SOM model, externally validated on a set of compounds sharing only partially the chemical scaffolds of the training set. A percentage of 78 of the actives in the test set were correctly predicted and six out of seven chemical scaffolds (which did not participate in training) were identified. We conclude that SOM classification modeling is an efficient tool for the discovery of chemically diverse MEK1 inhibitors.



INTRODUCTION

Drug discovery programs pursuit the identification of novel compounds active against a therapeutically relevant biological target. Novelty, in terms of scaffolds (molecular frameworks) provides the bases for lead-identification and optimization.[1] Typically, computational methods are employed to first prepare a chemical library of structurally diverse compounds and secondly to narrow the search by virtual screening (VS) against the target of interest. The most promising representatives of different scaffolds can be further submitted for experimental activity determination.

Ligand-based VS methods, *e.g.*, similarity search, pharmacophore search, supervised learning methods etc, are valuable tools for rational hit and lead-identification.[2–4] A successfully applied algorithm, in a variety of fields,[5] is Kohonen's self-organizing map (SOM).[6]

SOM is a topographic mapping pattern recognition algorithm based on a neural network architecture by which objects of a multi-dimensional space are mapped into a regular predefined grid of units (neurons). During training, each object, repeatedly presented to the map is assigned to a so-called winning unit, which is the one most similar to the training object. Every unit

* Corresponding author: alina.bora@gmail.com
** Supporting Information on http://web.icf.ro or http://revroum/lew.ro

(and eventually winning unit) is repeatedly updated together with the units in its immediate neighborhood. Thus, neighboring units tend to map similar objects. In the case of classification, the dependent variable can be modeled and predictions can be obtained.[5]

In drug discovery, SOM (Kohonen network) modeling has been used for the design of screening library, scaffold hopping and drug repurposing.[7] Pharmacophoric molecular representation were successfully used for compiling small, structurally diverse, activity-enriched sets of chemicals.[7] Compounds mapped in the receptive field of a neuron exhibit similar pharmacophoric features and often various scaffolds. A common VS procedure, as reported by several successful studies, is to map a chemical library into a SOM and to depict, for activity measurements, compounds from the neuron/neurons that clustered most of the known actives embedded into the library.[7,8]

To our knowledge, the scaffold hopping potential of SOM has not been investigated for the identification of MEK1 (dual specificity mitogen-activated protein kinase kinase 1) inhibitors. MEK1 belongs to the family of MAPKKs (known as mitogen-activated protein kinase kinases), which are dual specificity enzymes that phosphorylate threonine and tyrosine residues within the activation loop of their MAP kinase substrates.[9] Signaling pathways such as mitogen-activated protein kinase (MAPK) cascades regulate normal cell proliferation, survival and differentiation[10] but abnormalities within this signaling pathway play a critical role in the development and progression of cancer, especially in those that suffer genetic mutations.[11,12] Many MEK1 (and MEK2) inhibitors have failed in clinical trials, but several are still investigated as promising antitumor agents.[9] However, new chemical scaffolds are still desirable to provide new leads with improved pharmacological potential and reduced toxicity.

Melanoma is one of the most aggressive forms of skin cancer, associated with a heterogenic molecular pattern determined by progressive genetic mutation.[13] New molecular therapeutic approach represented by the MAPK inhibitor vemurafenib and more recently, trametinib[14] are being used for the 50% of the cases with harboring mutation on the BRAF gene, with effects on cellular growth, survival and angiogenesis, but still with poor overall survival rates. For 15-20% of the cases of melanoma that express NRAS mutation, still no specific treatment has been established.

In this study, we attempt to find a prediction model capable to enrich novel scaffolds of MEK1 inhibitor. We assess the diversity of the chemical frameworks among already known MEK1 inhibitors and use these results to train and test an efficient SOM model for virtual screening.

