



King Saud University

Saudi Pharmaceutical Journal

www.ksu.edu.sa
www.sciencedirect.com



ORIGINAL ARTICLE

Alignment independent 3D-QSAR, quantum calculations and molecular docking of Mer specific tyrosine kinase inhibitors as anticancer drugs



Fereshteh Shiri ^{a,*}, Somayeh Pirhadi ^b, Jahan B. Ghasemi ^b

^a Department of Chemistry, University of Zabol, P.O. Box 98615-538, Zabol, Iran

^b Drug Design in Silico Laboratory, Chemistry Faculty, K.N. Toosi University of Technology, Tehran, Iran

Received 14 February 2015; accepted 13 March 2015

Available online 31 March 2015

KEYWORDS

Mer receptor tyrosine kinase;
GRIND;
Enhanced replacement
method;
Fractional factorial design;
Molecular docking

Abstract Mer receptor tyrosine kinase is a promising novel cancer therapeutic target in many human cancers, because abnormal activation of Mer has been implicated in survival signaling and chemoresistance. 3D-QSAR analyses based on alignment independent descriptors were performed on a series of 81 Mer specific tyrosine kinase inhibitors. The fractional factorial design (FFD) and the enhanced replacement method (ERM) were applied and tested as variable selection algorithms for the selection of optimal subsets of molecular descriptors from a much greater pool of such regression variables. The data set was split into 65 molecules as the training set and 16 compounds as the test set. All descriptors were generated by using the GRIND INdependent descriptors (GRIND) approach. After variable selection, GRIND were correlated with activity values (pIC₅₀) by PLS regression. Of the two applied variable selection methods, ERM had a noticeable improvement on the statistical parameters of PLS model, and yielded a q^2 value of 0.77, an r^2_{pred} of 0.94, and a low RMSEP value of 0.25. The GRIND information contents influencing the affinity on Mer specific tyrosine kinase were also confirmed by docking studies. In a quantum calculation study, the energy difference between HOMO and LUMO (gap) implied the high interaction of the most active molecule in the active site of the protein. In addition, the molecular electrostatic potential energy at DFT level confirmed results obtained from the molecular docking. The identified key features obtained from the molecular modeling, enabled us to design novel kinase inhibitors.

© 2015 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author. Tel.: +98 5424822186; fax: +98 5424822180.

E-mail address: Fereshteh.shiri@gmail.com (F. Shiri).

Peer review under responsibility of King Saud University.

1. Introduction

Receptor tyrosine kinases (RTKs) are a large family of cell-surface transmembrane receptors having an important function in both normal and malignant cells to signal



Production and hosting by Elsevier

<http://dx.doi.org/10.1016/j.jsps.2015.03.012>

1319-0164 © 2015 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

transduction (Verma et al., 2011). One subfamily of RTKs is referred to as the TAM family, containing Tyro-3, Axl, and Mer. The Mer (Mertk, Nyk, c-Eyk) protein consists of an extracellular domain with 2 immunoglobulin-like and 2 membrane proximal fibronectin III motifs, a transmembrane region, and an intracellular tyrosine kinase domain (Graham et al., 1995, 1995). Mer is expressed in hematopoietic lineages such as natural killer (NK) cell monocytes, dendritic cells, macrophages, megakaryocytes, and platelets (Angelillo-Scherrer et al., 2001). Mer *in vivo* regulates macrophage activation, promotes apoptotic cell engulfment, and supports platelet aggregation and clot stability. Mer overexpression has been reported in neoplastic progression of several human cancers and has been correlated with poorer prognosis. The growth arrest specific protein 6 (Gas6) as the biological ligand for Mer is a member of the vitamin K dependent protein family (Chen et al., 1997; Stitt et al., 1995). Gas6 is the common ligand among TAM family and interaction of Gas6 with Mer, Axl, and Tyro-3 is important in platelet degranulation and aggregation in response to known agonists.

In the QSPR/QSAR theory, the ultimate goal is to develop mathematical models for the estimation of relevant properties and chemical and biological activities of interest (Pirhadi et al., 2014). It is important especially in cases where they cannot be experimentally determined. The molecular descriptors as the most important building block of this process can be obtained experimentally or computed through mathematical formulas obtained from different theories, such as quantum mechanics, chemical graph theory, and information theory. Herein, we used GRid INdependent descriptors (GRIND) to drive valid and predictive 3D-QSAR models that are more interpretable and efficient. Structural alignment of compounds has a significant effect on the accuracy of the related models and the most important feature of GRIND is that there is no need for alignment of compounds. Wherein, the obtained descriptors are not sensitive to the coordinate frame of the space, and are free of error due to the alignment. GRIND are also easily interpreted by going back to the compounds. So, the original information can be obtained. Fractional factorial design and enhanced replacement method approaches were used to select descriptors for PLS model building. The validation of models was carried out through a prediction set, and leave one-out cross-validation. Furthermore, the principles of organization for economic cooperation and development (OECD) for regulatory acceptability of QSARs were considered (OECD principles for the validation). Also, docking study was used to analyze the interaction pattern between inhibitors and protein, and confirm the obtained results of ERM-PLS and FFD-PLS models.

The electronic effects play a key role in the identification of important interactions between potential drugs and desired targets in the drug discovery process (Feng et al., 2005). In the present work, docking studies and quantum chemistry calculations were done on the basis of DFT theory in order to explore those amino acids involved in the binding site of the Mer tyrosine kinase. The special attention was paid to electronic effects, which are related to HOMO and LUMO energies and molecular electrostatic potential map.

