J Mol Model (2006) 12: 855–868
DOI 10.1007/s00894-006-0106-2

ORIGINAL PAPER

Nelilma Correia Romeiro · Magaly Girão Albuquerque · Ricardo Bicca de Alencastro · Malini Ravi · Anton J. Hopfinger

Free-energy force-field three-dimensional quantitative structure–activity relationship analysis of a set of p38-mitogen activated protein kinase inhibitors

Received: 11 November 2005 / Accepted: 23 January 2006 / Published online: 16 March 2006

Abstract The p38-mitogen-activated protein kinases (p38-MAPKs) belong to a family of serine–threonine kinases activated by pro-inflammatory or stressful stimuli that are known to be involved in several diseases. Their biological importance, related to the release of inflammatory pro-cytokines such as tumor necrosis factor-α (TNF-α) and interleukin-1 (IL-1), has generated many studies aiming at the development of selective inhibitors for the treatment of inflammatory diseases. In this work, we developed receptor-based three dimensional (3D) quantitative structure–activity relationship (QSAR) models for a series of 33 pyridinyl imidazole compounds [Liverton et al. (1999) 42:2180], using a methodology named free-energy force-field (FEFF) [Tokarski and Hopfinger (1997) 37:792], in which scaled intra- and intermolecular energy terms of the Assisted Model Building Energy Refinement (AMBER) force field combined with a hydration-shell solvation model are the independent variables used in the QSAR studies. Multiple temperature molecular-dynamics simulations (MDS) of ligand–protein complexes and genetic-function approximation (GFA) were employed using partial least squares (PLS) as the fitting functions to develop FEFF-3D-QSAR models for the binding process. The best model obtained in the FEFF-3D-QSAR receptor-dependent (RD) method shows the importance of the van der Waals energy change upon binding and the electrostatic energy in the interaction of ligands with the receptor. The QSAR equations described here show good predictability and may be regarded as representatives of the binding process of ligands to p38-MAPK. Additionally, we have compared the top FEFF-3D-QSAR model with receptor independent (RI) 4D-QSAR models developed in a recent study [Romeiro et al. (2005) 19:385].

Keywords FEFF · 3D-QSAR · p38-MAPK inhibitors · Pyridinyl imidazole · SB-203580 · Inflammation · Anti-inflammatory drugs

N. C. Romeiro · M. G. Albuquerque · R. B. de Alencastro
Laboratório de Modelagem Molecular (LabMMol),
Departamento de Química Orgânica, Instituto de Química,
CCMN, Universidade Federal do Rio de Janeiro, CT,
Bloco A, Sala 609, Ilha do Fundão,
Rio de Janeiro 21949-900, Brazil

N. C. Romeiro (✉) · M. Ravi
Laboratory of Molecular Modeling and Design (LMMD),
M/C 781, Department of Medicinal Chemistry
and Pharmacognosy, College of Pharmacy,
The University of Illinois at Chicago,
833 South Wood Street,
Chicago, IL 60612-7231, USA
e-mail: nelilma@far.fiocruz.br
Tel.: +55-21-33485064
Fax: +55-21-2560-2518

A. J. Hopfinger
College of Pharmacy, MSC09 5360,
1 University of New Mexico,
Albuquerque, NM 87131-0001, USA

Introduction

Molecular modeling using structure-based design methods, in which the three-dimensional (3D) structures of the receptor or ligand–receptor are known, have been widely used in drug design [1–5]. Such methods aim at a better understanding of the features involved in the ligand–receptor binding process. A few methods that use structural information to build receptor-dependent 3D-quantitative structure–activity relationships (RD-3D-QSAR) have also been developed in the past years. They represent an effort to shed light on the physicochemical properties involved in the binding of a ligand to its receptor, by means of a representation of the thermodynamics of binding [6–8].

Hopfinger and coworkers have described a new approach to RD-3D-QSAR formalism termed free-energy force-field (FEFF) 3D-QSAR analysis [8–10]. The FEFF methodology computes the enthalpic and entropic terms for ligand–receptor complexes in a solvent medium. The set of enthalpic and entropic contributions to binding is treated as the independent variables that are used in 3D-QSAR to generate QSAR models. Optimum FEFF-3D-QSAR models are constructed using a Genetic Algorithm coupled to the Partial Least Squares [genetic-function approximation–

partial least squares (GFA-PLS)] method [11]. The ligand–receptor binding states are sampled by means of molecular dynamics simulations (MDS) of reduced models of the ligand–receptor complexes, in an attempt to keep a good compromise between geometric and energetic stability of the system while retaining the reliability of the models.

This paper reports the use of FEFF-3D-QSAR analysis of a set of inhibitors of p38-mitogen-activated protein kinases (p38-MAPK), a member of the large family of protein kinases [12–17] whose function has been related to many diseases, such as diabetes, cancer, and rheumatoid arthritis [12–17]. Due to the biological importance of this protein family in the release of inflammatory pro-cytokines, such as the tumor necrosis factor-α (TNF-α) and interleukin-1 (IL-1), there has been a great effort in the search for new potent and selective inhibitors of this protein for the treatment of inflammatory diseases [18–27].

Materials and methods

The FEFF formalism

The free-energy force-field (FEFF) methodology has been described in full detail elsewhere [8]. Briefly, it represents the ligand–receptor interaction as follows:

$$\underset{\text{(unbound)}}{\text{LM}} + \text{RM} \overset{K}{\rightleftarrows} \underset{\text{(bound)}}{\text{LRM}} \tag{1}$$

where L is the ligand, R is the receptor, M is the solvent medium, and K is the equilibrium or binding constant. The binding free energy, ΔG, of a ligand, L, to a receptor, R, in a solvent medium, M, is equal to the difference in the free energy of the bound and unbound states and can be expressed as:

$$\Delta G = G_{\text{LR}} - (G_{\text{L}} + G_{\text{R}}) = -kT\text{Ln}K \tag{2}$$

where G_{LR} is the free energy of the bound or complex state, G_{L} is the free energy of the unbound ligand, G_{R} is the free energy of the unbound receptor, k is the gas constant, and T is the temperature of the system. In spite of the fact that the free energy of formation of the complex, LRM, cannot actually be broken into a set of component terms, this kind of approximation is thought to be reasonable and has thus, been adopted as part of the empirical nature of the FEFF-3D-QSAR formalism. Based on these theories, the free energy of a ligand–receptor complex can be decomposed into its component terms:

$$\begin{aligned} G_{\text{LR}} =& [G_{\text{LR}}(\text{LL}) + G_{\text{LR}}(\text{RR}) + G_{\text{LR}}(\text{MM}) + G_{\text{LR}}(\text{LR}) \\ &+ G_{\text{LR}}(\text{LM}) + G_{\text{LR}}(\text{RM})] \end{aligned} \tag{3}$$

in which $G_{\text{LR}}(XY)$ is related to the interaction between X and Y for LR (bound state).

Each term in Eq. 3 can be further partitioned in enthalpic (H_{LR}) and entropic (S_{LR}) contributions, according to the equation:

$$G_{\text{LR}} = H_{\text{LR}} - TS_{\text{LR}} \tag{4}$$

Moreover, a representation of the enthalpic terms, $H_{\text{LR}}(XY)$, can be obtained through the internal energy contribution, $E_{\text{LR}}(XY)$, as the work term, $P\Delta V$, of binding, can be neglected for small organic molecules (solutes) complexed with macromolecules at low concentrations. Therefore, we have:

$$\begin{aligned} H_{\text{LR}} =& E_{\text{LR}} = [E_{\text{LR}}(\text{LL}) + E_{\text{LR}}(\text{RR}) + E_{\text{LR}}(\text{MM}) \\ &+ E_{\text{LR}}(\text{LR}) + E_{\text{LR}}(\text{LM}) + E_{\text{LR}}(\text{RM})] \end{aligned} \tag{5}$$

The entropic contribution can be described as follows:

$$\begin{aligned} S_{\text{LR}} =& [S_{\text{LR}}(\text{LL}) + S_{\text{LR}}(\text{RR}) + S_{\text{LR}}(\text{MM}) \\ &+ S_{\text{LR}}(\text{LR}) + S_{\text{LR}}(\text{LM}) + S_{\text{LR}}(\text{RM})] \end{aligned} \tag{6}$$

Finally, the free energies of the isolated ligand, G_{L}, and the receptor, G_{R}, can also be partitioned into the following components:

$$G_{\text{L}} = [G_{\text{L}}(\text{LL}) + G_{\text{L}}(\text{LM}) + G_{\text{L}}(\text{MM})] \tag{7}$$

$$G_{\text{R}} = [G_{\text{R}}(\text{RR}) + G_{\text{R}}(\text{RM}) + G_{\text{R}}(\text{MM})] \tag{8}$$

As we have demonstrated with the other terms, the enthalpic contributions of L, R, and LR at low solute concentrations can also be represented by the change in their individual free-energy force-field terms:

$$\begin{aligned} \Delta G =& \alpha_1 \Delta E_{\text{stretch}} + \alpha_2 \Delta E_{\text{bend}} + \alpha_3 \Delta E_{\text{torsion}} \\ &+ \alpha_4 \Delta E_{\text{vdW}} + \alpha_5 \Delta E_{\text{electrostatic}} \\ &+ \alpha_6 \Delta E_{\text{hydrogenbonding}} + \alpha_7 \Delta E_{\text{solvation}} \\ &+ \alpha_8 T \Delta S \end{aligned} \tag{9}$$

In Eq. 9, $\Delta E_{\text{stretch}}$ corresponds to unbound to bound change in internal energy for bond stretching, ΔE_{bend} corresponds to the change in bond angle bending energy, $\Delta E_{\text{torsion}}$ represents the change in the torsional energy, ΔE_{vdW} is the change in the van der Waals interaction energy, $\Delta E_{\text{electrostatic}}$ is the change in the electrostatics interaction energy, $\Delta E_{\text{hydrogen bonding}}$ is the change in the hydrogen bonding energy, $\Delta E_{\text{solvation}}$ is the change in the solvation energy, and ΔS is the change in the conformational and alignment entropy of the L, R, M system, which can also be partitioned into its individual components.