MATERIALS AND METHODS

Data sets

Inhibitors of the dual specificity mitogen-activated protein kinase kinase 1 (MEK1; target ID CHEMBL3587; UNIPROT Q02750) were downloaded from ChEMBL database (https://www.ebi.ac.uk/chembl/; accessed on June 30, 2014).[15,16] A number of 159 ligands were kept showing $IC_{50} \leq 1$ μM, activity assessed on human kinase and confidence score of 9. In cases of multiple activity results available for a single compound, we computed the average value. In order to serve as decoys, we selected randomly a set of 10000 compounds from PubChem Compounds (https://pubchem.ncbi.nlm.nih.gov/; last accessed on June 30, 2014).

Data sets processing. The ionization state of all compounds was set to pH 7.4. The molecules were filtered against the *BlockBuster* filter available in FILTER (version 2.0.2, www.eyesopen.com).[17] All 159 MEK1 inhibitors (see *Supporting Information Table S1*, *Fig.S1*, *Fig. S2*) and 8671 PubChem compounds passed the filter and were used for model optimization and validation.

Molecular framework

MEK1 inhibitors were submitted first to Chemaxon Standardizer to clear stereoisomery, neutralize and remove hydrogens, JChem 6.1.0, 2013, ChemAxon (http://www.chemaxon.com) and second to Chemaxon Calculator *cxcalc*, to generate Bemis-Murcko[18] molecular frameworks, Calculator 6.1.0, 2013, ChemAxon (http://www.chemaxon.com).

Training and Test sets

We split the actives into two sets: 80% of the compounds for training and 20% for testing. In order to explore the scaffold hopping capabilities of SOMs, the test set has been designed to contain 60% of the compounds sharing scaffolds different from those encountered among the training set

molecules. The other 40% consist of randomly chosen derivatives of the first seven most representative chemical frameworks, *i.e.*, BMFs **1** (4), **2-4** (2), **5-7** (1) (see Fig. 1). It is commonly considered that the number of actives embedded into a screening library is much smaller compared to the number of inactives. Thus, we randomly chose a 200 times larger set of inactives for validation and the remaining compounds were used for training the SOM (see Table 1).

Table 1

Description of the training and the test set composition

Classes	Training	Test
Number of actives	127	32
Number of inactives	2271	6400
A:I[a]	1:18	1:200
Number of BMFs	20	17

[a] the active to inactive ratio

Molecular descriptors

We generated the Moreau–Broto autocorrelation (Autocorrelation of a Topological Structure, ATS)[19] descriptors (path of length 1 to 5; atomic properties were centered by subtracting the average property value in the data sets)[20] for charge, mass and polarizability using PaDEL-Descriptor.[21] Each variable has been scaled by subtraction of means and division by the corresponding standard deviation (unit variance scaling).

Self-Organizing map

SOM was applied using function *xyf* (supervised SOM: two parallel maps) of package *kohonen*[5] available in R statistical software.[22] Default parameter values were maintained *(i.e.*, learning rate, independent variable weight). Only the size of the grid has been varied and subsequently evaluated in terms of the number of actives in the test set correctly predicted by the model.

Evaluation

The SOM models generated in this study have been explored: (i) visually, for the training set to cover the grid and keeping the number of empty units to a minimum; (ii) numerically, in terms of accuracy, *Acc* (*i.e.*, the percentage of molecules in the test set correctly classified by the model), sensitivity, *Se* (*i.e.*, the ratio of the number of active molecules found by the model to the number of all active test set compounds) and specificity, *Sp* (*i.e.*, the ratio of the number of inactive compounds that were not selected by the model to the number of all inactive molecules in the test set).[23] A further analysis has been performed concerning the diversity of the molecular frameworks and activity among the true positives.