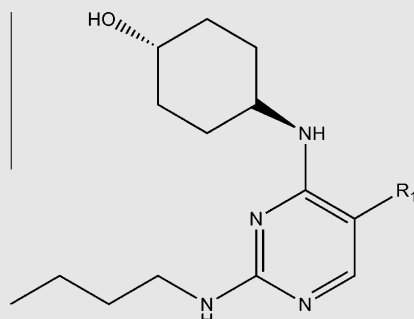
2. Materials and methods

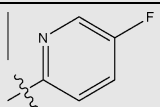
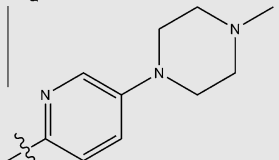
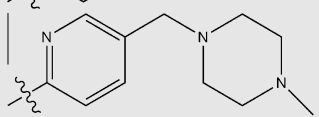
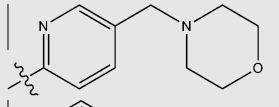
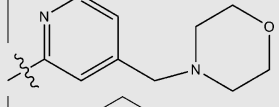
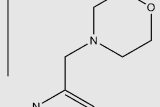
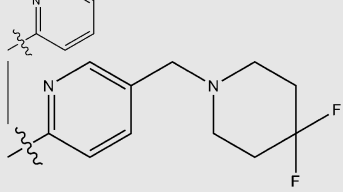
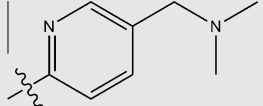
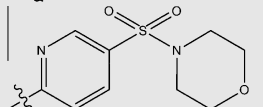
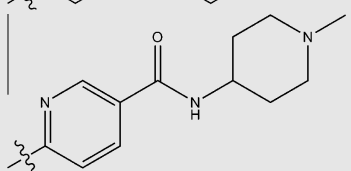
2.1. Data set

All 81 small molecule Mer inhibitors and their biological activities (IC_{50} values) were collected from two references (Zhang et al., 2013a,b). For the QSAR analysis, IC_{50} values were taken in molar range and were expressed in negative logarithmic units, pIC_{50} ($-\log IC_{50}$). The chemical structures and biological activity values of all compounds are shown in Table 1. The selection of training and test sets was done by considering both structural diversity and activity. The training set of 65 molecules was used to adjust the parameters of models and 16 rests of molecules were used to evaluate the model prediction ability. In November 2004, the OECD member countries regarded five principles for the validation of (Q)SAR models for regulatory purposes, now known as the OECD principles for (Q)SAR validation. According to these principles, a (Q)SAR model should be associated with the following information: a defined endpoint; an unambiguous algorithm; a defined domain of applicability; appropriate measures of goodness-of-fit, robustness and predictivity, and a mechanistic interpretation if possible. Wherein, pIC_{50} was regarded as the endpoint, two PLS models based on FFD and ERM variable selection methods were constructed. Applicability domain of the best predictive model along with r^2 , q^2 , and r_{pred}^2 of the models was determined. Then, based on effective GRIND a mechanistic interpretation was done.

2.2. Calculation of descriptors

GRid-INdependent descriptors are a new class of molecular descriptors developed by Pastor et al. (2000). Pentacle program has been proposed as a tool to build models. GRid (Goodford, 1985) molecular interaction fields (MIFs) of nodes are computed by four GRid probes, and a pair of nodes (GRid MIF minima) are used as descriptors (variables). Only those pairs of nodes (for the same or different probe types) with the highest product of interaction energy (IE), at the given distance range, were used for the PLS analysis. For the derivation of MIFs, four most recommended probes were used. To represent steric and hydrophobic interactions, hydrogen bond acceptor, and hydrogen bond donor groups, we used DRY (hydrophobic probe), O (carbonyl oxygen), and N1 (amide nitrogen), respectively. These probes stand for strong non-covalent interactions between molecules and receptor. Moreover, to regard molecular shape effects in the receptor ligand interaction process, and as complementary to point interaction based information, a supplementary probe, called TIP (shape probe), was applied that extracts each ligand's isosurface at 1 kcal/mol from the field of a normal GRid calculation. AMANDA algorithm as implemented in the software (Durán et al., 2008) was applied for the filtering. Maximum auto and cross-correlation (MACC2) algorithm was applied for the encoding. The grid spacing was set to 0.5 Å and the smoothing windows to 0.8. The MACC2 analysis output is usually represented directly in correlograms where each point represents the product of two particular nodes within the distance bin separating the nodes of a certain compound.

Table 1 Structures of Mer specific tyrosine kinase inhibitors along with their pIC₅₀ values.


Compound	R ₁	pIC ₅₀
M01		8.2
M02		8.55
M03		8.41
M04 ^a		8.77
M05		7.92
M06 ^a		7.74
M07		8.96
M08		8.77
M09		9.15
M10		9.16

(continued on next page)

Table 1 (continued)

Compound	R_1	pIC_{50}
M11		8.47
M12 ^a		9.16
M13		8.89
M14		9.09
Compound	R_2, R_2', N	pIC_{50}
M15		5.9
M16		8.74
M17 ^a		8.35
M18		7.74
M19		7.47

Table 1 (continued)

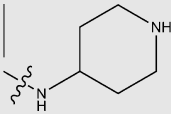
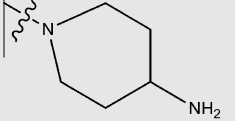
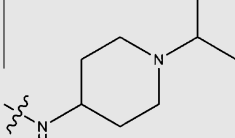
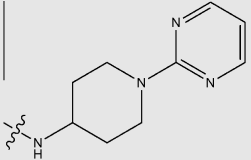
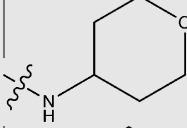
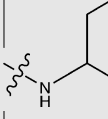
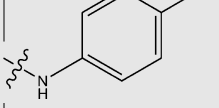
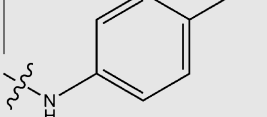
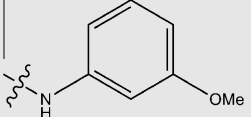
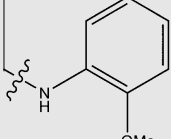
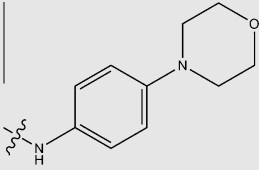
Compound	R_1	pIC_{50}
M20		7.14
M21		6.8
M22		6.7
M23		5.95
M24		6.22
Compound	$R_3R'_3N$	pIC_{50}
M25 ^a		5.57
M26		6.27
M27		6.77
M28		7.3
M29		7.72
M30		7.36
M31		6.42
M32		4.78
M33		7.08

(continued on next page)

Table 1 (continued)

Compound	R_1	pIC_{50}	
M34		6.96	
M35 ^a		7.85	
M36 ^a		6.49	
M37		6.66	
M38		7.36	
Compound	R_1'	NHR_2	pIC_{50}
M39			7.4
M40			4.9
M41			7.64
Compound	R_1	pIC_{50}	
M42 ^a		7.17	
M43		5.23	