Most of these terms are included in a classical molecular mechanics force field, and their retrieval from molecular dynamics sampling is the basis of a FEFF-3D-QSAR. The solvation energy may be obtained using the hydration shell method proposed by Hopfinger and Koehler [28]. However, only the L and R components to the free energy of aqueous solvation can be extracted from this model. Furthermore, if a common binding mode in a series of structurally related molecules is generally assumed, the entropic terms may be neglected.

Biological data set

The co-crystal structure of p38-MAPK with SB-203580 (4-(4-fluorophenyl)-2-(4-methylsulfinylphenyl)-5-(4-pyridinyl)-1H-imidazole) (**1a**, Fig. 1), the prototype inhibitor of the pyridinyl imidazole class and other analogous compounds, shows that these inhibitors partially occupy the ATP binding site, preventing further phosphorylation of p38-MAPK [29, 30]. Liverton and coworkers [22] have reported the synthesis and the anti-inflammatory activity evaluation of a series of potent and selective substituted imidazole inhibitors of p38-MAPK [16] that are improved analogues of compound **1a**.

Figure 2 and Table 1 show the structural features and the biological activities (pIC_{50}) of **1a** and the 32 imidazole p38-MAPK inhibitors (**2a**–**t**, ***3a***–**d**, ***4a***–**g**, and **5a**) reported by Liverton and coworkers [22]. The ΔG values of binding for these compounds are not available but it has been assumed that they scale to the measured 50% nanomolar inhibition values, IC_{50} (nM), using the purified enzyme [22], whose values range from 0.11 to 2,100 nM. In this series, the aminobenzyl-substituted pyrimidine compounds demonstrate higher activity against this enzyme than the corresponding pyridine compounds [22]. The IC_{50} (nM) values were converted to molar units and then expressed in negative logarithmic units, $-\log IC_{50}$ or pIC_{50} (Table 1).

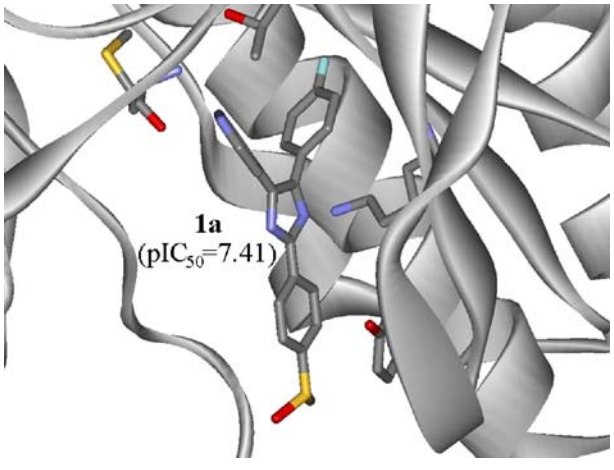
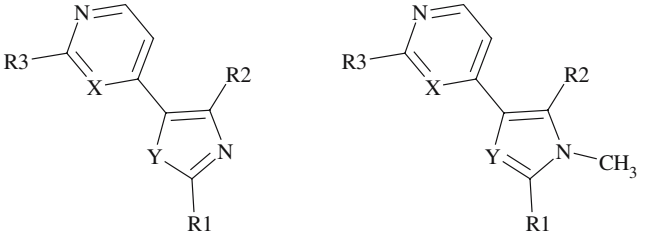


Fig. 1 Compound **1a** (SB-203580, pIC_{50}=7.41), the prototype p38-MAPK inhibitor of the pyridinyl imidazole class, is used as a reference compound for the alignments of the training and test set compounds



1a, 2a-2t, 3a-3d, 4a-4g **5a**

Fig. 2 General structure of compounds **1a**, **2a**–**2t**, ***3a***–**3d**, ***4a***–**4g**, and **5a**, developed by Liverton et al. 1999 [22]. See Table 1 for further structural details

The FEFF-3D-QSAR models were developed and cross-validated (internal validation) using a set of 28 compounds (training data set), selected from the original 33 compounds. The best models were externally validated with the use of five compounds (***2e***, ***2i***, ***2l***, ***3a***, and ***4a***) (evaluation or test data set), selected from the original 33 compounds. Care was taken to include all classes of compounds in both the training and test sets.

Building and energy minimization of the ligands

The three-dimensional (3D) structures of the training and test set (Fig. 2) were constructed using the Builder module available in the Insight II molecular modeling package [31], using the structure of **1a** (SB-203580) co-crystallized with p38-MAPK [30] as a template. This structure is the bound conformation to the enzyme active site, or the active conformation of compound **1a**. The crystallographic structure of human p38-MAPK co-crystallized with **1a** is available in the Protein Data Bank (PDB) [32] under the PDB code 1A9U [32]. After building the molecules of the training and test set, each structure was subsequently energy-minimized and the partial atomic charges were assigned using the AM1 semiempirical method [33], available in the Insight II package [31].

Protein model size determination

The crystallographic structure of p38-MAPK complexed with **1a** contains 2,834 atoms, considering only the heavy atoms from the 354 amino acid residues [30]. As it is less time-consuming to restrict the MDS analysis to a region around the active site, protein binding-site models were constructed using the "pruning" method developed by Hopfinger and colleagues [8] (Fig. 3), which consists in the pruning of the protein using spheres of different diameters, which in this work were of 10, 12, and 14 Å radii, centered around the center of the active site.

This procedure was performed using Hyperchem version 5.01 [34], under the option "select sphere" and the selection mode on residues, so that we would include the whole residues in the sphere. Missing peptide linkages left

Table 1 Structural features and pIC_{50} values of compounds **1a**, **2a**–**2t**, ***3a***–**3d**, ***4a***–**4g**, and **5a** [22]. Training set compounds are in bold and test set compounds are in bold italics

Cmpd Nr.	Structural features					
	R_1	R_2	R_3	X	Y	pIC_{50}
1a	4-$SOCH_3$-Ph	4-F-Ph	H	CH	NH	7.41
2a	Ph	Ph	H	CH	NH	6.74
2b	Ph	3-CF_3-Ph	H	CH	NH	7.07
2C	Ph	4-CF_3-Ph	H	CH	NH	6.70
2d	Ph	2-Cl-Ph	H	CH	NH	6.82
2e	Ph	3-Cl-Ph	H	CH	NH	*6.89*
2f	Ph	4-Cl-Ph	H	CH	NH	7.32
2g	Ph	4-CN-Ph	H	CH	NH	6.01
2h	Ph	4-OCH_3-Ph	H	CH	NH	6.52
2i	Ph	4-OH-Ph	H	CH	NH	*5.85*
2j	Ph	4-Ph-Ph	H	CH	NH	6.00
2k	Ph	3-OH-Ph	H	CH	NH	6.85
2l	Ph	3-F-Ph	H	CH	NH	*6.31*
2m	Ph	3-OCH_3-Ph	H	CH	NH	6.26
2n	Ph	3,4-Cl-Ph	H	CH	NH	7.55
2o	Ph	3,5-Cl-Ph	H	CH	NH	6.86
2p	Ph	*i*-Pr	H	CH	NH	5.68
2q	H	4-F-Ph	H	CH	NH	6.77
2r	cyclohexyl	4-F-Ph	H	CH	NH	7.15
2s	piperidin-4-yl	4-F-Ph	H	CH	NH	8.05
2t	piperidin-3-yl	4-F-Ph	H	CH	NH	7.72
3a	Ph	Ph	H	N	NH	*6.15*
3b	Ph	Ph	CH_3NH	N	NH	6.92
3C	Ph	Ph	4-$CH_3COPhCH_2NH$	N	NH	7.42
3d	Ph	Ph	NH_2	N	NH	7.24
4a	piperidin-4-yl	3-CF_3-Ph	4-$CH_3COPhCH_2NH$	N	NH	*8.12*
4b	piperidin-4-yl	3-CF_3-Ph	NH_2	N	NH	8.15
4C	piperidin-4-yl	3-CF_3-Ph	(*S*)-$PhCH(CH_3)NH$	N	NH	9.12
4d	piperidin-4-yl	3-CF_3-Ph	(*R*)-$PhCH(CH_3)NH$	N	NH	9.01
4e	piperidin-4-yl	3-CF_3-Ph	(*S*)-$PhCH(CH_3)NH$	N	NCH_3	9.96
4f	piperidin-4-yl	3-CF_3-Ph	(*R*)-$PhCH(CH_3)NH$	N	NCH_3	8.82
4g	piperidin-4-yl	3-CF_3-Ph	(*S*)-$PhCH(CH_3)NH$	CH	NCH_3	9.72
5a	piperidin-4-yl	3-CF_3-Ph	(*S*)-$PhCH(CH_3)NH$	N	N	5.91

in this procedure were filled and methyl groups added to "mimic" the missing α-carbons, keeping the original geometries of the crystal structure. After that, we docked the largest inhibitor in the training set (**4e**, Fig. 2) using **1a** as a template for the alignment, to create the model of a suitable size for further MDS analysis.