RESULTS AND DISCUSSION

Chemical diversity analysis of MEK1 inhibitors

We assessed the chemical diversity of the 159 MEK1 inhibitors in terms of chemical frameworks. A simple way to encode chemical scaffolds was proposed by Bemis and Murcko.[18] A Bemis-Murcko framework (BMF) is a representation of the ring systems and linkers of a molecule (side chains are ignored). BMFs with cleared stereochemistry, aromaticity and atom-types were generated for the 159 MEK1 inhibitors. A number of 30 distinct BMFs were found as described in Fig. 1. With very few exceptions *(i.e.*, **11**, **21**, **24**, **30**), one can observe a pattern consisting of one atom-long linkers connecting two single (five/six atom dimension) ring systems. In some cases, the second ring-system is composed of two fused rings. Often a third (or a fourth) ring and longer linkers can be observed, but surprisingly the BMF of the most active compound consists of a fused system (of a seven and a five-atom ring) linked directly to a five-atom ring.

BMF **1** is shared by 39 compounds and together with the next three frameworks they account for 55% of the structures. Among these, BMF **2** shows the highest activity among its members, *i.e.*, the mean IC_{50} is 0.085 μM. The three most active clusters (scaffolds), as suggested by their representatives, are singletons: BMF **21** (IC_{50} of 0.006 μM), BMF **22** (IC_{50} of 0.008 μM) and BMF **23** (IC_{50} of 0.015 μM).

MEK1 inhibitors indicate remarkable molecular framework diversity. At an early stage of drug discovery programs, the design of chemical libraries to be screened for MEK1 inhibition can rely on the patterns described above to maximize true positive outcomes. However, the span of the activity values among the members of each framework is heavily influenced by the consistency of the side-chains. This analysis can assist medicinal chemists in decision-making in later stages such as lead-development and optimization.

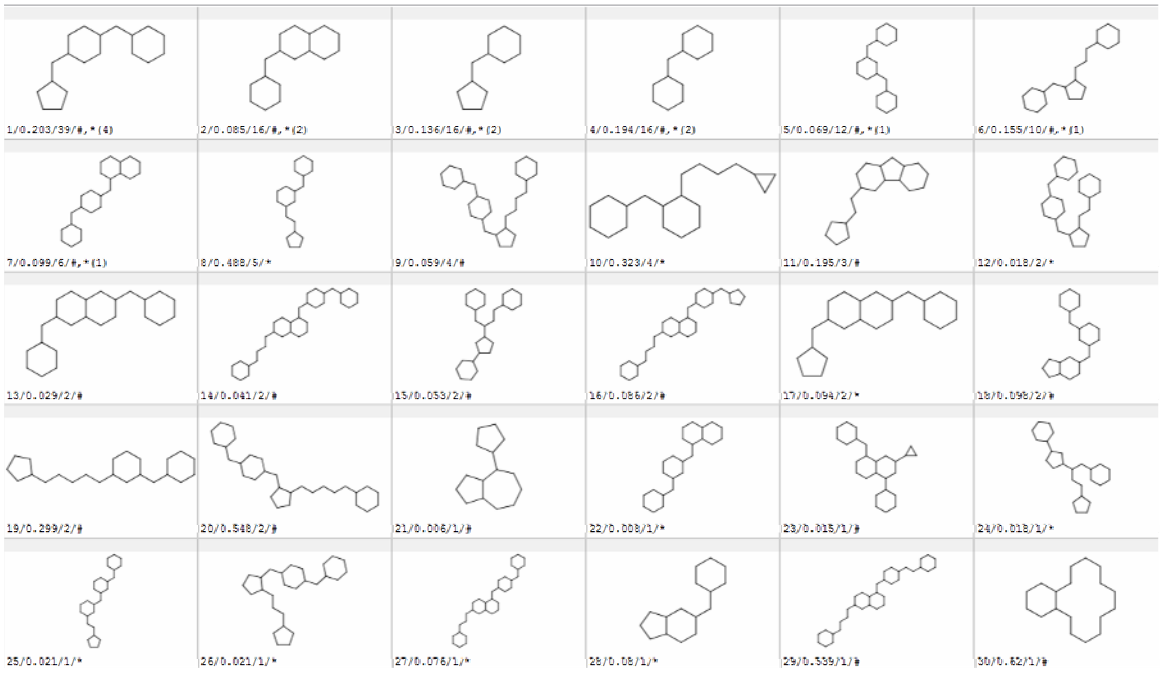


Fig. 1 – Representation of the Bemis-Murcko frameworks of the MEK1 inhibitors used for SOM modeling and validation; we report, for every BMF, the identifier/the average IC50 (μM) / the number of inhibitors sharing the BMF / used in the "#" training and "*" validation of the SOM (for the validation set only: in parentheses we report the number of randomly selected compounds; if not specified all compounds sharing the BMF were used in this set).

