



universität
wien

DISSERTATION

Titel der Dissertation

„IN SILICO SCREENING ON THE HERG POTASSIUM CHANNEL“

Verfasser

Mag. Andrea Schiesaro

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien, 2011

Studienkennzahl lt. Studienblatt:	A 091 449
Dissertationsgebiet lt. Studienblatt:	Pharmazie
Betreuerin / Betreuer:	Univ.-Prof. Dr.Gerhard F. Ecker

Don't worry, we are in the hands of God

Frère Roger

*To God,
to my wife Oyunbayar,
to the kids that will come,
to my grandmother Gemma in the paradise,
to my grandfathers in the sky,
to my mother Michela,
to my father Francesco,
to my sister Alessandra and her husband Daniele,
to my father-in-law Dorj Chuluun,
to my mother-in-law Luvsandorj Bud,
to my sister-in-law Oyuntuul,
to my new family in Mongolia
to all my nephews
and to all my friends.*

Danksagungen

The present work was carried out at Pharmacoinformatics Research Group, Department of Medicinal Chemistry, Faculty of Life Sciences, University of Vienna, Austria.

I warmly thank the University of Vienna that provided the financial support under the PhD program "Molecular Drug Targets".

I wish to express my deepest and sincere gratitude to my advisor Prof. Gerhard F. Ecker for your generosity with which you accepted me in your lab. I am happy that you gave me the opportunity to join a so high-level group.

I wish my gratitude to Prof. Hering, Prof. Timin and Mag. Andreas Windisch, Department of Pharmacology and Toxicology, Faculty of Life-Sciences, University Vienna, Austria, for your collaboration.

I wish my sincere gratitude to Univ. Ass. Dr. A. Weinzinger for the models and the helpful discussions.

I am grateful to Dr. Khac-Mihn Thai for your help to introduce me in the topic.

I wish my sincere gratitude to Prof. Gabriele Costantino, Pharmaceutical Department, University of Parma, Italy, for your generosity with which you gave me opportunity to join your group for three months.

I want to thank my colleague Daniela Digles for suggestions to improve my work, and for the nice discussions in "our" room. I wish my gratitude to my colleague Lars Richter for the SVL scripts and helpful suggestions in programming. I warmly thank my colleagues: Yogesh Aher, Ishrat Jabeen, Rene Weissensteiner, Rainer Dangl, Andreas Jurik, Sabine Mydza, Barbara Zdrzil, Freya Klepsch, Regina, Daniels, Vasanthanathan Poongavanam, Michael Demel, Rita Schwaha, Wenwen Huang, Christoph Waglechner, Wolfgang Heith, Martin, Johanna, Maria Koeppner and Kamil Medschedow for many different reasons.

For all the nice moments that we had I want to thank my colleagues from the graduate training program (Initiativkolleg, IK) "Molecular Drug Targets": Markus Aleschko, Renate Baumgartner, Waheed Shabbir, Gowri Shankar Bagavananthem Andavan, Wu Shengqian, Judith Singhuber, Andreas Windisch, Chonticha Suwattanasophon and Barbara Taferner.

I want to thank also my colleagues in Parma Agostino Bruno, Daniele Pala, Luigi Caposella and Claudia Beato for the helpful discussions and all the nice moments we had together.

I want to thank Prof. Stefano Moro for your support during my thesis and your support to come here.

I want to thank my wonderful wife Oyunbayar. She is all for me.

I want to thank my mother Michela. She gave me a lot of love and supported me in many difficult situations. I want to thank my sister Alessandra with whom I shared many special moments, and to the husband Daniele. A special thanks to my grandparents.

I want to thank my new family: in primis my father-in-law Dorj Chuluun and my mother-in-law Luvsandorj Bud who were so kind with me. I want to thank my sister-in-law Oyungerel for all the nice moments that we had in Vienna. I want to thank my little love Sudaregshiglen and the father Gambat. I need to mention also the „little“ sister Oyuntuul for all the fantastic moments we had in these months. Thank also to the rest of the family Oyunerdene, Oyunsuren, Butensain, Zolboot, Butenbayar and Khuslen with who I am sure I will have fantastic moments in Mongolia.

A special thanks to all my friends in Vicenza. Nothing will break our friendship.

At least, but not last, a special thanks to all my friends in Vienna Riccardo, Beatrice e la piccola Elisabetta, Beatrice (the inox one) Gigi, Dennis, Orsoloya, Giulio, Ullly, Norbert, Sara, Sofia, Elisa, Francesco, Carlo, Ion, Manuela, Ahmed and Domenico for all the coffee after the pray and for all the incredible moments we had in these years. A special thanks to my friends Tutik and Nabil for all the nice dinners. Of course, thanks also to Padre Adalberto.

CONTENTS

LIST OF FIGURES	IV
LIST OF TABLES	V
ABBREVIATIONS	VI
English Abstract	X
Abstract Deutsch	XII
Chapter 1: INTRODUCTION	1
1.1 Long QT syndrome	2
1.2 Structure of the hERG channel	3
1.3 Ligand based approaches	4
<i>1.3.1 Pharmacophore models</i>	4
<i>1.3.2 3D-QSAR</i>	8
<i>1.3.3 2D-QSAR</i>	12
<i>1.3.4 1D-QSAR</i>	18
1.4 Classification models	19
<i>1.4.1 Decision trees</i>	27
1.5 Matched molecular pairs	30
1.6 Structure based approaches	31
<i>1.6.1 Homology models of the hERG channel</i>	31
<i>1.6.2 hERG inhibition and drug trapping</i>	32
<i>1.6.3 Case studies: propafenone derivatives trapping</i>	33
<i>1.6.4 Amino acids involved in hERG inhibition</i>	34
<i>1.6.5 Hydrogen bonds with Ser624, Thr623 or Val625?</i>	36
<i>1.6.6 Role of Gly648</i>	37
<i>1.6.7 Which subunits are involved in drug binding?</i>	37
<i>1.6.8 Influence of para-substituents on the phenyl ring</i>	39
<i>1.6.9 Two or three binding interactions?</i>	39
<i>1.6.10 Docking studies and prediction of hERG binding affinity</i>	40

1.6.11 Case studies: docking studies and improvement of the selectivity	41
1.6.12 Orthogonal binding site?	44
1.6.13 Docking results	44
1.7 Conclusions	49
Chapter 2: THE TRAPPING PHENOMENA	51
2.1 The trapping phenomenon in the propafenone and derivatives	53
2.2 How is possible to investigate in silico the trapping phenomena?	55
Chapter 3: COMPUTATIONAL BACKGROUND	57
3.1 Docking	58
3.1.1 Search algorithms	58
3.2 Scoring functions	60
3.3 Compounds investigated	61
Chapter 4: AIM OF THE STUDY	63
Chapter 5: RESULTS	65
5.1 Docking into the homology model of hERG channel in the closed state	65
5.1.1 Material and methods	65
5.1.1.1 Docking protocol	65
5.1.1.2 The symmetry problem in the hERG channel	67
5.1.1.3 Alignment	67
5.1.1.4 RMSD matrix calculation	69
5.1.1.5 Duplicate poses deletion	69
5.1.1.6 Identification of the common binding modes	70
5.1.1.7 Analysis of ligand interactions	71
5.1.2 Results	71
5.1.2.1 Alignment and deletion of duplicate poses	71
5.1.2.2 Pose analysis	76

5.1.3 Conclusions	83
5.2 Docking studies into a homology model of the hERG channel in the open state	84
5.2.1 Material and methods	84
5.2.1.1 Docking protocol	84
5.2.1.2 Alignment	85
5.2.1.3 RMSD matrix calculation and duplicate poses deletion	85
5.2.1.4 Identification of the common binding modes	85
5.2.1.5 Analysis of ligand interactions	87
5.2.2 Results	87
5.2.2.1 Poses analysis	87
5.2.3 Conclusions	99
5.3 Docking studies into models of hERG channel in the open state obtained from molecular dynamics simulations	101
5.3.1 Material and methods	101
5.3.1.1 Docking protocol	101
5.3.1.2 Identification of the common binding modes	102
5.3.1.3 Analysis of ligand interactions	102
5.3.2 Results	102
5.3.2.1 Pose analysis	102
5.3.3 Conclusions	111
REFERENCES	112
CURRICULUM VITAE	122

LIST OF FIGURES

Fig. 1 Chemical scaffolds of the combinatorial libraries	20
Fig. 2 Fragments related to strong hERG inhibition	22
Fig. 3 Fragments found in hERG blockers	31
Fig. 4 2D structure of the two compounds docked into the hERG channel	43
Fig. 5 Pose alignment	68
Fig. 6 Docking results of propafenone docked with MOE	72
Fig. 7 Common scaffold used to calculate the RMSD matrix	77
Fig. 8 Centroids of the four selected clusters	78
Fig. 9 First binding mode	79
Fig. 10 Second binding mode	80
Fig. 11 Glide result	81
Fig. 12 First and second binding modes	82
Fig. 13 Centroid of the first cluster	89
Fig. 14 Centroid of the second cluster	90
Fig. 15 Centroid of the third cluster	91
Fig. 16 Position of R1 in the poses of cluster1	93
Fig. 17 Position of R1 substituents in the poses of cluster2	94
Fig. 18 Position of R1 in the poses of cluster3	95
Fig. 19 Centroid of cluster four	104
Fig. 20 Centroid of cluster five	105
Fig. 21 Position of the substituents in the poses of cluster four	107
Fig. 22 Lateral view of the substituents R1 position in cluster five	108

LIST OF TABLES

Table 1 Published homology models of the hERG potassium channel in the closed and open state	50
Table 2 Propafenone derivatives	62
Table 3 Distribution of similar poses in MOE in the first 100 poses selected according to the scoring function	73
Table 4 Distribution of similar poses in Glide	73
Table 5 Number of similar poses in MOE in the first 100 poses selected according to the scoring function	74
Table 6 Number of similar poses in Glide	74
Table 7 Deletion of the poses from MOE docking in the first 100 poses selected according to the London dg scoring function	75
Table 8 Deletion of the poses from Glide docking	75
Table 9 Poses of cluster1	97
Table 10 Poses of cluster2	98
Table 11 Poses of cluster3	99
Table 12 Structural information on the eight hERG models	103
Table 13 Poses of cluster4	109
Table 14 Poses of cluster5	110

