JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

An Automated Docking Protocol for hERG Channel Blockers

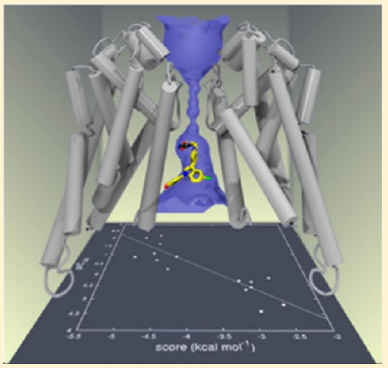
Giovanni Paolo Di Martino,[†] Matteo Masetti,*[,†] Luisa Ceccarini,[†] Andrea Cavalli,[†,‡] and Maurizio Recanatini[†]

[†]Department of Pharmacy and Biotechnology, Alma Mater Studiorum, Università di Bologna, Via Belmeloro 6, 40126 Bologna, Italy

[‡]Department of Drug Discovery and Development, Istituto Italiano di Tecnologia, via Morego 30, 16163 Genova, Italy

Supporting Information

ABSTRACT: A docking protocol aimed at obtaining a consistent qualitative and quantitative picture of binding for a series of hERG channel blockers is presented. To overcome the limitations experienced by standard procedures when docking blockers at hERG binding site, we designed a strategy that explicitly takes into account the conformations of the channel, their possible intrinsic symmetry, and the role played by the configurational entropy of ligands. The protocol was developed on a series of congeneric sertindole derivatives, allowing us to satisfactorily explain the structure−activity relationships for this set of blockers. In addition, we show that the performance of structure-based models relying on multiple-receptor conformations statistically increases when the protein conformations are chosen in such a way as to capture relevant structural features at the binding site. The protocol was then successfully applied to a series of structurally unrelated blockers.



INTRODUCTION

The voltage-gated hERG channel (or Kv11.1) is expressed in several organs and tissues, and in the heart, it is responsible for the rapid component of a delayed potassium current (I_{Kr}), which plays a central role for the repolarization phase of the cardiac action potential.[1] During the past decade, a remarkable number of studies has been focused on this channel, as alterations in its functionality have been associated to a potentially lethal proarrhythmic condition known as long QT syndrome-type 2 (LQTS2).[2] Of particular relevance is the hERG dysfunction caused by an accidental block by drugs.[2] Albeit rare in occurrence, the seriousness of this condition has raised severe concerns about drug safety, and the need for a detailed knowledge of the molecular features at the basis of the channel block has therefore become of primary importance in drug discovery. In this scenario, assessing the blockade activity at the early stages of a drug discovery process is highly desirable, as it would prevent waste of time and resources in the development of compounds potentially carrying a hERG-mediated toxicity.[3]

From a computational point of view, since no crystallographic structures of hERG are presently available, ligand-based approaches have represented the most suited choice to perform predictions. In this respect, a number of quantitative structure−activity relationship (QSAR) and classification models built either using 2D[4−9] or 3D[10−15] descriptors have been developed and utilized in academia and industry. Notwithstanding the predictive power achieved by these models, they suffer from common limitations of ligand-based approaches: 2D descriptors lead to a difficult chemical interpretation, whereas 3D models remarkably depend on the conformations chosen to generate the alignment.

Over the years, a considerable effort has also been spent in the development of structure-based models for hERG blockers. Unlike ligand-based approaches, structure-based models benefit (at least in principle) from a straightforward physicochemical interpretation, while the molecular alignment is an outcome of the docking procedure. As previously mentioned, the major difficulty in undertaking a structure-based modeling approach is due to the lack of a crystal structure, and the low percentage of sequence identity between hERG and available templates, makes homology modeling particularly challenging. To date, both open and closed states of the channel have been generated and utilized for structure-based studies. For instance, Österberg and Boukharta docked a series of sertindole derivatives in an open state model of the channel,[16,17] whereas a more general set of blockers was employed along with a closed hERG conformation by Farid[18] and Coi.[19] Furthermore, either multiple[20] or "hybrid"[21] states have also been conceived to improve the agreement between predictions and experimental data. In the absence of structural information, structure-based models are usually validated by their ability to reproduce the observed trend in binding free energy over the considered set of blockers.[16,17,20] Despite these remarkable efforts, the lack of a consistent binding mode for the most potent blockers entails an incomplete comprehension of the phenomenon under study.

The hERG cavity involved in the interaction with blockers represents an uncommon binding site.[22] The intrinsic symmetry of the channel, the conformations of key aromatic residues, and the importance of unspecific interactions in ligand binding are issues that strongly challenge docking simulations.

Received: July 11, 2012
Published: December 21, 2012

To the best of our knowledge, the impact of these aspects on the quality of the derived structure-based models has not been explicitly addressed yet. For instance, it is customary to perform docking on channel models generated by satisfying the C4 point group symmetry,[18,21] whereas when using the relaxed complex scheme (docking studies carried out on an ensemble of protein configurations sampled by means of molecular dynamics simulations),[23] the symmetry of the channel is naturally broken.[16,17,24]

In this work, we present a docking protocol aimed to address some of the issues reported above. Starting from a well grounded homology model of the open state channel,[25] we built several putative hERG-blocker models employing different protein conformations. The quality of each model was evaluated using as a figure of merit the squared correlation coefficient (r^2) between experimental activities and docking scores obtained over a selected set of compounds. The proposed protocol consists of the following steps (see Figure 1): (1) Extensive conformational sampling of the binding site's

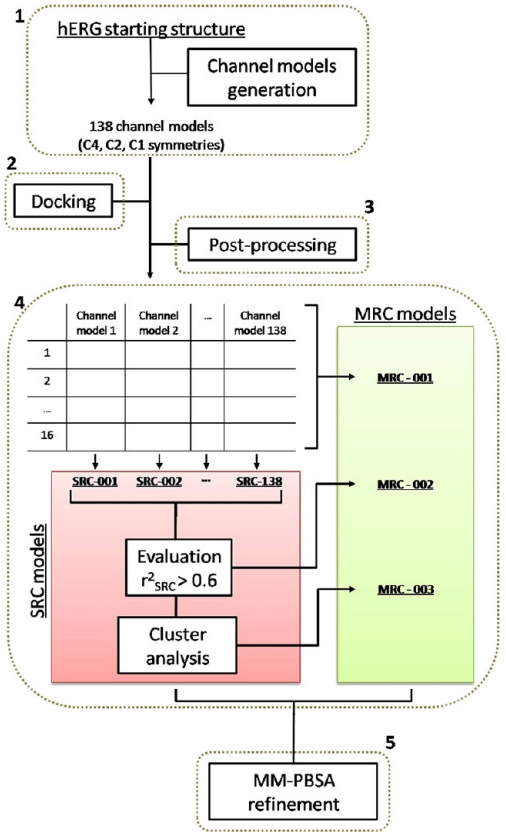


Figure 1. Flowchart describing the entire procedure: (1) channel models generation; (2) automated docking of ligands against the 138 hERG channel models; (3) automated postprocessing; (4) construction of the hERG-blocker models; (5) refinement of the fittest models.

amino acids was performed imposing the C4, C2, or C1 point group symmetries, thus leading to as many families of channels (C4, C2, and C1, respectively). A total number of 138 models passed a filter criterion meant to assess the stereochemical quality of the protein and to prune possible conformational redundancy. (2) A series of congeneric sertindole derivatives (n = 16, series 1 in Table 1), for which experimental blocking activity is available and measured in controlled and consistent conditions,[14] was docked into the channel models as obtained in step 1. (3) The poses obtained in step 2 were postprocessed to resolve redundancy due to protein symmetry (in case of C4 and C2 families of models) and rescored by taking into account their configurational entropy.[26,27] (4) The docking outcome was finally evaluated via both a single-receptor and a multiple-receptor conformation approach (hereafter referred to as SRC and MRC, respectively).[28] Specifically, each model relying upon an SRC description consisted of a single channel model associated to a unique binding solution for every ligand belonging to the set, thus leading to a total number of 138 models (SRC-001 to SRC-138). An MRC model was instead built to account for protein flexibility and induced fit phenomena, and it was obtained using information provided by the whole set of channels (MRC-001). Then, by exploiting the information achieved through a careful analysis of the SRC models, two additional MRC models were obtained (MRC-002 and MRC-003). Finally, (5) the fittest hERG-blocker models (either SRC or MRC) were further refined to better describe solvation effects upon binding via a single-configurational MM-PBSA rescoring scheme performed using a multiple dielectric description.[29,30] Steps 2−4, and their input/output flows, were automatized and generalized to be used with any kind of series of blockers (see text and Figure S1 in the Supporting Information (SI)).

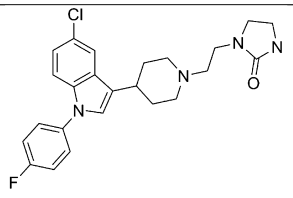
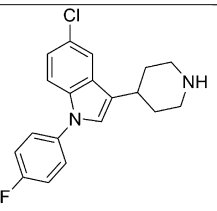
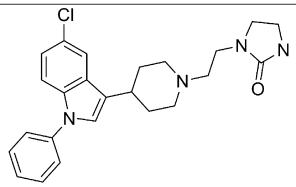
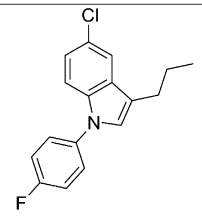
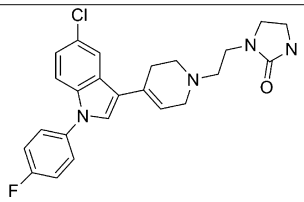
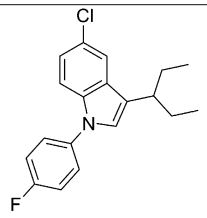
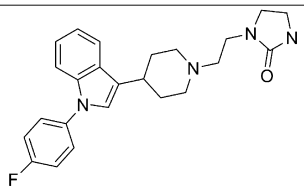
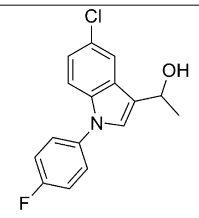
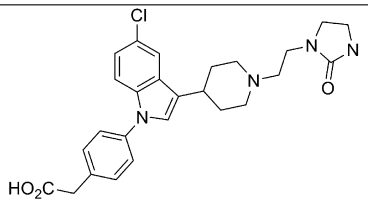
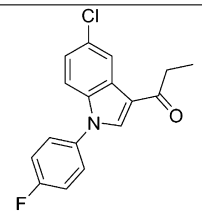
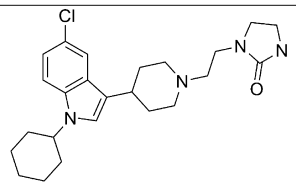
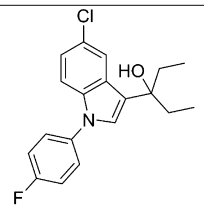
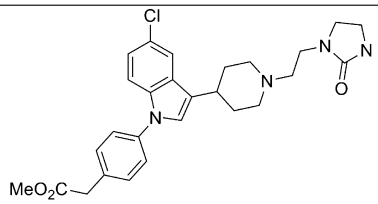
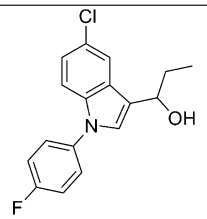
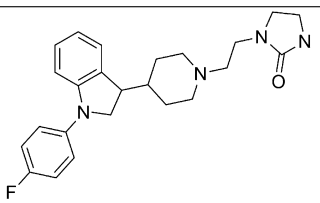
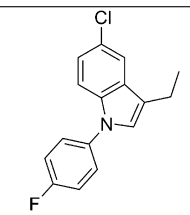
The main result of this work is the obtainment of a strategy to achieve a small set of putative hERG-blocker models to quantitatively relate docking scores with blockers activity in a fully automated way. We show that the use of a limited amount of knowledge-based information to derive a minimal subset of channel models is required to significantly improve the quality of the MRCs compared to a more simplistic SRC description. As a corollary of the effectiveness of the protocol, we demonstrate that the symmetry of the channel conformations has a non-negligible impact on the performance of the structure-based models derived. Moreover, we also show that among the many possible channel models, there is a high percentage of nonvaluable conformations to derive structure−activity relationships. In retrospect, we highlight seven channel conformations as a subset of relevant and structurally diverse candidates to perform MRC-based docking studies efficiently and without relevant loss of information for the sertindole series of analogues. The protocol was then validated using a series of structurally unrelated blockers (series 2 in Table 2).