Table 1 (continued)

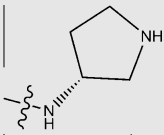
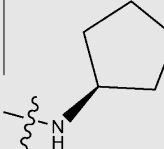
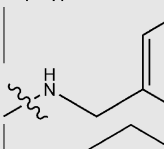
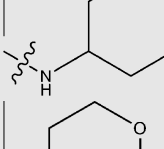
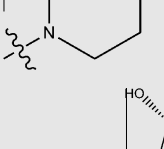
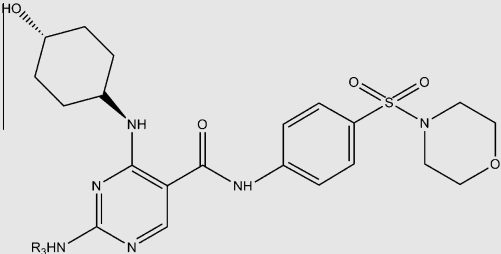
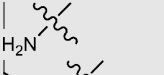
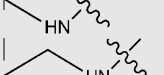
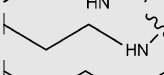
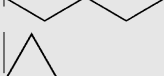
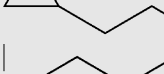
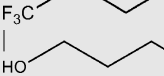
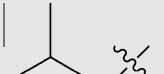
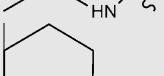
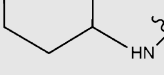

Compound	R_1	pIC_{50}
M44		8.05
M45		5.6
M46		8.08
M47		8.39
M48 ^a		7.89
M49 ^a		8.1
M50		6.64
M51		7.82
M52		7.54
M53 ^a		7.41
M54		7.85

(continued on next page)

Table 1 (continued)

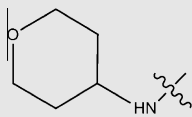
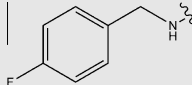
Compound	R ₁	pIC ₅₀
M55		8.17
M56		8.2
M57		7.85
M58		8.28
M59		8.37
Compound	NHR ₂	pIC ₅₀
M60 ^a		8.68
M61 ^a		6.23
M62		7.43
M63		6.85
M64		7.21

Table 1 (continued)

Compound	R_1	pIC_{50}
M65		6
M66		6.96
M67		6.21
M68 ^a		6.77
M69		5
		
Compound	NHR_3	pIC_{50}
M70		5.62
M71		6.82
M72		7.43
M73		7.96
M74		7.57
M75		8.14
M76		7.07
M77		7.23
M78 ^a		7.62
M79		7.33

(continued on next page)

Table 1 (continued)

Compound	R_1	pIC ₅₀
M80 ^a		6.74
M81		6.92

^a Test set.

2.3. Variable selection and modeling

Several methods can be used to reduce the original pool of descriptors and to extract the most informative variables with an appropriate size. Uninformative descriptors that have either no information or those redundant with other descriptors present in the descriptor pool would be eliminated (Ghasemi and Tavakoli, 2012). We used FFD and ERM methods as the variable selection approaches to extract the most informative GRIND. Then, partial least squares method was applied for model building. FFD was applied several times on the generated set of descriptors to reach the stable and reliable results. The optimum number of PLS components was chosen by monitoring changes in the model's predicting index (q_{too}^2 , leave one-out) evaluated by applying the cross-validation procedure available in the pentacle. The PLS coefficients stand for the contribution of each single variable to the model only with respect to the Y . Positive value coefficients increase the inhibition potency of the compound, and vice-versa. Before using the ERM approach, all zero value variables were deleted. Also, correlations of each two variables were checked and between those highly correlated (above 0.95), the one with lower correlation with Y was discarded. The replacement method (RM) has been evolved from stepwise algorithm. The first report was by Duchowicz et al. for the QSPR study on normal boiling points of some organic compounds (Duchowicz et al., 2005). To minimize the standard deviation (S), it systematically searches the pool of D ($N \times D$) variables according to the MLR procedure, to find d optimal descriptors:

$$S = \frac{1}{(N - d - 1)} \sum_{i=1}^N \text{res}_i^2 \quad (1)$$

where N is the number of molecules in the training set and res_i is the difference between the experimental and the predicted properties for the molecule i . The RM first chooses a vector of d descriptors at random and establishes a linear regression (Duchowicz et al., 2006). Then, in the resulting set, each time a variable with the highest standard deviation value in its coefficients is substituted with all of the remaining $D - d$ descriptors, one by one (without considering the one(s) changed previously). This procedure is repeated until the standard deviation value does not decline by more replacements. Then, descriptors having the smallest value of S (in Eq. (1)) as the final optimal sets of d are retained. In the modified replacement

method (MRM) version, the descriptor with the largest error is substituted even if that replacement is not accompanied by a smaller value of S . The order of RM–MRM–RM is named ERM. It judiciously removes the noisy variables from informative ones in a semi-full search manner (Mercader et al., 2010).

2.4. The Applicability Domain (AD)

Having a defined domain of applicability is one of the five principles for the validation of (Q)SAR models for regulatory purposes, now referred to as the OECD principles for (Q)SAR validation. This need is based on the fact that (Q)SARs have unavoidably limitations in terms of types of physicochemical properties, chemical structures, and mechanisms of action for which they can make reliable predictions. In the assessment of a new compound, without any experimental value, it is not possible to calculate the standardized residual, so the decision can only be based on the leverage. The space of domain of applicability is generated by the descriptors of the training set and corresponding biological activities. If the predicted biological activity for a compound falls within this domain, it may be considered as reliable (Gramatica, 2007). The Williams plot is sometimes used to identify compounds that are outside the AD on the basis of both leverages and standardized residuals.