ABBREVIATIONS

ADMET	Absorption, distribution, metabolism, elimination and toxicity
hERG	Human ether-a-go-go-related-gene K ⁺ channel
QSAR	Quantitative structure-activity relationship
LQTS	Long QT syndrome
TdP	Torsade de Pointes
eag	Ether-a-go-go
S	Transmembrane domain
ECG	Electrocardiogram
CSD	Cambridge Structural Database
PCA	Principal Component Analysis
PLS	Partial Least Squares
LMS	Least Median Squares
CoMFA	Comparative Molecular Field Analysis
CoMSIA	Comparative Molecular Similarity Indices Analysis
SVR	Support Vector Regression
RF	Random Forest
CCR	Chemokine receptor
SOM	Self-Organizing Maps
LSER	Linear Solvation Energy Relationship
KNN	<i>k</i> Nearest Neighbor
PNN	Probabilistic Neural Network
SVM	Support Vector Machine
CPG-NN	Counter-propagation neural network

LDA	Linear Discriminant Analysis
QA	Quaternary ammonium
PLS-DA	Partial Least Squares Discriminant Analysis
LIE	Linear interaction energy
MD	Molecular dynamics
SAR	Structure-activity relationship
SA	Simulated annealing
ISA	Iterative simulated annealing
ROC	Receiver Operating Characteristic
AUC	Area under the curve
SVL	Scientific Vector Language
RMSD	Root Mean Square Deviation
G	Binding free energy
H	Enthalpy
T	Temperature
K _i	Binding constant
R	Gas constant
MOE	Molecular Operating Environment
PDB	Protein Data Bank
AMBER	Assisted Model Building with Energy Refinement
MMFF	Molecular Mechanical Force Field
ΔS	Difference in entropy between two states
E	Potential energy
MIF	Molecular Interaction Field
MED	Molecular Electrostatic Potential

IEL	Local Ionization Energy
EAL	Local Electron Affinity
POL	Local Polarizability
SHANI and SHANE	Shannon entropies
GRIND	Grid-INdependent Descriptors
HQSAR	Hologram QSAR
EVA	Eigen Values
HEK	Human Embrionic Kidney
CHO	Chinese Hamster Ovary
QSTR	Quantitative Structure-Toxicity Relationship
BLR	Baseline Lipophilicity Relationship
<i>a</i>	Hydrophobicity factor of the binding site
<i>D</i>	n-octanol/water partition
ΔG_{inr}	Intrinsic binding energy
ETA	Extended Topological Atom indices
E-state	Electro-Topological-state index of an atom in the molecule
H-bond	Hydrogen-bond
P_VSA	Property labelled Van der Waals surface area
VSA	Van der Waals surface area
FCFP	Functional class fingerprint
CF	Comprehensive fragments
CP_*	C-pharmacophore descriptors
SIBAR	Similarity-based descriptors
DIPDENS	Dipolar density
CHBBA	Covalent hydrogen bond basicity

MR	Molar refractivity
QSUM	Sum of atomic charges on an atom
QSUM-	Sum of negative ESP charges
MGHBD	Minimal geometric distance between two hydrogen-bond donor atoms
ECFP	Extended connectivity fingerprint
Δ_{ele}	Difference in the electrostatic Energy values
Δ_{vdw}	Difference in the van der Waals energy values
Glide	Grid-based ligand docking with energetics
GOLD	Genetic Optimisation for Ligand Docking
KDR	Kinase domain receptor
CYP2D6	Cytochrome P450 2D6
I _{kr}	Rapid delayed rectifier K ⁺ current

English abstract

During the drug development process, almost 35% of the compounds fail due to poor absorption, distribution, metabolism, excretion and toxicity (ADMET). An important role on these failures is played by improper interactions with antitarget proteins, such as cytochrome P450, P-glycoprotein and the hERG potassium channel.

The hERG potassium channel is expressed in various cells and tissues, such as heart, neurons and smooth muscle. In the heart, the hERG channel plays an important role in the third phase of heart repolarization, due to the conduction of the rapid delayed rectifier K^+ current (I_{Kr}). A delay of this phase of repolarization leads to a syndrome called Long QT syndrome (LQTS) which might cause a potentially fatal arrhythmia called Torsade de Pointes (TdP). Many different classes of compounds were withdrawn from the market in the past decade due to their interaction with the hERG channel. Similar to other antitarget proteins, the hERG channel is polyspecific in the ligand recognition, hence it can interact with many classes of compounds, such as psychiatric, antihistaminic, antiarrhythmic and antimicrobial drugs. Several studies show that some molecules do not dissociate during the channel gating and are trapped in the closed state of the hERG channel.

In this study, propafenone and derivatives were docked into homology models of the hERG channel in the closed and open states to shed more light on hERG inhibition and on drug trapping.

With the aim to investigate the interactions between the hERG channel in the closed state and the compounds investigated, a series of trapped propafenone derivatives were docked into the homology model of the hERG channel in the closed conformation using Dock, the docking tool of MOE, and Glide, the docking tool of Schrödinger. A svl script called ROTALI was used to generate RMSD matrices with which the duplicate poses lying in different directions of the central cavity were detected and deleted, thus allowing to identify possible binding modes through agglomerative hierarchical clustering. This analysis led to the identification of two possible binding modes.

The same process was applied to the poses obtained by docking the propafenones into a homology model of the hERG channel in the open state. Three possible binding modes were

selected through agglomerative cluster analysis of the RMSD matrix generated taking into account the propafenone derivatives' common scaffold and the amino acids that might interact.

Finally, in order to take into account protein flexibility, nine propafenone derivatives were docked into eight models of the hERG channel in the open state obtained from snapshots of molecular dynamics simulations. Clustering both according to the common scaffold RMSD and the RMSD matrix of the amino acids interacting with the poses, two binding modes were selected. Biological studies suggest that non-trapped propafenones hinder the hERG channel gating with a mechanism called "foot in the door". In four out of the five selected clusters, it is possible to explain the "foot in the door" mechanism.

Interestingly, ranking the poses of the five clusters above-mentioned according to the potential energy values of the R1 substituent, and according to this value divided by the number of heavy atoms, it is possible to distinguish between trapped and non-trapped propafenones. In the non-trapped compounds, this value is always higher than in the trapped ones. The fact that it works also in cluster five, where the R1 substituents are placed under the ring formed by the four Phe656, might indicate that drug trapping phenomena depend more on intrinsic properties of the R1 substituent rather than on its conformation when it interacts with the hERG channel. Hence, this might indicate that the rigidity and the bulkyness of the substituent determines whether a propafenone derivatives is trapped or not independently of the binding mode in the hERG channel.

Abstract Deutsch

Während des Arzneistoffentwicklungsprozesses scheitern fast 35% der Arzneistoffe wegen schlechter Absorption, Verteilung, Metabolismus, Ausscheidung und Toxizität (ADMET). Ein wichtiger Bestandteil dieses Scheiterns ist die Interaktion mit Anti-Target Proteinen wie Cytochrom P450, P-glycoprotein und dem hERG Kaliumkanal.

Der hERG Kaliumkanal ist in vielen verschiedenen Zellen und Geweben wie dem Herz, Nerven und glatten Muskelzellen vorhanden. Im Herzen spielt der hERG Kanal während des Aktionspotentials in der dritten Phase der kardialen Repolarisierung wegen der Weiterleitung des schnellen Kalium Ausstroms (I_{Kr}) eine wichtige Rolle. Ein Verzögern dieser Phase führt zum Long QT Syndrom (LQTS), das eine potenziell tödliche Arrhythmie verursachen kann. Viele Klassen von Medikamenten wurden wegen ihren Wechselwirkungen mit dem hERG Kanal in den letzten zehn Jahren vom Markt zurückgezogen. Wie auch andere Anti-Target Proteine, ist der hERG Kanal in der Ligandenerkennung unspezifisch, weshalb er mit vielen Klassen von Arzneistoffen wie Psychopharmaka, Antihistaminika, Antiarrhythmika und Antibiotika interagieren kann. Viele Studien zeigen, dass eine erhebliche Anzahl von Molekülen während der Schließung des Kanals nicht dissoziieren und im geschlossenen Zustand des hERG Kanals gefangen bleiben.

In dieser Studie wurden Propafenon und dessen Derivate in ein Homologie-Modell des hERG Kanals im geschlossenen und geöffneten Zustand gedockt, um die hERG Hemmung und das „drug trapping“ besser verstehen zu können.

Ziel war es, die Wechselwirkungen zwischen dem hERG Kanal im geschlossenen Zustand und den Liganden zu untersuchen. Aufgrund dessen wurde eine Serie von „trapped“ Propafenon-Derivaten im hERG Kanal, welcher sich im geschlossenen Zustand befand, mit Dock, einem Docking Modul des Programms MOE, und GLIDE, dem Docking-Programm von Schrödinger, gedockt. Es wurde ein svl-Skript, genannt ROTALI, verwendet, um RMSD Matrizen zu erstellen, mit welchen die Duplikate unter den Posen, die in Bezug auf die Central Cavity unterschiedlich positioniert waren, zu erkennen und zu löschen. In weiterer Folge wurden die möglichen binding modes durch agglomeratives hierarchisches Clustering identifiziert. Die Analyse der Posen führte zur Identifizierung von zwei möglichen Binding Modes.

Derselbe Prozess wurde angewandt, um eine Serie von Propafenon-Derivaten in ein Homologie-Modell des hERG Kanals im geöffneten Zustand zu docken. Drei mögliche Binding Modes wurden durch die agglomerative Cluster Analyse der RMSD Matrix identifiziert, welche durch das gemeinsame Gerüst der Propafenon Derivate und jenen Aminosäuren generiert wurde, die mit den Molekülen interagierten. Um die Flexibilität des Proteins zu berücksichtigen wurden die Propafenon Derivate zusätzlich in acht verschiedene Schnappschüsse einer Moleküldynamik des Homologie-Modelles des hERG Kanals im geöffneten Zustand gedockt. In diesem Fall wurden zwei Binding Modes selektiert.