RESULTS

Development of the Protocol. *1. Channel Models.* To account for flexibility, the local conformational space of the cavity of the open state channel was extensively explored leading to a set of channel models (Figure 2). Under the assumption that the starting configuration represented an optimal geometric assembly of the protein in agreement with experimental data,[25] only the side chains of key amino acids (see Methods) were relaxed. In particular, we focused on the conformations explored by Tyr652 and Phe656, since their importance in ligand binding has largely been discussed in the literature.[31] The χ_1/χ_2 plot for these residues before and after applying the stereochemical quality filter (shown in Figure S2A and B, respectively, in the SI) makes us confident that, taking into account all the models, these amino acids sampled the most of their conformational space. Notably, the χ_1/χ_2 plots are in good agreement with those recently obtained by Knape and co-workers and derived from 50 ns of molecular dynamics

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159−175

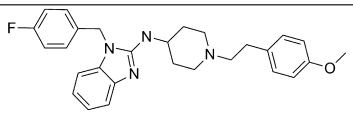
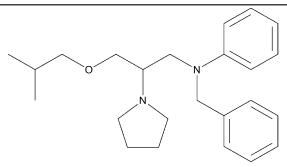
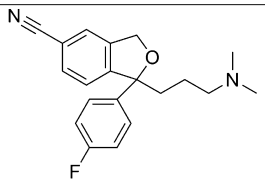
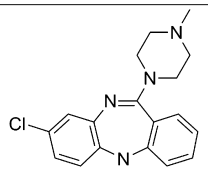
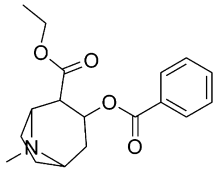
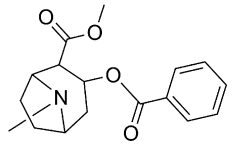
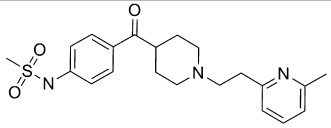
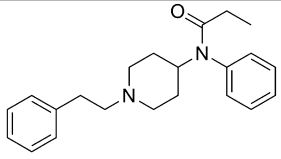
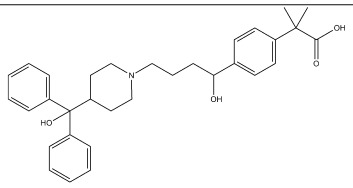
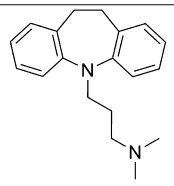
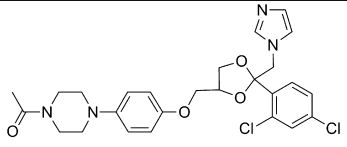
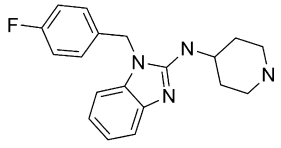
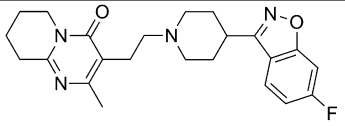
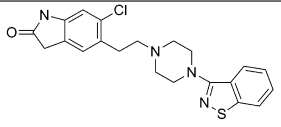
Table 1. Set of Compounds Used in the Development of the Docking Protocol (Series 1)[a]

Compd	Structure	IC_{50} (nM)	Compd	Structure	IC_{50} (nM)
Sertindole		3	**14**		204
1		88	**16**		26,000
2		10	**17**		1480
3		7	**18**		4550
4		579	**19**		1947
6		137	**20**		15,700
7		131	**21**		2200
13		23.5	**22**		3500

[a]Sertindole derivatives and their experimental IC_{50} (nM). The ligand numbering was adopted according to the work of Pearlstein.[14]

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159−175

Table 2. Set of Structurally Unrelated Blockers Used in the Validation of the Docking Protocol (Series 2)[a]

Compd	Structure	IC_{50} (nM)	Compd	Structure	IC_{50} (nM)
Astemizole		1	Bepridil		501
Citalopram		3981	Clozapine		199
Cocaethylene		1259	Cocaine		7943
E-4031		8	Fentanyl		1995
Fexofenadine		19,953	Imipramine		3162
Ketoconazole		1585	Norastemizole		25
Risperidone		158	Ziprasidone		158

[a]The structurally unrelated compounds and their experimental IC_{50} (nM).[38]

simulation.[32] In Figure 3A, the shape of the cavity measured as maximum pore radius plotted against channel axis (z) is reported for all the channel models. It clearly emerged that most of the variability in shape was in proximity of the phenyl ring of Phe656. From a docking standpoint, the four Phe656 residues could constitute a potential pore restriction that might critically affect the generation of reliable hERG-blocker models. Indeed, in this region of the pore, in some cases we observed a maximum radius as small as about 1 Å, underlying protein conformations virtually unproductive to accommodate the majority of the larger blockers. For sake of clarity, we refer to these conformations as "narrow channel models" to be distinguished by the "wide channel models", bearing in this region a maximum radius greater than 3 Å. Conversely, Tyr652 protruded much less toward the cavity, and the sampled conformational space was significantly more limited (see SI Figure S2B).

In order to classify the channel conformations and to prune their possible redundancy, a cluster analysis was performed. Because of the channel overall symmetry and the internal symmetry of amino acids' side chains such as phenylalanine and tyrosine, the RMSD metric appeared to be particularly ill suited to address the problem. We therefore used as a measure of the dissimilarity among channel models a metric describing the local difference in shape calculated over the pore axis (see Methods for details), although we acknowledge that a cluster analysis in the dihedral space might also be successfully applied. The proposed metric turned out to be quite effective for our

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159−175

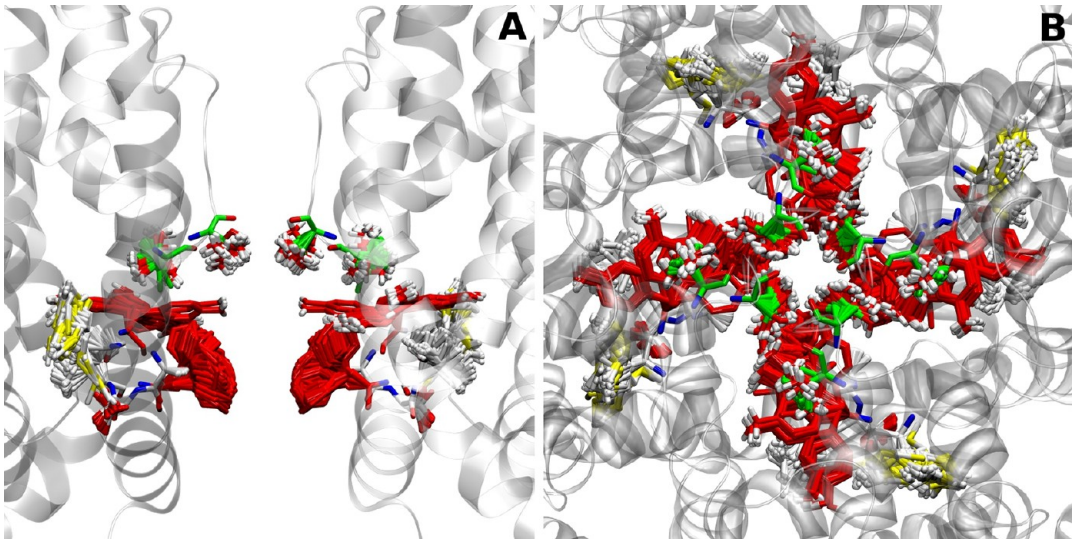


Figure 2. (A) Side view (only two out of four subunits are shown for clarity) and (B) top view of the 296 channel models generated. The conformational search was performed on the side chains of amino acids of the inner cavity: Thr623, Ser624 shown in green; Tyr652, Phe656 shown in red; and Met651, Ala653, Ser654, Ile655 shown in yellow.

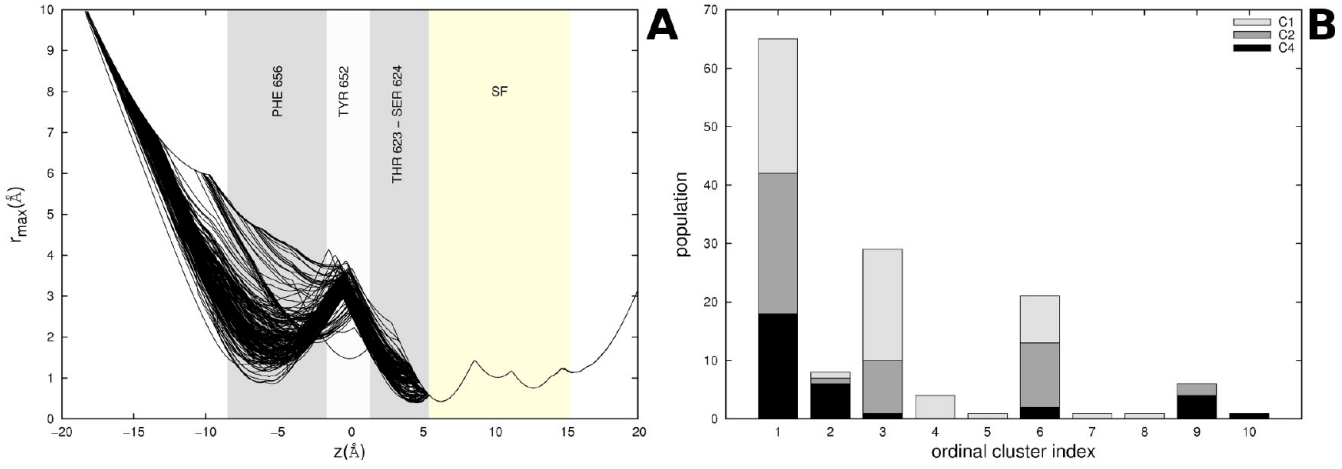


Figure 3. (A) Shape profile of the pore of the 138 channel models. The important areas of the pore are shown as grayish rectangles (the Phe656, Tyr652, and Thr623−Ser624 rings). For clarity, the region spanned by the selectivity filter is also shown (in beige). (B) Population of the clusters represented in terms of the different symmetry families of the channel models.

purposes and provided an intuitive picture of the classification of channel models into clusters. In Figure 3B, the population of the 10 clusters along with their internal population in terms of symmetry families is reported. As it can be seen, the symmetry of the channel models was transversally parted on the clusters. In other words, there did not existed a direct relationship between the symmetry family and a particular shape of the cavity, implying that a similar shape could be obtained by imposing different symmetries during the generation of the channel models (see Discussion).