Docking of the ligands into the protein models

The energy minimized ligands were "docked" into the active site of the protein using the common imidazole ring of the bound inhibitor, **1a**, as a reference for further root-mean square-deviation (RMSD) evaluations. The super-imposition procedures were performed with the option "overlay" available in Hyperchem 5.01 [34]. The resulting bad contacts between the docked ligands and protein residues in the binding sites were manually relieved. After the docking procedures, each complex was energy minimized using the modified Assisted Model Building Energy Refinement (AMBER) force field available in MOLSIM [35].

Molecular-dynamics simulations of the ligand-protein models

After the energy minimization of the ligand–protein models, molecular-dynamics simulations (MDS) were performed using the software MOLSIM [35] with a modified AMBER force field [36], at a temperature of 303 K for 10 ps, using a step size of 0.001 ps and a temperature bath with a relaxation time of 0.01 ps, coupled to the system. The nonbonded cutoff used was 10 Å and a distance-dependent dielectric function ($\varepsilon_r = D \times r_{ij}$) of $3.5 \times r_{ij}$. The bound ligand, **1a**, was used as the template for

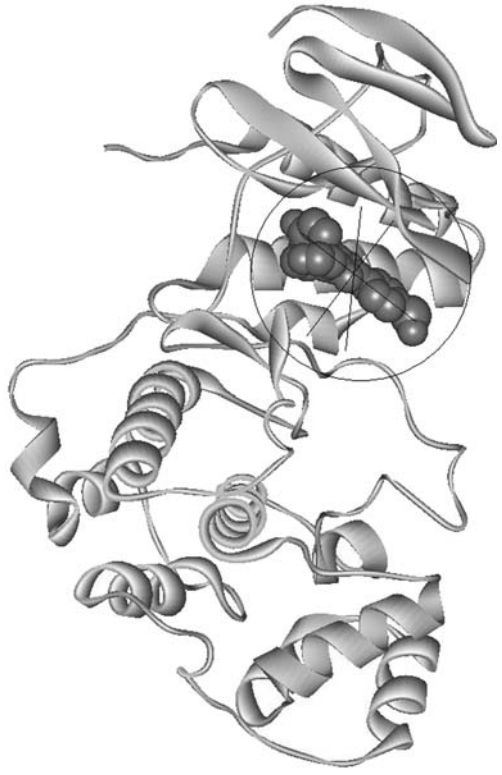


Fig. 3 Schematic representation of the pruning sphere geometry used in the protein model size determination

comparison of the root mean square (RMS) fit of the trial models to the initial crystal structure.

This procedure is aimed at evaluating the conformational and binding integrity of the models, which was also made through the comparison of the total intermolecular van der Waals and electrostatic energies for each of the three protein trial models. It is important to notice that each model was also evaluated in relation to assigned fictitious masses of 2,500 or 5,000 au to the α-carbon atoms of the protein backbone, to avoid larger deviations from the initial crystallographic geometry. Also, the methyl atoms added to the models were constrained with masses of 2,500 or 5,000 au, preventing the explosion of some unconnected residues.

Molecular dynamics temperature sampling

As the temperature of a MDS can only be approximated to the corresponding real temperature of the system, the way the force field represents it may influence the relationship between actual and simulation temperatures [8]. For this reason, temperature is taken as a scaling variable in FEFF-3D-QSAR analyses by means of a temperature-sampling scheme [8] for the bound and unbound states of the system under study: that is, bound and unbound ligand and protein (protein model). In this work, we adapted this scheme and the MDS of the chosen ligand–protein models were initiated at 303 K, the temperature at which the biological tests were performed. A simulation length of 20 ps with a time step of 0.001 ps were used, and the system was coupled to a temperature bath with a relaxation time of 0.01 ps. Next, the lowest-energy conformation of the complex was chosen and the same procedure was repeated at 200, 100, 50, 25, and 10 K. At 10 K, both the ligand and the protein were isolated from the complex. The unbound protein was subjected to MDS for 10 ps, using the same conditions described above for the complexes, except that the temperature was increased to 25, 50, 100, 200, and 303 K.

In a similar fashion, the unbound ligands were submitted to MDS for 50 ps, using the same temperature scheme used for the protein. Solvation energies of the complexes and unbound ligand and protein were calculated using the hydration-shell model proposed by Hopfinger and colleagues [28]. All the internal energy contributions and their corresponding variations for both the bound and unbound states were collected for each simulation temperature and represent the independent variables of the lowest energy states used in the construction of the FEFF-3D-QSAR models.

Construction of the FEFF-3D-QSAR models

The independent (the force field and hydration-shell energy terms and their variations, collected in the previous step) and the dependent variables (the biological activity values, pIC_{50}) correspond to the original Data Base (DB), which was submitted to GFA–PLS [11] calculations using the software Wolf version 5.5 [37]. The optimization of each DB was initiated using 100 randomly generated models, each of them having initially four variables. Mutation probability over the crossover optimization was set to a rate of 10% and the smoothing factor (it controls the number of independent variables in the models) was changed to find equations with three to six terms. We have used four to five components for the PLS regression and 10,000 to 60,000 crossover operations throughout this step of the methodology. The ten best FEFF-3D-QSAR models found by GFA–PLS analysis [11] were ranked according to "lack-of-fit" factor of Friedman (LOF) values [37] and submitted to the "leave-one-out" cross-validation method. The principal statistical parameters that were evaluated were the following: R^2 (squared correlation coefficient), Q^2 (cross-validated R^2), Q^2_{adjusted} (adjusted cross-validated Q^2) [38], standard error of estimate (SE), residual of fit (predicted minus observed pIC_{50} values), and number of outliers.

Results

Selection of the best ligand–protein model

To select the best ligand–protein model size and the most appropriate fictitious mass assignment for the MDS temperature sampling, the pruned complexes of compound **1a** with the protein have been analyzed geometrically and

energetically, considering the 10, 12, and 14 Å radii protein models and the fictitious masses of 2,500 and 5,000 au, adding up to six complexes.

The lowest-energy conformation of each one of the six complexes was used for the geometric and energetic comparisons, by means of the average values of the RMSD obtained with the superimposition with the initial complexes and the intermolecular energy values (Coulomb and van der Waals) of the selected ligand–protein models from the previous procedure, respectively.

In the comparison of the three models to which were assigned masses of 2,500 au, the lowest RMSD value (1.2 Å) for atomic movement was obtained for the 12-Å radius model (Table 2). However, the other values were very close, 1.4 and 1.3 Å for the 10 and 14 Å radii models, respectively. Therefore, it may be concluded that these RMSD values did not allow the unequivocal selection of the best protein model size. As for the analysis of the intermolecular energy values, it was observed that the smaller energy value (−31.60 kcal mol^{-1}) was obtained for the 12-Å radius model, although the 10-Å radius model has a close value (−30.92 kcal mol^{-1}). However, the intermolecular energy value increases abruptly in the 14 Å radius model (−17.5 kcal mol^{-1}), which seems to indicate that, in the case of this model, the simulation was not done long enough to permit the protein model to reach equilibrium, which is presumably due to its larger number of atoms (Table 2).

When the three models with fictitious masses of 5,000 au were compared (Table 2), a similar behavior to the previous ones with masses of 2,500 au was observed. The lower RMSD value (1.2 Å) was obtained for the 10- and 12-Å radii models, while the 14-Å radius model also showed a close RMSD value (1.3 Å). Moreover, analyzing the intermolecular energy values, it was seen that the lowest energy value, −32.29 kcal mol^{-1} was obtained for the 12-Å radius model, although the 10- and 14-Å radii models also have close values, −31.02 and −31.21 kcal mol^{-1}, respectively (Table 2).

Table 2 Root mean square deviation (RMSD) and intermolecular complex energy (E_{LR}(LR)) related to the three models obtained using **1a** co-crystallized with p38-MAPK, with 10, 12, and 14 Å radii and assigning 2,500 and 5,000 au masses (in atomic units) to the atoms in the enzyme backbone

Protein backbone atoms	Fictitious mass 2,500 au		Fictitious mass 5,000 au	
Radius[a] (Å)	RMSD[b] (Å)	E_{LR}(LR)[c] (kcal mol^{-1})	RMSD[b] (Å)	E_{LR}(LR)[c] (kcal mol^{-1})
10	1.4	−30.92	1.2	−31.02
12	1.2	−31.60	1.2	−32.29
14	1.3	−17.50	1.3	−31.21

[a]The center of mass of **4e** complexed to the protein is used as a reference
[b]RMSD between the complexed protein after energy minimization and the crystal structure using the protein backbone atoms
[c]Coulomb and van der Waals energies

These data show that there is a greater stability in the ligand–protein interaction, either in a geometric (close RMSD values) or in an energetic sense (close intermolecular energy values), when fictitious masses of 5,000 au were assigned to the protein backbone atoms, due to the restriction of the protein degrees of freedom. For that reason, the 5,000-au mass assignment scheme was selected as the best one for the FEFF-3D-QSAR studies.