Interessanterweise war es durch das Einordnen der Posen der fünf oben genannten Cluster nach der potenziellen Energie des R1 Substituenten, geteilt durch die Anzahl an Schweratomen, möglich, zwischen den „Trapped“ und „non-Trapped“ Propafenon-Derivaten zu unterscheiden. Dieser Wert war bei den „non-Trapped“ Substanzen immer höher als bei den „Trapped“ Molekülen. Der Umstand, dass dies auch bei den Vertretern des fünften Clusters möglich ist, bei denen der R1 Substituent unterhalb der vier Phe656 zum Liegen kommt, deutet darauf hin, dass das Phänomen des Drug-Trappings mehr auf die inhärenten Eigenschaften des R1 Substituenten als auf seine Konformation zurückzuführen ist, wenn er mit dem hERG Kanal interagiert. Dies könnte bedeuten, dass die Starrheit und die Sperrigkeit der Substituenten bestimmt ob Propafenon und dessen Derivate „Trapped“ sind oder nicht, unabhängig vom Bindemodus im hERG Kanal.

1

INTRODUCTION

¹The ensemble of properties which describe absorption, distribution, metabolism, elimination and toxicity (ADMET) are of utmost importance for the drug discovery process. In light of the studies on antitarget proteins responsible for poor ADMET properties of drug candidates, many high-throughput methods have been developed for early identification of compounds with a bad ADMET profile. In the past decade many strategies were developed to screen libraries of compounds and to assess the risk of a bad ADMET profile. There is overwhelming evidence that

¹The introduction is taken from: Schiesaro A, Ecker GF, "Prediction of hERG channel inhibition using *in silico* techniques", chapter of the book of Dr. Gupta SP " Ion channels and their inhibitors" edited by Springer (submitted).

the inhibition of the hERG potassium channel is directly correlated with a lethal arrhythmia called Torsade de Points. Thus, the discovery of unwanted interactions between compounds under development and the hERG channel is a killing criteria. One of the main challenges of antitarget proteins such as hERG, cytochrome P-450, P-glycoprotein and serum albumin is their polyspecificity. Even though the binding sites of many hERG blockers were indentified through docking into homology models, the molecular basis of the hERG polispecificity is not clear yet. Due to the lack of crystal structures, ligand-based approaches were developed in order to predict the hERG activity of candidate compounds. Recent studies show that a combination of different methodologies, such as 2D- and 3D-QSAR with pharmacophore modeling and classification algorithms was quite successful and might be a powerful tool in the in silico screening of compound libraries. Within this chapter we will present selected in silico models developed to predict hERG channel blockers.

1.1 Long QT syndrome

The cardiac action potential is composed of five phases. Phase 0 is characterized by rapid depolarization due to an inward Na^+ current. This phase is followed first by an initial short repolarization (phase 1) due to an outward K^+ current and then by a plateau (phase 2), which is characterized by a slow repolarization rate. During the plateau phase an inward Ca^{2+} current is balanced by an outward K^+ current. In phase 3 of the cardiac action potential the rapid and the slow delayed rectifier K^+ (I_{Kr} and I_{Ks} , respectively) determine the repolarization of the myocyte with its membrane potential to the resting state (phase 4). A long cardiac action potential is a safeguard mechanism that prevents the onset of potentially dangerous arrhythmias. However, this mechanism is not perfect, as a delay of cardiac repolarization prolongs the QT time.

Through the electrocardiogram is possible to register the electrical activity of the heart. The P wave registers the atrial depolarization, followed by the QRS complex, which registers the ventricular depolarization. The T wave finally corresponds to the ventricular repolarization. During the interval between the Q and T waves (QT interval), which is a measure of repolarization duration, the heart is refractory to new excitations. A delay in ventricular repolarization is registered in the ECG as a prolonged QT interval. The Long QT syndrome (LQTS) is associated with increased risk of occurrence of potentially lethal arrhythmias called Torsade de Pointes (TdP). The LQTS is estimated to affect 1 person every 5000-10000 people.[1]

The LQTS is mostly due to mutations of the hERG channel, a potassium-selective voltage-gated channel that conducts the I_{Kr} current, and to the KCNQ1 channel, that encodes the Kv7.1 channel which is responsible for the I_{Ks} current. Up to now there are almost 300 mutations known to affect the hERG potassium channel.

Many classes of compounds known to block the hERG potassium channel prolong the QT interval and thus also might be involved in potentially dangerous arrhythmias such as TdP.[2-21]

1.2 Structure of the hERG channel

The name of the hERG potassium channel originates from the human homologous gene ether-a-go-go (*eag*) found in *Drosophila melanogaster*. The human related ether-a-go-go related gene is also called KCNH2 according to the new nomenclature, and the protein encoded is often referred as hERG, even if the official annotation according to the new nomenclature name is Kv11.1.

The hERG potassium channel has a tetrameric architecture, where each subunit is formed by six transmembrane domains (S1-S6). Up to date two forms of hERG are known, hERG1a and hERG1b, which can form a homo- or hetero-tetrameric structure with different kinetic properties. Each subunit is formed by a voltage-sensing transmembrane domain (S1-S4), and the K^+ -selective pore (S5-S6), which is responsible for the K^+ conduction.

The K^+ channel can be divided into three regions: the K^+ -selectivity filter, the central cavity and the inner pore.[22] The K^+ -selectivity filter is a narrow cylinder built to mimic perfectly an aqueous environment and to conduct specifically K^+ ions. The highly conserved signature sequence in the Kv channels (Thr/Ser-Val-Gly-Tyr/Phe-Gly) at the C-terminal end of the selectivity filter is responsible for the creation of the aqueous environment. The hydroxyl groups of Thr as well as the carbonyl groups of the other four amino acids form several octahedral binding sites which coordinate the K^+ ions analogous to a water filled environment. The diameter of the cylinder is too large for the Na^+ ions and thus they cannot be well coordinated, so they stay in the aqueous environment where the water molecules coordinate the ions.

The central cavity is a wide water filled pore located below the selectivity filter, whose size changes in the open and in the closed conformation.

The inner pore is located at the cytoplasmic side of the membrane and connects the central cavity with the cytoplasm of the cell. It consists of two concentric rings made by four Tyr652 and

four Phe656. The four Phe656 form the first ring located at the cytoplasmatic side of the hERG channel. The four Tyr652 build the second ring which faces the central cavity. These two concentric rings are considered the main feature responsible for the polyspecificity of the hERG channel, mainly due to the possibility to make multiple and compound specific interactions. In the closed conformation the S6 domains intersect at the cytoplasmatic side of the channel, closing the inner pore and preventing the flow of K⁺ ions. During channel activation the S6 domains rotate outwards, opening the inner pore and allowing ions to flow.

The hERG channel and other K⁺-channels contain a PAS domain on the cytoplasmatic N-terminus,[23] which is involved in protein-protein interactions. Although the function of the PAS domain in the hERG channel is not understood, mutagenesis studies show that disruption of this domain reduces the outward current and prolongs the QT interval.[24]

Also the role of the C-terminus domain is not yet completely clarified, although a cAMP binding domain is located there. cAMP accelerates the deactivation kinetics of the hERG channel.[25]

1.3 Ligand based approaches

1.3.1 Pharmacophore models

Several pharmacophore models have been published up to date. They provide a powerful tool for the identification of potential hERG blockers by virtual screening of compound databases. Some pharmacophore models were already reviewed elsewhere.[26, 27]

Morgan and Sullivan[28] analyzed a set of class III antiarrhythmic drugs. In the pharmacophore model the charged nitrogen is linked by 1-3 atoms to two features which might be an aromatic ring or an alkyl moiety. A third feature, consisting of a para-substitued ring, is linked to the charged nitrogen through a chain of 1-4 atoms.

The model derived by Ekins et al.[29] was generated from a training set of 15 compounds taken from the literature. The pharmacophore model contains one positive ionizable feature connected with four hydrophobic moieties.

A “constructionist approach” was used by Cavalli et al.[30] to develop a pharmacophore model. The crystal structure of astemizole was used as template on which molecules with similar

geometry were aligned, leading to the addition of new pharmacophoric features which were not present in the previous pharmacophore. To the initial pharmacophore consisting of a basic nitrogen (N) and two aromatic features (C0 and C1), a fourth feature defined by the aromatic ring (C2) of an astemizole derivative was added. Thus, the charged nitrogen group (N) is connected to three aromatic groups (C0-C2). The authors noted that all the hERG blockers tested have the C0 feature, while some molecules lack the C1 or C2 moieties. They also found that a polarizable function in C0, such as a carboxylic or a sulfonamidic group, may also influence the activity.

Pearlstein et al.[21] developed a “drain plug” model through a combination of hERG homology modeling and the CoMSIA analysis of 22 sertindole derivatives and 10 structurally diverse hERG inhibitors. The analysis of the hERG homology model reveals a “drain plug” picture which is complementary to the shape of the cavity, where the amino acids responsible for the interactions are located. The “drain plug” pharmacophore model consists of two aromatic groups that can interact with Phe656, and a basic nitrogen that might interact with Tyr652. They also suggest that a hydrophobic or an aromatic substituent in this portion of the molecule might improve the potency of hERG blockers through the interaction with Tyr652.

Aronov et al.[31] used a set of 85 hERG blockers to develop three hypotheses of three point pharmacophores. The first two hypotheses are in agreement with the C0-N-C2 and C0-N-C1 pharmacophores published by Cavalli et al.[30] In the third pharmacophore model there is a hydrogen bond acceptor feature placed between the charged nitrogen and the aromatic ring, at a distance of 1.8-3.7 Å from the latter one. Similarly Testai et al.[32] found that an acceptor feature is normally placed within 4-6 Å from an ionizable center.