2. SRC Models. The performance of the hERG-blocker models was evaluated by calculating the correlation (in terms of r^2) between the docking scores and the experimental blocking activities. The r^2 values for all the SRC models along with the information concerning the symmetry family of the corresponding channel model are reported in SI Table S1. As explained in the Methods section, r^2 values were calculated employing the Autodock[33] score (AD score) after applying a pose rescoring, which took into account the ligands' configurational entropy. This approach was based on the assumption that a better description of binding would be obtained by reweighting the score associated to a given pose by its relative probability to be achieved during extensive docking simulations (i.e., the so-called Colony Energy, CE).[27] Within the CE formalism, the weight assigned to a specific binding mode statistically incorporates (in an approximate way) the contribution of the ligand configurational entropy in a rescoring scheme-like fashion.[27]

In SI Figure S3, the performance of the SRC models is reported for each symmetry family (C1, C2, and C4 in panel A, B, and C, respectively). In the same figure, we also compare the r^2 values calculated before (i.e., AD solutions) and after applying the CE rescoring. The rescoring procedure either slightly or significantly decreased the fitness of the SRC model. However, for a limited number of channel conformations, an interesting performance increasing was recorded. By classifying the fitness of the hERG-blocker models as bad ($r^2 < 0.4$), intermediate ($0.4 < r^2 < 0.6$), and good ($r^2 > 0.6$), the performances of the SRCs were reported as histograms for each symmetry family. Figure 4 shows that C1 channel models

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159−175

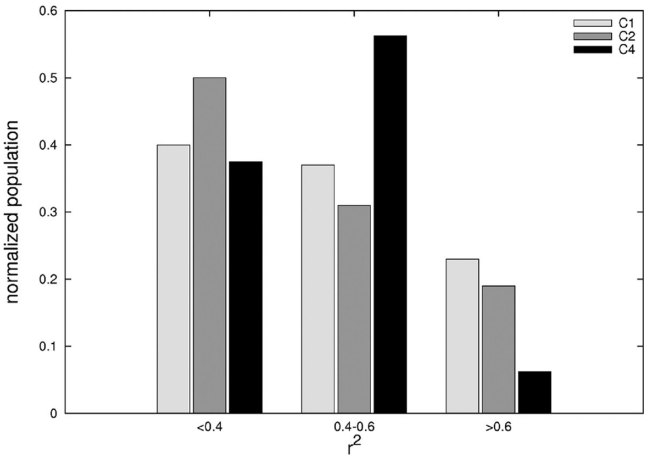


Figure 4. Performance evaluation of the SRC models for series 1 based on the squared correlation coefficient and reported in terms of the symmetry family displayed by the corresponding channel model.

performed on average better than those belonging to other families. However, the best SRC models of the entire ensemble belonged to the C2 family (SRC-001 and SRC-002, r^2 of 0.83 and 0.79, respectively, see SI Figure S3B), even though, at the same time, this family owned the largest amount of poorly performing models. Generally, the C4 family of models performed much worse.

Since no clear-cut connections between symmetry and cavity shape, or symmetry and performance of the derived SRC models were obtained, we investigated whether any relationships between shape and fitness could be derived. In Figure 5A, the clusters of the channel models along with the quality of the corresponding SRCs are presented. Apart from the trivial singleton clusters, a certain correspondence could be inferred. From a hERG-blocker model standpoint, clusters 2, 3, and 4 clearly underlined effective channel model shapes, whereas cluster 6 definitely did not. The shape profiles of the channel conformations associated to the fittest SRC models are shown in Figure 5B for the nonsingleton clusters. The channel model representative of cluster 6 exhibited an extremely narrow cavity conformation, thus explaining the poor performances of SRC models within this cluster. Indeed, the restriction provided by the Phe656 ring hampered the access of ligands to the upper portion of the binding site, thus preventing a proper binding and in turn an efficient channel block. We would like to comment here that such results could be an indirect confirmation of the health of the protocol. In fact, because the AD docking box was generously extended toward the intracellular side of the protein, ligands were allowed to bind even in the presence of the restriction. Therefore, the fact that the narrowest channels systematically provided much worse SRC models might be interpreted as a signal of a correct quantitative response of our protocol. Notably, while narrow channels always provided poor SRC models, the opposite did not hold true. In fact, it was possible to identify moderately wide channel models providing a poor correlation, even though in general they rather returned average or good SRC models.

3. MRC Models. Even though analyzing the performance of SRC models is an informative exercise, it is clear that the most relevant picture of binding should be achieved through a correlative hERG-blocker model based on an MRC description. The first MRC model (MRC-001), obtained using the information coming from all the docking simulations in a standard ensemble docking approach, turned out to be rather poor. In fact, either using the best scores and the average scores data treatment, an unsatisfactory correlation was obtained (r^2 of 0.57 and 0.53, respectively; see Table 3). Only when using the Boltzmann weighted averages an acceptable correlation was reached (r^2 = 0.65). Although disappointing, these results were not completely unexpected, as it is well-known that using a large number of conformations might have a detrimental effect on the performances of the ensemble docking approach.[34] The second MRC model generated (MRC-002) was built employing only the information coming from the docking simulations performed on a worthwhile subset of channel models yielding the best correlations between calculated and experimental data in the SRC description (r^2 > 0.60). The rationale of this procedure was based on the assumption that protein conformations responsible for the fittest SRC models, on average, would carry more relevant information to describe the binding of blockers than the others. In other words, the identification of a minimal subset of channel models would statistically improve the performance of the MRC description by reducing the noise possibly related to scoring function inaccuracies or limitations in sampling. In this respect, our approach is similar in spirit (although slightly less rigorous but

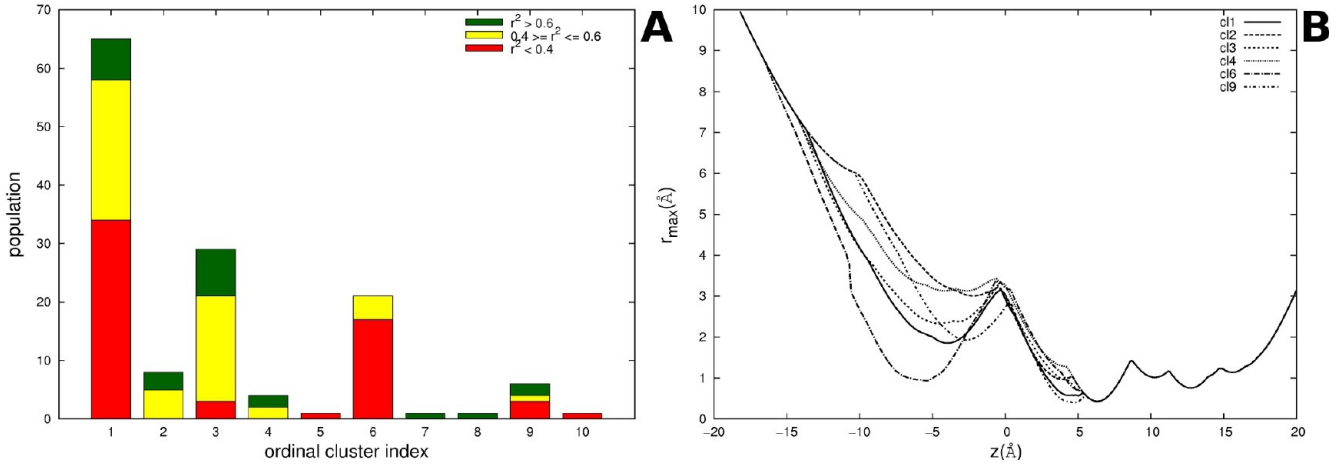


Figure 5. (A) Population of the clusters represented in terms of the associated SRCs' fitness calculated for series 1. (B) Shape profiles of the best SRC models for each cluster. Singleton clusters are not shown.

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159−175

Table 3. Fittest Correlative Structure-Based Models for Series 1[a]

				MRC-001			MRC-002			MRC-003		
compd	SRC-001	SRC-002	SRC-008	best	avg	bz	best	avg	bz	best	avg	bz
Sertindole	−5.07	−5.33	−5.41	−5.41	−3.56	−4.60	−5.41	−4.66	−4.93	−5.41	−4.93	−5.11
1	−4.42	−5.23	−4.85	−5.23	−3.47	−4.48	−5.23	−4.43	−4.76	−5.23	−4.68	−4.89
2	−4.79	−5.23	−5.31	−5.34	−3.58	−4.67	−5.31	−4.67	−4.95	−5.31	−4.91	−5.05
3	−4.81	−4.92	−4.84	−5.05	−3.30	−4.39	−5.01	−4.46	−4.64	−4.92	−4.62	−4.72
4	−4.02	−4.36	−4.36	−5.93	−3.91	−4.80	−5.55	−4.38	−4.75	−4.97	−4.38	−4.54
6	−4.82	−5.43	−5.44	−5.73	−3.86	−4.88	−5.59	−4.71	−5.09	−5.44	−5.03	−5.22
7	−4.29	−4.35	−4.50	−5.21	−3.19	−4.15	−5.21	−4.03	−4.47	−4.80	−4.24	−4.43
13	−4.49	−4.86	−4.06	−5.38	−3.83	−4.60	−5.26	−4.47	−4.78	−4.94	−4.60	−4.73
14	−4.50	−4.76	−4.98	−5.63	−3.78	−4.79	−5.10	−4.40	−4.63	−4.98	−4.60	−4.68
16	−2.71	−2.83	−2.69	−3.71	−2.61	−2.99	−3.34	−2.91	−3.00	−3.28	−2.91	−2.98
17	−2.51	−2.67	−2.65	−3.29	−2.14	−2.68	−3.10	−2.62	−2.78	−2.76	−2.64	−2.65
18	−3.31	−2.97	−2.89	−3.63	−2.72	−3.01	−3.47	−2.99	−3.14	−3.42	−3.07	−3.16
19	−2.99	−3.41	−2.85	−3.83	−2.74	−3.24	−3.68	−3.18	−3.28	−3.43	−3.15	−3.23
20	−2.28	−2.62	−2.24	−3.66	−2.08	−2.81	−3.35	−2.58	−2.88	−2.81	−2.55	−2.62
21	−3.35	−2.84	−2.39	−3.68	−2.48	−2.99	−3.57	−2.95	−3.16	−3.35	−2.95	−3.12
22	−2.99	−2.91	−2.70	−3.58	−2.63	−2.96	−3.24	−2.92	−3.00	−3.22	−2.97	−3.02
r^2	**0.83**	**0.79**	**0.74**	**0.57**	**0.53**	**0.65**	**0.67**	**0.76**	**0.73**	**0.73**	**0.77**	**0.76**

[a]Docking score (kcal mol^{-1}) of the 16 compounds used to develop the protocol for the three fittest SRC (SRC-001, SRC-002, and SRC-008) and MRC models. For the MRC models, the results obtained using the three data treatment methods employed are also shown (best scores, arithmetic mean, and Boltzmann-weighted average). The performance of each model measured in terms of r^2 is reported at the bottom of the table.

