Combining structure- and ligand-based approaches for studies of interactions between different conformations of the hERG K+ channel pore and known ligands

Alessio Coi [a], Anna Maria Bianucci [b],*

[a] INSTM (Consorzio National Interuniversity Consortium of Materials Science and Technology), Via Giusti 9, 50121 Firenze, Italy
[b] Dipartimento di Farmacia, Università di Pisa, Via Bonanno 6, 56126 Pisa, Italy

ARTICLE INFO

ABSTRACT

Drug-induced insurgence of cardiotoxic effects signaled by the prolongation of the QT interval in the electrocardiogram, has the potential to evolve into a characteristic arrhythmic event named Torsade de Pointes (TdP). Although several different mechanisms can theoretically lead to prolonged QT interval, most of drugs showing this side effect, prolong the cardiac repolarization time through the inhibition of the rapid component of the delayed repolarizing current (I_{Kr}) which in humans is carried by a K+ channel protein encoded by hERG. In this study, four 3D-models, representing different conformational states of hERG K+ channel, were built by a homology-based technique. A dataset of 59 compounds was collected from the literature and rationally selected according to the availability of IC_{50} values derived from whole-cell patch clamp performed at 37 °C on HEK cells.

Molecular docking was carried out on each one of the four conformations of the channel, hundreds of docking-based molecular descriptors were obtained and used, together with other 2D and 3D molecular descriptors, to develop QSAR models. The statistical parameters describing the accordance between predicted and experimental data and the interpretation of the QSAR models enabled us to assess the reliability of the four 3D-models of the channel pore, thus allowing to look in more depth at binding modes and key features of the interactions occurring between the hERG K+ channel and ligands endowed of blocking activity.

© 2013 Published by Elsevier Inc.

1. Introduction

The recent years have been characterized by the withdrawal from the market of several drugs (e.g. cisapride, terfenadine, astemizole, grepafloxacine, sertindole) belonging to diverse pharmacological classes because their use in patients has been associated with unexpected occurrence of cardiotoxic effects and, in particular, early cardiac repolarization. This effects is signaled by the prolongation of the QT interval on electrocardiogram which has the potential to evolve into a characteristic arrhythmic event named Torsade de Pointes (TdP). TdP is a potentially life-threatening polymorphic ventricular arrhythmia characterized by QRS complex twisting around the isoelectric line of the ECG. This type of arrhythmia is usually self-limiting, but in 25% of the cases it may degenerate into irreversible ventricular fibrillation culminating in sudden cardiac death.

Although several mechanisms can theoretically lead to prolonged QT interval, most of the drugs used in clinics have been recognized to prolong the cardiac repolarization time through the inhibition of the rapid component of the delayed repolarizing current (I_{Kr}), which in humans is carried by a repolarization K+ channel protein encoded by hERG (human Ether-ago-go-Related Gene) [1,2].

Most drugs associated with Long QT Syndrome (LQTS) and/or TdP have been retrospectively demonstrated to interact with the α-subunit of the I_{Kr} channel and hereby reducing the amplitude of the K+ current.

Since hERG K+ channel inhibition is the main mechanism correlated to LQTS and TdP, the potency of a drug to block hERG current is generally expressed as the concentration of the drug that reduces the hERG tail current by 50% (IC_{50}) [3].

The hERG K+ channel belongs to the K_V channel family and has an architecture showing a tetrameric structure. Each subunit contains six transmembrane helices (S1–S6). S1–S4 helices comprise the voltage-sensor domain, while the ion permeation pore is constituted by both the inner cavity, made by S5–S6 helices, and the selectivity filter (SF), in the loop between S5 and S6.

* Corresponding author. Tel.: +39 050 2219564.
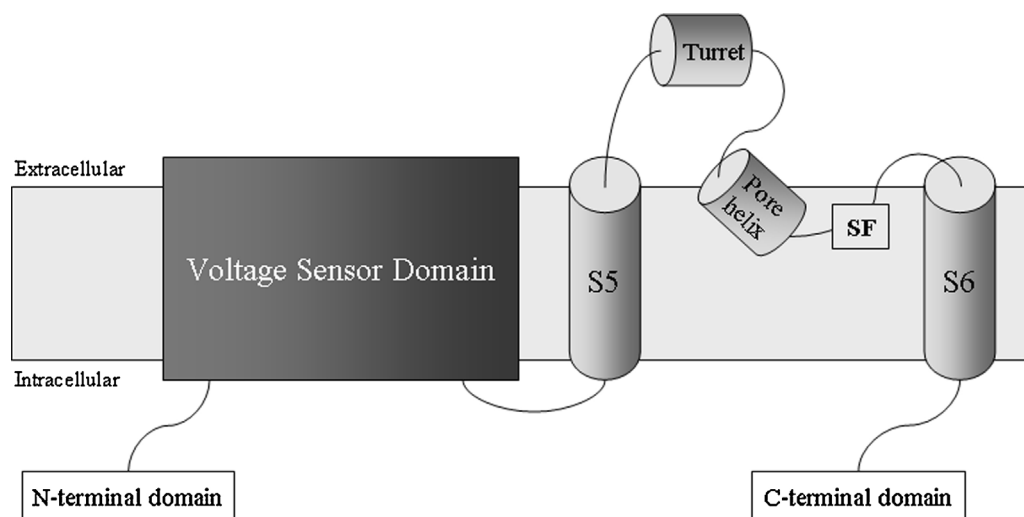E-mail address: anna.bianucci@farm.unipi.it (A.M. Bianucci).

Fig. 1. Schematic representation of a hERG subunit. The voltage-sensor domain is constituted by S1–S4 helices; the ion permeation pore is constituted by helices S5–S6 and the selectivity filter (SF). The turret (about 40 residues long) links the helix S5 to the pore helix.

A long chain of about 40 residues, linking the S5 helix to the SF, the *turret* (or S5-P linker), represents a fragment characteristic of hERG (Fig. 1). Mutagenesis studies enabled identifying some residues as the ones involved in drug binding: Thr623, Ser624, Val625 [2,4–7], Gly648 [2,4–6], Val659 [2,4–6], Tyr652 [2,4–10], Phe656 [2,4–10].

Variations in the membrane potential induce the channel pore to take the proper open/closed state. The pore must be in its active *open* state in order to allow the ligands to bind it, since the inner cavity of the pore is not accessible in the *closed* state. Looking at the X-ray-determined structures which mostly resemble the *closed* and *open* states of K+ channels, it can be hypothesized that a conformational change involving the C-terminal moiety of the S6 helices leads to the pore widening with consequent opening of the ion channel.

In the literature, several 3D-models of hERG K+ channel, built thanks to its homology with other K+ channels of known crystal structures, are described. The structure of KcsA from *Streptomyces lividans*, which takes a conformation similar to the *closed* form of the hERG K+ channel, was obtained first [11] and widely used as a template for building several 3D models of the hERG K+ channel by homology based techniques [2,6,9,12,13].

In other attempts, several models were built using, as template, the structure of MthK (from *Methanothermobacter thermautotrophicus*), which takes a conformation similar to the *open* form of the channel [8,12,14,15]. The bacterial voltage-dependent K+ channel, KvAP from *Aeropyrum pernix* [16], was also used as the structural template for homology-based modeling of the *open* state of hERG [17,18].

In this work four 3D-models corresponding to different conformations of hERG K+ channel pore (*open*, *closed*, *rotated* and *intermediate* states) were built with the aim of analyzing their ability to simulate the ligand-channel interactions. A suggestion about the most plausible 3D theoretical model of the channel explaining the interaction with ligands is given. An indirect validation of such a 3D model was made with the aid of QSAR models based on 1D-, 2D-, and 3D-molecular descriptors. The QSAR models can be used in turn for predicting potential inhibition properties of molecules, independently from structure-based studies.

3D-type molecular descriptors were computed after molecular docking where a dataset of 59 compounds were screened against each one of the four states of the pore (workflow in Fig. 2).

Docking-based molecular descriptors were extrapolated from the analysis of the most plausible poses of the ligands in the binding site and additional 1D-, 2D- and 3D-descriptors were calculated. This complex protocol enabled us to take into account different conformations of the channel and to identify the state which better fits the experimental data, so allowing us to look in more depth at the ligand-channel interactions.
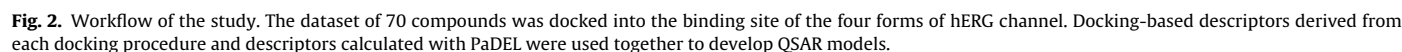
2. Methods

2.1. Homology modeling

PSIPRED and MEMSAT (at the PSIPRED server) were used to predict the secondary structure and the transmembrane helix topology, respectively [19–21].

The hERG aminoacidic sequence was obtained from SwissProt [22] (accession number Q12809). PSI-Blast [23] was used to scan a part of the hERG sequence, corresponding to the region that extends from the beginning of the pore helix to the C-terminal of helix S6, in order to identify 3D-structures available from Protein Data Bank (PDB) [24] and homologous to hERG.

Among the structures selected by PSI-Blast, MthK (PDB ID: 1LNQ) [14] and KcsA (PDB ID: 1BL8) [11] were chosen to build the *open* and *closed* forms of the hERG K+ channel, respectively. The sequences of the template structures were then aligned to the sequence of hERG, using ClustalW (Fig. 3) [25].

Modeller 9v8 [26] was used to build the homology models of the *open* and *closed* forms of the hERG K+ pore, based on the alignment reported in Fig. 2. Since mutagenesis experiments showed that the binding site of most drugs is located inside the pore cavity [2], only the channel portion including S5 and S6 (which are the helices surrounding the pore) was built.