Aronov et al.[33] also used the pharmacophore elucidation module implemented in MOE to analyze an in-house dataset of 194 uncharged compounds. At the beginning two point queries are generated and used to make more complex n-point queries through a “build up” method. Two five point pharmacophore models containing three hydrophobic/aromatic features and two hydrogen bond acceptors were selected from all generated pharmacophore models. This two five point pharmacophore models, which differ only in the position of one of the hydrogen bond acceptors, were merged into a six point pharmacophore. The two five point pharmacophores were able to correctly classify 78% and 69% of potent hERG blockers of the dataset. The six point pharmacophore matched 21% ($IC_{50} < 10 \mu M$) and 44% ($IC_{50} < 30 \mu M$) of hERG blockers

and 4% of non-blockers. Applying a ClogP > 1 cutoff, the number of false positives was reduced in all three models.

Crumb et al.[34] developed a qualitative pharmacophore model through the analysis of 11 antipsychotic drugs. The model, which consists of one aromatic query surrounded by three hydrophobic moieties, is matched by the most potent hERG blockers present in the dataset.

A collection of 1,075 hERG inhibitors was used by Johnson et al.[35] to develop a series of pharmacophore and QSAR models. As seen in previous studies, the basic nitrogen, which is placed 6-9 Å apart from a centroid of an aromatic ring, was found to be an important feature for hERG blockers. In the second pharmacophore model the effect on the activity of the basic nitrogen is attenuated by the presence of two hydrogen bond acceptors and a lipophilic group. This might be a useful hint for development of compounds with reduced hERG blocking activity. In the third pharmacophore the aromatic rings are placed at the opposite side of the molecule and are separated by 14 bonds. In the fourth model the aromatic ring feature is coupled with a hydrogen bond moiety. The fifth pharmacophore model consists of two aromatic rings placed within 6-13 Å from two hydrogen bond donors.

With the aim to subdivide specific and nonspecific hERG blockers in a dataset of 113 compounds, Kramer et al.[36] developed a method where pharmacophore and QSAR techniques are combined. SA15 (sertindole analogue), clemastine, tolterodine and haloperidol were used to build the first pharmacophore model. Similarly to other published models, it contains a positive ionizable nitrogen feature connected to two aromatic/hydrophobic features, one of which is in close proximity with a hydrophobic spot. The model was tested against the entire dataset and 51 compounds matched the pharmacophore. The potent hERG blockers astemizole, cisapride, flunarizine and sertindole, which match the first pharmacophore model, were used to develop a more specific model. Differently from the first pharmacophore, in this model the hydrophobic queries are directly connected with the charged nitrogen. The second model is similar to the one obtained by Cavalli et al.[30]

Through the analysis of 56 compounds using Catalyst Garg et al.[37] generated seven models. The most predictive pharmacophore model consists of one hydrophobic group (HP), one aromatic ring (RA) and one hydrogen bond acceptor lipid group (HBAI), which are three important features for potent hERG blockers. This model was able to find 22 of the 25 potent blockers in the dataset, showing to be capable to distinguish between potent and non-potent

blockers.

With the aim to avoid the potentially lethal hERG blockage of chemokine receptor antagonists, Shamovsky et al.[38] analyzed the influence of four classes of fragments on the hERG inhibition. They obtained two pharmacophore models. The first pharmacophore model consists of one aromatic ring, one basic center, two hydrogen bond donors and one hydrogen bond acceptor. The fragments that match this pharmacophore increase the hERG blocking potency even if they decrease the lipophilicity of the compounds. In the second pharmacophore model there are two aromatic rings connected with a basic nitrogen. Here the fragments that match this pharmacophore increase the hERG potency by increasing the lipophilicity of the compounds.

Coi et al.[39] generated a “toxicophore” using the docking poses of compounds docked in the hERG channel in the closed state. The analysis of the interactions of the lowest-energy poses with the hERG channel shows that there are several hot spots in the binding site: Ser624 (E), Gly657 (I), the region around Phe656 (H), and four cavities in the region around Tyr652 (C). The docking pose of the potent hERG blocker astemizole was used as “template” to generate the toxicophore. The analysis of the poses of astemizole, haloperidol, ritanserin, R59022, cisapride, spiperone, 8-hydroxy-DPAT, sotalol, quinidine, trifluoperidol, and tetracaine, allowed the identification of the features needed to interact with the hERG channel: (a) the optimal distance between the protonated nitrogen and a hydrogen-bond acceptor is 4.5 Å; (b) aromatic rings located in I or in E. Moreover, the authors suggest that also hydrophobic interactions with the amino acids located in C and/or H should be avoided. To evaluate the toxicophore 18 known hERG blockers/non-blockers were docked into the homology model of the hERG channel in the closed state. The results confirmed that the toxicophore is able to distinguish between hERG inhibitors and non-inhibitors.

In summary, the pharmacophore models suggest that the hERG blockers are characterized by the presence of a protonated nitrogen linked with two or three hydrophobic and/or aromatic moieties. The only exception is the pharmacophore model obtained by Aronov et al., which was generated solely by uncharged molecules. This indicates that the development of uncharged compounds is not a safe way to avoid unwanted inhibition of the hERG channel. The charged nitrogen atom, the hydrophobic and the aromatic features might interact with the amino acids Tyr652 and Phe656. Some pharmacophore models suggest also that hydrogen-bond donor and/or acceptor groups might play an important role for hERG inhibition, probably by interactions with

Thr623, Ser624, or Ser649.

1.3.2 3D-QSAR

The first 3D-QSAR model for prediction of hERG blockers was developed by Ekins et al.[29] through the analysis of 15 molecules with Catalyst. These compounds are characterized by the presence of one ionizable feature and four hydrophobic moieties. The pharmacophore shows the presence of one ionizable feature surrounded by four hydrophobic features. The 3D-QSAR model shows a high correlation with an r^2 of 0.90, and also a good performance in the prediction of the activity of the external test set, with an r^2 of 0.83. The ability of the model to correctly rank the hERG blockers according to their IC_{50} values achieved a Spearman's rank coefficient of 0.76 and 0.77 for the training and the test set, respectively. The excellent performance of the model for both qualitative prediction and quantitative ranking of hERG inhibitors indicates that it is a suitable tool to discover potential hERG blockers.

Cavalli et al.[30] developed a 3D-QSAR model through the analysis of 31 hERG blockers using the CoMFA technique. For most of the molecules the 3D structure was retrieved from the Cambridge Structural Database (CSD) or by adding substituents to the crystallographic structure. 3D structures of additional molecules were generated with SYBYL. The alignment of the molecules was performed using the previously generated pharmacophore. The model shows a good predictive performance with $r^2 = 0.95$ and $q^2 = 0.74$. In a further validation using a test set of compounds not involved in the model generation, an r^2 of 0.74 was achieved. The comparison of the pharmacophore and the CoMFA models shows that the pharmacophoric features C1 and C2 are sterically favorable regions, while C0 is influenced by the steric and electrostatic properties of the compounds. In particular an increased volume in C0 will decrease the activity, while an opposite effect is predicted for charged groups.

Through the application of the CoMSiA technique, 22 sertindole analogues and a set of 10 structurally different hERG blockers were analyzed by Pearlstein et al.[21] The best model reached a q^2 of 0.571. Docking studies performed in a homology model of the hERG channel in the open state could explain the pharmacophore and the CoMSiA models.

In total 882 compounds were used by Cianchetta et al.[40] to develop a GRIND-based model. Four probes representing hydrophobic interactions (DRY), hydrogen bond acceptor (sp^2 carbonyl oxygen), hydrogen bond donor (neutral flat amide NH) and the molecular shape (TIP) were used

to calculate the GRIND descriptors. The correlation between the GRIND descriptors and the pIC_{50} values of the hERG blockers was analyzed through multivariate techniques such as principal component analysis (PCA) and partial least squares (PLS). The dataset was subdivided into two subsets characterized by the presence or absence of a basic nitrogen atom. The 338 hERG blockers which form the subset of compounds without a charged nitrogen, were subdivided into a training set of 322 compounds and a test set of 16 molecules. The model obtained resulted in four latent variables and showed an r^2 of 0.76 and a q^2 of 0.72. The 544 compounds that constitute the charged nitrogen database were split into a training set of 518 molecules and in a test set of 26 compounds. The model obtained from the PLS regression analysis had three latent variables yielding an r^2 of 0.77 and a q^2 of 0.74. The descriptors involved in the two models were practically identical. The authors suggest that this might indicate that the charged and noncharged compounds share the same binding mode. The two models differ in terms of distance between the edge of the molecule and the space between a hydrophobic MIF and a hydrogen bond donor group. In the first case the distance for the optimum space between the two fields generated is 25 Å for the noncharged molecules and 29 Å for the charged compounds. In the second case the DRY/hydrogen bond distance for the noncharged compounds is 14 Å, while for the charged molecules it is 21 Å. Statistical analysis shows that in both models a hydrogen bond moiety close to the edge of the molecules plays an important role.

Johnson et al.[35] analyzed 1,075 compounds through a combination of physicochemical and pharmacophoric descriptors. Least median squares (LMS) regression was used to analyze the 925 compounds of the training set and the resulting model shows an r^2 of 0.65 and a q^2 of 0.66. In a further test using 1,679 compounds the model showed an r^2 of 0.54. These compounds were then clustered based on Daylight Fingerprint Tanimoto using the average linkage method and a similarity cut off of 0.7. In the largest cluster the model achieved an r^2 of 0.32. However, when only the compounds in the training set with a Tanimoto similarity greater than 0.65 were considered, the r^2 for the validation set increased to 0.72.