SOM model

Chemical libraries are typically preprocessed to remove undesirable compounds *(e.g.*, low solubility, high chemical reactivity, high toxicity etc) before proceeding to virtual screening. Here, we employed OpenEye's *BlockBuster* filter,[17] which aims to profile drug-like compounds. This simple rule-based method is grounded on 141 best-selling, non-antibiotic, prescription drugs, evaluated in terms of physical properties and functional groups. A number of 159 MEK1 inhibitors (actives) and 8671 inactives (decoys; see *Training and Test sets* in Materials and Methods) passing this filter is assumed to exert drug-like features.

The training set, consisting of 127 actives and 2271 decoys, was submitted to supervised SOM modeling based on 15 ATS topological autocorrelation descriptors (see *Molecular descriptors*, Material and Methods). A series of square SOM models were explored by varying the size of the grid between 10x10 and 20x20. The 12x12 grid SOM provided a good covering of the training set (see Fig. 2a), convergence and the highest recovery of actives in the test set. A number of 25 out of 32 actives in the test set were predicted properly (recovery 78%) by 5 units, i.e., 121,122, 133, 134 and 135 (see Fig. 2b). In the attempt to mimic a typical virtual screening scenario the number of inactives (decoys) in the test set has been chosen 200 times larger compared to the number of actives.

The active prediction units, 121, 122, 133, 134, 135, modeled 15 scaffolds covering 25 compounds. These units correctly predicted 78.1% of the actives in the test set, as reported in Table 2. In addition to this high sensitivity (*Se*), we calculated a specificity (*Sp*) of 0.990, which suggests very good discrimination capacities. The accuracy (*Acc*) of the predictions reached 0.998, however it is mostly influenced by the high class imbalance, *i.e.*, active to inactive ratio of 1 to 200 (see Table 1). As stated above, here we sought to set up a VS scenario, where, typically, the number of actives is much smaller compared to the number of inactives. These, high class-discrimination capacities qualify SOM as a promising approach to enrich actives in VS.

A search in the literature reveals that many potent MEK1 hits have been identified in high-throughput screenings and were further submitted to structure-activity relationship exploration, *e.g.*, coumarin derivatives (BMF **2**),[24] 4-anilino-3-cyano-6,7-dialkoxyquinolines (BMF **7**),[25] 3-hydroxy-5-arylamino-isothiazoles (BMF **3**),[26] isothiazole-4-carboxamidines (BMF **1**)[27] etc. Some series have been later employed for 2D and 3D quantitative structure-activity relationship (QSAR) modeling and docking.[28,29]