Ten 3D-structures for both the *open* and *closed* form were built based on the secondary structure prediction previously performed by using PSIPRED. In each case, the model for the pore helix was extended by four residues at its N-terminal end (see Fig. 3) by applying α-helical restraints. The hERG *turret* was omitted (according to other Authors) and the proper 3D arrangement of the remaining moiety to be modeled was ensured by inserting a bond between Glu575 and Pro605 (Fig. 2). Loops optimization was performed with Modeller. In each case, the homology model with the best discrete optimized protein energy (DOPE) score [27] was selected.

Fig. 2. Workflow of the study. The dataset of 70 compounds was docked into the binding site of the four forms of hERG channel. Docking-based descriptors derived from each docking procedure and descriptors calculated with PaDEL were used together to develop QSAR models.

The *open* and *closed* forms of the channel pore so obtained were refined by using validation and structure-improvement tools at the Molprobity web service and related utilities [28]. Molprobity evaluates each side chain conformation against an updated rotamer library and smoothed multidimensional χ-angle database distributions, each φ, ψ main chain torsion pair against updated Ramachandran-plot distributions, and the deviation of each Cβ atom vs. the ideal position calculated from local backbone. All-atom contact criteria are then addressed in Molprobity by first using Reduce [29] to add and optimize all hydrogen atoms in their local H-bond networks, with automatic resolution of ambiguous Asn/Gln/His orientations, producing a modified and commented PDB-format coordinate file.

Looking at the structure of the closed form it could be observed that it does not show a proper orientation of side chains of key residues Tyr652 and Phe656. In order to avoid this problem, a rotation of the S6 helix was investigated starting from the *closed* form of the channel, from which a further structure (here referred as the *rotated* form) of the hERG K$^+$ channel was built, according to a protocol similar to the one described by Stansfeld et al. [30]. An initial *rotated* model was so obtained with Modeller by "shifting" the sequence of the S6 helix one residue toward the C-terminus. After that, the *closed* and the initial *rotated* models were processed at Morph server [31] in order to produce a series of 30 intermediate structures and the one showing the best orientation of the four Tyr652 and Phe656 residues within the inner cavity of the pore, was chosen.

The opening of the hERG K$^+$ channel was simulated with the normal mode analysis using the elNémo server [32]. This approach was applied to the *rotated* model obtained in the previous step. The server produced five structures describing the transition from the closed to the open state of the channel. The transition form, showing key-residues (i.e. the ones known from mutagenesis studies to be implicated in drug binding) arranged in an optimal orientation within the inner cavity, was selected and conventionally called the *intermediate* state of the channel. The final conformations chosen for both the *rotated* and the *intermediate* state of the channel were refined with Molprobity.

All four models resembling the *open*, *closed*, *rotated* and *intermediate* forms of hERG channel were subjected to energy minimization with the UCSF Chimera package v1.6 [33] and the optimized structures were checked again with Molprobity prior to molecular docking. (For Molprobity outputs see Supporting Material.)

2.2. Dataset

A dataset of 59 compounds was selected from literature according to the availability of IC$_{50}$ values derived from electrophysiological measurements (whole-cell patch clamp), performed at 37 °C on HEK cells (Table 1).

The selection of experimental data has a great importance in determining the performances of QSAR model when accurate values of the target property (in our case IC$_{50}$ values) have to be predicted. The possibility of using values of the biological indicator
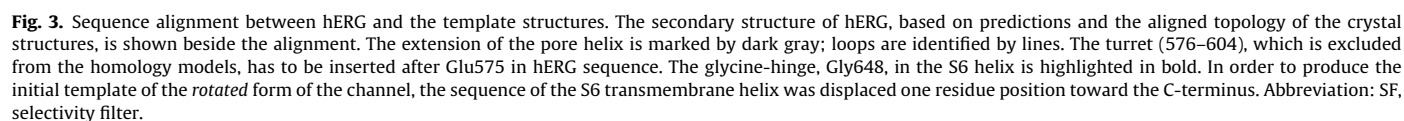
Fig. 3. Sequence alignment between hERG and the template structures. The secondary structure of hERG, based on predictions and the aligned topology of the crystal structures, is shown beside the alignment. The extension of the pore helix is marked by dark gray; loops are identified by lines. The turret (576–604), which is excluded from the homology models, has to be inserted after Glu575 in hERG sequence. The glycine-hinge, Gly648, in the S6 helix is highlighted in bold. In order to produce the initial template of the *rotated* form of the channel, the sequence of the S6 transmembrane helix was displaced one residue position toward the C-terminus. Abbreviation: SF, selectivity filter.

Table 1
Dataset of compounds collected from literature and selected according to the availability of IC_{50} values (expressed in μM) derived from electrophysiological measurements (whole-cell patch clamp techniques) performed at 37 °C on HEK cells. Abbreviations: Ref., references; K^+, potassium concentration in bath solution; t, duration of the depolarization; measurement voltage: membrane potential during repolarization step used for current measurement.

Compound	IC_{50} (μM)	Ref.	K^+ (mM)	t (s)	Holding potential (mV)	Depolarization potential (mV)	Measurement voltage (mV)
2-Hydroxymehtyl-olanzapine	11.6	[34]	4	1	−80	20	−50
4-Aminopyridine	4400	[35]	4	1	−40	30	−40
Alosetron	3.2	[36]	4	0.4	−75	10	−40
Amiodarone	0.0259	[37]	4	0.5	−75	10	−40
Astemizole	0.022	[38]	4	0.5	−75	10	−40
b-Estradiol	1.3	[39]	4	4	−80	50	−50
BMCL-202220-10d	85	[40]	4	1	−80	20	−40
BMCL-202220-10e	55	[40]	4	1	−80	20	−40
BMCL-202220-10f	59	[40]	4	1	−80	20	−40
BMCL-202220-10g	18	[40]	4	1	−80	20	−40
Cisapride	0.0067	[36]	4	0.4	−75	10	−40
Clemastine	0.012	[41]	4	2	−80	20	−40
Clomiphene	0.18	[42]	4	2	−40	30	−40
Clomipramine	0.13	[43]	5.4	4.8	−80	20	−60
Clozapine	0.191	[38]	4	0.5	−75	10	−40
Clozapine-N-oxide	133.3	[38]	4	0.5	−75	10	−40
Desipramine	1.39	[36]	4	0.4	−75	10	−40
Desmethylastemizole	0.001	[44]	4	4	−80	0	−50
Desmethylolanzapine	14.2	[36]	4	0.4	−75	10	−40
Diphenhydramine	2.6	[45]		2	−80	20	−50
Dofetilide	0.0092	[46]	4	1	−80	20	−80
Doxepin	6.5	[47]	4	2	−80	20	−40
Dronedarone	0.0591	[48]	4	1	−40	30	−40
E-4031	0.0181	[36]	4	0.4	−75	10	−40
Fluvoxamine	3.8	[49]	4	2	−80	20	−40
Haloperidol	0.0281	[38]	4	0.4	−75	10	−40
Imipramine	0.616	[37]	4	0.5	−75	10	−40
JMC-132931-03	0.31	[46]	4	1	−80	20	−80
JMC-547030-47	7.9	[34]	4	1	−80	20	−50
Ketanserin	0.121	[45]		2	−80	20	−50
Ketoconazole	1.9	[36]	4	0.4	−75	10	−40
Lapatinib	0.8	[50]	5.4	2	−80	20	−40
Lidocaine	262.9	[51]	4	0.5	−40	30	−40
Lidoflazine	0.0166	[52]	4	2	−80	20	−40
Loratadine	0.173	[38]	4	0.4	−75	10	−40
Maprotiline	0.13	[53]	5.4	4.8	−80	20	−60
(R)-Methadone	7	[54]	5	2.5	−80	40	−50
(S)-Methadone	2	[54]	5	2.5	−80	40	−50
Mizolastine	0.441	[38]	4	0.4	−75	10	−40
Moxifloxacin	74.7	[55]	4	0.5	−75	10	−40
Norastemizole	0.0277	[44]	4	4	−80	0	−50
Norclozapine	4.49	[36]	4	0.4	−75	10	−40
Olanzapine	0.181	[38]	4	0.4	−75	10	−40
Papaverine	0.58	[56]	5.4	4.9	−80	20	−60
Pentamidine	252	[57]	4	1	−80	20	−80
Pilsicainide	20.4	[58]	4	4	−60	50	−50
Pimozide	0.0497	[38]	4	0.4	−75	10	−40
Procainamide	7.45	[37]	4	0.5	−75	10	−40
Propiverine	10	[59]	4	5	−80	0	−50
Protriptyline	1.18	[60]	5.4	4.8	−80	20	−60
Quinidine	0.86	[45]		0.5	−80	20	−50
Risperidone	0.148	[36]	4	0.4	−75	10	−40
Sertindole	0.126	[38]	4	0.4	−75	10	−40
Sildenafil	33.3	[61]	4	0.5	−75	10	−40
Tamoxifen	0.198	[37]	4	0.5	−75	10	−40
Tolterodine	0.0096	[62]	5	3	−80	0	−50
Trazodone	0.69	[63]	4	1	−80	20	−80
Vardenafil	12.8	[61]	4	0.5	−75	10	−40
Ziprasidone	0.152	[40]	4	0.4	−75	10	−40

coming from analogous experimental conditions is a major point in this kind of studies. Even though, generally speaking, for high throughput screening of virtual molecules through the use of QSAR models, a high accuracy of predicted target property values may not be required, in this particular case it was crucial. Indeed the fit of QSAR models obtained by using 3D-type molecular descriptors was expected to give a critical insight in the biologically active conformation of the channel.