2.5. Docking studies

All molecule structures were sketched in the ChemDraw program. After that, they were transferred into the Discovery Studio 2.5 (Accelrys Inc, San Diego, CA, USA). Compounds were then typed with a CHARMM force field and partial charges were calculated by the Momany–Rone option (Momany and Rone, 1992). A smart minimizer algorithm was used to minimize the resulting structures, which performs 1000 steps of the steepest descent with a RMS gradient tolerance of three which was followed by a conjugate gradient minimization. The X-ray crystal structure of the tyrosine-protein kinase MER in complex with the inhibitor UNC1917 was taken from PDB (4M3Q) and was used for docking studies. All protein preparation and minimization were done using tools and protocols in the Discovery Studio 2.5. A CHARMM force field was used to type the complex. Using the protein preparation protocol, hydrogen atoms were added to the complex, all water molecules were removed and the pH of the protein was set to almost neutral value, 7.4. A sphere binding site with a nine Å radius was

defined around the bonded ligand to identify the binding site of the protein structure. The CDOCKER (CHARMm-based DOCKER) program was chosen as the docking algorithm to dock the most active compound 12 in the set into the tyrosine-protein kinase MER binding pocket. CDOCKER is a CHARMm-based docking tool that considers the receptor rigid and generates several prime random orientations for ligand within the receptor active site followed by the MD-based simulated annealing and the final refinement by minimization (Wu et al., 2003).

2.6. Quantum chemical calculations

In this study, the geometry optimization and calculations for the bioactive conformer obtained in the docking studies were performed at the density functional theory (DFT) level on a personal computer (PC) by energy optimization, using the Gaussian 09 (Frisch et al., 2009) program package. The highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) energies at B3LYP/6-31G (d,p) level were calculated for the bioactive conformer obtained in docking studies. The molecular electrostatic potential (MEP) was also studied at the same level, and was created using Molekel (Varetto, 2009).

3. Results and discussion

3.1. Docking results

Docking computations were carried out to explore the probable binding conformations of potential inhibitors of the receptor and to inspect significant interactions with the protein. Root-mean-square distance (RMSD) value was calculated between the co-crystal (UNC1917) and the re-docked one

using the CDOCKER to validate the docking reliability which was found to be 1.39 Å. The value shows a high reliability of the docking method to reproduce the experimentally binding mode of inhibitors. The best docked conformation of compound 12 is shown in Fig. 1. In compound 12, a cyclopropylamino serves as a R_1 substituent at the 5-position of the pyridine ring, which is exposed mostly to the solvent. There are some van der Waals contacts between cyclopropylamino and Thr681, and Tyr685. The hydroxyl group in the cyclohexyl substituent forms a hydrogen bond with the side chain of Asp741. The cyclohexyl ring itself may contain favorable hydrophobic interactions with residues such as Val601, and Met730. Also the NH from the butylamino side chain forms a hydrogen bond with the backbone carbonyl of Pro672. Another hydrogen bond between compound 12 and Mer protein is formed between pyrimidine-amino group and the backbone NH of Met674.

3.2. Results of the ALMOND model

DRY, O, N1, and TIP probes were used to compute PLS coefficients for the models (Ghasemi and Davoudian, 2014). The FFD variable selection method chose a set of 370 descriptors from a pool of 680. The best statistical PLS model was selected with five latent variables based on the highest squared correlation coefficient (r^2) of 0.91, and cross-validated squared correlation coefficient, q^2 , of 0.55, and interpretability and simplicity of the model. The model also resulted in an r^2_{pred} of 0.80. The FFD-PLS coefficient histogram showing the contribution of each single variable to the model versus the value of Y is shown in Fig. 2. Positive values of the coefficients represent a direct correlation to the Y , and the negative ones show an inverse correlation to it. As can be seen, the most important variables that have a positive effect on the biological activity are DRY-O: 279, O-O: 81, TIP-TIP: 211, and O-TIP: 586. In contrast, the analysis of all the distances at higher PLS

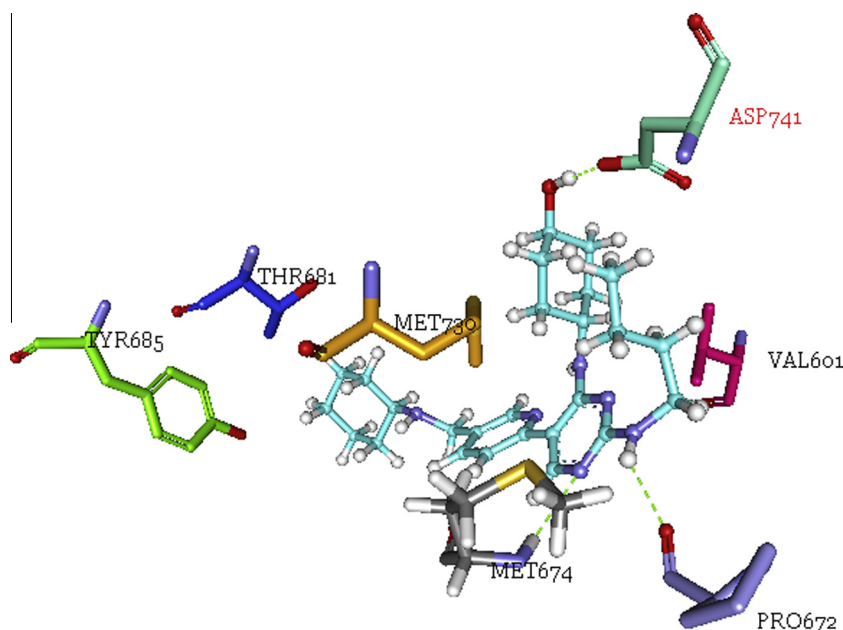


Figure 1 The most active molecule 12 docked into the cavity of Mer kinase.