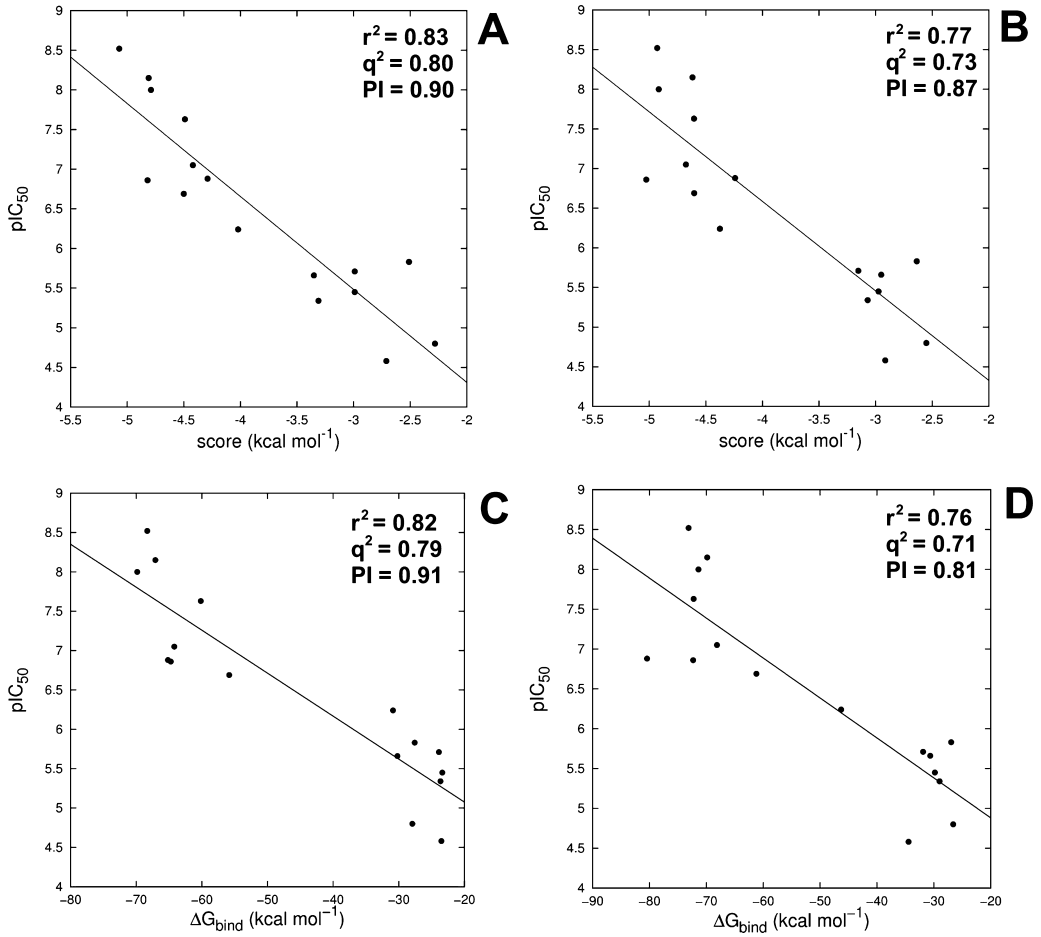


Figure 6. Correlation plots between the experimental pIC_{50} and the calculated binding energies for series 1 before (upper panels) and after (lower panels) applying the MM-PBSA refinement. Panels A and C refer to SRC-001, whereas panels B and D show the results for MRC-003.

much simpler) to the method proposed by Yoon and Welsh.[35] Accordingly, the MRC-002 statistics significantly improved (see Table 3). Notably, only 24 channel models out of 138 were used in this description. In pursuing the definition of the minimal subset of channel conformations, MRC-003 was built using only the channel models belonging to the previous subset

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159−175

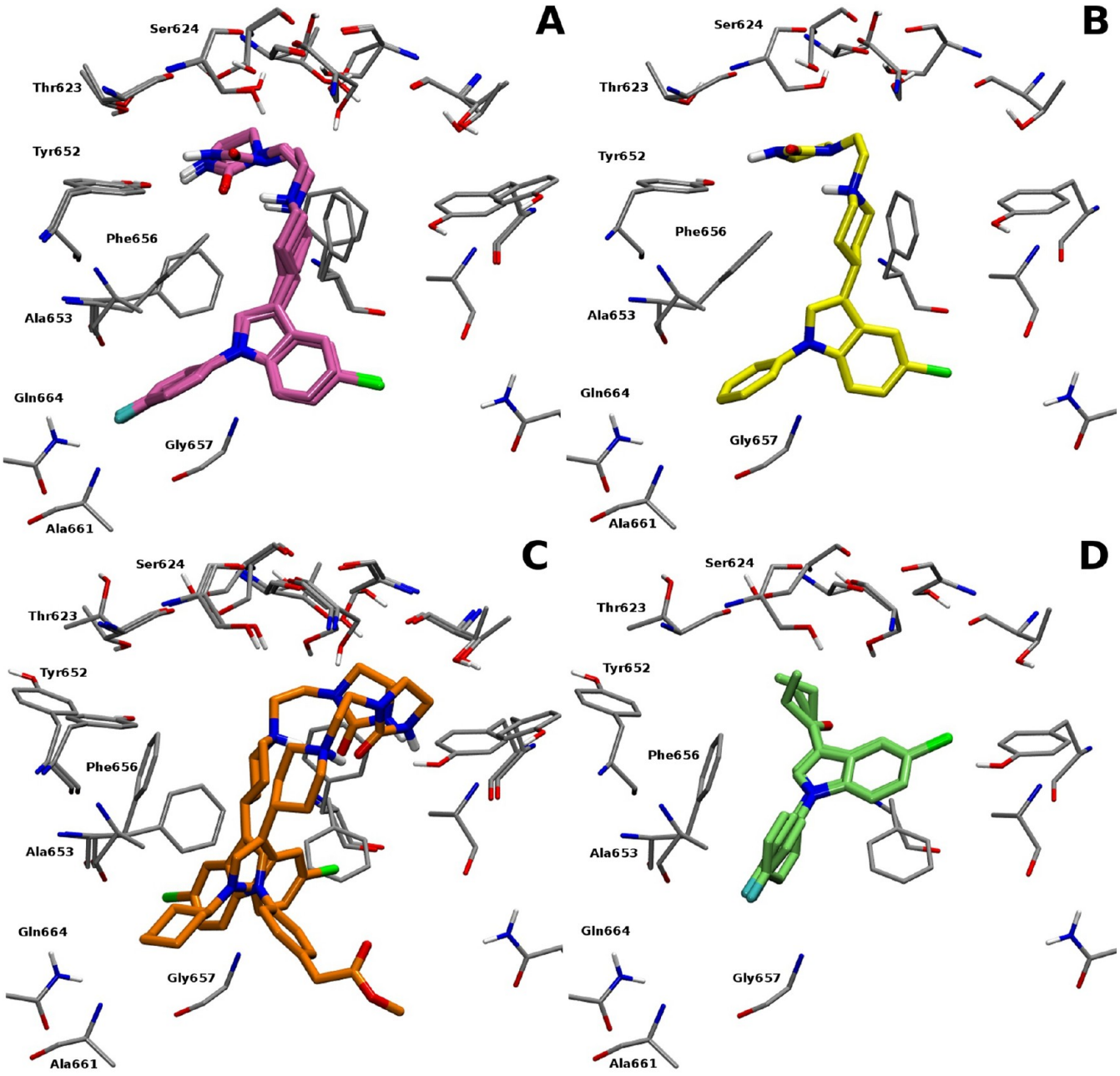


Figure 7. Most relevant binding modes displayed by MRC-003 in series 1: (A) sertindole, **2**, and **3** (mauve); (B) compound **1** (yellow); (C) compounds **6** and **7** (orange); and (D) compounds **16**, **17**, **19**, and **22** (lime). The multiple conformations adopted by the protein are shown in gray.

leading at the same time to the fittest SRCs along different clusters. Compared to MRC-002, MRC-003 displayed better statistics. For example, for the best scores data treatment, the r^2 increased from 0.67 to 0.73 (see Table 3). This result demonstrates that a good MRC model for the hERG channel can be successfully obtained with an extremely limited subset of channel conformations. In this respect, the analysis of SRC models turned out to be instrumental in defining a strategy to select the MRC channel conformations.

Examining the methods used for the data treatment, averaging docking scores across the channel conformations turned out to provide better statistical performances rather than using exclusively the best energy scores. However, no substantial differences were noticed whether a simple arithmetic mean or a Boltzmann-weighted mean were used.

In the best case scenario, MRCs would be the only hERG-blocker models to be considered, since their derivation is solely driven by the physics of the scoring function, while SRCs are obviously conditional models.

4. Structure–Activity Relationships. Here, we present the binding modes adopted by the sertindole analogues in the fittest structure-based models: SRC-001 (r^2 = 0.83, see Figure 6A) and MRC-003 (r^2 = 0.77, adopting the arithmetic mean data treatment method, Figure 6B). Despite the slightly better performance of SRC-001, a more consistent binding mode was achieved using the MRC approach.

In MRC-003, all the compounds displayed a binding mode similar to that proposed by Österberg and Åqvist,[16] by fitting into the cavity parallel to the channel axis, adopting an extended conformation, and pointing the imidazolidinone

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159−175

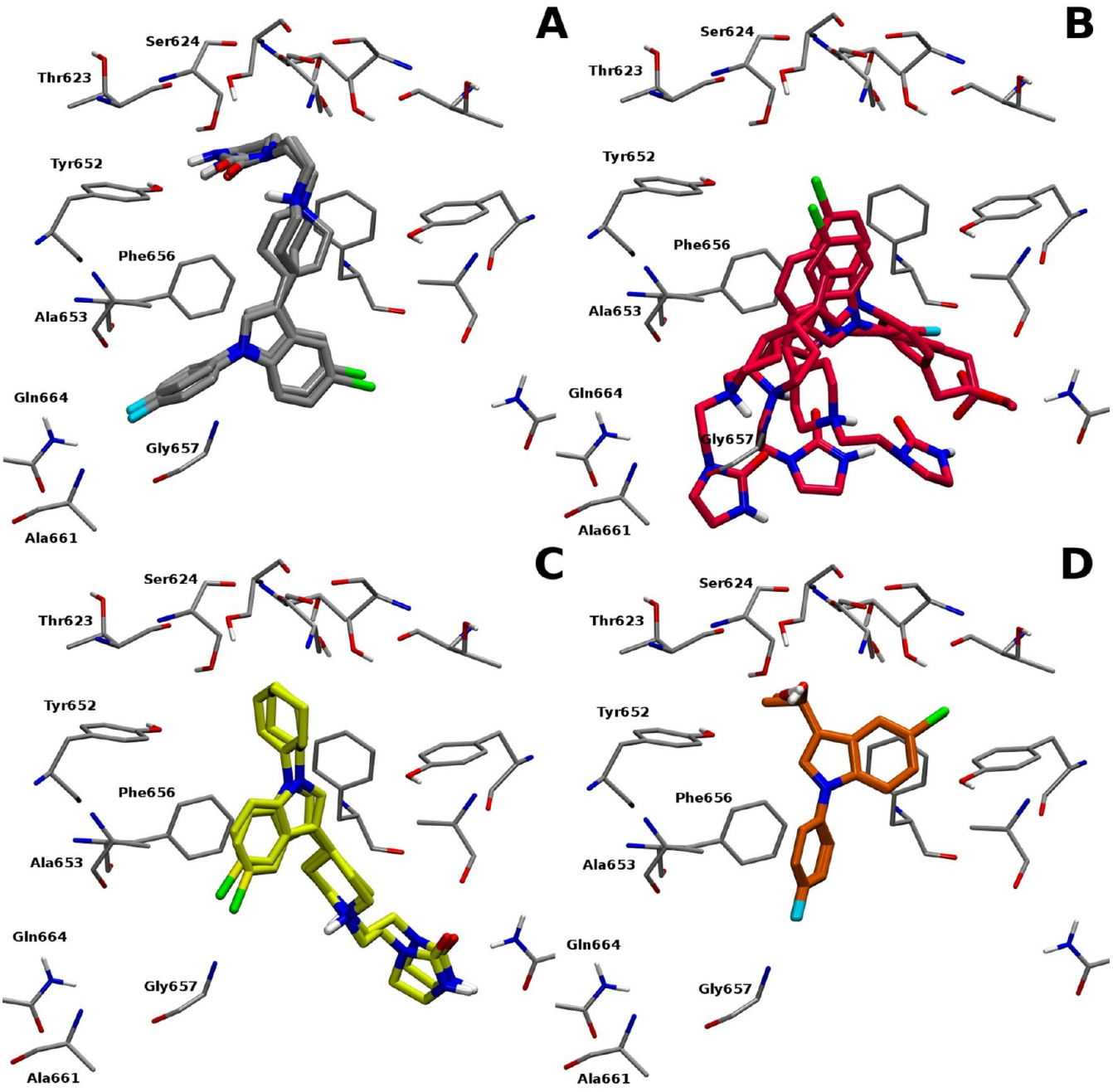


Figure 8. Most relevant binding modes displayed by SRC-001 in series 1: (A) sertindole, **2**, and **3** (silver); (B) compounds **13**, **7**, and **4** (magenta); (C) compounds **1** and **6** (yellow); and (D) compounds **18** and **21** (orange).