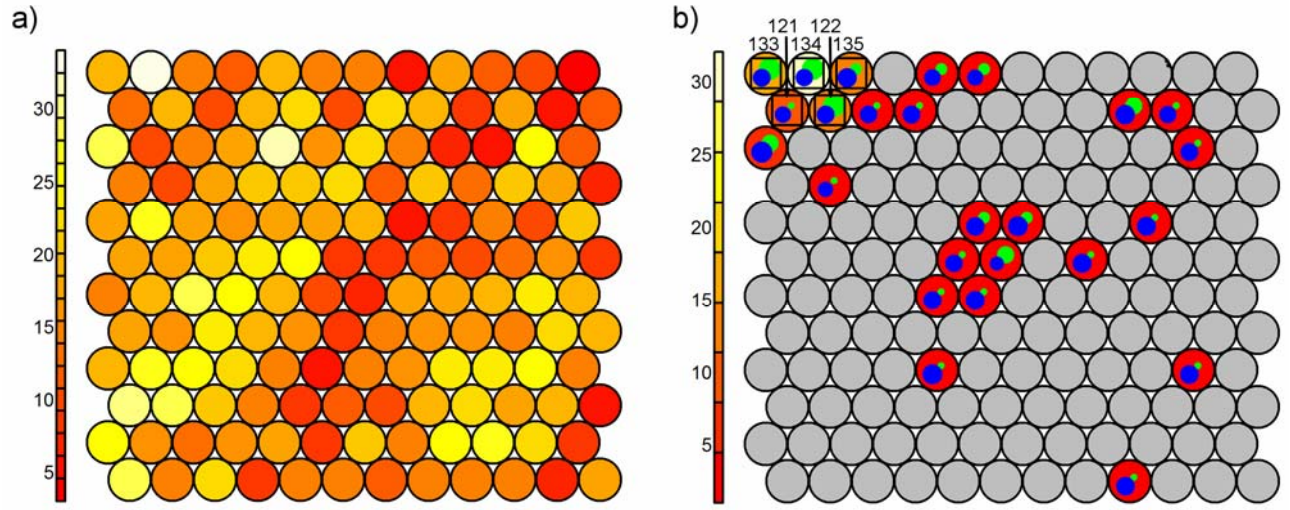


Fig. 2 – (a) Count plot of the SOM model mapping all training data; (b) Count plot of the 12x12 SOM model mapping the actives in the training data; the diameter of the green circles correspond to the number of different BMFs (ranging from 1 to 4) covering the members assigned to the units, the diameter of the blue circles suggest pIC50 values (ranging from 2.404 to 3.756) of the members assigned to the units; units containing squares, accompanied by identifiers, are predictors of MEK1 inhibitors; units colored grey are empty.

Table 2

Validation results

SOM grid	*Se*[a]	*Sp*[b]	*Acc*[c]
12x12	0.781	0.990	0.989

[a] Sensitivity; [b] Specificity; [c] Accuracy (see *Evaluation* in Materials and Methods)

Table 3

Scaffold hopping results

UNIT ID[a]	BMF ID[b]	Number of actives	IC_{50}[c]
121	**10**	4	0.178
122	n.a.[d]	n.a.	n.a.
133	**8**	5	0.287
	25	1	0.002
134	**12**	2	0.002
	26	1	0.002
135	**17**	2	0.016

[a] identification number for units in SOM (see Fig. 2b);[b] identification number for Bemis-Murcko molecular frameworks (see Fig. 1);[c] average IC_{50} (μM) per unit;[d] n.a. – not available.

Virtual screening studies aiming to identify MEK1 inhibitors are scarcely encountered in the literature. In a swift search we found successful structure-based docking experiments,[30,31] but surprisingly no ligand-based VS model. Thus, in addition to the available tools to identify MEK1 inhibitors, the SOM model derived herein demonstrates that also ligand-based VS can be efficiently applied to enrich MEK1 ligands.

Scaffold hopping of the SOM model

The detection of diverse chemical scaffolds exhibiting an inherent specific biological activity against a protein-target of interest widens the horizons of find lead or drug-candidates with enhanced pharmacological properties. Effective VS methods should be able to identify new scaffolds of active compounds.[4] In this sense, we

challenged the SOM classification model by providing a chemically diverse test set. As described in subsection *Training and Test set* (in Material and Methods) we used for training 20 scaffolds and for testing 17, with only 10 BMFs shared by the two sets. Actives derived from the other 7 scaffolds in the test set are expected to be predicted by the model.

In Table 3 we describe the prediction units (see Fig. 2b) and the actives properly predicted by the model, but sharing different BMFs from those employed for training. One can see that except the empty unit 122, the other four identified one or two "new" scaffolds (these can be examined in Fig. 1). Interestingly, the actives covered by BMF **25**, **12**, **26** and **17** are very potent MEK1 inhibitors, *i.e*., in the nano-molar range.