For this reason, molecules were selected by looking carefully at the overall experimental protocol used for collecting the IC_{50} values. Similar conditions in terms of cell model, temperature, patch clamp technique, K^+ concentration in bath solution, holding potential, depolarization potential, depolarization voltage, duration of the depolarization, membrane potential during repolarization step used for current measurement, were chosen.

IC_{50} values range from 0.001 μM of desmethylastemizole to 4.4 mM of 4-aminopyridine, so ensuring a large variability of the biological indicator. In order to handle data in a more convenient way, IC_{50} values were transformed through a minus logarithm function (pIC_{50}).

2.3. Molecular docking

AutoDockTools 1.5.6 [64] was used to prepare the pdbqt files for all compounds constituting the dataset and for the *open*, *closed*, *rotated*, and *intermediate* forms of hERG K$^+$ channel. The polar hydrogens and united atom Kollman charges were assigned to the four forms of hERG K$^+$ channel during the preparation of the input files. For what concerns the ligands, partial atomic charges were determined by the Gasteiger method with modification to ensure integer charges or no charge (0, +1, −1) on each residue. Rotatable bonds were also assigned with AutoDockTools.

All compounds of the dataset were docked into each form of hERG K$^+$ channel using Autodock 4.2 [65].

Prior to AutoDock, AutoGrid was exploited for the preparation of the grid map for each form of the hERG channel using grid boxes with a npts (number of points in *xyz*) of 62–62–62, 48–46–60, 44–58–44, and 44–58–44 Å for the *open*, *closed*, *rotated* and *intermediate* form respectively, which defines the simulation space. The box spacing was 0.375 Å and the grid was set in order to cover the entire space of the binding site.

AutoDock was run using a maximum number of energy evaluation retries and generations of 2500000 and 27,000, respectively. The Lamarckian genetic algorithm (LGA) with the pseudo-Solis and Wets modification (LGA/pSW) method was used with default parameters for calculation of the docking possibilities [66]. For each ligand-channel pair, 100 Autodock runs were performed.

The clusters of the docking poses and their populations were analyzed by the integrated clustering method in Autodock. Cluster analysis used a rmstol = 2.0 Å for atoms in the ligands. In this study, a cluster was considered to be meaningful for the purpose of properly identifying an ideal binding site, only when more than 5% of the total poses where assigned to the same cluster. A good representative binding mode was defined as the lowest-energy complex of the cluster with the largest population.

2.4. Molecular descriptors

QSAR models were developed for each form of the hERG K$^+$ channel by using docking-based molecular descriptors (i.e. descriptors derived from optimized orientations and conformations of ligands obtained after molecular docking) and additional molecular descriptors computed by PaDEL, a software enabling us to obtain several types of fingerprints and 1D-/2D-molecular descriptors [67].

For what concerns the docking-based molecular descriptors, the binding modes of the 59 selected ligands were analyzed in terms of both number and type of contacting residues and related properties (Table 2), thus obtaining 231, 199, 179, and 205 docking-based molecular descriptors for the *open*, *closed*, *rotated*, and *intermediate* form of the hERG channel, respectively.

Interface surface and *interface volume* (respectively the surface and the volume of interaction between the ligand and the pore) were calculated for the lowest-energy conformation of each ligand obtained from molecular docking by using the Intersurf tool in the UCSF Chimera package [33]; a prune distance of 10 Å was fixed. These descriptors were considered important to define the volume and surface of interaction between the ligand and the pore cavity after the molecular docking. The UCSF Chimera package was also used for the visualization of the molecular structures as reported in figures.

Energy terms, H-*bonds lengths*, θ *angles*, ϕ *angles*, H-*bond energies*, π–π *stacking* and *t-shaped interactions* are obtained from the Analyze docking tool of AutodockTools. The *ligand efficiency* is the ratio between the Autodock free energy of binding (intermolecular + torsional energy) and the number of ligand heavy atoms,

Table 2

List of molecular descriptors derived from molecular docking analysis of the representative pose of ligand in the *open*, *closed*, *rotated*, and *intermediate* form of the hERG K$^+$ channel. Definition of donor/acceptor atom types of the ligand: values 0/1 (for NO/YES) were used for —N acceptor, —O acceptor, —OH donor, —NH donor groups involved in each specific H-bond.

Type	Docking-based molecular descriptor
Surface and volume	Interface surface (Å^2)
	Interface volume (Å^3)
Energy terms	Free energy of binding (kcal/mol)
	Mean binding energy in the most populated cluster (kcal/mol)
	Ligand efficiency
	Intermolecular energy (kcal/mol)
	van der Waals/H-bond/desolvation energy (kcal/mol)
	Electrostatic energy (kcal/mol)
H-bond (generic)	Total number of H-bonds
	Number of H-bond acceptors
	Number of H-bond donors
H-bond (specific for each residue involved)	Donor/acceptor atom type of the ligand forming the H-bond with the binding site residue
	H-bond length
	θ angle
	ϕ angle
	H-bond energy
π–π and t-shaped interaction	Total number of π–π interactions
	Specific π–π interactions
	Total number of t-shaped interactions
	Specific t-shaped interactions
Contacts (specific for each residue involved)	Specific contacts with residues of chain A
	Specific contacts with residues of chain B
	Specific contacts with residues of chain C
	Specific contacts with residues of chain D
	Total number of contacts with a specific residue
Contacts (generic)	Total number of contacts
	Total number of contacts with key-residues defined from mutagenesis studies (Gly648, Tyr652, Phe656, and Val659)
	Total number of contacts with residues of chain A
	Total number of contacts with residues of chain B
	Total number of contacts with residues of chain C
	Total number of contacts with residues of chain D
	Distribution Index 1 (contacts with chains A and B/contacts with chain C and D)
	Distribution Index 2 (contacts with chains A and C/contacts with chain B and D)
	Total number of no-H-bond contacts
	H-bonds/no-H-bond contacts

simply called ΔG/number of heavy atoms; small molecules usually show high values.

Distribution indices (*Index 1* and *Index 2*) are docking-based descriptors defined by us. If the distribution is 1, the ligand is centered in the pore according to the four-fold symmetry of the hERG K$^+$ channel. If Distribution Index 1 is $\gg 1$, the ligand is shifted toward the chains A–B, if it is $\gg 1$ it is shifted toward the chains C–D. If Distribution Index 2 is $\gg 1$, the ligand is shifted toward the chains A–C, if it is $\gg 1$ it is shifted toward the chains B–D.

In order to develop the QSAR models, hundreds of 1D-/2D-molecular descriptors, Pubchem fingerprints, and substructure fingerprints (Table 3) were also computed for the 59 compounds constituting the dataset by means of the above mentioned PaDEL program.

For each form of the hERG K$^+$ channel, docking-based and PaDEL-computed molecular descriptors were used all together in order to develop QSAR models where the ligand properties were

Table 3
Molecular descriptors calculated by PaDEL.

Descriptor type	Descriptor ID	Class
AcidicGroupCount	nAcid	2D
Aromatic atoms count	naAromAtom	2D
Aromatic bonds count	nAromBond	2D
Atom count	nAtom, nHeavyAtom, nH, nB, nC, nN, nO, nS, nP, nF, nCl, nBr, nI	2D
BasicGroupCount	nBase	2D
BondCount	nBonds, nBonds2, nBondsS, nBondsS2, nBondsS3, nBondsD, nBondsD2, nBondsT, nBondsQ	2D
Carbon types	C1SP1, C2SP1, C1SP2, C2SP2, C3SP2, C1SP3, C2SP3, C3SP3, C4SP3	2D
HBondAcceptorCount	nHBAcc, nHBAcc2, nHBAcc3, nHBAcc_Lipinski	2D
HBondDonorCount	nHBDon, nHBDon_Lipinski	2D
LargestChain	nAtomLC	2D
LargestPiSystem	nAtomP	2D
LongestAliphaticChain	nAtomLAC	2D
McGowanVolume	McGowan_Volume	2D
Ring count	nRing, n3Ring, n4Ring, n5Ring, n6Ring, n7Ring, n8Ring, n9Ring, n10Ring, n11Ring, n12Ring, nG12Ring, nFRing, nF4Ring, nF5Ring, nF6Ring, nF7Ring, nF8Ring, nF9Ring, nF10Ring, nF11Ring, nF12Ring, nFG12Ring, nTRing, nT4Ring, nT5Ring, nT6Ring, nT7Ring, nT8Ring, nT9Ring, nT10Ring, nT11Ring, nT12Ring, nTG12Ring	2D
Rotatable bonds count	nRotB	2D
Rule of five	LipinskiFailures	2D
Topological polar surface area	TopoPSA	2D
van der Waals volume	VABC	2D
Weight	MW	2D
Charged partial surface area	PPSA-1, PPSA-2, PPSA-3, PNSA-1, PNSA-2, PNSA-3, DPSA-1, DPSA-2, DPSA-3, FPSA-1, FPSA-2, FPSA-3, FNSA-1, FNSA-2, FNSA-3, WPSA-1, WPSA-2, WPSA-3, WNSA-1, WNSA-2, WNSA-3, RPCG, RNCG, RPCS, RNCS, THSA, TPSA, RHSA, RPSA	3D
Pubchem fingerprint	Hierarchal element counts	Fingerprint
	Rings in a canonic Extended Smallest Set of Smallest Rings (ESSSR) ring set	
	Simple atom pairs	
	Simple atom nearest neighbors	
	Detailed atom neighborhoods	
	Simple SMARTS patterns	
	Complex SMARTS patterns	
Substructure fingerprint		Fingerprint
Substructure fingerprint count		

computed also taking into account their 3D arrangement in the binding site, which can reasonably be considered their biologically active conformation.