moiety toward the selectivity filter. The basic nitrogen was found to occupy a central position in the area surrounded by Phe656−Tyr652 and, in agreement with other studies,[16,17] no cation-π interactions with Tyr652 were observed. The most potent blockers of the set of compounds, sertindole (IC_{50} = 3 nM), **3** (IC_{50} = 7 nM), and **2** (IC_{50} = 10 nM) showed similar binding modes though different channel models were chosen (channel model C1_43 for sertindole and **3** and C2_48 for **2**; Figure 7A). In particular, the imidazolidinone group was placed at the bottom of the filter. The amidic nitrogen of this group was involved in an H-bond interaction with the hydroxyl group of Tyr652, whereas the aliphatic cyclic moiety showed hybrophobic interactions with the adjacent Tyr652−Phe656 and with the backbone of Thr623−Ser624. In addition, the indole moiety formed hydrophobic or π-stacking interactions with at least three neighboring Phe656, while the fluorobenzyl group took electrostatic contacts with the side chain of Gln664 and hydrophobic interactions with Phe656 and Ala653. In the resulting binding modes, Gln664 together with Tyr652 seemed to play a key role in determining the blocking affinity. This was evident by the poses adopted by compounds **13** (IC_{50} = 23.5 nM, SI Figure S4A) and **1** (IC_{50} = 88 nM, Figure 7B), where the absence of the H-bond either with Gln664 (in the case of **1**, lacking the fluorine) or with Tyr652 (in the case of **13**) reflected the reduction of affinity. The presence of a methyl phenylacetate instead of the fluorobenzyl group, together with the absence of polar contact with Tyr652 in **7** (IC_{50} = 131 nM) or the total substitution of the halo-aromatic ring with a

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159−175

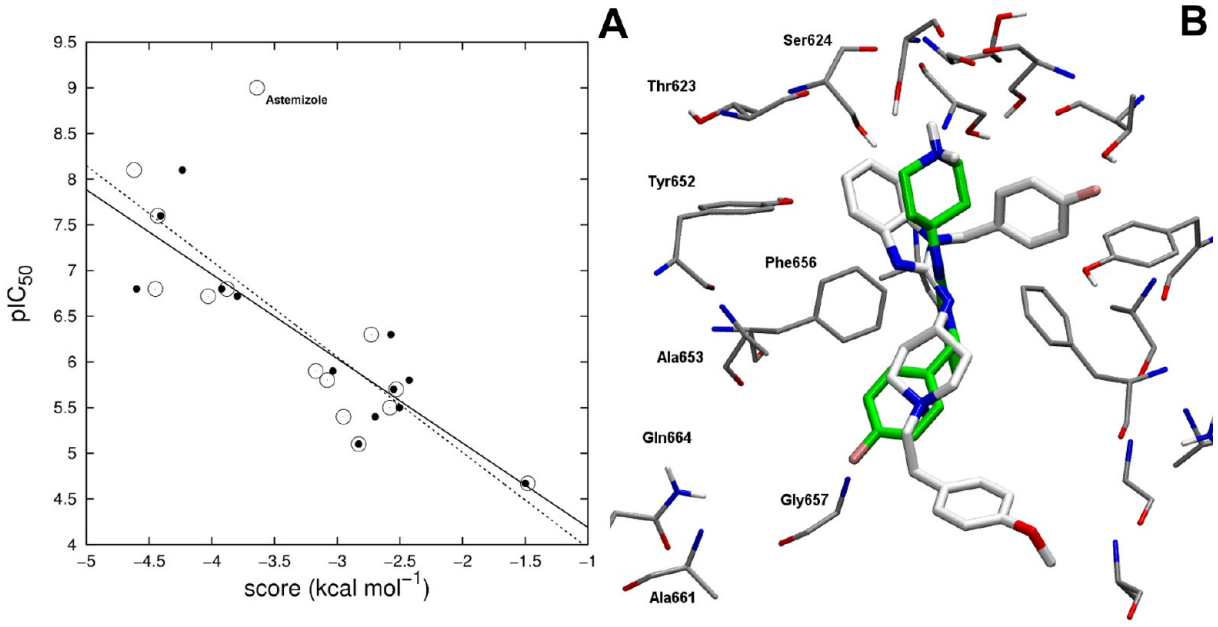


Figure 9. (A) Correlation plots between the experimental pIC_{50} and the calculated binding energies for series 2 before (n = 14, empty circles, dotted line) and after (n = 13, filled circles, solid line) the removal of the outlier astemizole. (B) Representative binding modes for selected blockers: astemizole (white) and norastemizole (green).

cyclohexyl moiety in **6** (IC_{50} = 137 nM), were found to be further detrimental for the blocking activity (Figure 7C). Although the H-bond with Gln664 was preserved, **4** (IC_{50} = 579 nM) did not interact with Tyr652 and fewer hydrophobic contacts between the indole group and Phe656 were observed (SI Figure S4B). Concerning the less potent binders, which were also the smaller molecules of the set (**14**–**22**), the indole moiety replaced the piperazine group in the binding site, occupying the accessible volume delimited by the four Phe656. The fluorobenzyl portion of the ligands was mainly located in the lower side of the cavity in between Ala653 and two copies of adjacent Phe656 (Figure 7D). An exception is represented by **18** (IC_{50} = 4550 nM), which displayed the fluorobenzyl toward the selectivity filter and was involved in nonpolar contacts with two neighboring Thr623, and in a π–π T-shaped interaction with one copy of Phe656 (SI Figure S4C). Compound **14** (IC_{50} = 204 nM) was the most potent blocker among the smaller compounds, and the only one showing a positively charged nitrogen sequestered by the polar area consisting in Thr623 and Ser624 at the bottom of the filter. The alkyl portions were mainly located in the hydrophobic pocket formed by two adjacent subunits consisting in the residues Thr623, Ser624, Tyr652 and Phe656, while the presence of the hydroxyl functional groups, in **20** (IC_{50} = 15,700 nM) and **21** (IC_{50} = 2200 nM), resulted in additional polar contacts with Ser624 (SI Figure S4D).

As previously mentioned, the binding picture derived from the SRC-001 model showed less consistency when compared to MRC-003. In particular, three different binding modes could be observed among diverse group of molecules. In analogy with the MRC-003 representation, the highest affinity compounds, sertindole, **3**, and **2**, retained an Österberg-like binding mode[16] with the imidazolidinone group pointed toward the filter, and the fluorobenzyl moiety connected to Gln664 through an H-bond interaction (Figure 8A). Together with a progressive reduction of affinity, different binding modes emerged. Similarly to the binding mode observed by Boukharta and co-workers,[17] compounds **13**, **7**, and **4**, placed the indole moiety in the central area of the cavity surrounded by Phe656, which was also involved (in the case of **7** and **4**) in π-stacking interactions with the chlorine oriented toward the selectivity filter (Figure 8B). Furthermore, the fluorobenzyl, the phenylacetate, or the methyl phenylacetate groups in compounds **13**, **4**, and **7**, respectively, were accommodated in the area beneath the Phe656, taking favorable polar interactions with Gln664. However, unlike the binding mode described by Boukharta, the imidazolidinone moiety of these compounds was not placed close to the bottom of the filter, but rather it contacted the C_α of Gly657 and the side chain of Ala661 via multiple hydrophobic interactions. A different scenario was provided by compounds **1** and **6**, which showed a binding mode overall in agreement with the one commented by Stansfeld and co-workers.[21] As shown in Figure 8C, these compounds adopted an extended conformation, parallel to the channel axis, and pointing the phenyl group (**1**) or the cyclohexyl ring (**6**) toward the bottom of the filter, while the imidazolidinone interacted via polar contacts with the Gln664 side chain. The indole moiety of these compounds was sequestered in the central core of the cavity by π–π interactions involved with the four Phe656. In contrast with the MRC-003 description, the binding modes adopted by smaller compounds showed less consistency. For most of the molecules (compounds **16**, **17**, **19**, **20**, and **22**), the alkyl portions were located in the hydrophobic pocket defined by two copies of adjacent Phe656 and Ala653, in the opposite side with respect to the one occupied by the fluorophenyl group (SI Figure S5A). The additional presence of a hydroxyl functional group (**18** and **21**, shown in Figure 8D) or a protonated nitrogen (**14**, SI Figure S5B) was responsible to drive the compounds toward the polar area at the bottom of the filter.

5. Assessment of the Solvation Free Energy Contribution. To further confirm the reliability of the fittest hERG-blocker models, an additional force-field based rescoring scheme was employed. In SI Table S2, we reported the estimated MM-

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159–175

PBSA free energies of binding (kcal mol^{-1}) for the docked compounds in the fittest structure-based models: SRC-001 and MRC-003. For each model, the molecular mechanics (ΔG_{MM}) and both the polar (ΔG_{PB}) and nonpolar (ΔG_{SA}) contributions to the solvation free energy are also reported (see Methods for details). As it can be seen, the favorable formation of the complexes was driven by the molecular-mechanics energy ΔG_{MM} and by the nonpolar component of the solvation energy ΔG_{SA}, which assumed negative values, while ΔG_{PB} component showed positive values. Interestingly, while ΔG_{SA} contributions appeared always small and similar among the compounds (ranging from −2.79 to −4.62 kcal mol^{-1} for SRC-001 and from −2.71 to −5.00 kcal mol^{-1} for MRC-003), although somehow correlated with molecular size, ΔG_{MM} and ΔG_{PB} values were clearly different through the set and were directly dependent on the net charge of the compounds: higher contributions were observed for the positive charged sertindole and compounds **1**, **2**, **3**, **6**, 7, **13**, and **14**. In Figure 6C and D, the regression plots of the ΔG_{bind} versus pIC_{50} for SRC-001 and MRC-003 are respectively shown. In general, we observe that the MM-PBSA method did not significantly influence the scenario depicted by the docking scoring function. In particular, the r^2 changed from 0.83 (s = 0.51) to 0.82 (s = 0.52) in SRC-001, while it passed from 0.77 (s = 0.59) to 0.76 (s = 0.60) in MRC-003 before and after refinement. Concerning the predictive statistical parameters, while the q^2 slightly decreased, the predictive index (PI)[36] retained satisfactory values (larger than 0.80).