Prediction methods rely heavily on the type of descriptors (independent variables) employed to characterize the objects. SOM modeling grounded on molecular pharmacophoric description has been successfully employed for scaffold hopping in earlier studies.[8,32] Our study demonstrates that, by using autocorrelation descriptors, the SOM model is able to detect highly active MEK1 inhibitors with novel scaffolds.

In the future, such effective ligand-based approaches may assist in the identification and development of new synthetic or natural compounds, inhibitors of MEK1/2 as well as PI3K/mTOR kinases from the activated pathways in NRAS melanoma.[33] These can be subsequently tested for their activity on cancer progression and angiogenesis in simple, viable *in vivo* melanoma assay on the chorioallantoic membrane of the chick embryo.[34]

CONCLUSIONS

The discovery of new chemical scaffolds is of major interest in the pharma industry. In spite of long-term efforts, efficient screening of large databases for new lead compounds via *in silico* methods remains a challenging task. Here, we challenged the scaffold hopping capabilities of SOM by employing a set of known MEK1 inhibitors. In this study, we demonstrated that highly potent MEK1 inhibitors with scaffolds different from those employed in training can be identified by the SOM model established on 2D autocorrelation descriptors. Ligand-based algorithms such as SOM applied in hit or lead identification provide powerful tools for data analysis and scaffold hopping.

Acknowledgements: This work was supported by Project 1.2 of the Institute of Chemistry Timişoara of the Roumanian Academy and by an Internal grant at the Victor Babeş University of Medicine and Pharmacy Timişoara, PII-C2-TC-2014, CAMMelRasNa, obtained by Avram Stefana. We thank ChemAxon Ltd. and OpenEye Scientific Ltd. for the free academic licenses to their software.

REFERENCES

1. H. Sun, G. Tawa and A. Wallqvist, *Drug Discov. Today*, **2012**, *17*, 310-324.
2. S. I. Avram, L. Crisan, A. Bora, L. M. Pacureanu, S. Avram and L. Kurunczi, *Bioorg. Med. Chem.*, **2013**, *21*, 1268-1278.
3. S. I. Avram, L. M. Pacureanu, A. Bora, L. Crisan, S. Avram and L. Kurunczi, *J. Chem. Inf. Model*, **2014**, doi: 10.1021/ci5002668.
4. S. Avram, L. M. Pacureanu, E. Seclaman, A. Bora and L. Kurunczi, *J. Chem. Inf. Model.*, **2011**, *51*, 3169-3179.
5. R. Wehrens and L. M. C. Buydens, *J. Stat. Softw.*, **2007**, *21*, 1-19.
6. T. Kohonen, *Biol. Cybern.*, **1982**, *43*, 59-69.
7. P. Schneider, Y. Tanrikulu and G. Schneider, *Curr. Med. Chem.*, **2009**, *16*, 258-266.
8. D. Kaiser, L. Terfloth, S. Kopp, J. Schulz, R. de Laet, P. Chiba, G. F. Ecker and J. Gasteiger, *J. Med. Chem.*, **2007**, *50*, 1698-1702.
9. C. Frémin and S. Meloche, *J. Hematol. Oncol.*, **2010**, *3*, 8.
10. P. J. Roberts and C. J. Der, *Oncogene*, **2007**, *26*, 3291-3310.
11. A. S. Dhillon, S. Hagan, O. Rath and W. Kolch, *Oncogene*, **2007**, *26*, 3279-3290.
12. A. Akinleye, M. Furqan, N. Mukhi, P. Ravella and D. Liu, *J. Hematol. Oncol.*, **2013**, *6*, 27.
13. L. Finn, S. N. Markovic and R. W. Joseph, *BMC Med.*, **2012**, *10*, 23.
14. A. M. Menzies and G. V Long, *Clin. Cancer Res.*, **2014**, *20*, 2035-2043.
15. A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani and J. P. Overington, *Nucleic Acids Res.*, **2012**, *40*, D1100-1107
16. A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. a Krüger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos and J. P. Overington, *Nucleic Acids Res.*, **2014**, *42*, D1083-1090.
17. OMEGA 2.5.1.4: OpenEye Scientific Software, Santa Fe, NM. www.eyesopen.com.
18. G. W. Bemis and M. A. Murcko, *J. Med. Chem.*, **1996**, *39*, 2887-2893.
19. G. Moreau and P. Broto, *Nouv. J. Chim.*, **1980**, *359*, 359.
20. R. Todeschini, V. Consonni, in *Recent Adv. QSAR Stud.*, **2010**, pp. 29-103.
21. C. W. Yap, *J. Comput. Chem.* **2011**, *32*, 1466–1474.
22. R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, **2012**, ISBN 3–900051–07–0. http://www.R-project.org/.
23. J. Kirchmair, P. Markt, S. Distinto, G. Wolber and T. Langer, *J. Comput. Aided. Mol. Des.*, **2008**, *22*, 213-228.
24. S. Han, V. Zhou, S. Pan, Y. Liu, M. Hornsby, D. McMullan, H. E. Klock, J. Haugen, S. A Lesley, N. Gray, J. Caldwell and X. Gu, *Bioorg. Med. Chem. Lett.*, **2005**, *15*, 5467-5473.