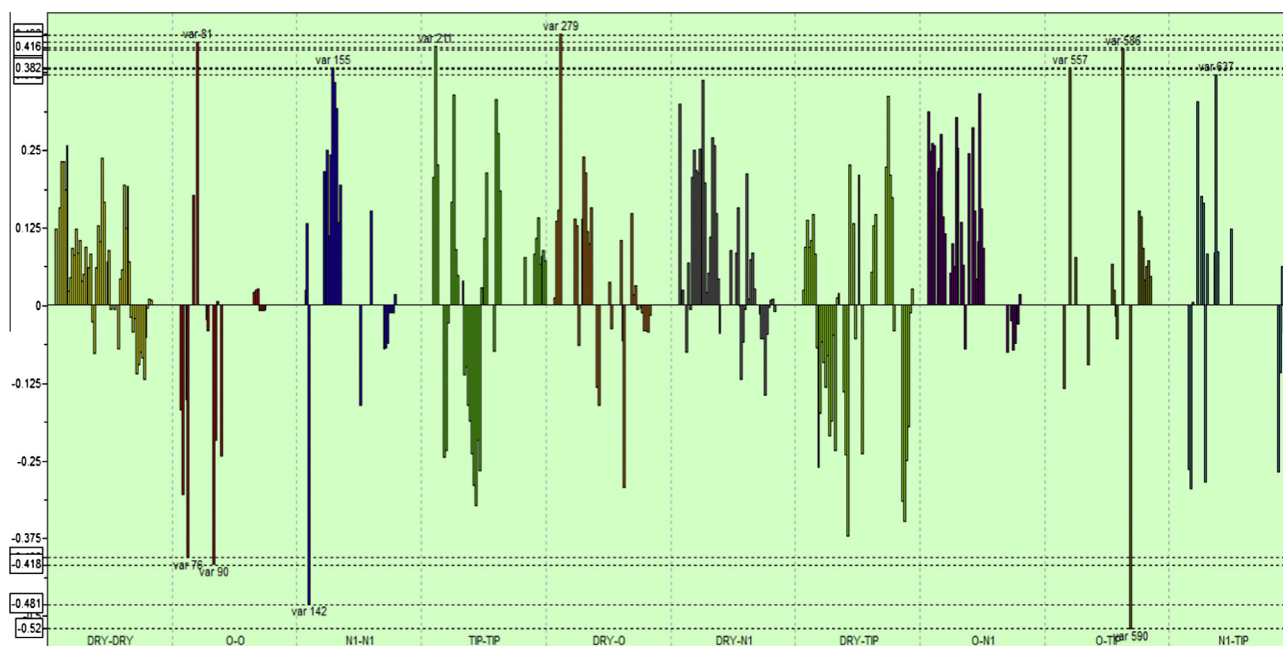


Figure 2 The PLS coefficient histogram of FFD-PLS, showing the importance of single descriptors to explain the Y : positive values of a coefficient indicate a direct correlation to the Y , and negative ones indicate an inverse correlation to the Y .

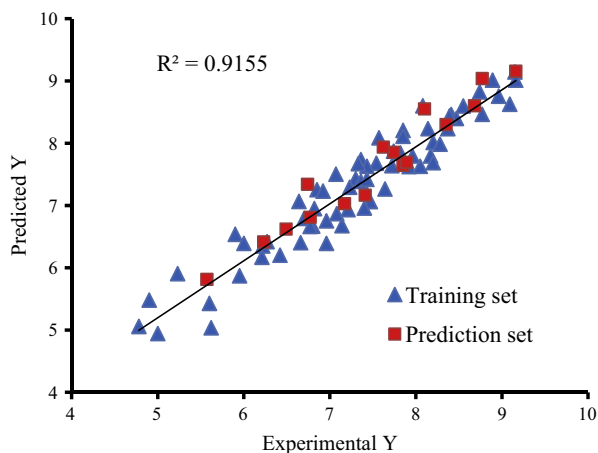


Figure 3 Plot of experimental versus predicted activities for the ERM-PLS model.

coefficients revealed that the variables O-TIP: 590, N1-N1: 142 and O-O: 90 correlate negatively with the activity.

ERM-PLS model was generated based on 20 descriptors selected by ERM approach. The best model was chosen with four latent variables and the lowest error. The statistical results of the model were more superior to the FFD-PLS model. The plot of experimental versus predicted values for this model is shown in Fig. 3. It yielded a $q^2 = 0.77$, a $r^2 = 0.92$, a RMSEC = 0.31, a RMSEP = 0.25 and a r^2_{pred} of 0.94. The most important descriptors (based on VIP scores greater than one) and with positive impacts on Y are O-TIP: 576, N1-TIP: 638, DRY-TIP: 460 and the ones with negative effects are TIP-TIP: 230 and DRY-N1: 358. Williams plot for defining the applicability domain for the best constructed model, using four latent variables for 65 chemicals was considered. The

parameters were defined as $p = 5$, $n = 65$, and $h^* = 3 \times 5 / 65 = 0.23$. With moving from the origin toward the x direction, the unreliability of the predicted values will be increased, and going toward the y direction the predictivity of the model will be decreased (Fig. 4). This Figure shows that the ERM selected variables are so successful that no molecules were identified as an outlier in the ERM-PLS model.

3.3. GRIND interpretation

FFD and ERM were used to choose the most important descriptors generated by AMANDA. According to Fig. 3, the chemical interpretation of the generated FFD-PLS models was carried out by choosing ten most important descriptors: three O-O node pairs, two O-TIPs, two N1-N1, one

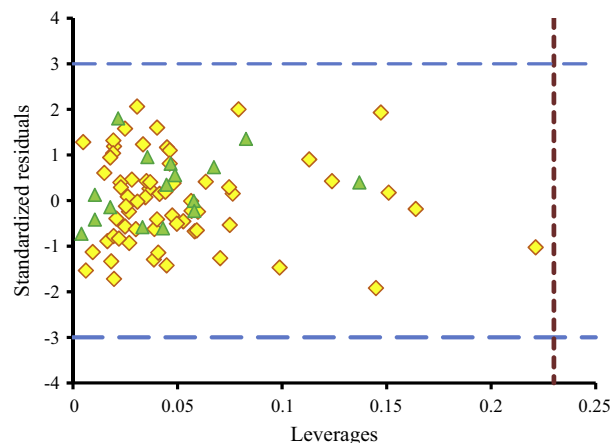


Figure 4 The domain of applicability of ERM-PLS on the GRIND MIFs.