Validation of the Protocol. In order to show the general applicability of the proposed docking strategy, the above-reported protocol was applied to a series of structurally unrelated blockers (series 2). The compounds were those of the test set employed by Cavalli et al. for the validation of a 3D-QSAR model.[10] To achieve a uniform distribution of experimental activities, other classical hERG blockers were also taken into account, leading to a validation set of 14 structurally unrelated compounds (see Table 2). Series 2 covered a range of more than four pIC_{50} units of activity (from 4.67 to 9.00). The correlation between experimental activity and predicted docking score for the fittest SRC model turned out to be fairly good (r^2 = 0.60; see Figure 9A). Figure 9A clearly showed that the major outlier was astemizole, and when excluded from the set, very good correlations were obtained. In particular, the fittest SRC displayed an r^2 of 0.83, while the corresponding MRC-003 showed an r^2 of 0.77 (for the average score data treatment, see Figure 9A and SI Table S3). In Figure 9B, the binding mode of some selected compounds belonging to series 2 is reported.

DISCUSSION

In this work, we have presented a docking protocol aimed to predict the putative binding mode for a series of hERG blockers. The protocol was first developed using a series of sertindole analogues that has become a standard reference for structure-based models of hERG block.[16,17] Then, the same strategy was applied to a more challenging set of structurally unrelated molecules. The protocol consisted in the automatic generation of an ensemble of binding modes for the considered set of compounds and in their evaluation according to the r^2 calculated between docking scores and experimental blocking activities. Equivalently, the proposed procedure generated and compared an ensemble of possible correlative models to quantify the hERG blocking activity. These models were obtained using both SRC and MRC descriptions. In the latter, different conformations of the pore amino acid side chains were taken into account. Specifically, the statistical parameters of the MRC models were improved by using the information obtained by an in depth analysis of the relevant SRCs.

Pore Shape of the hERG Channel Models. As previously stated, the features and the statistical performance of the MRC models were strongly dependent upon both the results achieved by the SRCs and the analysis strategy undertaken. In particular, the employed description of the shape of the cavity turned out to be a simple but successful approach to address the problem of an uncommon binding site such as that of the hERG channel. In addition, we were able to point to a relationship between the shape of the pore and the fitness of the SRC model derived from series 1. While a posteriori it was not surprising that narrow channel models always led to poor SRCs, it was indeed interesting to note that the fitness of the hERG-blocker models did not monotonically improve with the pore width at the constriction region. Despite the importance of the punctual arrangement of amino acids in determining a certain binding mode, and the fact that a given shape could be obtained with different side chain conformations, it is tempting to advance an explanation of the general performance of the SRC models in terms of the shape of the binding site. In Figure 5B, we show that the best SRCs for series 1 were generated basing on wide channel models, even though the widest one (C4_25, belonging to cluster 2; see SI Table S1) displayed a slightly worse performance than the others (r^2 = 0.52). We could speculate that an optimal SRC model would be obtained with a moderate amount of restriction in the channel pore. A possible explanation could be that a right compromise between wide and narrow channel volume and shape had to be obtained to allow the present series of compounds to properly dock into hERG. Indeed, the Phe656 ring had to achieve an optimal ensemble of conformations to interact productively with the most potent ligands. Concerning this point, it is worth commenting on the overwhelming population of cluster 1 and the wide range of r^2 displayed by the related SRC models. Cluster 1 entailed a quite broad range of conformations of the channel, whose shape could be crudely described as encompassing both narrow and slightly open pore conformations, showing a maximum radius in the restriction region lower and greater than 2 Å, respectively. Notably, the maximum radius never exceeded the value of 2.2 Å in this cluster. For most of the channel models belonging to cluster 1, the Phe656 restriction prevented the achievement of effective SRC models. However, because of the rearrangement of these amino acids, whenever the pore radius exceeded a critical threshold (of about 2 Å), highly productive SRC models became within reach. These considerations hold for both the series of compounds, and as such, they can be considered of general significance. Clearly, taking into account Phe656 conformations in binding is a necessary but not sufficient condition to explain at best the experimental blocking activities, as the conformations of several other residues comes into play at a molecular level.

As already pointed out in the Results section, no straightforward relationships between the shape of the pore and the symmetry family were identified. In this respect, an interesting exception was provided by cluster 2. As previously discussed, cluster 2 represented the ensemble of widest channel models, and it was mostly constituted by C4 models. Interestingly, the channel conformation leading to the fittest

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159−175

SRC within this cluster belonged to the C1 point group symmetry for both the series of compounds (C1_20, see SI Table S1), demonstrating that the shape, rather than the symmetry used in the channel model generation, is relevant to build effective structure-based models based on an SRC description. Indeed channel model C1_20 virtually displayed C4 symmetry, whereas no symmetry restraints were imposed in its derivation. In this context, the fact that C4 channels apparently performed slightly worse than those belonging to the C2 and C1 class (see Figure 4 for series 1) should not be overestimated. Indeed, the worse performance displayed by C4 channel models is only apparent and is due to the fact that, because of the high number of restraints that must be satisfied in their derivation, there is a greater chance to incur either in narrower or wider pore shapes compared to other symmetries, which were generally associated to less fit SRCs.

Binding Modes of the Sertindole Analogues. From the perspective of the development of the protocol, the MRC-003 model represents the main achievement of this work, as it provided at the same time a consistent binding mode for structurally similar blockers and satisfying structure−activity relationships. Even though the strategy employed to reach this description of binding was not free from a certain degree of knowledge-based subjectivity,[35] it must be stressed that the resulting model was obtained in a completely unbiased way. The fact that the performance of some SRC models turned out to be comparable (and in few cases better) to MRC-003 could be interpreted as an evidence of the existence of alternate or multiple binding modes for hERG channel blockers. Although such a perspective is undoubtedly attractive, we think it is too speculative to be embraced without a reasonable skepticism. Indeed, further studies relying on free energy calculations should be carried out to properly address such a hypothesis.

In line with the survey carried out by Zachariae and co-workers,[37] it is interesting to compare the binding modes obtained in our fittest structure-based model with those previously presented in the literature and attempt to relate them with the available pharmacophores for the hERG blocking activity. Updating the scenario depicted by our previous work,[22] several binding modalities for potent hERG blockers have been proposed so far. Considering sertindole as a reference compound, we mention (1) the perpendicular (with respect to the channel axis) curled solution (Farid-like binding mode),[18] where sertindole adopts a crown shaped conformation delimited by the underneath Phe656 ring, (2) the parallel extended solution where the fluorophenyl group interacts with Ser624 (Stansfeld-like binding mode),[21] (3) the parallel extended solution opposite to the previous one, where the imidazolidine group interacts with Ser624 (Österberg-like binding mode),[16] and (4) the perpendicular extended solution where the imidazolidine group interacts with Ser624 (Boukharta-like binding mode).[17] Accordingly, the binding mode displayed by MRC-003 agrees with the Österberg-like binding solution, which can be at some extent consistent with the pharmacophore model proposed by Cavalli et al.[10,38] Furthermore, along the many possibilities found in the SRC models, several Stansfeld-like and Boukharta-like binding modes could be identified, whereas no Farid-like solutions were obtained.

Performance of the Structure-Based Models. The automated docking protocol was purposely developed using a series of congeneric derivatives, in an attempt of capturing the relevant features of the channel-blocker interactions that might lead to meaningful structure−activity relationship models. A series of analogue compounds should reduce the scoring function uncertainty related to an approximate estimation of entropic and/or solvation contributions. Because of the large amount of unspecific interactions between hERG channel and blockers, a pose reranking based on a statistical (approximated) evaluation of the ligands' configurational entropy was here utilized. The good correlations achieved by the fittest SRCs and MRC-003 (r^2 more than 0.7), together with the consistency of the binding modes found in the latter model, made us confident about the feasibility of the approach. The fact that SRC-001 and MRC-003 retained good performances after the MM-PBSA rescoring was a further confirmation about the performance of our protocol. One could also envision a protocol where each binding mode is rescored with the MM-PBSA method. However, it should be underlined that such strategy can be highly CPU demanding.

The protocol was then validated using a series of structurally unrelated blockers, covering several different classes of drugs. Indeed, obtaining a good correlation with a general set of blockers is a much more challenging than using a series of analogous compounds. As expected, the docking protocol returned less satisfying results, even though an acceptable SRC model could be obtained (r^2 = 0.60). Figure 9A clearly showed that astemizole was a major outlier, and its removal led to a significantly improved correlation with an r^2 more than 0.8 (channel C1_53, belonging to cluster 3). The relatively high energies attributed to astemizole by the AD score, which in turn led to poor correlations, was likely due to the fact that the docking algorithm was unable to properly pose a relatively large drug into a crowded binding site. In spite of this, the encouraging correlations pointed to an acceptable general applicability of the present protocol to series of compounds in order to achieve predictive qualitative and quantitative insights about their hERG blocking ability.

CONCLUSIONS

Achieving a reliable description of the molecular interactions responsible for the hERG channel blockade by drugs represents one of the hottest topics in drug discovery. In recent years, a number of docking studies attempting to unravel the most likely binding modes for several series of hERG blockers have appeared in the literature. In most of these studies, the reliability of the docking solutions has been assessed by the ability to explain the experimental activity either directly, by using docking scoring functions, or indirectly, by relying on more sophisticated rescoring schemes. Eventually, these procedures implicitly provide different structure-based models that might be useful in prospect to predict the blockade potency of newly designed compounds, thus accelerating the identification of LQTS liabilities. Notwithstanding the importance of these docking studies, it is nonetheless surprising to notice the low agreement between binding modes proposed by different authors, especially considering the fact that the overall features of the starting channel models appear to be quite similar. The latter observation, together with the consideration that to the best of our knowledge no systematic docking studies on multiple conformations of the hERG channel has been undertaken yet, prompted us to develop a specific protocol aimed to overcome several modeling difficulties associated to the problem under investigation.[22] On the one hand, the main achievement of the present work is the unbiased obtainment of a consistent binding modality for a

congeneric series of sertindole derivatives which supports one of the most accredited pictures of binding (Österberg-like binding mode). On the other hand, the protocol itself proved to be successful and allowed us to elucidate some methodological aspects of docking so far not explicitly addressed. Concerning this point, we found that the conformational symmetry imposed to the amino acids of the binding site can have an impact on the reliability of the structure-based models derived. However, since similar shapes of the channel cavity can be achieved with different symmetries, the shape of the pore turned out to be a simpler and better descriptor to analyze the relevant structural features of the channel. Since channel models belonging to the C1 class on average provided the fittest SRCs, we suggest that there is no real need to impose symmetry restraints in the channel models generation. We consider this finding as a general validity result for symmetrical binding sites, not only in the context of hERG docking. In the attempt to overcome the tendency to achieve geometrically sparse binding modes for structurally similar compounds, we used a rescoring method suited to reweight the docking score by the probability to achieve a given binding mode. In the present case, the approach turned out to successfully increase the consistency of the docking outcomes and the possibility to rationalize structure−activity relationships for the SRC description. Finally, we have shown the possibility to obtain a minimal subset of channels spanning the relevant conformations of the cavity that are suitable for predicting hERG liability on a series of analogues compounds. These channel conformations would be in turn instrumental to reduce the computational cost in view of the lead optimization phase of drug discovery.