In this way, four modeling sets constituted by 59 instances (compounds) and 842, 805, 785, and 823 attributes (docking-based and PaDEL-computed molecular descriptors) were obtained for the *open*, *closed*, *rotated*, and *intermediate* states of the hERG channel, respectively.

Descriptors showing an inter-correlation higher than 0.95 (a threshold commonly used for this kind of computation) were discarded from each modeling set, thus reducing the number of molecular descriptors to 421, 426, 412, and 425 for the *open*, *closed*, *rotated*, and the *intermediate* form, respectively.

2.5. QSAR modeling

The multi-linear regression algorithm available in the WEKA software [68] was used to develop the QSAR models.

The number of molecular descriptors was progressively and rationally reduced by using different methods. In the first step, the CfsSubsetEval attribute evaluator [69] within WEKA, was applied and both the BestFirst and the Genetic Algorithm were chosen as the search methods.

The Genetic Algorithm method evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features which are highly correlated with the class of compounds while having low inter-correlation are preferred. The Best First method searches the space of attribute subsets by greedy

hill-climbing augmented with a backtracking facility. A forward search direction was chosen among the parameters of the method.

Using both a Leave-Ten-Out (L-10-O) and a Leave-One-Out (LOO) cross-validation, the number of molecular descriptors was reduced and four subsets of molecular descriptors (two for each search method) were selected according to the averaged percentages of a ranking value.

In the first step the number of molecular descriptors was reduced from hundreds to tens.

In the second step, one QSAR model was developed for each subset selected for the *open*, *closed*, *rotated*, and *intermediate* forms (for a total of 12 models) by applying the greedy method for the attribute selection within the linear regression algorithm of WEKA. Each model was then refined by removing molecular descriptors which do not appear to be significant, by following the *p*-value criterion (with $p < 0.05$). In this way, all models were developed with a maximum of 8 molecular descriptors; thus, an optimal ratio between the number of descriptor exploited and molecules comprised in the dataset is ensured and the risk of obtaining over-correlated equations is sufficiently limited [70].

Two statistical parameter, R^2, the square of the coefficient of regression, and q^2, the Leave-One-Out Cross-Validated (LOO-CV) R^2, were used to assess the goodness-of-fit of the QSAR models. Threshold values for R^2 and q^2 we set to 0.6 and 0.5, respectively, as already done in other studies [71,72]. In a following step, a test set, the size of which is about 1/3 of the whole dataset, was randomly selected and the R^2 value on the test set was used to estimate the predictive power of each model.
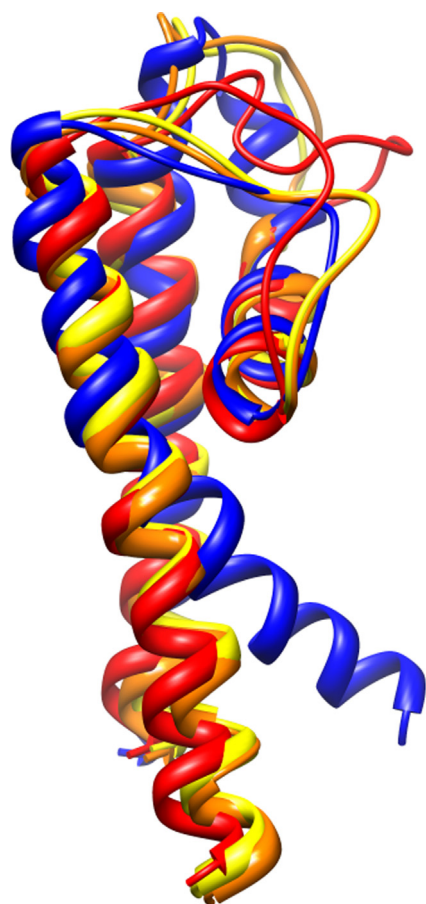
Fig. 4. Bending of the S6 helix at Gly648 in the chain A of *open* state (in blue) of hERG channel in comparison with the chain A of the forms derived from the closed state: *closed*, *rotated* and *intermediate* forms are reported in red, orange and yellow, respectively.

Other than the direct use of good QSAR models which could be made for the purpose of predicting properties of new chemical entities, we were also expecting that an indirect validation of one or more 3D-models of the channel pore, out of the four ones generated in this study, could be inferred after the quality of the QSAR models obtained was ranked. Indeed we strongly believe that the quality of the QSAR models is related to ligand properties computed on putative biologically active conformations suggested by molecular docking.

3. Results and discussion

3.1. Analysis of the homology models for the channel pore

In the homology model representing the *closed* form and in the other models derived from it (the *rotated* and *intermediate* forms), the S6 helices cross over to each other leading to a narrow region below the central cavity. In the *open* form, this point of constriction is enlarged through a kink which implies a rotation at Gly648. Such a rotation tends to position the helices in order to allow the access to the pore. The glycine-hinge residue (Gly648), commonly used for the alignment of the S6-transmembrane helix, may be responsible of the conformational transition required for the channel to open [73]. The conformational transition between the *open* and the *closed* forms of the hERG channel is due to a bend of about 40° at the Gly648 level (located in the S6 helix) (Fig. 4).

Mutagenesis studies suggested that interactions with Tyr652 and Phe656, in particular, are crucial for the drug blocking activity
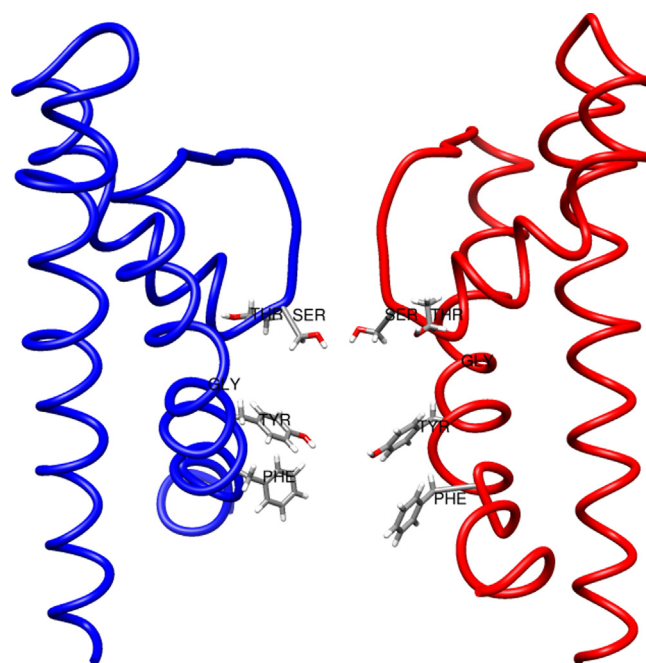
Fig. 5. Stereoview of chains A (in blue) and C (in red) of the homology model of the *open* state of hERG channel. The putative binding site with Thr623, Ser624, Gly648 and the key residues Tyr652, and Phe656 is explicitly displayed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

[2,4–10]. Looking at the four states of the hERG K$^+$ channel, the lumen of the pore generally consists of two layers of aromatic side chains contributed by such residues.

In the *open* form, Tyr652 and Phe656, as well as Thr623 and Ser624, appear to be well exposed in the pore cavity, which is in accordance with mutagenesis studies (Fig. 5).

In this form, the aromatic ring of Tyr652 residue is oriented almost perpendicularly to the axis of the channel which is in contrast with what is observed for the closed state.

In the *closed* form of the channel, Phe656 residues of chains A and C did not show plausible orientations and result to be not accessible for the drug binding (Fig. 6).