With the aim to develop a model able to discriminate between hERG inhibitors and non inhibitors, Li et al.[41] combined pharmacophore-based GRIND descriptors and support vector machine techniques. Four GRIND probes representing hydrophobic interactions (DRY), hydrogen bond acceptors (sp^2 carbonyl oxygen), hydrogen bond donors (NH neutral flat amide)

and molecular shape descriptors (TIP) were used. From a library of 495 compounds, only the 192 molecules with IC_{50} values lower than 40 μM were considered. The model generated from the PLS analysis consists of three latent variables and shows a rather poor r^2 of 0.34 and a q^2 of 0.07. The analysis of the predicted versus the experimental pIC_{50} shows that the IC_{50} of sertindole derivatives is overestimated. A better model with $r^2= 0.57$ and $q^2= 0.41$ was obtained considering only the hERG blockers with an IC_{50} value lower than 32 nM and by removing the sertindole derivatives. The analysis of the influence of the descriptors highlights that the descriptor related to two hydrogen donor atoms placed at a distance of 9 Å has the highest principal component coefficient. This descriptor is encoded by the charged nitrogen atom. Also the descriptors related to the presence of a hydrogen bond acceptor and a hydrogen bond donor (6 Å or 9.5 Å apart), to the presence of the hydrophobic moiety and a hydrogen bond donor (9 Å or 16 Å apart) and to the hydrogen bond donor and one of the edges of the molecule (10 Å or 17.5 Å apart), show a high correlation with biological activity. These data are in agreement with the model published by Cianchetta et al.[40]

In a recent study Kramer et al.[36] reported a predictive QSAR model able to distinguish between specific and non specific binding. 113 compounds from the literature were split into six groups of equal activity range. From the 113 compounds 15 molecules were randomly selected as validation set. The remaining 98 compounds were divided into a training set of 75 molecules and a test set of 23 molecules. The uncharged forms of the compounds was used to calculate 1D-, 2D- and 3D-descriptors such as the molecular electrostatic potential (MEP), the local ionization energy (IEL), the local electron affinity (EAL), the local polarizability (POL), and the Shannon entropies (SHANI and SHANE). The first model was obtained using a combination of ϵ -SVR and multiple linear regression descriptor selection. An r^2 of 0.81 for the training set was obtained by using EALmax, EALmin, POLmin, SHANIbar, Naryl, shapeQ2 and shapeQ4 descriptors. The model shows a good performance also for the validation set with an r^2 of 0.70, while the q^2 is 0.50. In the second model the Naryl, shape and EALmin descriptors were selected. The statistical analysis of this model resulted in an r^2 for the training, test and validation set of 0.64, 0.61 and 0.62, respectively. The Naryl descriptor encodes for the number of aromatic rings in the molecule, and its selection highlights that the potency of the blockers is correlated with the number of aromatic rings. The shape descriptors are a measure of the shape similarity with astemizole, cisapride and sertindole. They have a positive sign, indicating that the more the shape of the molecule is similar to those of one of the three compounds cited above, the higher is

the probability of the compound to be a potent blocker. The EALmin descriptor has a positive value, which might suggest that the potency of the hERG blockers is decreased if the minimum electron affinity decreases. Additional six models were generated, all of them showing quite satisfactory performance. These six models describe a specific or an unspecific binding type. In general, the models suggest that the affinity of the blockers is correlated to the similarity of the compounds with the three hERG inhibitors cited above, and with the number of aromatic rings present. To decrease the hERG potency it is necessary to introduce electronegative moieties, such as carbonyl groups, as indicating by the EALmin descriptor. The positive coefficient of the ClogP indicates that decreasing the lipophilicity of the compounds might also lead to a hERG affinity decrease. In addition, the number of hydrogen-bond donors is negatively correlated with the potency, probably due to the desolvation penalty that cannot be compensated by the hydrogen-bond interactions with the amino acids facing the central cavity of the hERG channel.

Ermondi et al.[42] reported a GRIND-based 3D-QSAR model generated by the ALMOND software. They used the 31 hERG blockers that Cavalli et al.[30] used to develop the CoMFA model, and six compounds as a test set. The descriptors were calculated using probes that represent the hydrophobic interaction (DRY), hydrogen bond acceptor properties (sp^2 carbonyl oxygen), hydrogen bond donor properties (neutral flat amide NH) and molecular shape (TIP probe). The PLS analysis resulted in a model with three latent variables, which shows an r^2 of 0.93 and a q^2 of 0.69. The analysis of the model suggests that the presence of aromatic rings on the edges of the molecule, hydrogen bond donor moieties not related to the basic nitrogen, hydrogen bond acceptors far from the aromatic rings and placed at the same distance of the hydrogen bond donors from the aromatic group, increase the potency of the hERG blockers. The 37 molecules were also analyzed using the DRY-DRY GRIND descriptors. Interestingly, the PC1 discriminates between the potent blockers and the weak blockers. In particular, it highlighted that the potent blockers have two or more hydrophobic regions far away from each other. The ALMOND model was compared with the CoMFA model obtained by Cavalli et al.[30] Both models show a comparable predictivity and have problems to predict the activity of lipophilic hERG blockers. In the ALMOND model there is a second hydrophobic feature which is missing in the CoMFA model, probably due to the different calculation methods for hydrophobic interactions.

The 3D-QSAR techniques, such as Catalyst, CoMFA, CoMSiA and ALMOND have proven to

be a powerful tool for hERG potency prediction. Together with the pharmacophore models, 3D-QSAR shed light on the molecular determinants that characterize hERG inhibitors, and provided insights for the potential binding mode of blockers. The 3D-QSAR models further underline the importance of a basic nitrogen, of a hydrophobic/aromatic moieties and of a hydrogen bond donor and acceptor groups.

1.3.3 2D-QSAR

Although 3D-QSAR techniques are a powerful resource in the drug discovery process and 3D-models can be very useful to discover compounds which potentially block the hERG channel, they suffer from limitations due to the requirement of conformational sampling alignment (CoMFA, CoMSiA).

Aptula et al.[43] developed a 2D-QSAR model based on the stepwise regression analysis using the hydrophobicity corrected for ionization ($\log D$) and the maximum diameter for the molecules (D_{\max}) as descriptors. The model, based on 19 molecules, shows a good internal and external validation with an r^2 of 0.87 and a q^2 of 0.73. The analysis of 81 hERG blockers highlights that the most active compounds have a $D_{\max} > 18 \text{ \AA}$, indicating that the cavity of the hERG channel is big enough to accommodate large compounds.

Keserü et al.[44] developed a QSAR model for a series of 55 hERG inhibitors. The model consists of five descriptors, such as $ClogP$, molar refractivity, partial negative surface area, polarizability (Volsurf W2) and hydrophobicity (Volsurf D3). It shows a good performance with an r of 0.97. The five descriptors indicate that lipophilic, polarizable compounds with a basic moiety and a large size might interact with the hERG channel. The model was evaluated with a test set of 13 compounds, and gained an r value of 0.75. In a second test the model was evaluated in the ability to classify 82 active and nonactive compounds selected from the World Drug Index. The molecules of the dataset with an $IC_{50} < 1 \mu$ were considered as actives. The model showed a good performance classifying correctly 83% of the blockers and 82% of the non-blockers. Also a hologram QSAR model was generated using the same data set. The model shows an excellent performance with an r value of 0.98. The test set achieved an r value of 0.90, indicating that the model has also a high predictive power. For the active compounds of the World Drug Index, the model was able to correctly classify 81% of the molecules. The HQSAR model was further tested against 743 compounds approved in the World Drug Index, where it achieved a low rate of

false positive (18%).

Five QSAR models with a q^2 ranging from 0.65 to 0.90 were developed by Fioravanzo et al.,[45] where 29 compounds from a training set and 30 compounds from a test set were analyzed through PLS and PCA techniques with EVA and DRAGON descriptors.

Song et al.[46] used a combination of fragment based descriptors, support vector regression (SVR), partial least squares (PLS) and random forest (RF) to develop a QSAR model to predict the activity of hERG blockers. 71 compounds were initially used to calculate and identify fragment descriptors correlated with hERG inhibition. Subsequently, they were used to build SVR, PLS and RF models. 19 compounds served as test set. The best performance was achieved by the SVR model with an r^2 of 0.91 and a q^2 of 0.64. The model showed also a good performance for the test set with an r^2 of 0.85. The analysis of the fragment descriptors highlights that lipophilic fragments have a positive impact on the hERG activity, while hydrophilic moieties generally decrease the binding affinity. Fluorine and methane sulfonamide are two exceptions on this general rule. The two moieties, despite their hydrophilic nature, are positively related with the hERG activity. The model also reveals that tertiary amines are important for drug binding.

The role of the nitrogen atom on the hERG activity of compounds was studied by Zolotoy et al.[47] They found that a tertiary amino group is present in 84% of the hERG channel blockers with $IC_{50} < 1 \mu M$. In 73% of the inhibitors with $1 \mu M < IC_{50} < 10 \mu M$ the charged nitrogen is located at the periphery of the compounds. In 84% of the weak blockers the amine is primary, secondary or neutral.

The CODESSA program was used by Coi et al.[48] to study a series of hERG blockers. Two experiments were performed. In the first experiment the compounds of the dataset were divided into 55 and 27 molecules for the training and test set, respectively. The first model showed an r^2 of 0.77 using 12 descriptors. The training and the test set used to generate the second model contained 64 and 18 compounds, respectively. The best model selected for this sets showed an r^2 of 0.74 using 9 descriptors. The analysis of the descriptors selected in the two models highlighted that the relative number of double bonds (RNDB), the factorized molecular volume (MV/XYZB), the relative number of carbon atoms (RNCA) and the relative negative charge (RNCG) are the most important descriptors. The first two descriptors are negatively correlated with the pIC_{50} , highlighting that the hydrophobicity and the properties related to the volume of the molecules increase the potency of the hERG blockers. These two descriptors might explain

why the ideal hERG blocker candidates are hydrophobic molecules with a large size, as in the case of some potent hERG blockers such as MK-499 and astemizole. On the opposite, small molecules with high globularity lacking hydrophobic moieties normally show a lower ability to block the hERG channel.