METHODS

Channel Models Generation. Since no crystal structures of the hERG channel are currently available, a rigorously validated homology model of the channel was used.[25] In order to characterize the channel cavity flexibility, an exhaustive sampling of the amino acids side chains (residues 623−624 of the selectivity filter and 651−656 of the S6 helix) was carried out by means of Modeller 9v8.[39] The refinement routine consisted of multiple optimization and molecular dynamics cycles. Specifically, an initial conjugate gradient optimization was followed by a heating−cooling phase of molecular dynamics with simulated annealing. An additional step of conjugate gradient completed the cycle. A set of 296 channel models (Figure 2A and B) was therefore generated by satisfying different point group symmetries (C4, C2, and C1). The stereochemical quality of the models was assessed by comparing parameters such as bond lengths, bond angles, torsion angles, and chirality with those derived by high resolution protein structures[40,41] with the Procheck software.[42] A total number of 215 channel models turned out to satisfy the stereochemical quality filter.

Shape-Based Cluster Analysis. Rather than using a well established but problematic clustering analysis in the RMSD space, a shape-based cluster analysis was employed. The accessibility profile of the channel pore for ligand binding (here simply referred to as shape profile) was described by measuring the maximum radius (r_{max}) along the channel axis (z) using the HOLE program (see Figure 3A).[43] The local dissimilarity between each channel model i and j was defined by taking the modulo difference of the maximum radius of the internal cavity sampled at a given position of the pore axis. Then, by summing these local dissimilarities along z, the pairwise distance used for the clustering was obtained:

$$d_{i,j} = \sum_{z} |r_{i_{max}}(z) - r_{j_{max}}(z)| \quad (1)$$

The cluster analysis was performed with the R software environment[44] using the average linkage method.[45] We stress that the metric introduced is a description of the accessible space available for ligand binding, rather than a direct and detailed measure of the shape of the internal cavity. Moreover, we note that the metric adopted, simply based on the maximum radius along the pore axis, can be employed in virtue of a substantially cylindrical binding site and because of the 4-fold symmetry displayed by the tertiary structure of potassium channels, which was not altered during the protein conformational analysis procedure. In other words, we implicitly deal with the assumption that different channel conformations, showing similar internal accessible cavities, would also display similar chemical environments, which is indeed reasonable in light of the above-reported peculiar features of the binding site.

Two clustering thresholds were employed in this work, depending on the purpose of the analysis performed. A height cutoff of 15 Å turned out to be effective in pruning the redundancy in the channel models, which were therefore considerably reduced from a number of 215 to 138. Such a cutoff was chosen in a heuristic way, by carefully evaluating the trade-off between the ability of the method to detect redundancy (either due to conformational duplication or symmetry multiplicity) and the inevitable loss of information. Besides such a fine-grained clustering, a coarser clustering analysis was actually employed to classify the channel models for the subsequent analysis of the SRCs. Rather than explicitly using a height cutoff, in this case the cluster dendogram was cut so as to obtain a convenient number of clusters able to describe the elementary shape of the pore cavity. A number of 10 clusters was arbitrarily chosen as a compromise between manageability of the outcome and accuracy of the description.

Docking. An automated docking protocol was applied to a set of 16 congeneric sertindole derivatives, listed in Table 1, for which experimental hERG blocking activity is available.[14] These compounds were selected in order to uniformly cover a quite broad range of activity (about 4 pIC_{50} units) and to span a significantly large chemical space (although necessarily limited by the congenericity relationship).

The docking simulations were performed using the Autodock 4.2 software.[33] The ligands were treated using Gasteiger partial charges,[46] whereas Kollman charges[47] were used for the protein. An all-atoms representation was adopted. The *autogrid* box was built so as to include the cavity of the channel models which were considered as rigid bodies during simulations. The search was carried out with the Lamarckian genetic algorithm, which allowed an exhaustive sampling of the conformational space, and the docking runs were set to 250 with the initial population of 150 individuals and a mutation rate of 0.02. Maximum number of generations and energy evaluations were set to 27 000 and 2.5×10^6, respectively.

The same parameters were used for series 2 (Table 2).

Postprocessing of the Docking Outcome. To cope with the docking poses redundancy arising as a consequence of the C2 and C4 channel symmetries, the docking outcome was subject to rotation along the channel axis (z) by 90, 180, and 270°. Then, among the different solutions, only the orientation

showing the lowest RMSD with respect to a reference structure was kept.

The pose rescoring was performed using the CE (Colony Energy) method proposed by Xiang et al.[26] in the context of protein loop modeling and later extended to protein−ligand docking problems.[27] The central assumption of the method is that each sampled conformation represents a colony of states on the potential energy surface and that similar configurations (binding poses) belong to the same basin. The size of the colonies, depending on the density of poses which are close in the configurational space (in the limit of an exhaustive and uniformly distributed sampling), represents a statistical assessment of the configurational entropy for a given docking solution. The CE score assigned to *i*th configuration is expressed as[26,27]

$$\mathrm{CE}_i \equiv -k_B T \ln\left[\sum_{j=1}^{N} e^{-\mathrm{RMSD}_{i,j}^3/t^3} e^{-1/k_B T E_j}\right] \quad (2)$$

where k_B is the Boltzmann constant, and T is the temperature (300 K). As it can be noticed from eq 2, the CE score has the form of a free energy and the argument of the logarithm plays the role of a partition function where the energy (score) of the *j*th docking solution is weighted by a function that takes into account the similarity of the given docking solution to the colony leading configuration (*i*th pose). In the weighting function, the t parameter was set to 1.8 Å.[48]We note that to properly employ the CE approach both the symmetry around the channel axis and the internal symmetry of the ligands must be taken into account. To this aim, all the symmetric atom pairs were swapped in coordinates, and the lowest RMSD was used in the CE calculation. The use of 250 runs in the docking procedure was motivated by the need to obtain a large ensemble of binding modes for each ligand, so as to meet the statistical requirements of an exhaustive and uniform sampling.

For each ligand, the AD score corresponding to the lowest (top-ranked) CE pose was used to describe the binding.

Building and Evaluation of hERG-Blocker Models. Two kinds of correlative structure-based models were built using the information provided by the docking procedure: the more naïve SRCs and the more physically sound MRCs.

In the SRC approach we considered the results obtained by the docking performed against each individual receptor conformation, so that all the 138 channel models were treated independently. For each channel, the AD score obtained for the top-ranking binding modes highlighted by the CE method was stored and employed in evaluating the performance of the derived structure-based model. As figure of merit, the squared correlation coefficient r^2 between the ligands' docking scores and experimental hERG blocking activities expressed as pIC_{50} was used. To simplify the analysis of the hERG-blocker models, a classification of the fitness depending on the observed r^2 was adopted. Accordingly, structure-based models showing an r^2 lower than 0.4 were classified as "bad" performing models, whereas those displaying a squared correlation coefficient greater than 0.6 were considered as "good" performing models.

Three MRC models were built by differently exploiting the information provided by the docking procedure and the analysis of the channel models. The MRC-001 model was derived combining all the information obtained by the docking procedure against the 138 channel conformations. The MRC-002 was obtained upon a selection of the channel models leading to the fittest SRCs (r^2 > 0.6: 24 channel conformations). Finally, the MRC-003 was built by combining the latter selection with the results of the cluster analysis, leading to a total number of 7 channel structures. Unlike SRCs, in the MRCs evaluation, three statistical data treatments across the ensemble members were used:[49] the best scores, the arithmetic mean, and the Boltzmann-weighted average (see Table 3 for the sertindole series). In the latter, the Boltzmann distribution function was applied to the ligands' scores within the ensemble of conformations:

$$p = \frac{N_i}{\sum N_i} = \frac{e^{-E_i/kT}}{\sum e^{-E_i/kT}} \quad (3)$$

Besides the performance of the hERG-blocker models, we also examined their predictive ability through two different statistical approaches: the widely used leave-one-out q^2 and PI.[36] The q^2 was estimated by a correlation between ligands' scores and related predicted activities, derived from the leave-one-out technique: each molecule was in turn removed from the initial set of compounds, and then, its activity was predicted using the remainder of the data. While the q^2 proved the robustness of a model, the PI was employed to quantify the ability of a model to rank the compounds according to their binding affinities. The PI was calculated with the equation:

$$\mathrm{PI} = \frac{\sum_{j>i}\sum_i \omega_{ij} C_{ij}}{\sum_{j>i}\sum_i \omega_{ij}} \quad (4)$$

with

$$\omega_{ij} = |E(j) - E(i)| \quad (5)$$

and

$$C_{ij} = \begin{cases} 1 & \text{if } [E(j) - E(i)]/[P(j) - P(i)] < 0 \\ -1 & \text{if } [E(j) - E(i)]/[P(j) - P(i)] > 0 \\ 0 & \text{if } [P(j) - P(i)] = 0 \end{cases} \quad (6)$$

where $E(i)$ is the experimental pIC_{50} and $P(i)$, the score related to the reference *i*th compound. The PI assumes values ranging from −1 to +1: a value of +1 indicates perfect predictions; a value of −1 indicates wrong predictions; a PI = 0, suggests completely random predictions.

The procedure consisting in docking, postprocessing of the docking outcome, and building and evaluation of the correlative models was entirely automated through the Unix BASH-shell scripting language with multiple Tcl and Awk calls, see SI text. The automated procedure will be available upon request.

MM-PBSA Refinement. In order to better treat solvation effects, the fittest hERG-blocker models were refined using a single-configurational MM-PBSA[29] rescoring scheme. Accordingly, the ΔG_{bind} for each ligand was computed as

$$\Delta G_{\mathrm{bind}} = G_{\mathrm{complex}} - G_{\mathrm{protein}} - G_{\mathrm{ligand}} \quad (7)$$

where each term was evaluated as the sum of the force field energy (G_{MM}) and the polar (G_{PB}) and nonpolar (G_{SA}) contributions to solvation free energy. For sake of simplicity, no entropic contributions were taken into account.[50,51] The force field energy was calculated after 1000 steps of minimizations (500 steps of steepest descent followed by 500 steps of conjugate gradient) of the complexes by means of the *sander* module of the Amber11 package.[52] The generalized Born model of Hawkins, Cramer, and Truhlar was used as solvation

dx.doi.org/10.1021/ci300326d | J. Chem. Inf. Model. 2013, 53, 159−175

model.[53,54] The AMBER ff99SB-ildn force field[55] was used for the protein, and the General Amber Force Field (GAFF)[56] together with RESP charges[57,58] was adopted to treat ligands. RESP charges were calculated with the G09 package[59] at the B3LYP/6-31G*//B3LYP/6-31G* level of theory. The polar contributions to solvation free energy were estimated by solving the linearized Poisson−Boltzmann (PB) equation for the minimized structures using APBS.[60] In doing so, a multiple dielectric description of the systems was adopted: water, solute, and membrane were treated as different environments using ε values of 80, 4, and 2, respectively. We used a 3-level of focusing approach which starts by solving the PB equation on a coarse grid of large size, then on a medium grid, and finally on a fine grid. We performed two types of calculations: the first, in a multidielectric environments using an ionic strength of 0.1 M of both +1/−1 ions with a radius of 2.0 Å in the aqueous environment, and the second in vacuum (no membrane, homogeneous dielectric $\varepsilon = 1$ for solute and solvent, and null ionic strength). The solvation energy was estimated by subtracting the latter term (the Coulombic contribution, in vacuum) to the first calculation. Finally, the nonpolar components of the solvation free energy were computed with APBS according to the following equation[61]

$$G_{SA} = \gamma \mathrm{SASA} + b \tag{8}$$

where SASA is the solvent-accessible surface area estimated using a probe with radius of 1.4 Å, γ is the solvent surface tension parameter (0.00542 kcal mol^{-1} Å^{-2}), and b is the free energy of nonpolar solvation for a point solute (0.92 kcal mol^{-1}).[50,62]

■ ASSOCIATED CONTENT

Supporting Information

Text detailing some technical features of the automated procedure and related references; list of the SRC models for series 1, Table S1; table showing the MM-PBSA free energy contributions for series 1, Table S2; list of the fittest hERG-blocker models for series 2, Table S3; the input/output flows and the dialogue box of the automated protocol, Figure S1; χ_1/χ_2 plots, Figure S2; plot showing the r^2 for all the models, Figure S3; additional binding modes, Figures S4 and S5. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: matteo.masetti4@unibo.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Federico Falchi for useful discussions and also acknowledge funding from Istituto Italiano di Tecnologia, Genova, Italy (Seed project: HERGSYM) and the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 241679 (ARITMO project).