25. D. M. Berger, M. Dutia, D. Powell, M. B. Floyd, N. Torres, R. Mallon, D. Wojciechowicz, S. Kim, L. Feldberg, K. Collins and I. Chaudhary, *Med. Chem.*, **2008**, *16*, 9202-9211.
26. C. V. N. S. Varaprasad, D. Barawkar, H. El Abdellaoui, S. Chakravarty, M. Allan, H. Chen, W. Zhang, J. Z. Wu, R. Tam, R. Hamatake, S. Lang, and Z. Hong, *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 3975-3980.
27. H. El Abdellaoui, C. V. N. S. Varaprasad, D. Barawkar, S. Chakravarty, A. Maderna, R. Tam, H. Chen, M. Allan, J. Z. Wu, T. Appleby, S. Yan, W. Zhang, S. Lang, N. Yao, R. Hamatake, and Z. Hong, *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 5561-5566.
28. G. Melagraki, A. Afantitis, H. Sarimveis, O. Igglessi-Markopoulou, P. Koutentis and G. Kollias, *Chem. Biol. Drug Des.*, **2010**, *76*, 397-406.
29. B. M. Reddy, K. Tanneeru, P. A. Meetei and L. Guruprasad, *Chem. Biol. Drug Des.*, **2012**, *79*, 84-91.
30. H. Park, S. Lee and S. Hong, *Bioorg. Med. Chem. Lett.*, **2012**, *22*, 4946-4950.
31. K. Amaning, M. Lowinski, F. Vallee, V. Steier, C. Marcireau, A. Ugolini, C. Delorme, F. Foucalt, G. McCort, N. Derimay, C. Andouche, S. Vougier, S. Llopart, N. Halland and A. Rak, *Bioorg. Med. Chem. Lett.*, **2013**, *23*, 3620-3626.
32. T. Noeske, A. Jirgensons, I. Starchenkovs, S. Renner, I. Jaunzeme, D. Trifanova, M. Hechenberger, T. Bauer, V. Kauss, C. G. Parsons, G. Schneider and T. Weil, *Chem. Med. Chem.*, **2007**, *2*, 1763-1773.
33. C. Posch, H. Moslehi, L. Feeney, G. a Green, A. Ebaee, V. Feichtenschlager, K. Chong, L. Peng, M. T. Dimon, T. Phillips, A. I. Daud, T. H. McCalmont, P. E. LeBoit and S. Ortiz-Urda, *Proc. Natl. Acad. Sci. U.S.A.*, **2013**, *110*, 4015-4020.
34. C. A. Dehelean, S. Feflea, D. Gheorgheosu, S. Ganta, A. M. Cimpean, D. Muntean and M. M. Amiji, *J. Biomed. Nanotechnol.*, **2013**, *9*, 577-589.