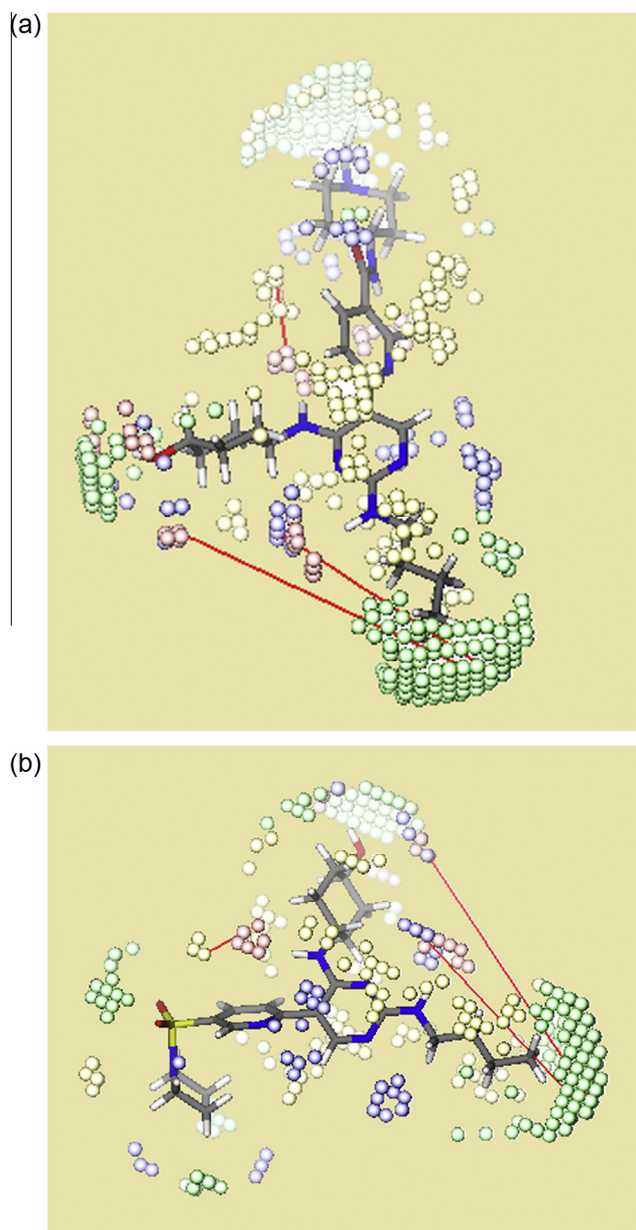


Figure 5 The graphical display of GRIND variables with variables 279 of the DRY–O block, 576 of O–TIP and 638 of N1–TIP block for the selected compounds (a) 10 and (b) 9.

N1–TIP, one TIP–TIP and one DRY–O. The variable 279 with the highest positive effect on Y has a distance of 2.8–3.2 Å and is between DRY hydrophobic node and O acceptor node. It is among the carbonyl of the amide group and the carbon of N-methylpiperazine in the compound 10 (Fig. 5a). This variable can be observed obviously in active compounds such as compounds 9, 14, 7, 13 and 8. The DRY node shows the importance of the hydrophobic interactions between analogues and Mer cavity. The second positive effective descriptor is O–O: 81, (5.2–5.6 Å) that shows the importance of hydrogen bonding acceptor groups. So, for this dataset, hydrogen bonding is very important. As a case in point, in compound 9, one node is directed to the hydroxyl group attached to the cyclohexyl substituent. As mentioned earlier, the polar group on the R^2 site is critical and can have a hydrogen bond with

the carbonyl of Asp741 (Zhang et al., 2013b). For example, removing the hydroxyl group from the cyclohexyl ring to generate ligand 15 significantly declined Mer activity (74-fold), while introducing an amino group at that position boosted the activity of the analogue 16. Remarkably, the PLS coefficient plot reveals that most of the variables of O–N1 correlogram that have positively correlated bars are situated on the right side of the correlogram with the larger node–node distance. There is an inverse relationship between the variable 590 O–TIP at a distance of 18.4–18.8 Å, and the biological activity. This predictor does not occur in some of active compounds such as 9 and 13.

Variable importance in projection (VIP) scores estimate the importance of each variable in the projection used in a PLS model. The PLS–VIP method showed best in identifying relevant variables and outperformed the other approaches. Herein, a general ‘greater than one rule’ is used to choose relevant descriptors. Namely, a variable with a VIP score greater than one can be regarded important in a given model (Chong and Jun, 2005). Among 20 predictors chosen by ERM approach, nine of them had VIP score more than one. The largest VIP scores were related to correlograms TIP–TIP, O–TIP, N1–TIP, DRY–TIP, DRY–N1, O–O, DRY–O and O–N1. The predictor O–TIP: 576 had the highest positive effect on the activity. This variable indicated a significant distance of 12.8–13.2 Å between O and TIP nodes. In the most active training compounds 10, 9, 14, 7, and 13, the acceptor node was seen on the hydroxyl group of the R_2 substituent on the cyclohexyl ring, as indicated before, and this substituent forms a hydrogen bond with the side chain of Asp741, Fig. 5b. The TIP shape node is located on the butyl amino side chain. This flexible lipophilic butyl side chain is fully confined to the relatively small adenine pocket. The second positive predictor is the N1–TIP pair. The N1 variable in compounds 10, 9, 7, 13 and 14 is located on the NH attached to the cyclohexyl ring. A TIP shape node is again situated on the butyl aminoside chain. Another important descriptor is a DRY–TIP. In compound 10, a TIP is situated on the hydroxyl group attached to the cyclohexyl ring, and a DRY is seen on the methyl of N-methylpiperazine. This DRY probe in compound 9 is located on the morpholine substituent.

3.4. Frontier molecular orbital

According to the frontier molecular orbitals theory, HOMO and LUMO energies are two significant indicators of the chemical reactivity. Electron donor and electron acceptor characters of a compound are measured by the HOMO and LUMO energies respectively. The energy difference between the HOMO and LUMO (gap) is also a simple important measure of the molecular stability. A small gap value implies the high reactivity of molecules in reactions while a large gap value implies the high stability of molecules and the low reactivity in reactions.

$$\text{GAP} = E_{\text{HOMO}} - E_{\text{LUMO}} \quad (2)$$

HOMO represents the ability to donate an electron, while LUMO as an electron acceptor represents the ability to obtain an electron. In Fig. 6, HOMO and LUMO of the conformer obtained from docking, together with the HOMO–LUMO gap are given. As can be seen, Fig. 6 reveals that the