In addition, as the analysis of the three models with fictitious masses of 5,000 au showed close RMSD and intermolecular values, the 12-Å radius model was chosen, which kept reasonable geometric conservation in relation to the crystallographic structure, while maintaining the Coulomb and van der Waals intermolecular energy values close to those obtained for the 10- and 14-Å radii models (Table 2).

Selection of the best FEFF-3D-QSAR models according to the MDS temperature sampling

MDS temperature is one of the parameters in the FEFF-3D-QSAR methodology. Therefore, considering the six temperature values that were evaluated in the MDS studies, namely, 303, 200, 100, 50, 25, and 10 K, six different Data Bases were generated: DB1, DB2, DB3, DB4, DB5, and DB6, respectively.

The analysis of the best models obtained for each simulation temperature showed $Q^2_{adjusted}$ values in a wide range (0.30–0.77), as can be seen in plot I (Fig. 4).

The models with four to six terms obtained from the temperatures of 303 K (DB1) and 50 K (DB4) have the best $Q^2_{adjusted}$ values (0.56 to 0.77). Except for the models derived from the temperatures of 200 and 100 K, the

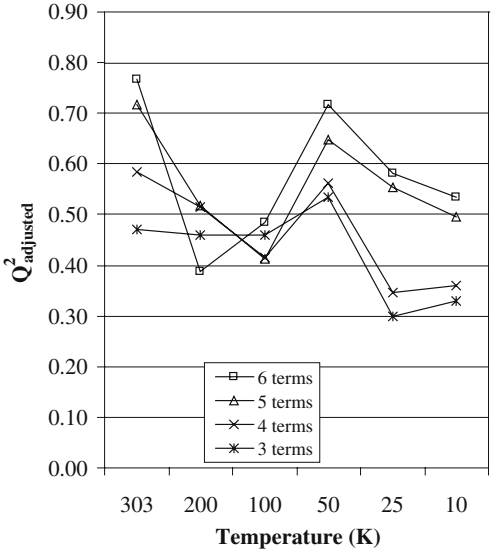


Fig. 4 $Q^2_{adjusted}$ values obtained for the best models with 3 to 6 terms for the FEFF MDS simulation temperatures after GFA–PLS optimization

$Q^2_{adjusted}$ values tend to decrease with decreasing simulation temperatures. This behavior may be associated to the decrease in the vibrational motion due to the reduction of the kinetic energy of the system as the energy terms show a lower variability (data not shown). In other words, as the system freezes, the models become less predictive, probably because the system tends to sample less states that are significant to the binding process.

An interesting feature was observed for the temperatures 200 and 100 K, which showed $Q^2_{adjusted}$ values unexpectedly lower in an average than the values obtained for the other temperatures of simulation. These temperatures likely correspond to transient, unstable states, located between the highest temperature of evaluation (303 K) and lower temperatures (50, 25, and 10 K).

As the models with three and four terms have the lowest $Q^2_{adjusted}$ values and the models with six terms may represent an over fitting of the data in relation to the number of compounds in the training set, we considered for analysis only models with five terms from this point forward. The criteria used for the selection of the best models were as follows: (1) models with five terms; (2) models with $Q^2_{adjusted}$ values ≥0.60; and (3) models with the highest $Q^2_{adjusted}$ values.

Based on these criteria, models A and B (Table 3) were selected as the most representative of the set of models. Model A corresponds to the five-term model obtained from DB1 (303 K) and model B corresponds to the five-term model obtained from DB4 (50 K). The definitions of the force field energy terms present in the best FEFF-3D-QSAR models obtained from DB1 (model A) and DB2 (model B) are given in Table 4.

Analysis of the best FEFF-3D-QSAR models (models A and B)

The best FEFF-3D-QSAR models, namely, models A (303 K) and B (50 K), were previously selected as the most representative of the set of models. The analysis of the cross-correlation matrix between the residuals of models A and B, showed that they are not meaningfully correlated, as the value of R is 0.23, which is very low. This means that the two models are providing different explanatory information about the data set.

In addition, one needs to analyze the cross-correlation matrix of the descriptors of the models individually. In general, R values ≥0.75 indicate a high degree of correlation, which is not desirable, as each descriptor should contribute in a unique way to the models. Consequently, the hypothesis of excluding such models has to be taken into consideration. However, in the FEFF-3D-QSAR ligand–protein models, there is always the chance of some degree of correlation [8], and it is necessary to investigate each case, judging if the correlation is fortuitous, that is, if it is purely mathematical (numerical), and in this case the model should not be invalidated, or if the correlation is real, that is, if the descriptors supply the same type of information about the ligand-receptor system. In this case the model should be excluded. Moreover, even when the correlation is real, the descriptors may supply, partially, the same type of information, or they may be indirectly correlated. In this case, the descriptors likely give some essential insight on the mechanistic nature of the binding process and the model is valid.

Analysis of the FEFF-3D-QSAR model A (303 K)

In the case of model A, an analysis of the cross-correlation matrix of the descriptors (Table 5) showed no significant correlation, except for an inverse correlation between the pairs $E_{LR,ele}$(RR) and $\Delta E_{L,ele}$(LL) (R=−0.50) and $\Delta E_{L,ele}$(LL) and $E_{L,bend}$(LL) (R=−0.52).

The first pair of correlated terms: $E_{LR,ele}$(RR), that represents the total electrostatic energy of the protein in the complex, and $\Delta E_{L,ele}$(LL), which represents the total electrostatic energy of the ligand upon binding, suggest that both are giving individual contributions to model A but there is some partial or indirect correlation between them, as the electrostatic energies of ligand and protein are mutually dependent. Also, both terms have a negative coefficient in the model (Table 3), meaning that the energies must be negative (therefore favorable) to enhance the activity. This observation suggests that electrostatic energy is playing a very important role in the total binding energy.

The second pair of correlated terms corresponds to $E_{L,bend}$(LL), which represents the bending energy of the unbound ligand and $\Delta E_{L,ele}$(LL), which represents the total variation of the electrostatic energy of the ligand upon binding. It is clear in this case that the variation of the electrostatic energy of the ligand upon binding is dependent on the bending energy of the isolated ligand. The latter term is presumably responsible for variations in interatomic distances upon which the electrostatic energy is dependent.

Table 3 Top FEFF-3D-QSAR models obtained from MDS at 303 K (DB1, model A, 5 terms) and 50 K (DB4, model B, 5 terms) for the training set of 28 compounds, inhibitors of p38-MAPK [22]

Model	Equation
A	$pIC_{50}=6.00-0.016\times E_{LR}-0.006\times E_{LR,ele}(RR)-0.080\times \Delta E_{L,ele}(LL)+0.069\times E_{L,bend}(LL)-0.057\times E_{R,vdW}(RR)$ N=28; LOF=0.45; SE=0.187 R^2=0.848; Q^2=0.761
B	$pIC_{50}=7.118+0.134\times \Delta E_{torsion}-0.091\times E_{LR,ele}(LL)+0.237\times \Delta E_{L,vdW}(LL)+0.072\times \Delta E_{R,vdW}(RR)+0.041\times E_{LR}$ N=28; LOF=0.49; LSE=0.203; R^2=0.835; Q^2=0.697

Table 4 Definition of the free-energy force-field (FEFF) energy terms present in the best FEFF-3D-QSAR equations

FEFF	Definition
E_{LR}	The total electrostatic and van der Waals intermolecular energy of the complex
$E_{LR,ele}$(RR)	The intramolecular electrostatic energy of the bound receptor
$\Delta E_{L,ele}$(LL)	The change in the intramolecular electrostatic energy of the ligand on binding
$E_{L,bend}$(LL)	The bending energy of the unbound ligand
$E_{R,vdw}$(RR)	The intramolecular van der Waals energy of the unbound receptor
$\Delta E_{torsion}$	The change in torsional energy on binding
$E_{LR,ele}$(LL)	The electrostatic energy of the bound ligand
$\Delta E_{L,vdw}$(LL)	The change in the van der Waals energy of the ligand upon binding
$\Delta E_{R,vdw}$(RR)	The change in the van der Waals energy of the receptor upon binding

In spite of being different energy terms, there may be a partial or indirect dependence between them, leading both terms to contribute for a better understanding of the ligand–protein binding process.

Interestingly, the energy term, $E_{L,bend}$(LL), in model A, has a positive coefficient (Table 3), suggesting that the energy associated with angle deformation (bending) of the unbound ligand should be positive (unfavorable), if we want to increase the activity. In fact, due to the evolutive model generation process with the GFA–PLS method, we may suggest that this term represents, statistically, the total intramolecular energy variation of the ligand in the binding event. Thus, this term may be a "survivor" of the expression "E_{LR}(LL)−E_{L}(LL)" in the model optimization process, that includes the $E_{L,bend}$(LL) term [10]. This definition makes physical sense in the classic ligand–receptor binding mechanism, as the electrostatic energy of the ligand–protein complex should be more negative (favorable) than the sum of the electrostatic energies of unbound ligand and protein, to favor ligand–protein complex formation energetically. Besides, it is important to emphasize that the main objective in the drug development process is not to find a ligand with the absolute intramolecular energy at a minimum, to achieve optimum activity but to have an energy variation between the bound and unbound states at a minimum.