Du-Cuny et al. overcame this problem by manually adjusting the rotameric conformations so that both Tyr652 and Phe656 pointed toward the selectivity filter, followed by molecular dynamics for the structure optimization [74]. In this study, we had preferred to build and refine a *rotated* model in order to overcome such discrepancy between what has come out from the modeling results for the *closed* form and the mutagenesis studies. The rotation of the S6 helix was obtained by "shifting" the sequence of the S6 helix one residue toward the C-terminus according to a protocol similar to that described by Stansfeld et al. [30] and Chen et al. [75]. In this way, an initial rotated structure was obtained, where residues of Tyr652 (chains A-D) are packed in the central cavity of the hERG pore and Val659 define a hydrophobic gate of the channel. After that, the Morph server was used to create thirty structures corresponding to different torsion angle of the S6 helix between the *closed* and the initial rotated form, both used as the starting conformations submitted to the Morph server. The intermediate structure (among the closed and the initial rotated forms) #25 was chosen as the *rotated* form of the hERG pore. Such a choice was made because the four Tyr652 and Phe656 residues (chain A–D), known from mutagenesis studies to be implicated in drug binding, showed an optimal orientation within the inner cavity, pointing toward the cytoplasmic opening of the pore (Fig. 7). However, Phe656 residues still appeared to be tightly packed together, thus constituting a
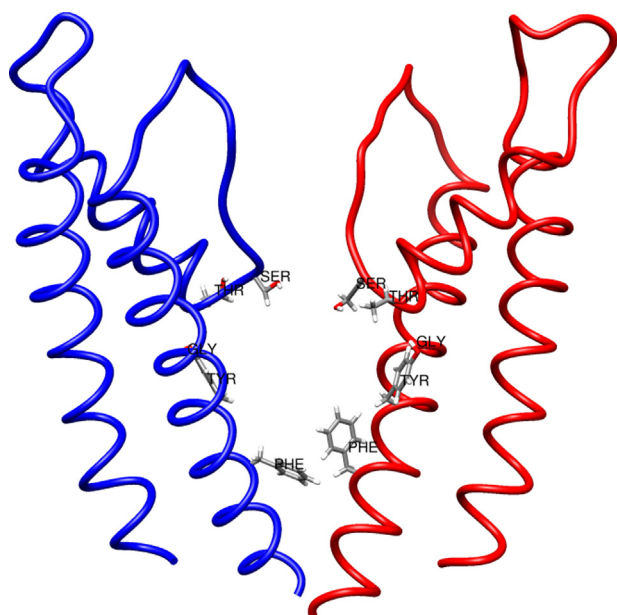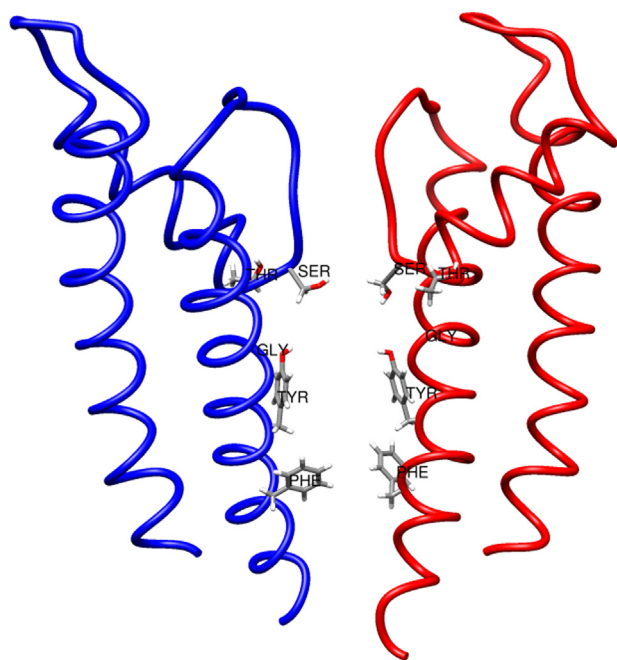
Fig. 6. Stereoview of chains A and C of the homology model of the *closed* state of hERG channel. The putative binding site with Thr623, Ser624, Gly648 and th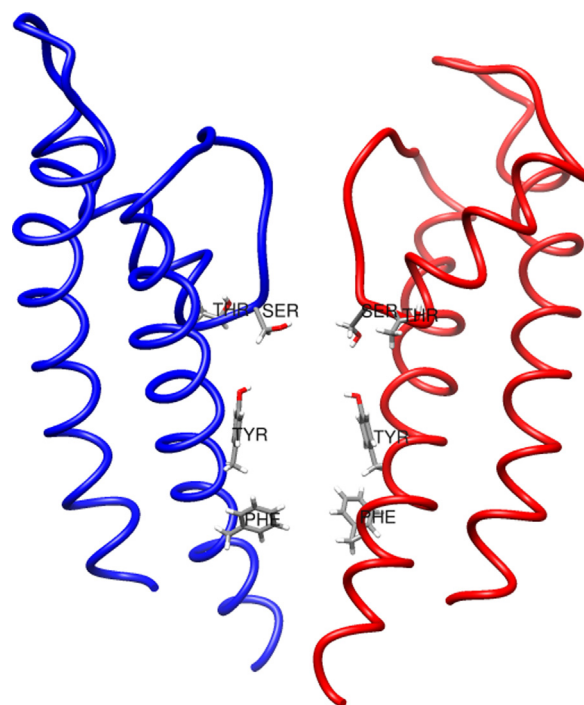e key residues Tyr652, and Phe656 is explicitly displayed. Phe656 in chain A appear to have an unreliable orientation and, together with Phe656 in chain A (not displayed) result to be not accessible for drug binding.

rather inaccessible region to the binding of ligands, which is unrealistic and is not in full agreement with mutagenesis studies.

For this reason, the channel opening was simulated by normal mode analysis as described in Section 2.1, starting from the *rotated* form (derived from and resembling a closed state of the channel) in order to obtain an *intermediate* structure between the closed and the open state of the hERG K$^+$ channel (Fig. 8).

This procedure created five intermediate structures describing the transition from the closed to the open state of the channel. The

Fig. 7. Stereoview of chains A and C of the homology model of the *closed* state of hERG channel. The putative binding site with Thr623, Ser624, Gly648 and the key residues Tyr652, and Phe656 is explicitly displayed.

Fig. 8. Stereoview of chains A (in blue) and C (in red) of the homology model of the *intermediate* state of hERG channel. The putative binding site with Thr623, Ser624, Gly648 and the key residues Tyr652, and Phe656 is explicitly displayed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

transition form, showing key-residues (i.e. the ones known from mutagenesis studies to be implicated in drug binding) arranged in an optimal orientation within the inner cavity, was selected and conventionally called the *intermediate* state of the channel.

3.2. Molecular docking and QSAR models

All molecules belonging to the selected dataset were docked into the models for the four forms of hERG K$^+$ channel (for a total of 236 docking runs) and hundreds of molecular descriptors were derived from the optimized orientation/conformation of each compound. These docking-based descriptors together with the ones calculated by PaDEL were used to develop QSAR models for each channel form.

According to the protocol described in Section 2, four criteria for features selection were used in order to select reduced subsets of molecular descriptors on which the models were developed. After that, the greedy mode for attribute selection was applied within the linear regression algorithm of WEKA. Finally, molecular descriptors which turned out to be less significant, in terms of *p*-value, were removed. Four QSAR models were so obtained for each channel form, as reported in Table 4.

3.2.1. Open form

The binding site of the *open* form was defined by a total of 40 residues, identified on the basis of the possibility that at least one compound of the data set was able to interact with them.

Ser624 (chains A, C, and D) and Tyr652 (chains A, B, and C) residues were found to be involved as H-bond donors.

Leu622 (chains B, C, and D), Thr623 (chains A, B, C, and D), Ser624 (chain B) Met645 (chains A, B, C, and D), Ser649 (chain D), Tyr652 (chains A, B, C, and D), Ala653 (chains A, C, and D) residues were found to be involved as H-bond acceptors.

Tyr652 (chains A, B, C, and D) and Phe656 (chain B) residues gave π–π stacking interactions with aromatic rings of several

Table 4
QSAR models developed on the basis of docking-based molecular descriptors and descriptors calculated by PaDEL. Models which passed the validation steps (thresholds for the training set: $R^2 > 0.6$, $q^2 > 0.5$; for the test set: $R^2 > 0.6$) are reported in bold. FS1, FS2, FS3, and FS4 refer to the protocol employed for the Feature Selection tool of WEKA package: FS1, BestFirst method with a Leave-Ten-Out (L-10-O) cross validation; FS2, BestFirst method with a Leave-One-Out (L-O-O) cross validation; FS3, Genetic search method with a Leave-Ten-Out (L-10-O) cross validation; FS4, Genetic search method with a Leave-One-Out (L-O-O) cross validation. Abbreviations: FS selection is the number of descriptors obtained after reduction with the FS criteria; Greedy selection is the number of descriptors reduced in the second step after the application of the Greedy method within the linear regression function of WEKA; D is the number of molecular descriptors involved in the final model.

Model	FS selection	Greedy selection	D	R^2	q^2	R^2_{test}
QSAR models derived from the *Open* form of hERG channel						
open-FS1	16	9	8	0.73	0.49	0.71
open-FS2	17	11	8	0.76	0.57	0.67
open-FS3	25	9	8	0.55	0.38	0.18
open-FS4	29	10	8	0.61	0.44	0.19
QSAR models derived from the *Closed* form of hERG channel						
cl-FS1	11	6	6	0.48	0.16	0.27
cl-FS2	21	10	8	0.68	0.38	0.48
cl-FS3	21	7	7	0.61	0.23	0.32
cl-FS4	22	11	8	0.61	0.50	0.51
QSAR models derived from the *Rotated* form of hERG channel						
rot-FS1	24	8	8	0.55	0.16	0.17
rot-FS2	21	8	8	0.55	0.25	0.28
rot-FS3	37	7	7	0.57	0.10	0.12
rot-FS4	30	11	8	0.61	0.38	0.46
QSAR models derived from the *Intermediate* form of hERG channel						
int-FS1	38	11	8	0.74	0.64	0.48
int-FS2	37	11	8	0.75	0.68	0.69
int-FS3	22	9	8	0.63	0.49	0.39
int-FS4	23	5	5	0.57	0.26	0.24

compounds of the dataset; Tyr652 (chains A, B, and C) residues are also involved in t-shaped interactions.

One of the QSAR models (**open-FS2**) satisfied the validation threshold set for both R^2 and q^2. The model was developed on the basis of 17 molecular descriptors which had been pre-selected, among the 842 available, by using the CfsSubsetEval attribute evaluator within the WEKA package through a BestFirst searching method with a Leave-One-Out Cross-Validation (LOO-CV) analysis (FeatureSelection, FS2). The greedy mode, within the linear regression function of WEKA, reduced the number of descriptors to 11. After that, the equation was refined by removing those descriptors that appeared to be not significant for the model (p-value > 0.05).