Yoshida et al.[49] analyzed 104 compounds collected from the literature. The model included the topological polar surface area (TPSA), the octanol/water partition coefficient (ClogP), the largest value in the distance matrix (diameter) and the summed surface area of atoms with partial charges from -0.25 to 0.20 (PEOE_VSA-4). They also introduced an indicator variable (Cell) indicating if the drugs were tested in the hERG channels expressed in Human Embryonic Kidney (HEK) cells (value of 0), or in Chinese Hamster Ovary (CHO) cells (value of 1). In the most significant model the statistical analysis showed an r^2 of 0.70 and a q^2 of 0.67. The interactions between the hERG channel and the blockers occur in the inner cavity. Many studies show that the size of the cavity is big enough to accommodate large compounds, and this is fully in agreement with the positive coefficient of the diameter descriptor. The positive coefficient for the PEOE_VSA-4 descriptor might be explained by the interactions with amino acids such as Thr623, Ser624, Ser649, and Tyr652, which are capable to form hydrogen bond interactions. The pore region has a hydrophobic nature due to the high number of hydrophobic amino acids, thus the potency of the drugs increase with the ClogP of the molecules. In contrast, increasing the hydrophilicity of the compounds, represented by the TPSA descriptor, decrease hERG potency.

Seierstad et al.[50] reported a combination of models based on neural network ensembles with different representations of the structural properties. For 439 compounds six sets of descriptors were calculated using DirectedDiversity, comprising 117 Kier-Hall (KH) topological indices, 142 Ghose-Crippen atom types, 166 Isis keys, 150 atom pair descriptors, 49 electrotopologic state descriptors, six common medicinal chemistry descriptors, and 146 2D descriptors calculated with the Molecular Operating Environment (MOE) package. To select the descriptors different feature selection algorithms were employed comprising four filter-based ones such as PCA, correlation with response variable, difference in distribution between actives and inactives, training error of single-feature models, and two wrapper-based such as Forward stepwise selection and Simulated annealing. The neural network and neural network ensemble models were obtained from 10 fold cross-validation procedures. The authors found that the “neural network ensembles have greater generalization ability and are less susceptible to the particular

choice of training and test sets” showing that this methodology is a powerful tool for hERG prediction. In the filter type technique the PCA showed the best performance, although the simulated annealing reached the highest r^2 value. The 2D model, which is based on 20 descriptors selected through the simulated annealing technique, shows an r^2 of 0.76. The model, tested with an external validation set of 40 compounds, achieves an r^2 of 0.52. Each of the 20 descriptors was used alone to build a model and the cross-validation r^2 values were calculated. As in many other models the top scoring descriptors, such as AlogP and L-10-L (which describe 2 hydrophobic moieties separated by a chain of the atoms), show that hydrophobic interactions and π -stacking are important for hERG blocking. The highest r^2 value was obtained with the model generated using only the descriptor F-C₁sp² (fluorines connected to sp² carbons). The importance of this group might depend to its ability to form strong π -stacking interactions. The Goose-Crippen atom type O in phenol, enol and carboxyl OH shows a negative correlation with the hERG inhibition. It is noteworthy that two of the three compounds which show the largest error contain a carboxylic moiety, thus the model seems to underestimate its contribution to the decrease of hERG inhibition.

Ekins et al.[51] showed that also recursive partitioning models can be applied for early discovery of potential hERG blockers. 99 compounds from literature sources were used to develop two models using the ChemTree software package. In the first model 99 compounds were used to generate 100 random models using 564 path length molecular descriptors. The model showed a good IC₅₀ correlation with an r^2 of 0.90, but the prediction of the 35 compounds of the validation test set resulted in an r^2 of 0.33. In the second model 134 compounds were used to generate 694 path length descriptors, which were employed in the generation of 100 random models. The second model showed a lower r^2 value (0.85).

Leong et al.[52] used a combination of pharmacophore modeling and support vector machines to predict the potency of hERG blockers. The model obtained from the 26 compounds of the training set had an r^2 of 0.97, while for the 13 compounds of the validation test set the q^2 value was 0.94.

Recently, Gavaghan et al.[53] studied 1,312 compounds through the combination of D-optimal onion design, partial least squares and principal component analysis techniques. The molecular descriptors calculated with the software packages Selma, DRONE and VolSurf were integrated with fragment based descriptors, which were used to calculate base level PLS models from each

descriptor set. The scores generated from the base level PLS models were combined and used as descriptors in the upper level hierarchical PLS models. The best hierarchical PLS model generated from 13 descriptors and a training set of 436 compounds showed a q^2 of 0.59. The model was tested with an additional external test set of 7,520 compounds, composed of 4,813 actives and 2,707 inactives. Unfortunately the model was unable to distinguish between potent, moderate and nonactive blockers.

Recently a dataset of 68 compounds collected from the literature was used by Garg et al.[37] to perform a 2D-QSTR analysis based on S_sNH2 (primary amines), JX (Balaban index), Kappa-3 (third order of Kier shape index), ADMET_PPB (tendency to bind to plasma protein), Atype_O_57 (atom type O in phenol, enol, carboxyl OH), Atype_O_59 (atom type O in Al-O-Al) and Atype_H_46 (atom type H attached to Csp³) descriptors. The model shows a good prediction ability achieving an r^2 of 0.84 and a q^2 of 0.78. In a validation with an external test set of 12 molecules, the QSTR model shows an r^2 value of 0.70. The importance of the Kappa-3 descriptor might indicate that the molecular shape can play a role in the hERG interaction. The S_sNH2 descriptor highlights that the presence of a primary amine group increases hERG inhibition. The Balaban index is inversely correlated with hERG potency, indicating that it might be possible to decrease the activity of hERG blockers by increasing the branching of the molecule. All the A-type descriptors are negatively correlated with the inhibition of the hERG channel. This suggests that hERG potency decreases by adding electronegative groups.

Shamovsky et al.[38] studied hERG selectivity of 4 classes of chemokine receptor (CCR) ligands. In this study, the baseline lipophilicity relationship (BLR) approach was used to increase the hERG selectivity of compounds. Mathematically the BLR is:

$$pIC_{50} = a \times \log D - k \times \Delta G_{intr} + \text{const} \quad 1$$

pIC_{50} = potency of the compound.

a = hydrophobicity factor of the binding site.

D = n-octanol/water partition coefficient.

ΔG_{intr} = intrinsic binding energy nonrelated with the desolvation.

k = coefficient equal to $(2.303 \times RT)^{-1}$.

The term $a \times \log D$ is the lipophilicity driven component of potency caused by desolvation. The

second term $-k \times \Delta G_{\text{intr}}$ is the intrinsic potency, which depends on the molecular interactions of the ligand with the target. When the hERG selectivity of the compound (C) is taken into consideration, the equation becomes:

$$\text{pIC}_{50}^{\text{C}} - \text{pIC}_{50}^{\text{hERG}} = (\text{pIC}_{50}^{\text{C}} - a^{\text{hERG}} \times \log D) + k \times \Delta G_{\text{intr}}^{\text{hERG}} + \text{const} \quad 2$$

The term $k \times \Delta G_{\text{intr}}^{\text{hERG}}$ does not depend on the interactions with the primary target. The term $(\text{pIC}_{50}^{\text{C}} - a^{\text{hERG}} \times \log D)$ (lipophilicity-adjusted primary potency) is independent on the hERG binding affinity. The lipophilicity-adjusted hERG potency $(\text{pIC}_{50}^{\text{hERG}} - a^{\text{hERG}} \times \log D)$ was used in a fragment-based QSAR analysis to identify moieties that form important non-hydrophobic interactions with the hERG channel. The fragment-based QSAR analysis in a given chemical series ranks the fragments according to their contribution to the hERG potency. The regression analysis using eq. 1 shows that a^{hERG} is close to one for all 4 classes of compounds. This means that if the lipophilicity increases by one unit, also the hERG potency increases. Further analysis of eq. 1 indicates that the lipophilicity has a stronger influence on hERG potency than on hERG selectivity. This renders it quite difficult to enhance the hERG selectivity by decreasing the lipophilicity, suggesting that it is easier to improve the hERG selectivity by increasing the lipophilicity-adjusted primary potency value.

In a follow up study Shamovsky et al.[54] used 464 CCR8 antagonists and 8 hERG mutants to increase the selectivity of the compounds towards the hERG channel. A 2D fragment-based QSAR analysis performed on a subset of 25 spirocyclic CCR8 antagonists revealed that bulky and rigid substituents decrease the hERG affinity due to van der Waals clashes with Phe656.

Based on a dataset of 67 compounds, 50 in the training set and 17 in the test set, Roy et al.[55] combined the extended topochemical atom (ETA) indices for diverse hERG blockers and non-ETA-descriptors with factor analysis followed by multiple linear regression, stepwise regression, and PLS, to develop QSTR models. The best model was obtained combining ETA and non-ETA descriptors (r^2 of 0.619 and a q^2 of 0.546). The model indicates that the hERG affinity is increased by enlarging the size of the molecule and increasing the electron richness, while the presence of a carboxylic group and an aliphatic tertiary nitrogen are detrimental for the hERG activity, which, under the light of all models discussed so far, seems quite unlikely.

Due to their simplicity, the 2D-QSAR were extensively applied to develop models able to predict hERG potency. Interestingly, some 2D-descriptors provided structural information of the central cavity of the hERG channel, suggesting that it can accommodate large molecules. The 2D-

descriptors also provided some interesting insight into the characteristics of hERG blockers, indicating that the nature and the charge of the nitrogen atom might have a strong influence on the potency.

1.3.4 1D-QSAR

A new QSAR technique was proposed by Diller et al.[56, 57] to analyze 230 compounds collected from the literature. In order to minimize the variation on the IC₅₀ values due to the use of different cell lines, the data were corrected by introducing a correction factor for each cell type to obtain a match of the data measured in HEK cells. Through the projection of the atoms of the molecules onto one dimension using multi dimensional scaling, the structures were described as a 1D string of atoms. 6 descriptors were employed to generate the model: Size (number of heavy atoms), C-Aliph-Estate (the electro-topological state of the atom if the atom is a carbon not in an aromatic ring), C-Arom-Estate (the electro-topological state of the atom if the atom is a carbon in an aromatic ring), N-Acc-Estate (10 minus the Estate of the atom if the atom is a nitrogen with a free lone pair), N-Don-Estate (the electro-topological state of a nitrogen with an attached hydrogen), and O-Estate (the electro-topological state of the atom if the atom is an oxygen). The statistical analysis of the model showed a correlation coefficient of 0.68 for the training (189 compounds) and of 0.76 for the test (41 compounds) set, respectively. The descriptors that mainly contribute to hERG inhibition were the N-Acc E-state and the C-Arom-Estate, revealing the importance of the basic nitrogen and the aromatic ring.