It should be noted that the energy terms E_{LR} and $E_{R,vdW}$(RR) in model A are effectively orthogonal to the other descriptor terms in the model. The term E_{LR}, which corresponds to the intermolecular energy (electrostatic and van der Waals) between the ligand and the protein has a negative coefficient, meaning that energy should be more negative to increase activity. In fact, the analysis of the Data Base obtained for the MD simulations at 303 K reveals that the values of E_{LR} for the most active compounds from series 3 and 4 (Fig. 2) tend to be more negative (data not labeled).

Finally, the presence of the term $E_{R,vdW}$(RR), that represents the intramolecular van der Waals energy of the protein, reinforces the interpretation that the electrostatic and van der Waals energy, either from the ligand or the protein, predominate in model A, indicating that the FEFF-3D-QSAR obtained fits a classic ligand–receptor model in which there is charge–charge interaction complementarity in the system and a predominance of nonbound energy terms. In agreement with this result, the analysis of the Data Base (data not labeled) shows a trend of decreasing (favorable) values with increasing activity for $E_{R,vdW}$(RR), thus, supporting the model.

Table 5 Linear cross-correlation matrix of the energy descriptors of model A, obtained from DB1 (5 terms) at 303 K, for the training set of 28 compounds, inhibitors of p38-MAPK [22]

	E_{LR}	$E_{LR,ele}$(RR)	$\Delta E_{L,ele}$(LL)	$E_{L,bend}$(LL)	$E_{R,vdW}$(RR)
E_{LR}	1.00				
$E_{LR,ele}$(RR)	−0.15	1.00			
$\Delta E_{L,ele}$(LL)	0.07	−0.50	1.00		
$E_{L,bend}$(LL)	0.00	0.13	−0.52	1.00	
$E_{R,vdW}$(RR)	−0.09	−0.29	−0.07	0.24	1.00

Analysis of the FEFF-3D-QSAR model B (50 K)

An analysis of the cross-correlation matrix of the descriptors of model B (Table 6), reveals that, in general, these descriptors are more cross-correlated than the descriptors in model A (Table 5). This is not an unexpected behavior because as temperature decreases, the system "freezes", and kinetic energy decreases. Consequently, the variance of the energy terms is low, increasing the chance of cross-correlation. The most correlated energy terms in model B (Table 6) are the pairs: $\Delta E_{L,vdW}$(LL) and $\Delta E_{R,vdW}$(RR) (R=−0.85); and $\Delta E_{R,vdW}$(RR) and E_{LR} (R=0.64).

The first pair of correlated energy terms are, $\Delta E_{R,vdW}$(RR), that represents the van der Waals intramolecular energy change of the protein between the bound and the unbound state and $\Delta E_{L,vdW}$(LL), that represents the same energy change for the ligand. These two descriptors show a high inverse correlation coefficient (R=−0.85). This correlation may indicate that the amount of energy that is gained by the protein (or the ligand) should be equivalent to the energy amount lost by the ligand (or the protein) in the binding process.

Table 6 Linear cross-correlation matrix of the energy descriptors in model B, obtained from DB4 (5 terms) at 50 K, for the training set of 28 compounds, inhibitors of p38-MAPK [22]

	$\Delta E_{torsion}$	$E_{LR,ele}$(LL)	$\Delta E_{L,vdW}$(LL)	$\Delta E_{R,vdW}$(RR)	E_{LR}
$\Delta E_{torsion}$	1.00				
$E_{LR,ele}$(LL)	−0.01	1.00			
$\Delta E_{L,vdW}$(LL)	0.31	−0.45	1.00		
$\Delta E_{R,vdW}$(RR)	−0.46	0.40	−0.85	1.00	
E_{LR}	−0.29	0.52	−0.46	0.64	1.00

It should be emphasized that both terms have positive coefficients in model B (Table 3), indicating that a positive energy change for both terms is activity- enhancing. It is not unreasonable that these energy changes are positive, as they are intramolecular terms, and may be related to negative entropy changes and also be balanced by favorable terms (negative), indirectly related to the intermolecular energy on the complex formation. It is also possible that these terms partially cancel each other, as both are positive and show a high inverse correlation coefficient value (R=−0.85) (Table 6).

The second pair of cross-correlated energy terms (R=0.64) correspond to E_{LR}, the intermolecular energy of the complex and $\Delta E_{R,vdW}$(RR), which has been previously described. This correlation shows that the intermolecular energy of the complex depends partially on the van der Waals energy change of the protein. It is also observed that both terms have positive coefficients in model B (Table 3). Considering the term, $\Delta E_{R,vdW}$(RR), a positive coefficient indicates that the increase of the van der Waals intramolecular energy of the protein (or the ligand) upon binding is activity-enhancing. This is explained by the fact that, at lower temperatures as is the case here, dispersion interactions such as van der Waals energy may not be favorable due to reduction in the vibrational energy.

However, the positive coefficient of the term, E_{LR}, seems unreasonable because it indicates that an increasing intermolecular energy of the complex is activity-enhancing. Presumably, this term is counterbalanced by the presence of other more energetically favorable terms in the model, such as the term $E_{LR,ele}$(LL) that corresponds to the intramolecular electrostatic energy of the bound ligand. Hence, the negative coefficient of the term, $E_{LR,ele}$(LL), of model B (Table 3) indicates that this term contributes to enhance the activity if it is negative (energetically favorable).

This counterbalancing of energy term descriptors is corroborated by the observation that the values of E_{LR} in the data base (data not labeled) tend to be more negative (favorable) for active compounds. It also reinforces the supposition that the electrostatic term has the greatest impact in balancing the intermolecular energy. Taking into consideration that there must be a good degree of charge distribution in the complex, it is not surprising that the term, $E_{LR,ele}$(LL), is one of the most correlated to activity (pIC_{50}), R=−0.55 (data not labeled), indicating that, if the values for this energy term increases, activity decreases, and vice versa. Thus, we may conclude, that this might balance sacrificed or gained energy in the ligand–protein binding process.

Additionally, the term $\Delta E_{L,vdW}$(LL), that is related to the variation in the intramolecular van der Waals energy of the ligand on binding, is observed to have the highest self-correlation with biological activity (R=0.64, data not labeled), which corroborates its importance in the analysis of the energy terms related to the biological activity for this series of compounds.

Finally, the last energy term, $\Delta E_{torsion}$, that corresponds to the change in the total torsional energy of the ligand and the protein upon binding indicates that if its values are positive, there is an increase in the activity. The intramolecular change in torsional energy correlates to conformational entropy (molecular flexibility) and so it seems reasonable that it increases upon ligand–protein complex formation, meaning that there is an entropic penalty for the "fitting" of the ligand into the protein cavity. It is interesting that this descriptor has been selected in the model obtained at 50 K, when we should expect this variation to be small as it has been collected at a low temperature.

Comparative analysis of the descriptors of models A and B

The analysis of the cross-correlation matrix of all the descriptors from models A and B showed some highly correlated pairs of descriptors (data not labeled), although the analysis of the residuals of the models result in a low correlation (R=0.23). However, the occurrence of high correlation between some of the energy descriptors from these models should not invalidate any of them, as the observed correlation occurs among descriptors of different models.

Among the most correlated descriptors, the terms, $\Delta E_{L,ele}$(LL), of model A and $E_{LR,ele}$(LL), of model B are most highly correlated (R=0.93). Both terms are related to the electrostatic energy of the ligand, showing equivalent coefficient values with the same sign (negative), indicating that these terms vary in the same proportion in both temperatures of evaluation (303 and 50 K). Thus, the electrostatic term seems not to be particularly dependent on simulation temperature.

The terms, $E_{L,bend}$(LL), of model A and $\Delta E_{L,vdW}$(LL), of model B are also highly correlated (R=0.84). The first term is related to the angular deformation energy (bending energy) of the unbound ligand and the second to the van der Waals energy change of the ligand upon complex formation. They have coefficients of the same sign (positive), leading to the same interpretation as given for the first pair of descriptors described earlier. In both cases, we can conclude that those energy terms, which are highly correlated, are probably behaving as substitutes for one

another in the equations. That is, these two descriptor terms are being interchanged in the model building process.

An interesting feature that was observed is that the term E_{LR} present in both models A (303 K) and B (50 K), which is related to the intermolecular energy between the ligand and the protein, does not show any significant cross-correlation (R=−0.30). This result suggests that the intermolecular energy between the ligand and the protein varies differently in both temperatures, and more, contributes differently to the activity in models A and B. Corroborating this idea, this term has a negative coefficient sign in model A, while in model B, the coefficient sign is positive (Table 3). Besides, it is possible that both equations show a different balance of the energy terms due to the different temperatures from which the descriptors have been determined.

Finally, these data show that both models, in spite of the fact of having been generated at two different simulation temperatures (303 and 50 K), can be representative of the FEFF-3D-QSAR methodology, as they have proven not to be mutually excluding but instead, complementary to one other.

Analysis of the internal and external predictivity of models A and B

The predictive power of models A and B were evaluated by using the pIC_{50} values for the compounds of the training and test set and the equations of models A and B. Figure 5 shows the plots of observed vs predicted pIC_{50} of models A and B obtained with the FEFF 3D-QSAR method for the p38-MAPK inhibitors [22]. The internal predictivity corresponds to the cross-validation procedure while the external predictivity corresponds to the calculated activity for the test set compounds, which were not part of the models' construction (the QSAR equations).