The final model is represented by equation (1) reported below:

Coefficients:

	Estimate	Std. error	*t* value	Pr(>|*t*|)
(Intercept)	−2.81472	0.63224	−4.452	4.78e−05
intermolE	−0.48683	0.08616	−5.650	7.61e−07
NH-OH_Thr623B bond energy	0.93030	0.28083	3.313	0.00172
NH-O_Ser624B	−2.34485	0.73128	−3.206	0.00234
NH-O_Thr623C bond energy	0.66591	0.23441	2.841	0.00649
Contact with Ala653A	0.64739	0.26805	2.415	0.01942
C1SP2	−0.46067	0.10798	−4.266	8.84e−05
PubchemFP601	−1.14465	0.35610	−3.214	0.00229
PubchemFP713	−1.61844	0.34886	−4.639	2.55e−05

This model is based on 8 molecular descriptors. The statistical parameters computed on the training set are: $R^2 = 0.76$ and $q^2 = 0.57$; the R^2 computed on the test set is 0.67.

Five out of the eight descriptors of the model were derived from molecular docking. The intermolecular energy from Autodock, the bond energy of the H-bond between a −NH group of the ligand and the hydroxylic oxygen of Thr623 (chain B), the bond energy of the H-bond between a −NH group the ligand and the hydroxylic oxygen of Thr623 (chain C), and the number of contacts with Ala653 (chain A) are docking-based molecular descriptors which positively correlate with pIC$_{50}$, leading to a high affinity of the ligand for the hERG K$^+$ channel.

Donor/acceptor atom types of the ligand were defined, within the modeling set matrix, as described in Table 2: values 0/1 (for

NO/YES) were used for −N acceptor, −O acceptor, −OH donor, −NH donor groups involved in each H-bond. In this way, the presence of a H-bond between a −O (or −N) acceptor and a specific donor group of the binding site was labeled with 1, otherwise it was set to 0. The same criterion was used for OH (or NH) donor groups of the ligands and an acceptor of the binding site.

Three PaDEL-computed molecular descriptors are involved in the equation; they negatively contribute to the pIC$_{50}$ value. While docking-based molecular descriptors refer to interactions between the ligand (whatever it is) and the binding site, molecular descriptors from PaDEL only "describe" the nature of the ligand. Description of C1SP2 and Pubchem fingerprints involved in the model **open-FS2** are reported in Supporting Material.

3.2.2. Closed form

The binding site of the *closed* form was defined by a total of 39 residues, identified by the same criterion described in Section 3.2.1.

Ser624 (chains A and D) residues were found to be involved as H-bond donors. Leu622 (chains A, B, and D), Thr623 (chains A, B, C, and D), Ser624 (chains A and D), Gly648 (chain C), Ser649 (chains A, B, C, and D), Tyr652 (chain A) residues were found to be involved as H-bond acceptors.

Tyr652 (chains A, B, C, and D) and Phe656 (chains B, C, and D) residues gave π–π stacking interactions with aromatic rings of several compounds of the dataset; this is in agreement with Mitcheson et al. who found strong π–π stacking interactions involving these residues in the *closed* state of the channel model built on the same template of KcsA [2] used by us. Tyr652 (chains A and B) and Phe656 (chains A, C, and D) residues are also involved in t-shaped interactions.

Three QSAR models have a R^2 higher than 0.6 but only one of them has a $q^2 = 0.5$ (**cl-FS4**).

The best model is based on 8 molecular descriptors. The statistical parameters computed on the training set are: $R^2 = 0.61$ and $q^2 = 0.50$; the R^2 computed on the test set is 0.51. The model is represented by equation (2) reported below:

Coefficients:

	Estimate	Std. error	t value	$Pr(>|t|)$
(Intercept)	−4.893920	1.012719	−4.832	1.32e−05
Group t-shaped interacting with Phe656D	1.722550	0.571092	3.016	0.00402
nN	−0.152766	0.083548	−1.828	0.07344
nBondsS	0.030810	0.020254	1.521	0.13451
C2SP2	0.103419	0.044217	2.339	0.02338
THSA	0.005405	0.002551	2.119	0.03907
PubchemFP187	0.760282	0.402097	1.891	0.06445
PubchemFP535	−0.501633	0.303032	−1.655	0.10411
SubFPC287	−0.124961	0.065049	−1.921	0.06044

where nN is the number of Nitrogen atoms, nBondsS is the number of single bonds (including bonds with Hydrogen atoms), THSA is the sum of solvent accessible surface areas of atoms with absolute value of partial charges less than 0.2. Description of C2SP2 and Pubchem fingerprints involved in the model **cl-FS4** are reported in Supporting Material. This model shows a large negative intercept; looking at the others coefficients of the equation, we can infer that, in order to have a pIC_{50} in the biological meaningful range, ligands should be characterized by almost one t-shaped group able of interacting with Phe656 (chain D) and, moreover, they should have several other characteristics which are represented by the molecular descriptors with positive coefficients (such as, nBondsS, C2SP2, THSA, and PubchemFP187). That can be considered an indication of the weakness of the model, which can also be potentially overfitted.

3.2.3. Rotated form

The binding site of the *rotated* form was defined by a total of 33 residues, identified by the same criterion described in Section 3.2.1.

Ser624 (chains A, C, and D) and Tyr652 (chains A, B, and C) residues of were found to be involved as H-bond donors. Leu622 (chains B, C, and D), Thr623 (chains A, B, C, and D), Ser624 (chain B), Met645 (chains A, B, C, and D), Ser649 (chain D), Tyr652 (chains A, B, C, and D), Ala653 (chains A, C, and D) residues were found to be involved as H-bond acceptors.

Tyr652 (chains A, B, C, and D) and Phe656 (chain B) residues gave π–π stacking interactions with aromatic rings of several compounds of the dataset; Tyr652 (chains A B, and C) residues are also involved in t-shaped interactions.

Only one QSAR model (**rot-FS4**) has a R^2 higher than 0.6 but it showed a $q^2 < 0.5$.

The model is based on 6 molecular descriptors. The statistical parameters computed on the training set are: $R^2 = 0.61$ and $q^2 = 0.38$; the R^2 computed on the test set is 0.56. The model is represented by equation (3) reported below:

Coefficients:

	Estimate	Std. error	t value	$Pr(>|t|)$
(Intercept)	−4.15163	0.58427	−7.106	4.10e−09
Group π–π interacting with Tyr652B	2.69145	0.95313	2.824	0.0068
nC	0.21246	0.02885	7.364	1.62e−09
PubchemFP24	−1.01553	0.38503	−2.638	0.0111
PubchemFP214	2.27215	1.06759	2.128	0.0383
PubchemFP308	0.90364	0.40668	2.222	0.0308
PubchemFP346	−0.53661	0.27625	−1.942	0.0577
PubchemFP380	−1.49808	0.56345	−2.659	0.0105
PubchemFP652	1.03330	0.67244	1.537	0.1307

A description of Pubchem fingerprints involved in the model **rot-FS4** is reported in Supporting Material.

3.2.4. Intermediate form

The binding site of the *intermediate* form was defined by a total of 36 residues, identified by the same criterion described in Section 3.2.1.

Ser624 (chains A, B, and D), Gly648 (chain B), and Tyr652 (chains A, B, and D) residues were found to be involved as H-bond donors. Leu622 (chain A), Thr623 (chains A, C, and D), Ser624 (chains A, B,

and C), Val644 (chain B), Met645 (chain B), Gly648 (chain A), Ser649 (chain B), Tyr652 (chains A, B, C, and D) residues were found to be involved as H-bond acceptors.

Tyr652 (chains A, B, C, and D) and Phe656 (chain C) residues gave π–π stacking interactions with aromatic rings of several compounds of the dataset; Tyr652 (chains B and C) and Phe656 (chain D) residues are also involved in t-shaped interactions.

Two QSAR models (**int-FS1** and **int-FS2**) satisfied the validation threshold set for both R^2 and q^2.

Looking at the statistical parameters, **int-FS2** appear to be the best one; it was developed on the basis of an initial set of 37 molecular descriptors which had been pre-selected, among the 823 available, by using the CfsSubsetEval attribute evaluator within the WEKA package through a BestFirst searching method with a Leave-One-Out Cross-Validation (LOO-CV) analysis (FeatureSelection, FS2). The greedy mode, within the linear regression function of WEKA, reduced the number of descriptors to 11. After that, the equation was refined by removing those descriptors that appeared to be not significant for the model (p-value > 0.05).

The best performing model is based on 8 molecular descriptors. The statistical parameters computed on the training set are: $R^2 = 0.75$ and $q^2 = 0.68$; the R^2 computed on the test set is 0.69.

The final model is represented by equation (4) reported below:

Coefficients:

	Estimate	Std. error	t value	$Pr(>|t|)$
(Intercept)	−4.14013	0.60594	−6.833	1.10e−08
OH-OH_Ser649B	1.92830	0.75657	2.549	0.013928
Group π–π interacting with Tyr652B	1.50552	0.53733	2.802	0.007209
nC	0.10629	0.02442	4.352	6.67e−05
PubchemFP599	2.34206	0.57092	4.102	0.000151
PubchemFP652	2.06563	0.56407	3.662	0.000604
PubchemFP717	0.91165	0.32145	2.836	0.006578
SubFP101	−1.22560	0.35898	−3.414	0.001277
SubFPC88	−0.80690	0.17439	−4.627	2.66e−05

Two out of eight descriptors are docking-based molecular descriptors that positively correlate with pIC_{50}, leading to a high affinity of the ligand for the hERG channel: the descriptor indicating the presence of a H-bond between a –OH donor group of the ligand and the hydroxylic oxygen of Ser649 (chain B) and the one indicating the presence of a π–π stacking interaction with Tyr652 (chain B).