The same QSAR technique based on 1D-descriptors was used by Diller et al.[58] The one dimensional representation of the molecules was achieved by multidimensional scaling from 2D topological descriptors. To generate the model 6 descriptors were used: the number of heavy atoms, E-state key for aliphatic carbons, E-state key for aromatic carbons, E-state key for nitrogen atoms with a free lone pair, E-state key for nitrogen atom with an attached hydrogen and E-state key for oxygen atoms. The IC₅₀ values of 230 compounds collected from the literature were corrected as previously described. The analysis of the contribution of the descriptors revealed the positive contribution of the E-state key of the aromatic carbon and of the nitrogen atom with a free lone pair. In particular, the model indicates that a separation of 9-10 bonds between two phenyl rings increases hERG affinity. This is consistent with previously published pharmacophore models, which indicates the importance of the presence of a tertiary amine

linked to two aromatic rings. The final model achieved the mean absolute errors of 0.62 and 0.62 for the training set (189 compounds) and test set (41 compounds). Differently from other publications, the authors used the mean absolute errors as performance measure, which makes it complicated to compare their model with the results already published.

The 1D-, 2D- and 3D- QSAR models discussed highlight the importance of the basic nitrogen and of the hydrophobic and aromatic groups for strong hERG blockers. To reduce hERG affinity, the QSAR models suggest to reduce the number of aromatic rings, to modulate the pKa of the basic nitrogen, and to reduce the lipophilicity of the molecule. It is noteworthy that the compounds with a carboxylic group are unlikely to be hERG blockers.

1.4 Classification models

To discover potential hERG blockers classification techniques became an important and powerful tool. Although they do not need accurate IC₅₀ values, it is necessary to set up a threshold which defines if the compounds are considered as active or as inactive.

Bains et al.[59] chose an IC₅₀ of 1 μM as threshold to define hERG blockers and non-blockers. Tobita et al.[60] used thresholds of 1 μM and 40 μM to classify active and inactive compounds. Roche et al.,[61] Dubus et al.,[62] Ekins et al.,[51] Thai et al.[63-65] and Chekmarev et al.[66] classified the hERG blockers into three different classes: low IC₅₀ (lower than < 1 μM, blockers), medium IC₅₀ (between 1 μM and 10 μM), and high IC₅₀ (higher than > 10 μM, non-blockers). Doddareddy et al.[67] used the cut off values of IC₅₀ < 3 μM, 6 μM and 10 μM to define hERG inhibitors, and of > 30 μM for the hERG inactives. O'Brien et al.[68] selected the value of 20 μM to define hERG inhibitors. Sun et al.[69] and Jia et al.[70] defined a compound as hERG blocker if it has an IC₅₀ value lower than 30 μM. Buyck et al.[71] used the threshold of 130 nM to define if the compounds are hERG blockers or not. Catana et al.[72] used an IC₅₀ of 40 μM as a cut off to define the separation between hERG blockers and non-blockers. Li et al.[41] used the threshold values of 1 μM, 5 μM, 10 μM, 20 μM, 30 μM and 40 μM. Some approved drugs have an IC₅₀ value between 1 μM and 10 μM. Typically a drug is considered safe if the IC₅₀ value is higher than 10 μM.[53] To assess the safety of compounds also the ration between the free therapeutic plasma concentration and the hERG IC₅₀ has to be considered. Generally, a molecule with a margin greater than 30-fold between the free therapeutic plasma concentration and the

hERG IC₅₀ is regarded as safe.[73]

Roche et al.[61] employed different techniques such as self-organizing maps (SOM), PCA, PLS and supervised neural networks to develop predictive models based on 1,258 descriptors in order to classify 472 compounds. The substructure analysis performed with LeadScope did not highlight any moieties exclusive for blockers or for non-blockers, although some weak trends could be noted. In 57% of the non-blockers and in 30% of the blockers 2 hydrogen-bond donors separated by six bonds were present. In 20% of the non-blockers and in 2% of the blockers a benzenesulfonyl groups was found. In 49% of hERG inhibitors the 1-R-4-alkyl-benzene moiety was found. The best prediction performance was achieved with the neural network model, which was able to correctly classify 93% of the non-blockers and 71% of the blockers (95 compounds validation test set). The prediction method was additionally used to analyze virtual combinatorial libraries to demonstrate its applicability for shaping compounds libraries towards low probability to contain potential hERG blockers. The structures based on scaffold one (fig. 1) have a structural moiety common to many hERG inhibitors, while the molecules based on scaffold two are designed to be non-blockers. The prediction highlights that in the library of compounds based on scaffold one there are 58% of potential hERG blockers, while in the second library the possibility to have hERG blockers decreases to 0.1%.

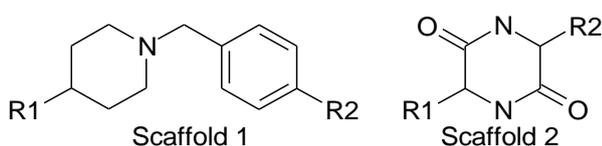


Fig. 1 Chemical scaffolds of the combinatorial libraries.

Bains et al.[59] applied evolutionary programming with fragment-based descriptors to predict hERG inhibition. The resulting model shows an accuracy of 85-90% for the classification of blockers and non-blockers. The model was generated calculating 618 fragment- and non-fragment-based descriptors on 124 compounds randomly partitioned in 70:30:24 for training, generalization and validation data sets. Ten different partitions and subsequently 10 runs for each partition were performed to generate a “consensus model” based on the average of the prediction of the 10 best generalizing models. The partition with the best performance over the data sets was selected. Through a meta-SAR analysis 30 descriptors related to hERG inhibition were identified. Inspection of the selected descriptors revealed that the presence of a secondary or tertiary amino group, of one or more aromatic rings and of a five-membered nitrogen heterocycle

increases the potency of hERG blockers. The model also shows that the presence of negatively ionizable groups such as COOH and of oxygens as H-bond acceptors is detrimental for hERG inhibition. Analysis of the 60 most correlated descriptors with hERG blockade provided a pharmacophore model similar to the already published ones. It consists of a nitrogen atom in the center to which an aromatic and a hydrophobic feature are attached, separated by a linker of 4-5 and 1-2 carbon atoms, respectively. This pharmacophore model highlights again the importance of the presence of a secondary or tertiary amino group linked to two hydrophobic or aromatic features, which might interact with the aromatic amino acids Phe656 and Tyr652. The model also suggests that the amino group should be located asymmetrically between the hydrophobic or aromatic features, in order to interact optimally with the hERG channel.

A set of linear solvation energy relationship (LSER) descriptors were used by Yap et al.[74] to develop an SVM-based classification using a training set of 271 compounds collected from the ArizonaCERT, Micromedex, *Drug Information Handbook*, Meyler's side Effect of Drugs, and from the work of De Ponti and the American Hospital Formulary Service. The model obtained was validated using leave-one-out and Y-randomization methods and model was further tested with an independent validation set of 78 compounds. Furthermore, the prediction accuracy was compared with those obtained from other classification methods such as *k* nearest neighbor (KNN), probabilistic neural network (PNN) and C4.5 decision tree. The SVM classification model shows a higher performance than the other three classification methods, reaching a prediction accuracy of 97.4% and 84.6% for the blockers and non-blockers and an overall accuracy of 91%.

Tobita et al.[60] also generated an SVM model which achieved a prediction accuracy of 90% and 95% with two different test sets. To build the discriminant models they used 73 compounds from the literature for which 57 2D descriptors from the MOE software package, and 51 molecular fragment-count descriptors taken from the MACCS key set were calculated. The support vector machine implemented in the WEKA software package was used to develop two different discriminant models using IC₅₀ values of 1 μM and 40 μM to define blockers and non-blockers. The accuracy of the two models to correctly classify the inhibitors and the non-inhibitors was evaluated through 10-fold cross validation. The model shows an accuracy of prediction of 70% when it was tested with an external dataset of 827 compounds using a threshold 1 μM. In both models the most accurate classification was achieved selecting 8

descriptors. For the model with a threshold of 40 μM the descriptors selected were five 2D descriptors (SlogP, PEOE_VSA6, PEOE_VSA + 1, SMR_VSA5, DIAMETER), and three fragment-count descriptors (number of NH_2 fragments which is correlated to the possible number of hydrogen bond sites, $\text{ACH}_2\text{CH}_2\text{A}$ and $\text{A\$A!A\$A}$ related to flexibility/hydrophobicity of the molecule and the presence of two rings connected by a bond respectively). The fragment $\text{ACH}_2\text{CH}_2\text{A}$ also suggests that the presence of a long chain might play an important role for potent hERG blockers. In the models with a threshold of 1 μM the descriptors identified as important were three 2D descriptors (VSA_BASE, PEOE_Vsa0, SMR_VSA0), and five molecular fragment-count descriptors (OAAAO, $\text{ACH}_2\text{AAACH}_2\text{A}$, Nnot\%A\%A , $\text{ACH}_2\text{AACH}_2\text{A}$ and 8-membered or larger rings). The descriptors selected by the two models are basically different. In the model with the IC_{50} value of 40 μM the 2D descriptors selected are measures of global properties of the compounds, while in the model with the IC_{50} value of 1 μM they are related to a particular structure of the hERG blockers. This highlights that non-potent hERG blockers need to satisfy some general properties such as the DIAMETER and the SlogP, while for potent blockers the presence of certain structural fragments is important. The analysis of molecular fragments shows that the descriptors selected for the threshold at 1 μM are specific cases of the descriptors selected for the threshold at 40 μM . The fragments $\text{ACH}_2\text{AACH}_2\text{A}$ and $\text{ACH}_2\text{AAACH}_2\text{A}$ are related to the global descriptor SlogP. The fragment Nnot\%A\%A is a specific case of the $\text{A\$A!A\$A}$ fragment. Two fragment patterns (pattern1 and pattern2) (fig. 2), common to many potent blockers, were discovered through the mapping of the descriptors selected with the threshold at 1 μM . The fragments of pattern1 are related to the $\text{ACH}_2\text{AACH}_2\text{A}$ and $\text{ACH}_2\text{AAACH}_2\text{A}$ fragments, while the fragments of pattern2 are characterized by a nitrogen atom bound to an aromatic ring.