The residuals of fit and the standard deviation of the residuals (SDres) were computed, taking into consideration only the 28 compounds of the training set and, subsequently, the values of the five compounds of the test set were included, making up a total of 33 compounds (see Fig. 2 and Tables 1 and 7).

The analysis of Table 7 shows that model A has two outliers, compounds **1a** and **2b**, based on the SDres of the training set, while model B exhibits only one outlier, compound, **2p**, again using the SDres obtained from the training set. Inclusion of the compounds of the test set in the calculation of the SDres values leads to an increase in the SDres values and consequently, the number of outliers also increases (Table 7).

Model A has three outliers; the first two from the training set (**1a** and **2b**) and one new outlier, compound ***4a*** of the test set. Model B has only two outliers, compounds ***2i*** and ***3a*** of the test set. Compound **2p** (training set) does not behave as an outlier in this analysis.

Thus, based upon the analysis of the data obtained at the temperature of 303 K (model A) and at 50 K (model B), model B has been selected as the most representative of the FEFF-3D-QSAR methodology. Model B shows a smaller number of outliers than model A, considering the whole set of 33 compounds and also the outliers of model B belonging to the test set, that is, the outliers are part of the data used in constructing the model.

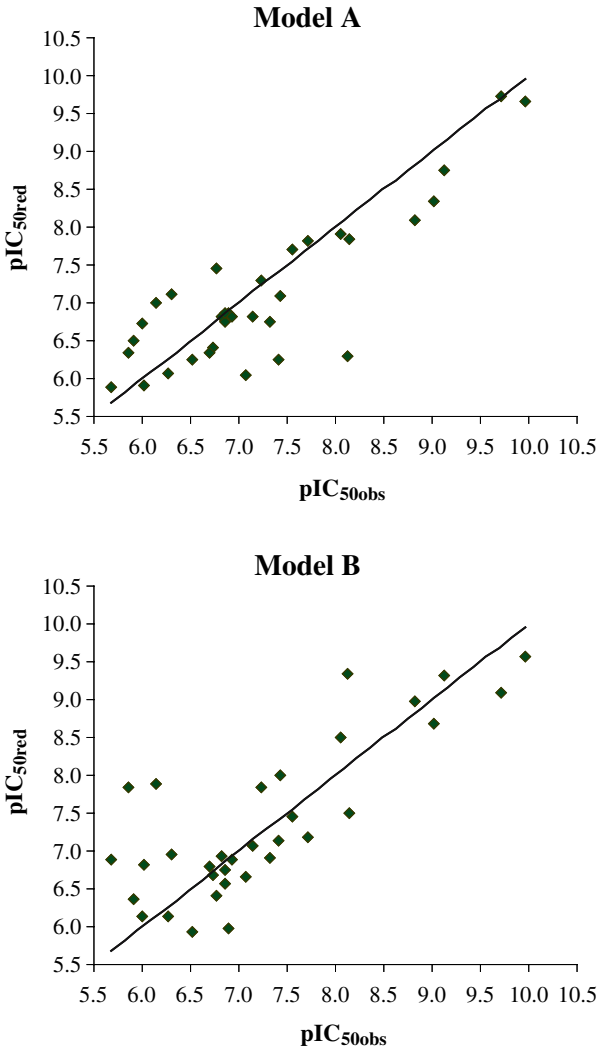


Fig. 5 Plots of observed vs predicted pIC_{50} of models A and B obtained with the FEFF 3D-QSAR method for the p38-MAPK inhibitors [22]

Table 7 $Q^2_{adjusted}$, standard deviation of the residuals (SD_{res}) and outliers observed in models A and B, with 5 terms, temperatures of 303 and 50 K, respectively, for the p38-MAPK inhibitors [22]

Model	$Q^2_{adjusted}$	SD_{res} [a]	Outliers[a]	SD'_{res} [b]	Outliers'[b,c]
A	0.72	0.44	**1a**, **2b**	0.57	**1a**, **2b**, ***4a***
B	0.65	0.46	**2p**	0.68	***2i***, ***3a***

[a] SD_{res} and outliers considering only the training set (28 compounds)
[b] SD'_{res} and outliers' including the test set (33 compounds)
[c] Test set compounds are in italics

Analysis of the outliers of model B

As can be seen in Table 7, when only the training set is considered, model B (50 K) has only one outlier, compound **2p** (the less active amongst the 33 compounds), whose activity has been overestimated ($pIC_{50pred=}$6.89 and $pIC_{50obs=}$5.68). As this is the only member of the training set which does not possess a phenyl ring in position 4 of the imidazole ring but has an alkyl (isopropyl) group as a substituent, it seems reasonable that this compound has a somewhat different structural profile from the other compounds of the training set, for instance, compound **2a** of series 2.

When the training and the test set (33 compounds) are jointly considered, model B has as outliers compounds ***2i*** and ***3a*** (Table 7). The structural analysis of compounds ***2i*** ($pIC_{50pred=}$7.85 and $pIC_{50obs=}$5.85), from the test set and **2k** ($pIC_{50pred=}$6.74 and $pIC_{50obs=}$6.85), from the training set, shows that these are the only members of the entire set of 33 compounds which possess a phenolic hydroxyl group in *para* and *meta* positions of the 4-phenyl ring, respectively. It is experimentally observed that ***2i*** is less potent, by one logarithmic unit, than **2k** (its regioisomer) and the reference compound, **2a** ($pIC_{50obs=}$5.85), a non-substituted member of this series, by one logarithmic unit.

In the crystal structure of **1a** bound to p38-MAPK (PDB code: 1A9U) [32], the amino group of residue Lys53 in the active site is oriented towards the nitrogen atom-3 of the imidazole ring of **1a**. This indicates a hydrogen bonding interaction between the $–NH_2$ group (donor) and atom N_3 of the imidazole ring (acceptor), or better, in the case of protonation of the amino group (Lys53) in the biological environment, there may be a hydrogen bonding strengthened by a positive charge $\left(-NH_3^+\right)$.

In the case of compound **2k**, the phenolic hydroxyl in position *meta* may reorient the Lys53 flexible chain in such a way that the hydroxyl group can act as the acid hydrogen donor to the amino group (Lys53), favoring an electrostatic interaction of the type "phenoxide–ammonium", even stronger than a classic hydrogen bond. Another possibility, in the case of protonation of the amino group (Lys53), is that there may be a hydrogen bond strengthened by a positive charge in which the hydroxy-phenyl group acts as a hydrogen bond acceptor. To counterbalance this new interaction, the reorientation of these groups in the active site might weaken, or even eliminate other interactions, reducing the intermolecular binding energy in a competitive fashion, which is corroborated by the fact that **2k** is less potent than **1a**, but somewhat more active than **2a**.

Compound ***2i*** has the hydroxyl group in the *para* position so that the same kind of reorientation in the active site is not likely to occur as Lys53 is more distant from the hydroxyl group. On the other hand, this compound may interact with other residues closer to the protein surface before entering the active site channel. This hypothesis could not be verified as MDS was not carried out to simulate the process of initial entrance of the ligand into the active site. This was only carried out in the case where the ligand was effectively bound in the active site. It has been assumed that the ligand was already bound in the active site, by means of docking procedures, before performing the MDS step. For this reason, additional ligand–receptor interactions are seen when compared to the unsubstituted compound, **2a**.

In fact, the visual inspection of Fig. 6, which is representative of the lowest energy conformations selected from the MDS of the ligand–protein complex for compounds ***2i*** and **2k** at 50 K, shows that both compounds keep the original orientation of **2a**. However, ***2i*** makes extra hydrogen bonds with the oxygen atom of the carbonyl group of the peptide backbone of Ala51 and with the carboxylate group of Asp168. This observation supports the suggestion that the overestimation of its potency is due to the formation of additional hydrogen bonds, which, interestingly, are not observed for the compound **2k** in the MDS.

Finally, for compound ***3a***, whose predicted potency is overestimated ($pIC_{50pred=}$7.88 and $pIC_{50obs=}$6.15), we could not rationalize a comprehensive explanation for its outlier behavior. It was experimentally observed that the compounds of series 3 show the following decreasing order of potency: **3c**, **3d**, **3b,** and ***3a***. This order probably corresponds to an additional interaction site in relation to ***3a***, the less active of the series, which is an analog of **2a**, except that a pyrimidine ring in ***3a*** replaces the pyridine ring of **2a**. Compound **2a** ($pIC_{50obs=}$6.74) is somewhat more potent than ***3a*** ($pIC_{50obs=}$6.15) in the biological assays, probably because the nitrogen atom of the pyridine moiety is more basic than the corresponding nitrogen atom in the pyrimidine ring of ***3a***, due to a second electronegative atom in the pyrimidine ring being a better hydrogen bonding acceptor.

The outlier behavior of ***3a*** could be rationalized using the argument that it is the only compound in series 3 that does not have substituents on the pyrimidine ring. Therefore, as it was not included in the training set, the model has not been "trained" for its type of structure. Nonetheless, this argument does not seem too strong, as this compound is well represented by **2a**, its pyridine analog, which is not an outlier.