Four PaDEL-computed molecular descriptors also positively contribute at increasing the pIC_{50} value: nC is the number of Carbon atoms, PubchemFP652 and PubchemFP599 are the simple SMARTS patterns O—C:C:C—N and H—C—C=C—H, respectively. PubchemFP717 is the complex SMART pattern Cc1ccc(Cl)cc1. All substructures defined by Pubchem fingerprints (simple and complex SMARTS patterns) are visually represented as chemical fragments in Table 5. Predicted vs. actual pIC_{50} values for the test set are reported in Fig. 9.
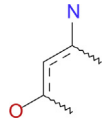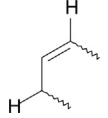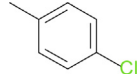
3.3. Comparison of the models

The QSAR model obtained from the *intermediate* form can be considered the best one in terms of statistical parameters, which means that it better fits the experimental data according to equation (4). Furthermore, the analysis of the molecular descriptors involved in the QSAR model is in accordance with the commonly accepted knowledge about the hERG blocking.

Several ligand features appear to be crucial for hERG binding, such as the presence of one or more aromatic rings involved in π–π stacking interactions with Tyr652 residues, as already observed by other Authors [29,18]. It also accounts for the importance of the hydrophobic effect as a major driving force for the ligand affinity to the hERG K^+ channel.

Table 5
Substructures defined by Pubchem fingerprints involved in the best QSAR model obtained for the *intermediate* form of the hERG channel and which positively correlate with the pIC$_{50}$.

Pubchem fingerprint	Substructure
PubchemFP652	
PubchemFP599	
PubchemFP717	

The fragment represented by the fingerprint PubchemFP652, resembling a portion of aromatic ring bearing a N atom placed at a defined distance from an O atom, is in agreement to what already found by us in previous studies [76,13]. The number of Carbon atoms, PubchemFP599 and PubchemFP717, which positively contribute to the pIC$_{50}$, suggest that ligands characterized by hydrophobic groups should have a high affinity for hERG. This is in accordance with the lipophilic nature recognized to be characteristic of most of the known hERG ligands.

QSAR models obtained from the *closed* and the *rotated* forms appear to be the less performing ones in terms of statistical parameters and it is not possible to make a reliable analysis of the molecular descriptors involved in the models and explain how they could affect the hERG binding.

Even though the *open* form do not seem to supply the best rationalization of the experimental IC$_{50}$ inhibition data, it allowed us to develop a QSAR model with computed values of statistical parameters which can be considered to be acceptable, so giving some interesting suggestions. In particular, the intermolecular energy obtained from Autodock, positively correlates with pIC$_{50}$: a low value of intermolecular energy (expressed in kcal/mol) means a high affinity of the ligand toward the hERG K$^+$ channel. This observation suggests that the molecular docking procedure may supply a plausible model for investigating binding modes of a ligand in the open state of the hERG pore. It has to be pointed out the behavior of cisapride, which has experimentally been proved to have affinity for the open state of the hERG channel [77]. The intermolecular energy value found for cisapride through our calculations was −8.27 kcal/mol, one of the lowest of the whole dataset;

that suggests high affinity of this compound for the open form of the channel, in accordance with literature data. However, only a slightly higher value of the Autodock intermolecular energy was found for cisapride in the binding with the intermediate form of the channel (−7.17 kcal/mol); this is also plausible: the *intermediate* form is expected to share some structural features with the open form, since it is an intermediate form between the closed and the open state of the hERG channel. Furthermore, the *intermediate* form is able to rationalize the experimental data of cisapride that contains three molecular features involved in equation (4) which positively correlate with pIC$_{50}$ (PubchemFP599, PubchemFP652, and PubchemFP717).

4. Conclusions

In the study presented here, four 3D theoretical models were built in order to simulate the hERG K$^+$ channel state when ligands are interacting with it, based on their homology with analogous channels with known 3D structure. They are labeled as *open*, *closed*, *rotated*, and *intermediate* forms of the channel. A dataset of 59 compounds was collected from literature and rationally selected according to the availability of IC$_{50}$ values derived from whole-cell patch clamp performed at 37 °C on HEK cells. Molecular docking was then carried out using each one of the models for the four forms of the channel. Hundreds of docking-based molecular descriptors were derived and used, together with other PaDEL-computed molecular descriptors, to develop QSAR models for each state of the channel.

Other than the direct use of good QSAR models which could be made for prediction properties (even if limited to the applicability domain defined by this small set of compounds), we were also expecting that an indirect validation of one or more out of the 3D models, representing the channel pore when it interacts with ligands, could be inferred from the ranked quality of the QSAR models obtained. Indeed their quality should be also related to ligand properties computed on the basis of optimized orientations/conformations supplied by molecular docking.

First of all, the overall results afforded QSAR models of high quality which could enable prediction of the capability of newly designed molecules to inhibit the hERG K$^+$ channel. Moreover the study allowed us to get insights into the binding modes and key features in the interaction between the hERG K$^+$ channel and ligands characterized by the unwanted blocking activity.

The form defined as *intermediate*, meaning an "intermediate" conformation between the *open* and the *rotated* form, led to the most reliable rationalization of the experimental IC$_{50}$ inhibition data. The interpretation of the QSAR model developed on docking-based and PaDEL-computed molecular descriptors suggests that the *intermediate* form is suitable to be used in order to obtain reliable information about the putative binding mode of a ligand in the hERG pore. Moreover, the importance of the hydrophobic interactions was once again highlighted as one of the most important driving forces for the affinity of ligands toward the hERG K$^+$ channel. Finally, this study points out the importance of analyzing different states of the channel which cannot be statically defined, in order to identify the most suitable form to be subjected to a molecular docking protocol.
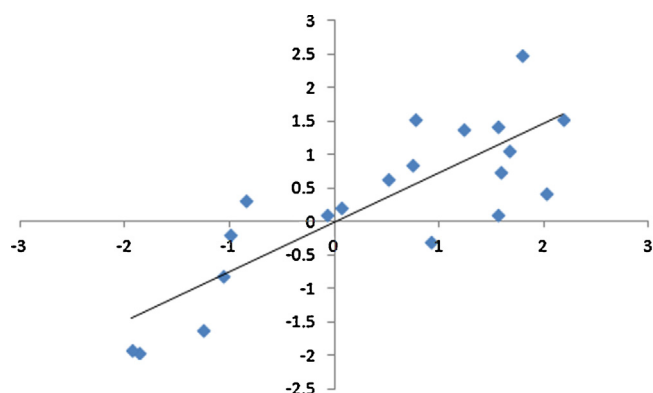
Fig. 9. Predicted vs. actual pIC$_{50}$ values for the test set of int-FS2 model.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2013.10.001.