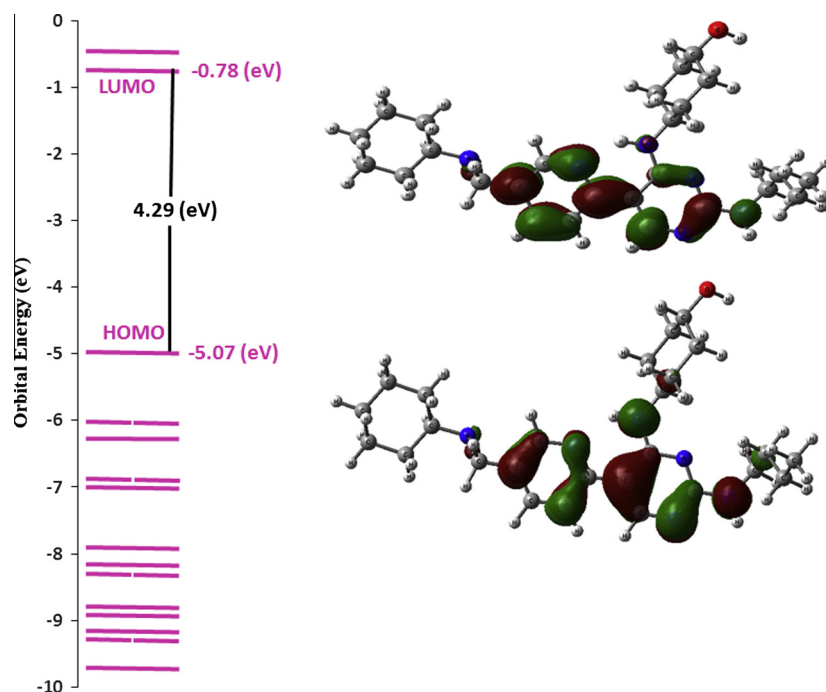


Figure 6 HOMO and LUMO and the energy levels for the most active molecule 12.

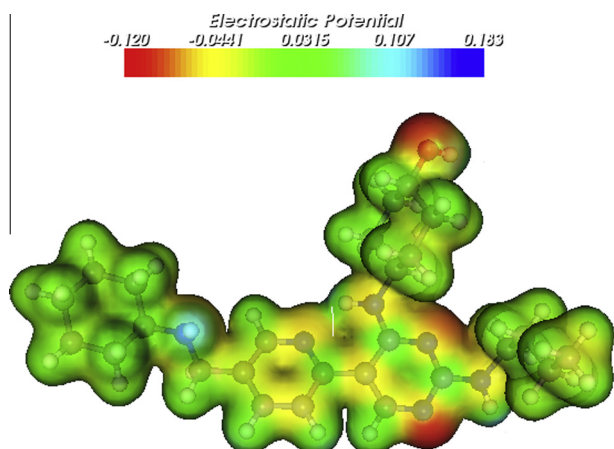


Figure 7 Molecular electrostatic potential map (in a.u.) of the most active molecule 12.

pyrimidine, the pyridine ring and the -NH group have extended HOMO and LUMO densities. These observations confirm the obtained results from the molecular docking.

3.5. Molecular electrostatic potential

The molecular electrostatic potential (MEP), $V(r)$, at a given point $r(x, y, z)$ located in the neighborhood of a molecule can be defined in terms of the interaction energy between the electrical charge generated from the molecule electrons and the nuclei as well as a positive test charge (a proton) placed at r . $V(r)$ values for the studied system were calculated employing the Eq. (2)

$$V(r) = \sum Z_A / |R_A - r| - \int \rho(r') / |r' - r| d^3r' \quad (3)$$

where Z_A is the charge of nucleus A located at R_A , $\rho(r')$ is the electronic density function of the molecule, and r' is the dummy integration variable.

The MEP is related to the electronic density and can also be used as a highly beneficial descriptor for the determination of sites for electrophilic attack and nucleophilic reactions as well as hydrogen-bonding interactions (Kaufman, 1979; Luque et al., 2001). The electrostatic potential $V(r)$ is also well-suited for analyzing processes based on the recognition of one molecule from another one as is in drug-receptor and enzyme substrate interactions. Since it is through the potentials two species first 'see' each other (Politzer et al., 1985). Defined as a real physical property, $V(r)$ can be determined experimentally by diffraction or computational methods (Politzer and Truhlar, 1981).

MEP was calculated at the B3LYP/6-31G (d,p) optimized geometry so that it was possible to anticipate reactive sites for electrophilic and nucleophilic attacks for the titled molecule. As shown in Fig. 7, two regions namely negative (red) and positive (blue) were related to electrophilic and nucleophilic reactivity respectively. As it is clear from Fig. 7, the negative region for the electrophilic attack is at the pyrimidine and pyridine ring with red and yellow color. The negative region is localized on the unprotonated nitrogen atom of the pyrimidine ring, N, with a maximum value of -0.12 a.u. that is an H-bond acceptor from the backbone NH group of Met674 in the docking. However, a maximum positive region is localized on the atom N near to the pyridine ring, probably due to the hydrogen, with a maximum value of 0.183 a.u. Furthermore, a positive region is localized on the atom N near to the pyrimidine ring that is an H-bond donor to the carbonyl

group of Pro672. The existence of an H-bond between –OH of the cyclohexanol and the carbonyl group of Asp741 was confirmed by the positive region of the H in the cyclohexanol. Also, it is in accordance with the GRIND study that there has been an O–TIP: 576 node pair between the –OH of the cyclohexanol and the chain of butyl amine in compound 10, Fig. 5a. Finally, we are very surprising from the agreement of the results of different molecular modeling methods, and the identified key features enabled us to design new potent kinase inhibitors.

4. Conclusions

In this work, we aimed to compare the results of two variable selection algorithms, namely FFD and ERM on a set of GRIND. The quality of results suggested that the ERM was more preferable to the FFD. Also, we followed the OECD guidelines by defining an appropriate end point of pIC_{50} , choosing unambiguous algorithms of FFD-PLS and ERM-PLS, defining the domain of applicability for the best statistical constructed model, having appropriate measures of goodness-of-fit, robustness and predictivity, and a mechanistic interpretation by using GRIND. The interpretation of selected descriptors showed that the polar group on the R^2 site is critical and can form a hydrogen bond with the carbonyl of Asp741. The predictor O–TIP: 576 had the highest positive effect on the activity and the importance of which was confirmed by both docking and MEP studies. Also, DRY nodes showed the importance of the hydrophobic interactions between analogues and the Mer cavity. The results of the molecular docking were in a good agreement with the MEP at a DTF level. Finally, it was possible to anticipate reactive sites of the electrophilic and nucleophilic attack for the titled molecule.