■ REFERENCES

(1) Tseng, G.-N. IKr: The hERG Channel. *J. Mol. Cell. Cardiol.* **2001**, *33*, 835−849.
(2) Sanguinetti, M. C.; Tristani-Firouzi, M. hERG potassium channels and cardiac arrhythmia. *Nature* **2006**, *440*, 463−469.
(3) Recanatini, M.; Poluzzi, E.; Masetti, M.; Cavalli, A.; De Ponti, F. QT prolongation through hERG K+ channel blockade: Current knowledge and strategies for the early prediction during drug development. *Med. Res. Rev.* **2005**, *25*, 133−166.
(4) Yoshida, K.; Niwa, T. Quantitative Structure-Activity Relationship Studies on Inhibition of HERG Potassium Channels. *J. Chem. Inf. Model.* **2006**, *46*, 1371−1378.
(5) Su, B.-H.; Shen, M.-y.; Esposito, E. X.; Hopfinger, A. J.; Tseng, Y. J. In Silico Binary Classification QSAR Models Based on 4D-Fingerprints and MOE Descriptors for Prediction of hERG Blockage. *J. Chem. Inf. Model.* **2010**, *50*, 1304−1318.
(6) Keserü, G. Prediction of hERG potassium channel affinity by traditional and hologram qSAR methods. *Bioorg. Chem. Med. Lett.* **2003**, *13*, 2773−2775.
(7) Song, M.; Clark, M. Development and Evaluation of an in Silico Model for hERG Binding. *J. Chem. Inf. Model.* **2005**, *46*, 392−400.
(8) Coi, A.; Massarelli, I.; Murgia, L.; Saraceno, M.; Calderone, V.; Bianucci, A. M. Prediction of hERG potassium channel affinity by the CODESSA approach. *Bioorg. Med. Chem.* **2006**, *14*, 3153−3159.
(9) Du-Cuny, L.; Chen, L.; Zhang, S. A Critical Assessment of Combined Ligand- and Structure-Based Approaches to hERG Channel Blocker Modeling. *J. Chem. Inf. Model.* **2011**, *51*, 2948−2960.
(10) Cavalli, A.; Poluzzi, E.; De Ponti, F.; Recanatini, M. Toward a Pharmacophore for Drugs Inducing the Long QT Syndrome: Insights from a CoMFA Study of HERG K+ Channel Blockers. *J. Med. Chem.* **2002**, *45*, 3844−3853.
(11) Ekins, S.; Crumb, W. J.; Sarazan, R. D.; Wikel, J. H.; Wrighton, S. A. Three-Dimensional Quantitative Structure-Activity Relationship for Inhibition of Human Ether-a-Go-Go-Related Gene Potassium Channel. *J. Pharmacol. Exp. Ther.* **2002**, *301*, 427−434.
(12) Aronov, A. M.; Goldman, B. B. A model for identifying HERG K+ channel blockers. *Bioorg. Med. Chem.* **2004**, *12*, 2307−2315.
(13) Aronov, A. M. Common Pharmacophores for Uncharged Human Ether-a-go-go-Related Gene (hERG) Blockers. *J. Med. Chem.* **2006**, *49*, 6917−6921.
(14) Pearlstein, R. A.; Vaz, R. J.; Kang, J.; Chen, X.-L.; Preobrazhenskaya, M.; Shchekotikhin, A. E.; Korolev, A. M.; Lysenkova, L. N.; Miroshnikova, O. V.; Hendrix, J.; Rampe, D. Characterization of HERG potassium channel inhibition using CoMSiA 3D QSAR and homology modeling approaches. *Bioorg. Chem. Med. Lett.* **2003**, *13*, 1829−1835.
(15) Durdagi, S.; Duff, H. J.; Noskov, S. Y. Combined Receptor and Ligand-Based Approach to the Universal Pharmacophore Model Development for Studies of Drug Blockade to the hERG1 Pore Domain. *J. Chem. Inf. Model.* **2011**, *51*, 463−474.
(16) Österberg, F.; Åqvist, J. Exploring blocker binding to a homology model of the open hERG K+ channel using docking and molecular dynamics methods. *FEBS Lett.* **2005**, *579*, 2939−2944.
(17) Boukharta, L.; Keränen, H.; Stary-Weinzinger, A.; Wallin, G.; de Groot, B. L.; Åqvist, J. Computer Simulations of Structure-Activity Relationships for hERG Channel Blockers. *Biochemistry* **2011**, *50*, 6146−6156.
(18) Farid, R.; Day, T.; Friesner, R. A.; Pearlstein, R. A. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg. Med. Chem.* **2006**, *14*, 3160−3173.
(19) Coi, A.; Massarelli, I.; Testai, L.; Calderone, V.; Bianucci, A. M. Identification of a "toxicophoric" features for predicting drug-induced QT interval prolongation. *Eur. J. Med. Chem.* **2008**, *43*, 2479−2488.
(20) Rajamani, R.; Tounge, B. A.; Li, J.; Reynolds, C. H. A two-state homology model of the hERG K+ channel: application to ligand binding. *Bioorg. Chem. Med. Lett.* **2005**, *15*, 1737−1741.
(21) Stansfeld, P. J.; Gedeck, P.; Gosling, M.; Cox, B.; Mitcheson, J. S.; Sutcliffe, M. J. Drug block of the hERG potassium channel: Insight from modeling. *Proteins: Struct. Funct. Bioinf.* **2007**, *68*, 568−580.
(22) Recanatini, M.; Cavalli, A.; Masetti, M. Modeling hERG and its Interactions with Drugs: Recent Advances in Light of Current Potassium Channel Simulations. *Chem. Med. Chem.* **2008**, *3*, 523−535.

(23) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational Drug Design Accommodating Receptor Flexibility: The Relaxed Complex Scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632−5633.
(24) Masetti, M.; Cavalli, A.; Recanatini, M. Modeling the hERG potassium channel in a phospholipid bilayer: Molecular dynamics and drug docking studies. *J. Comput. Chem.* **2008**, *29*, 795−808.
(25) Tseng, G.-N.; Sonawane, K. D.; Korolkova, Y. V.; Zhang, M.; Liu, J.; Grishin, E. V.; Guy, H. R. Probing the Outer Mouth Structure of the hERG Channel with Peptide Toxin Footprinting and Molecular Modeling. *Biophys. J.* **2007**, *92*, 3524−3540.
(26) Xiang, Z.; Soto, C. S.; Honig, B. Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7432−7437.
(27) Lee, J.; Seok, C. A statistical rescoring scheme for protein−ligand docking: Consideration of entropic effect. *Proteins: Struct. Funct. Bioinf.* **2008**, *70*, 1074−1083.
(28) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178−184.
(29) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(30) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221−5230.
(31) Fernandez, D.; Ghanta, A.; Kauffman, G. W.; Sanguinetti, M. C. Physicochemical Features of the hERG Channel Drug Binding Site. *J. Biol. Chem.* **2004**, *279*, 10120−10127.
(32) Knape, K.; Linder, T.; Wolschann, P.; Beyer, A.; Stary-Weinzinger, A. *In silico* Analysis of Conformational Changes Induced by Mutation of Aromatic Binding Residues: Consequences for Drug Binding in the hERG K+ Channel. *PLoS ONE* **2011**, *6*, e28778.
(33) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785−2791.
(34) Craig, I. R.; Essex, J. W.; Spiegel, K. Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *J. Chem. Inf. Model.* **2010**, *50*, 511−524.
(35) Yoon, S.; Welsh, W. J. Identification of a Minimal Subset of Receptor Conformations for Improved Multiple Conformation Docking and Two-Step Scoring. *J. Chem. Inf. Comput. Sci.* **2003**, *44*, 88−96.
(36) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J. Med. Chem.* **2001**, *44*, 3417−3423.
(37) Zachariae, U.; Giordanetto, F.; Leach, A. G. Side Chain Flexibilities in the Human Ether-a-go-go Related Gene Potassium Channel (hERG) Together with Matched-Pair Binding Studies Suggest a New Binding Mode for Channel Blockers. *J. Med. Chem.* **2009**, *52*, 4266−4276.
(38) Cavalli, A.; Buonfiglio, R.; Ianni, C.; Masetti, M.; Ceccarini, L.; Caves, R.; Chang, M. W. Y.; Mitcheson, J. S.; Roberti, M.; Recanatini, M. Computational Design and Discovery of "Minimally Structured" hERG Blockers. *J. Med. Chem.* **2012**, *55*, 4010−4014.
(39) Sali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M. Evaluation of comparative protein modelling by MODELLER. *Proteins* **1995**, *23*, 318−326.
(40) Morris, A. L.; MacArthur, M. W.; Hutchinson, E. G.; Thornton, J. M. Stereochemical quality of protein structure coordinates. *Proteins: Struct. Funct. Bioinf.* **1992**, *12*, 345−364.
(41) Engh, R. A.; Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst. A* **1991**, *47*, 392−400.
(42) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(43) Smart, O. S.; Neduvelil, J. G.; Wang, X.; Wallace, B. A.; Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph. Model.* **1996**, *14*.
(44) Ihaka, R.; Gentleman, R. R: A Language for Data Analysis and Graphics. *J. Comput. Graph. Stat.* **1996**, *5*, 299−314.
(45) Everitt, B. S.; Landau, S.; Leese, M. *Cluster Analysis*, Fourth ed.; Arnold: London, UK, 2001.
(46) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(47) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.
(48) Fogolari, F.; Tosatto, S. C. E. Application of MM/PBSA colony free energy to loop decoy discrimination: Toward correlation between energy and root mean square deviation. *Protein Sci.* **2005**, *14*, 889−901.
(49) Paulsen, J. L.; Anderson, A. C. Scoring Ensembles of Docked Protein:Ligand Interactions for Virtual Lead Optimization. *J. Chem. Inf. Model.* **2009**, *49*, 2813−2819.
(50) Ferrari, A. M.; Degliesposti, G.; Sgobba, M.; Rastelli, G. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 7865−7877.
(51) Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. Ligand Affinities Predicted with the MM/PBSA Method: Dependence on the Simulation Method and the Force Field. *J. Med. Chem.* **2006**, *49*, 6596−6606.
(52) Case, D. A.; Darden, T. A.; Cheatham, I., T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, A.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*; University of California: San Francisco, CA, 2010.
(53) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(54) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.* **1996**, *100*, 19824−19839.
(55) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct. Funct. Bioinf.* **2010**, *78*, 1950−1958.
(56) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(57) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(58) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.
(59) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima,

T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision B.01; Gaussian, Inc.: Wallingford, CT, 2009.

(60) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.

(61) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(62) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J. Med. Chem.* **2005**, *48*, 4040−4048.