References

[1] M.C. Trudeau, J.W. Warmke, B. Ganetzky, G.A. Robertson, Science 269 (1995) 92–95.
[2] J.S. Mitcheson, J. Chen, M. Lin, C. Culberson, M.C. Sanguinetti, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 12329–12333.
[3] A. Gintant, Z. Su, R.L. Martin, B.F. Cox, Toxicologic Pathol. 34 (2006) 81–90.
[4] K. Kamiya, J.S. Mitcheson, K. Yasui, I. Kodama, M.C. Sanguinetti, Mol. Pharmacol. 60 (2001) 244–253.
[5] K. Kamiya, R. Niwa, J.S. Mitcheson, M.C. Sanguinetti, Mol. Pharmacol. 69 (2006) 1709–1716.
[6] M. Perry, M.J. De Groot, R. Helliwell, D. Leishman, M. Tristani-Firouzi, M.C. Sanguinetti, J. Mitcheson, Mol. Pharmacol. 66 (2004) 240–249.
[7] M. Perry, P.J. Stansfeld, J. Leaney, C. Wood, M.J. de Groot, D. Leishman, M.J. Sutcliffe, J.S. Mitcheson, Mol. Pharmacol. 69 (2006) 509–519.
[8] D. Fernandez, A. Ghanta, G.W. Kauffman, M.C. Sanguinetti, J. Biol. Chem. 279 (2004) 10120–10127.
[9] J.A. Sanchez-Chapula, R.A. Navarro-Polanco, C. Culberson, J. Chen, M.C. Sanguinetti, J. Biol. Chem. 277 (2002) 23587–23595.
[10] J.A. Sanchez-Chapula, T. Ferrer, R.A. Navarro-Polanco, M.C. Sanguinetti, Mol. Pharmacol. 63 (2003) 1051–1058.
[11] D.A. Doyle, J. Morais Cabral, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait, R. MacKinnon, Science 280 (1998) 69–77.
[12] H.J. Witchel, C.E. Dempsey, R.B. Sessions, M. Perry, J.T. Milnes, J.C. Hancox, J.S. Mitcheson, Mol. Pharmacol. 66 (2004) 1201–1212.
[13] A. Coi, I. Massarelli, L. Testai, V. Calderone, A.M. Bianucci, Eur. J. Med. Chem. 43 (2008) 2479–2488.
[14] Y. Jiang, A. Lee, J. Chen, M. Cadene, B.T. Chait, R. MacKinnon, Nature 417 (2002) 515–522.
[15] R.A. Pearlstein, R.J. Vaz, J. Kang, X.L. Chen, M. Preobrazhenskaya, A.E. Shchekotikhin, A.M. Korolev, L.N. Lysenkova, O.V. Miroshnikova, J. Hendrix, D. Rampe, Bioorg. Med. Chem. Lett. 13 (2003) 1829–1835.
[16] Y. Jiang, A. Lee, J. Chen, V. Ruta, M. Cadene, B.T. Chait, R. MacKinnon, Nature 423 (2003) 33–41.
[17] F. Osterberg, J. Aqvist, FEBS Lett. 579 (2005) 2939–2944.
[18] R. Farid, T. Day, R.A. Friesner, R.A. Pearlstein, Bioorg. Med. Chem. 14 (2006) 3160–3173.
[19] D.T. Jones, J. Mol. Biol. 292 (1999) 195–202.
[20] D.T. Jones, FEBS Lett. 423 (1998) 281–285.
[21] L.J. McGuffin, K. Bryson, D.T. Jones, Bioinformatics 16 (2000) 404–405.
[22] C. O'Donovan, M.J. Martin, A. Gattiker, E. Gasteiger, A. Bairoch, R. Apweiler, Brief. Bioinform. 3 (2002) 275–284.
[23] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25 (1997) 3389–3402.
[24] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235–242.
[25] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673–4680.
[26] N. Eswar, M.A. Marti-Renom, B. Webb, M.S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali, Current Protocols in Bioinformatics, John Wiley & Sons, Inc., 2006, Supplement 15, 5.6.1–5.6.30.
[27] M.-Y. Shen, A. Sali, Protein Sci. 15 (2006) 2507–2524.
[28] V.B. Chen, W.B. Arendall III, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, Acta Crystallogr. D66 (2010) 12–21.
[29] J.M. Word, S.C. Lovell, J.S. Richardson, D.C. Richardson, J. Mol. Biol. 285 (1999) 1735–1747.
[30] P.J. Stansfeld, P. Gedeck, M. Gosling, B. Cox, J.S. Mitcheson, M.J. Sutcliffe, Proteins Struct. Funct. Genet. 68 (2007) 568–580.
[31] W.G. Krebs, M. Gerstein, Nucleic Acids Res. (2000) 1665–1675.
[32] K. Suhre, Y.H. Sanejouand, Nucleic Acids Res. 32 (2004) W610–W614.
[33] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, J. Comput. Chem. 25 (2004) 1605–1612.
[34] W.J. Crumb, S. Ekins, R. Dustan Sarazan, J.H. Wikel, S.A. Wrighton, C. Carlson, C.M. Beasley, Pharm. Res. 23 (2006) 1133–1143.
[35] J.M. Ridley, J.T. Milnes, Y.H. Zhang, H.J. Witchel, J.C. Hancox, J. Physiol. 549 (2003) 667–672.
[36] S. Ekins, W.J. Crumb, R.D. Sarazan, J.H. Wikel, S.A. Wrighton, J. Pharmacol. Exp. Ther. 301 (2002) 427–434.
[37] P.J.S. Chiu, K.F. Marcoe, S.E. Bounds, C.-H. Lin, J.-J. Feng, A. Lin, F.-C. Cheng, W.J. Crumb, R. Mitchell, J. Pharmacol. Sci. 95 (2004) 311–319.
[38] W.J. Crumb, personal communication (9 August 2000).
[39] F. Ando, A. Kuruma, S. Kawano, J. Membr. Biol. 241 (2011) 31–38.
[40] J.W. Benbow, K.A. Andrews, J. Aubrecht, D. Beebe, D. Boyer, S. Doran, M. Homiski, Y. Hui, K. McPherson, J.C. Parker, J. Treadway, M. VanVolkenberg, W.J. Zembrowski, Bioorg. Med. Chem. Lett. 19 (2009) 2220–2223.
[41] J. Ridley, J. Milnes, J. Hancox, H. Witchel, J. Mol. Cell. Cardiol. 40 (2006) 107–118.
[42] K. Yuill, J. Borg, J. Ridley, J. Milnes, H. Witchel, A. Paul, R. Kozlowski, J. Hancox, Biochem. Biophys. Res. Commun. 318 (2004) 556–561.
[43] S.H. Jo, H.K. Hong, S.H. Chong, K.H. Won, S.J. Jung, H. Choe, Eur. J. Pharmacol. 592 (2008) 19–25.
[44] Z. Zhou, V. Vorperian, Q. Gong, S. Zhang, C. January, J. Cardiovasc. Electrophysiol. 10 (1999) 836–843.
[45] G.E. Kirsch, E.S. Trepakova, J.C. Brimecombe, S.S. Sidach, H.D. Erickson, M.C. Kochan, L.M. Shyjka, A.E. Lacerda, A.M. Brown, J. Pharmacol. Toxicol. Methods 50 (2004) 93–101.
[46] D.H. Singleton, H. Boyd, J.V. Steidl-Nichols, M. Deacon, M.J. Groot, D. Price, D.O. Nettleton, K.K. Wallace, M.D. Troutman, C. Williams, J.G. Boyd, J. Med. Chem. 50 (2007) 2931–2941.
[47] R. Duncan, M. McPate, J. Ridley, Z. Gao, A. James, D. Leishman, J. Leaney, H. Witchel, J. Hancox, Biochem. Pharmacol. 74 (2007) 425–437.
[48] J. Ridley, J. Milnes, H. Witchel, J. Hancox, Biochem. Biophys. Res. Commun. 325 (2004) 883–891.
[49] J.T. Milnes, C.E. Dempsey, J.M. Ridley, O. Crociani, A. Arcangeli, J.C. Hancox, H.J. Witchel, FEBS Lett. 547 (2003) 20–26.
[50] H.-A. Lee, E.-J. Kim, S.-A. Hyun, S.-G. Park, K.-S. Kim, Bas. Clin. Pharmacol. Toxicol. 107 (2010) 614–618.
[51] A. Paul, H. Witchel, J. Hancox, Br. J. Pharmacol. 136 (2002) 717–729.
[52] J.M. Ridley, P.C. Dooley, J.T. Milnes, H.J. Witchel, J.C. Hancox, J. Mol. Cell. Cardiol. 36 (2004) 701–705.
[53] S. Jo, H. Hong, J. Sung, H. Seon, J. Yun, Y. Koh, H. Choe, Arch. Pharmacol. Res. 30 (2007) 453–460.
[54] C. Eap, S. Crettol, J. Rougier, J. Schläpfer, L. Sintra Grilo, J. Déglon, J. Besson, M. Croquette-Krokar, P. Carrupt, H. Abriel, Clin. Pharmacol. Ther. 81 (2007) 719–728.
[55] P. Lacroix, W. Crumb, L. Durando, G. Ciottoli, Eur. J. Pharmacol. 477 (2003) 69–72.
[56] Y.J. Kim, H.K. Hong, H.S. Lee, S.H. Moh, J.C. Park, S.H. Jo, H. Choe, J. Cardiovasc. Pharmacol. 52 (2008) 485–493.
[57] J.S. Cordes, Mol. Pharmacol. 68 (2005) 876–884.
[58] L. Wu, M. Orikabe, Y. Hirano, S. Kawano, M. Hiraoka, J. Cardiovasc. Pharmacol. 42 (2003) 410–418.
[59] T. Christ, E. Wettwer, M. Wuest, M. Braeter, F. Donath, P. Champeroux, S. Richard, U. Ravens, Naunyn–Schmiedeberg's Arch. Pharmacol. 376 (2008) 431–440.
[60] S.H. Jo, H.K. Hong, S.H. Chong, H. Choe, Life Sci. 82 (2008) 331–340.
[61] R. Dustan Sarazan, W.J. Crumb, C.M. Beasley, J.T. Emmick, K.M. Ferguson, C.A. Strnat, P.J. Sausen, Eur. J. Pharmacol. 502 (2004) 163–167.
[62] R.L. Martin, Z. Su, J.T. Limberis, J.D. Palmatier, M.D. Cowart, B.F. Cox, G.A. Gintant, J. Cardiovasc. Pharmacol. 48 (2006) 199–206.
[63] P. Tarantino, N. Appleton, K. Lansdell, Eur. J. Pharmacol. 510 (2005) 75–85.
[64] G.M. Morris, R. Huey, A.J. Olson, Current Protocols in Bioinformatics, Wiley, New York, 2008, Chapter 8, Unit 8.14.
[65] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, J. Comput. Chem. 30 (2009) 2785–2791.
[66] F.J. Solis, R.J.B. Wets, Math. Oper. Res. 6 (1981) 19–30.
[67] C.W. Yap, PaDEL-Descriptor, J. Comput. Chem. 32 (2010) 1466–1474.
[68] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, SIGKDD Explorations 11 (2009) 10–18.
[69] M.A. Hall, Correlation-based Feature Subset Selection for Machine Learning, Hamilton, New Zealand, 1998.
[70] J.G. Topliss, R.P. Edwards, J. Med. Chem. 22 (1979) 1238.
[71] A. Coi, I. Massarelli, M. Saraceno, N. Carli, L. Testai, V. Calderone, A.M. Bianucci, Chem. Biol. Drug Des. 74 (2009) 416–433.
[72] A. Coi, F.L. Fiamingo, O. Livi, V. Calderone, A. Martelli, I. Massarelli, A.M. Bianucci, Bioorg. Med. Chem. 17 (2009) 319–325.
[73] Y. Jiang, A. Lee, J. Chen, M. Cadene, B.T. Chait, R. MacKinnon, Nature 417 (2002) 523–526.
[74] L. Du-Cuny, L. Chen, S. Zhang, J. Chem. Inf. Model. 51 (2011) 2948–2960.
[75] J. Chen, G. Seebohm, M.C. Sanguinetti, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 12461–12466.
[76] L. Testai, A.M. Bianucci, I. Massarelli, M.C. Breschi, E. Martinotti, V. Calderone, Curr. Med. Chem. 11 (2004) 2691–2706.
[77] B.D. Walker, C.B. Singleton, J.A. Bursill, K.R. Wyse, S.M. Valenzuela, M.R. Qiu, S.N. Breit, T.J. Campbell, Br. J. Pharmacol. (1999) 444–450.