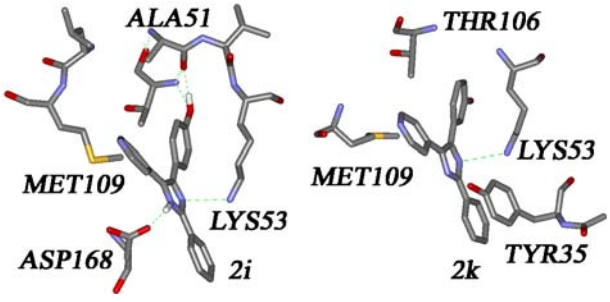


Fig. 6 Lowest energy conformations selected from the MDS of the ligand–protein complexes for compounds ***2i*** and **2k**, retrieved from the simulation at 50 K

Comparison between RD-FEFF-3D-QSAR and RI-4D-QSAR

One of the drawbacks of FEFF-3D-QSAR methodology is that the models are composed completely of thermodynamic or force-field terms as descriptors [8, 39]. Such models do not furnish any visual 3D pharmacophore information of the ligand–protein binding event. Thus, we decided to compare the results from the RD-FEFF-3D-QSAR (this work) and the RI-4D-QSAR (previous work) [40] methods. Therefore, the binding information in the model obtained in this work with the FEFF methodology (model B, Table 3) was compared with the grid cell occupancy descriptors (GCODs) of the best models obtained with the 4D-QSAR study (models B and D, Fig. 7 and Table 8) developed recently by our group using the same training and test sets [40].

In a strict sense, it is not correct to make a direct correlation between GCODs (4D-QSAR) and energy terms (FEFF) descriptors, because GCODs are localized molecular descriptors and the energy terms are global descriptors (thermodynamic) [41, 42]. However, we can make a qualitative correlation between the most significant descriptors from the best models obtained by the RI-4D-QSAR method and the energy terms obtained from the RD-FEFF-3D-QSAR analysis. These descriptors and energy terms are the nonpolar (np) GCODs descriptors and the van der Waals energy terms, the polar positive (p+) GCODs descriptors and the electrostatic energy descriptors and the any atom type GCODs descriptors and both of the mentioned energy terms.

In the best models obtained by RI-4D-QSAR methods, GCODs (−1,0,−1) and those (1,1,4) in model B, and GCODs (1,0,−1) and those (0,0,4) in model D, are among the most significant descriptors (Fig. 7, Table 8). In forming a comparison/correlation with the best RD-FEFF-3D-QSAR model, these nonpolar GCODs may be correlated to the descriptors, $\Delta E_{\mathrm{R,vdW}}$(RR) and $\Delta E_{\mathrm{L,vdW}}$(LL), that represent the van der Waals intramolecular energy change of the protein between the bound and the unbound state, and the same energy change for the ligand, respectively.

Also, it is reasonable to assume that the term $\Delta E_{\mathrm{torsion}}$ that corresponds to the change in the total torsional energy of the ligand and the protein upon binding may also be related to the nonpolar GCODs mentioned above, and this energy difference can be influenced by the fit of the ligands in the protein pocket. The difference will be largely dependent on the surface fit of nonpolar groups to the surrounding amino acid residues lining the active site and which can be influenced by conformational freedom of both ligand and receptor.

As there are descriptors (GCODs and energy terms) with positive and negative coefficients, this seems to correlate well with our assumption that these terms are balancing each other, and also reinforces the importance of hydrophobic interactions for ligand binding to p38-MAPK.

Additionally, the term $E_{\mathrm{LR,ele}}$(LL), which corresponds to the intramolecular electrostatic energy of the bound ligand, may correlate to GCOD(1,−2,5)(p+), of the polar-positive IPE atom type. Both descriptor terms represent electrostatic interactions, possibly related to a conformational sponsored change in ligand charge distribution after the binding to the protein binding pocket.

As to the predictivity of the two models, the Q^2 values obtained for the RI 4D-QSAR models (Table 8) are better than those of the FEFF-3D-QSAR models (Table 3). This can be visualized better in Fig. 8, which shows the plots of observed vs predicted pIC_{50} values obtained with the best model from the 4D-QSAR (model D) and FEFF 3D-QSAR methods (model B). However, model D, the best model of 4D-QSAR methodology, has two molecules from the training set as outliers (**2f** and **2n**) when only the training set (data not shown) is considered, and none from the test set. Conversely, the best model of the FEFF-3D-QSAR method has only one outlier when we consider the training set only (**2p**, Table 7) but two outliers (***2i*** and ***3a***) when the test set is also included. Thus, it is reasonable to conclude that even though the 4D-QSAR method may have better predictivity than FEFF-3D-QSAR approach in this particular study, the two methods are complementary to one another as they provide insights on the nature of the ligand–protein binding mechanism.

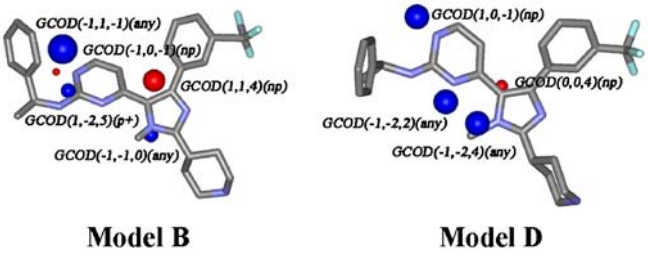


Fig. 7 Best models (B and D), obtained by the RI-4D-QSAR method using compound **4e** as a reference. *Blue spheres* indicate activity-enhancing pharmacophore sites, and red spheres indicate activity-decreasing pharmacophore sites [40]

Discussion and conclusions

The analysis of the models obtained by the receptor-dependent FEFF-3D-QSAR methodology showed that 50 K is the best simulation temperature for the construction of models using the compounds in this study. The best QSAR equation, model B, is shown to be predictive and to contain energy descriptors that represent the electrostatic and van der Waals contribution to the ligand–receptor interaction and ligand and receptor intramolecular energy terms that reinforce the importance of conformational changes, which include changes in entropy during the binding process.

As one of the reasons for performing the MDS at low temperatures was to capture solvent effects in the binding process, it is not too surprising that none of the solvation terms included in this work are present in the optimized models. Moreover, the intra- and intermolecular electrostatic and van der Waals energy terms may indirectly simulate solvent effects.

Table 8 Best models obtained with RI-4D-QSAR method for the training set of 28 inhibitors of p38-MAPK [22, 40]

Model	Equation
B	pIC_{50}=6.78+28.00(−1,1,−1)(any)+17.34(1,-2,5)(p+)−3.53(1,1,4)(np)−2.14(−1,0,−1)(np)+1.12(−1,−1,0)(any); N=28; R^2=0.89; SE=0.14; Q^2=0.85; F=0.43; LOF=0.25
D	pIC_{50}=6.67+3.30(1,0,−1)(np)+8.29(−1,−2,2)(any)−1.26(0,0,4)(np)+1.48(−1,−2,4)(any) N=28; R^2=0.88; SE=0.14; Q^2=0.86; F=0.61; LOF=0.24

N Number of compounds in the training data set, R^2 squared correlation coefficient, *SE* standard error, Q^2 leave-one-out cross-validated R^2, *F* Fischer's test of statistical significance, *LOF* Friedman's lack-of-fit score

We analyzed the individual correlations between the terms of model B and the solvation terms included in this work. A significant correlation (R=0.52) is seen between the term, ΔE_R(RM), which defines the solvation energy change of the receptor upon complex formation and the term $\Delta E_{L,vdW}$(LL), which defines the intramolecular van der Waals energy change of the ligand upon complexation. The correlation of these energy terms seemingly reflects the change in the protein solvent accessible surface upon complex formation, which depends directly on the surface of the ligand.

Moreover, the correlation of these FEFF-3D-QSAR terms with activity reveals a greater positive correlation (R=0.64) for the term, $\Delta E_{L,vdW}$(LL), in comparison to the term, ΔE_R(RM) (R=0.45), supporting the selection of the $\Delta E_{L,vdW}$(LL) term in the model optimization by the genetic algorithm approach.

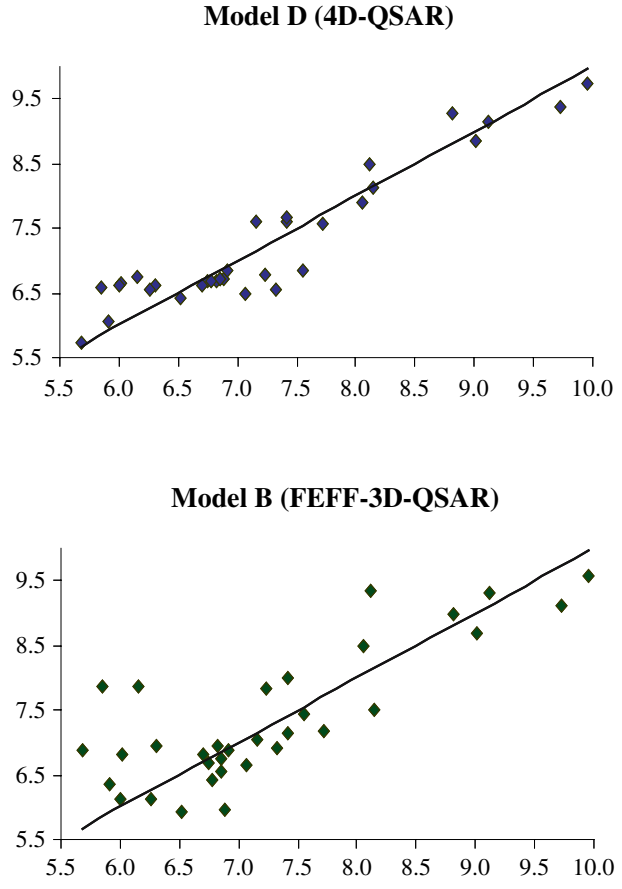


Fig. 8 Plot of observed vs predicted pIC_{50} values obtained with the best models from the 4D-QSAR (model D) and FEFF 3D-QSAR methods (model B)

The outlier behavior of compound **2i** in model B emphasizes the inherent limitation of 3D-QSAR methodologies, which assume that ligands in a congeneric series bind to the active site of the protein in a similar binding mode. The step that corresponds to the ligand reaching the binding site is left out of the QSAR analysis. Nonetheless, the FEFF-3D-QSAR methodology was able to distinguish this compound, among the others, as an outlier, as, in fact, it is one of the less potent compounds in the series.