References

- Angelillo-Scherrer, A., de Frutos, P.G., Aparicio, C., Melis, E., Savi, P., Lupu, F., Arnout, J., Dewerchin, M., Hoylaerts, M.F., Herbert, J.-M., 2001. Deficiency or inhibition of Gas6 causes platelet dysfunction and protects mice against thrombosis. *Nat. Med.* 7, 215–221.
- Chen, J., Carey, K., Godowski, P.J., 1997. Identification of Gas6 as a ligand for Mer, a neural cell adhesion molecule related receptor tyrosine kinase implicated in cellular transformation. *Oncogene* 14, 2033–2039.
- Chong, I.-G., Jun, C.-H., 2005. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* 78, 103–112.
- Duchowicz, P.R., Castro, E.A., Fernández, F.M., Gonzalez, M.P., 2005. A new search algorithm for QSPR/QSAR theories: normal boiling points of some organic molecules. *Chem. Phys. Lett.* 412, 376–380.
- Duchowicz, P.R., Fernández, M., Caballero, J., Castro, E.A., Fernández, F.M., 2006. QSAR for non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem.* 14, 5876–5889.
- Durán, A., Martínez, G.C., Pastor, M., 2008. Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in molecular interaction fields. *J. Chem. Inf. Model.* 48, 1813–1823.
- Feng, S., Wang, Z., He, X., Zheng, S., Xia, Y., Jiang, H., Tang, X., Bai, D., 2005. Bis-huperzine B: highly potent and selective acetylcholinesterase inhibitors. *J. Med. Chem.* 48, 655–657.
- Frisch, M., Trucks, G., Schlegel, H., Scuseria, G., Robb, M., 2009. Gaussian 09, Revisions B.1 and C.1. Gaussian, Inc., Wallingford CT.
- Ghasemi, J.B., Davoudian, V., 2014. 3D-QSAR and docking studies of a series of β -carboline derivatives as antitumor agents of PLK1. *J. Chem.* 2014.
- Ghasemi, J.B., Tavakoli, H., 2012. Improvement of the prediction power of the CoMFA and CoMSIA models on Histamine H3 antagonists by different variable selection methods. *Sci. Pharm.* 80, 547.
- Goodford, P.J., 1985. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28, 849–857.
- Graham, D.K., Bowman, G.W., Dawson, T.L., Stanford, W.L., Earp, H.S., Snodgrass, H.R., 1995. Cloning and developmental expression analysis of the murine c-mer tyrosine kinase. *Oncogene* 10, 2349–2359.
- Graham, D.K., Dawson, T.L., Mullaney, D.L., Snodgrass, H.R., Earp, H.S., 1994. Cloning and mRNA expression analysis of a novel human protooncogene, c-mer. *Cell Growth & Differ.: Mol. Biol. J. Am. Assoc. Cancer Res.* 5, 647–657.
- Gramatica, P., 2007. Principles of QSAR models validation: internal and external. *QSAR & Comb. Sci.* 26, 694–701.
- Kaufman, J.J., 1979. Quantum chemical and physicochemical influences on structure–activity relations and drug design. *Int. J. Quantum Chem.* 16, 221–241.
- Luque, F.J., López, J.M., Orozco, M., 2001. Perspective on “Electrostatic interactions of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects”. *Theoretical Chemistry Accounts*. Springer, pp. 343–345.
- Mercader, A.G., Duchowicz, P.R., Fernández, F.M., Castro, E.A., 2010. Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories. *J. Chem. Inf. Model.* 50, 1542–1548.
- Momany, F.A., Rone, R., 1992. Validation of the general purpose QUANTA® 3.2/CHARMM® force field. *J. Comput. Chem.* 13, 888–900.
- OECD principles for the validation, f.r.p., of (quantitative) structure–activity relationship models. <<http://www.oecd.org/dataoecd/33/37/37849783.pdf>> .
- Pastor, M., Cruciani, G., McLay, I., Pickett, S., Clementi, S., 2000. GRIND-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* 43, 3233–3243.
- Pirhadi, S., Shiri, F., Ghasemi, J.B., 2014. Pharmacophore elucidation and 3D-QSAR analysis of a new class of highly potent inhibitors of acid ceramidase based on maximum common substructure and field fit alignment methods. *J. Iranian Chem. Soc.*, 1–8
- Politzer, P., Laurence, P., Jayasuriya, K., 1985. Structure–activity correlation in mechanism studies and predictive toxicology. *Spec. Issue Environ. Health Perspect.* 61, 191.
- Politzer, P., Truhlar, D.G., 1981. Chemical applications of atomic and molecular electrostatic potentials.
- Stitt, T.N., Conn, G., Goret, M., Lai, C., Bruno, J., Radzlejewski, C., Mattsson, K., Fisher, J., Gies, D.R., Jones, P.F., 1995. The anticoagulation factor protein S and its relative, Gas6, are ligands for the Tyro 3/Axl family of receptor tyrosine kinases. *Cell* 80, 661–670.
- Varetto, U., 2009. Molekel 5.4. 0.8. Swiss National Supercomputing Centre, Manno, Switzerland.
- Verma, A., Warner, S.L., Vankayalapati, H., Bearss, D.J., Sharma, S., 2011. Targeting Axl and Mer kinases in cancer. *Mol. Cancer Ther.* 10, 1763–1773.
- Wu, G., Robertson, D.H., Brooks, C.L., Vieth, M., 2003. Detailed analysis of grid-based molecular docking: a case study of

- CDOCKER—A CHARMM-based MD docking algorithm. *J. Comput. Chem.* 24, 1549–1562.
- Zhang, W., McIver, A.L., Stashko, M.A., DeRyckere, D., Branchford, B.R., Hunter, D., Kireev, D., Miley, M.J., Norris-Drouin, J., Stewart, W.M., 2013a. Discovery of mer specific tyrosine kinase inhibitors for the treatment and prevention of thrombosis. *J. Med. Chem.* 56, 9693–9700.
- Zhang, W., Zhang, D., Stashko, M.A., DeRyckere, D., Hunter, D., Kireev, D., Miley, M.J., Cummings, C., Lee, M., Norris-Drouin, J., 2013b. Pseudo-cyclization through intramolecular hydrogen bond enables discovery of pyridine substituted pyrimidines as new mer kinase inhibitors. *J. Med. Chem.* 56, 9683–9692.