In conclusion, the FEFF-3D-QSAR models generated here are shown to incorporate physicochemical features that are thought to be involved in ligand–receptor interactions. Also, it seems that the balance or complementarity of energy terms in the ligand and protein alone, or in complex, are required and well represented in the optimized models. One interesting feature observed is the absence of solvation terms in the best QSAR equations. However, the presence of electrostatic terms linked to conformational variations in the active site, appear to capture the type of "packing" that happens on a solvated molecular system. Thus, we may conclude that reduction of the kinetic energy of the system (by lowering the temperature of sampling) is a reasonable tool to explore the nature of solvent-dependent interactions between ligand and protein.

Additionally, we have noticed the influence of simulation temperature in the representation of van der Waals interactions. This behavior should be analyzed very carefully given the accepted understanding of ligand–receptor intermolecular forces upon binding, when designing new p38-MAPK inhibitors. The best FEFF-3D-QSAR model obtained at the temperature of 50 K has been shown to be predictive and, therefore, may be a good tool to employ in structure-based design of new p38-MAPK inhibitors.

Finally, the binding information in the model obtained in this work with the RD FEFF-3D-QSAR methodology was compared with the GCODs of the best models obtained with a previous RI-4D-QSAR study, developed recently by our group using the same training and test sets [40]. We concluded that the best models obtained with these two methodologies comprise classic nonpolar and electrostatic contributions, and that each of the two types of models complement one another, providing insight as to the nature of the ligand–receptor binding mechanism.

Acknowledgements We thank the Brazilian agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) for their support. NCR is also grateful to the CNPq for fellowship support (Nr. 200650/2000-9) for financing her "sandwich" doctorate at the College of Pharmacy of the University of Illinois at Chicago to perform the current work and a previous one, named "Construction of 4D-QSAR Models for Use in the Design of Novel p38-MAPK Inhibitors" [40]. Resources of the Laboratory of Molecular Modeling and Design (UIC and UNM) were used to perform these studies. Support from The Chem21 Group is gratefully acknowledged. Finally, NCR would also thank Carlota Parreira for manuscript revision and her valuable suggestions.

References

1. Waszkowycz B (2002) Curr Opin Drug Discov Dev 5:407–413
2. Chen IJ, Neamati N, MacKerell AD Jr (2002) Curr Drug Targets Infect Disord 2:217–234
3. Lazoura E, Campbell W, Yamaguchi Y, Kato K, Okada N, Okada H (2002) Chem Biol 9:1129–1139
4. Williams DH, Mitchell T (2002) Curr Opin Pharmacol 2:567–573
5. Mao C, Sudbeck EA, Venkatachalam TK, Uckun FM (2000) Biochem Pharmacol 60:1251–1265
6. Ortiz AR, Pisabarro MT, Gago F, Wade RC (1995) J Med Chem 38:2681–2691
7. Holloway MK, Wai JM, Halgren TA, Fitzgerald PM, Vacca JP, Dorsey BD, Levin RB, Thompson WJ, Chen LJ, de Solms SJ (1995) J Med Chem 38:305–317
8. Tokarski JS, Hopfinger AJ (1997) J Chem Inf Comput Sci 37:792–811
9. Santos-Filho OA, Mishra RK, Hopfinger AJ (2001) J Comput Aided Mol Des 15:787–810
10. Venkatarangan P, Hopfinger AJ (1999) J Med Chem 42:2169–2179
11. Rogers D, Hopfinger AJ (1989) J Chem Inf Comput Sci 34:854–866
12. Martín-Blanco E (2000) Bioessays 22:637–645
13. Kyriakis JM, Avruch JJ (1996) Bioessays 18:567–577
14. Raingeaud J, Gupta S, Rogers JS, Dickens M, Han J, Ulevitch RJ, Davis RJ (1995) J Biol Chem 270:7420–7426
15. Lee JC, Kumar S, Griswold DE, Underwood DC, Votta BJ, Adams JL (2000) Immunopharmacol 47:185–201
16. Hale KK, Trollinger D, Rihanek M, Manthey CL (1999) J Immunol 162:4246–4252
17. Han J, Lee JD, Jiang Y, Li Z, Feng L, Ulevitch RJ (1994) Science 265:808–811
18. Boehm JC, Bower MJ, Gallagher TF, Kassis S, Johnson SR, Adams JL (2001) Bioorg Med Chem Lett 11:1123–1126
19. Lee JC, Kumar S, Griswold DE, Underwood DC, Votta BJ, Adams JL (2000) Immunopharmacol 47:185–201
20. Gallagher TF, Seibel GL, Kassis S, Laydon JT, Blumenthal MJ, Lee D, Boehm JC, Thompson-Fier SM, Abt JW, Sorenson ME, Smietana JM, Hall RF, Garigipati RS, Bender PE, Erhard KF, Krog AJ, Hoffman GA, Sheldrake PL, McDonnell PC, Kumar KF, Young PR, Adams JA (1997) Bioorg Med Chem 5:49–64
21. Badger AM, Bradbeer JN, Votta B, Lee JC, Adama JL, Griswold DE (1996) J Pharmacol Exp Ther 279:1453–1461
22. Liverton NJ, Butcher JW, Claiborne CF, Claremon DA, Libby BE, Nguyen KT, Pitzenberger SM, Selnick HG, Smith GR, Tebben A, Vacca JP, Varga SL, Agarwal L, Dancheck K, Forsyth AJ, Fletcher DS, Frantz B, Hanlon WA, Harper CF, Hofsess SJ, Kostura M, Lin J, Luell S, O'Neill EA, Orevillo CJ, Pang M, Parsons J, Rolando A, Sahly Y, Visco DM, O'Keefe SJ (1999) J Med Chem 42:2180–2190
23. Adams JL, Boehm JC, Kassis S, Gorycki PD, Hall R, Sorenson M, Lee JC, Ayrton A, Griswold DE, Gallagher TF (1998) Bioorg Med Chem Lett 8:3111–3116
24. Henry JR, Rupert KC, Dodd JH, Turchi IJ, Wadsworth SA, Cavender DE, Schafer PH, Siekierka JJ (1998) Bioorg Med Chem Lett 8:3335–3340
25. Lee JC, Kassis S, Kumar S, Badger A, Adams JL (1999) Pharmacol Ther 82:389–397
26. Boehm JC, Smietana JM, Sorenson ME, Garigipati RS, Gallagher TF, Sheldrake PL, Bradbeer J, Badger AM, Laydon JT, Lee JC, Hillegass LM, Griswold D, Breton JJ, Chabot-Fletcher MC, Adams JL (1996) J Med Chem 39:3929–3937
27. Redman AM, Johnson JS, Dally R, Swartz S, Wild H, Paulsen H, Caringal Y, Gunn D, Renick J, Osterhout M, Kingery-Wood J, Smith RA, Lee W, Dumas J, Wilhelm SM, Housley TJ, Bhargava A, Ranges GE, Shrikhande A, Young D, Bombara M, Scott WJ (2001) Bioorg Med Chem Lett 11:9–12
28. Koehler MG, Hopfinger AJ (1989) Polymer 30:116–126
29. Tong L, Pav S, White DA, Rogers S, Crane KM, Cywin CL, Brown ML, Pargellis CA (1997) Nat Struct Biol 4:311–316
30. Wang ZL, Canagarajah BJ, Boehm JC, Kassisa S, Cobb MH, Young PR, Abdel-Meguid S, Adams JL, Goldsmith EJ (1998) Structure 6:1117–1128
31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat, TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235
32. Insight II User Guide, Version 97.0 (1997) MSI, San Diego
33. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902–3909
34. Hyperchem Program Release 5.01 for Windows (1996) Hypercube I
35. Doherty DC (1997) MOLSIM User's Guide. The Chem21 Group, 1780 Wilson Dr, Lake forest, IL 60045
36. Weiner SJ, Kollman PA, Nguyen DTJ, Case DA (1986) Comput Chem 7:230–252
37. Rogers D (1994) Wolf Reference Manual, Version 5.5. Molecular Simulations
38. Livingstone D (1995) In Data analysis for chemists: applications to QSAR and chemical product design. Oxford University Press, New York
39. Pan D, Tseng Y, Hopfinger AJ (2003) J Chem Inf Comput Sci 43:1591–1607
40. Romeiro NC, Albuquerque MG, Alencastro RB, Hopfinger AJ (2005) J Comput Aided Mol Des 19:385–400
41. Robinson DD, Lyne PD, Richards WG (1999) J Chem Inf Comput Sci 39:594–600
42. Verli H, Albuquerque MG, Alencastro RB, Barreiro EJ (2002) Eur J Med Chem 37:219–229