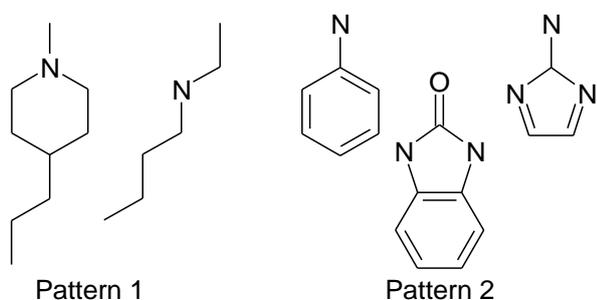


Fig. 2 Fragments related to strong hERG inhibition.

A series of 246 descriptors and the naïve Bayes classification technique were used by Sun et al.[69] A training set of 1,979 compounds from Roche and generic molecular descriptors and

fingerprint-based descriptors were used to generate a classification model. The model achieved a ROC accuracy of 0.87. The model, tested with an external dataset, predicted correctly 58 out of 66 molecules. The fingerprint-based naïve Bayes model was built using FCFP_6 (functional class fingerprints), a 2D-descriptor where each heavy atom of the molecule is described by a string of extended connectivity values, together with physicochemical descriptors such as AlogP, molecular weight, number of hydrogen bond donors and acceptors, and number of rotatable bonds. The model achieved a ROC accuracy of 0.93, indicating that the classification accuracy of the fingerprint-based model is higher than the one of the atom-typing model. The predictive accuracy of both models is similar, as shown by the predictions of the 66 compounds of the test set. Analysis of the most important atom-types indicates that some particular fragments might play an important role for hERG inhibition. The presence of acidic groups abolishes hERG blocking, while basic groups such as piperidines and piperazines are important for hERG blockage. They also observed that compounds which branch immediately after an aromatic moiety have the tendency to be hERG blockers.

Support Vector Machines (SVM) combined with the pharmacophore-based GRIND descriptors were used by Li et al.[41] to design a classification model. To generate the GRIND descriptors 495 compounds were docked into a homology model of the hERG channel in the open state. For every molecule the best scoring pose was selected to calculate the pharmacophoric GRIND descriptors, which were combined with the SVM to generate classification models using cut off values of 1 μM , 5 μM , 10 μM , 20 μM , 30 μM and 40 μM . Four probes were selected to calculate the pharmacophore-based GRIND descriptors: DRY (representing the hydrophobic interactions), O sp^2 carbonyl oxygen (representing H-bond acceptor), NH neutral flat amide (representing H-bond donor) and TIP probe (representing molecular shape descriptors). The model obtained using a threshold of 40 μM showed the best performance. It classified correctly 283 out of 343 non-blockers and 83 out of 152 blockers, with an overall accuracy of 74%. For the external test set comprised of 66 compounds from the WOMBAT-PK database, the model achieved an overall accuracy of 72%, with a correct prediction of 85% and 36% of blockers and non-blockers, respectively. In an additional test using 1,877 compounds from the PubChem database, the model correctly classified 107 out of 187 inhibitors and 1,271 out of 1,690 inactives.

Thai et al.[65] developed two binary QSAR models for the prediction of hERG blockers using two sets of descriptors, 32 P_VSA descriptors, and 11 relevant 2D descriptors such as

hydrophobic descriptors (SlogP, a_{hyd} , SlogP_VSA7, Q_VSA_HYD, PEOE_VSA_HYD), diameter, atom counts (a_{heavy}), bond counts (opr_{nrot}), subdivided surface areas (SMR_VSA5), as well as Kier and Hall connectivity indices (chi1v_C , chi0_C). A dataset of 313 compounds collected from the literature was divided into three classes based on the IC_{50} value: class 1 with $\text{IC}_{50} < 1 \mu\text{M}$ (low IC_{50}), class 2 with $\text{IC}_{50} \geq 10 \mu\text{M}$ (high IC_{50}) and class 3 with IC_{50} in the range 1-10 μM . In order to generate the training and test sets, 184 2D descriptors were calculated on the 313 molecules of the dataset and combined with the pIC_{50} to perform a diverse subset selection, which resulted in 240 compounds for the training set and 73 structures for the test set. A second dataset was generated removing the compounds containing carboxylic moieties. The best binary QSAR model with a cut off at 1 μM (MODEL I) was obtained using 11 relevant 2D descriptors and removing the compounds with carboxylic groups from the training and test sets. The model showed a total accuracy of 0.85 for the training set and 0.94 for the test set. The best binary QSAR model with a threshold at 10 μM (MODEL II) was also based on the dataset without compounds containing carboxylic groups. The model achieved a total accuracy of 0.83 for the training set and of 0.75 for the test set, respectively. Due to the difficulty to correctly classify compounds with IC_{50} values in the range of 1-10 μM , new training and test sets were generated omitting the molecules that belong to this class. The binary QSAR based on 11 relevant 2D descriptors (MODEL III) showed a total accuracy of 0.87 for the training set and of 0.93 for the test set. All three models were further tested with an external test set of 58 compounds taken from the literature and showed a good performance, with a total accuracy of 0.84, 0.78 and 0.86 for MODEL I, MODEL II and MODEL III, respectively.

Counter-propagation neural network (CPG-NN) were used by Thai et al.[64] to develop classification models using 285 compounds collected from the literature and 2 sets of 2D descriptors, one based on 32 P_VSA descriptors and the other one on 11 relevant descriptors. Based on the IC_{50} values the compounds were divided into three classes: class 1 ($\text{IC}_{50} \geq 10 \mu\text{M}$), class 2 ($10 \mu\text{M} \leq \text{IC}_{50} < 1 \mu\text{M}$) and class 3 ($\text{IC}_{50} < 1 \mu\text{M}$). The dataset was split into training and test sets by random division (80:20 and 50:50), or by diverse subset selection (80:20 and 50:50). The best CPG-NN classification performance, obtained with a 3D output layer combined with 11 selected 2D descriptors, reached a total accuracy of 0.93-0.95 for the training set and 0.83-0.85 for the test set, respectively. In particular, this model was able to correctly classify high, moderate and weak hERG inhibitors with an accuracy of 0.93 for class 1, 0.97 for class 2, and 0.96 for class 3. Using only one output layer, the CPG-NN was also used to predict hERG

affinity. The model based on 11 relevant descriptors showed highest performance with an r^2 of 0.87 for the training and the test set.

A new series of fragment/pharmacophore descriptors combined with Support Vector Machine (SVM) and Random Forest (RF) was applied by Catana et al.[72] to develop classification models for a dataset of 561 compounds. An external test set of 1,895 molecules from the PubChem was used to validate the model. Each molecule was hashed into different fragments. Subsequently, each fragment was mapped back onto the dataset compounds, numbered, and a C-fragment descriptor was calculated. The value of the “comprehensive fragment” descriptor (CF) is calculated by summing the contribution of each atoms of the fragment. This implies that the values of each CF descriptor differ for each molecule. The following descriptors were calculated: E-state (CF_E-state_*), AlogP (CF_AlogP_*), MR (CF_MR_*), positive Gasteiger partial charges (CF_GC_P_*), negative Gasteiger partial charges (CF_GC_N_*), the van der Waals surface area positively charged (CF_VSA_P_*), the van der Waals surface area negatively charged (CF_VSA_N_*), the van der Waals surface area (CF_VSA_*), the van der Waals surface area with a positive AlogP (CF_VSA_AlogP_P_*), the van der Waals surface area with a negative AlogP (CF_VSA_AlogP_N_*). In addition, pharmacophore fingerprints implemented in MOE were used to generate a new set of descriptors, called C-pharmacophore descriptors (CP_*) using the approach described above. The model generated by RF using the CF_Estate_*, CF_VSA_P_*, CF_VSA_N_* and CF_VSA_* and some 2D MOE descriptors showed an overall accuracy of 0.79 with a precision of 62.2% and 92.3% for true active and inactive compounds, respectively. For the external test set the model correctly predicted 105 of 193 inhibitors and 1,408 of 1,702 inactives. The model created using the SVM based on the C-pharmacophore descriptors showed a poor performance for the classification of hERG blockers (78 out of 193 compounds were correctly predicted). Conversely, the good result achieved in the prediction of hERG inactives indicates that this model might be used to select non-blockers.

In a recent study *in silico* binary QSAR and counter-propagation neural network were used by Thai et al.[63] to classify hERG blockers and non-blockers. The models were built using a dataset of 243 compounds with similarity-based (SIBAR) descriptors calculated on basis of four reference sets: 24 diverse drugs obtained from Sköld et al.,[75] 20 hERG blockers, 20 hERG non-blockers and 20 compounds divided in 10 blockers and 10 non-blockers. The SIBAR descriptors were calculated from 11 selected descriptors from a total number of 184 2D

descriptors, 86 VolSurf descriptors, 50 3D “inductive” QSAR descriptors (related to atomic electronegativity, covalent radii and intramolecular distances) and 32 P_VSA descriptors. The binary QSAR models were generated using 16 SIBAR descriptors and threshold values of 1 μM and 10 μM . The best classification model was obtained with the 11 hERG relevant descriptors using the 20 most diverse hERG blockers as reference set (total accuracy of 0.85-0.88 for the training set and 0.73-0.92 for the test set). The CPG-NN models were built using the dataset divided into class 1 (low hERG activity), class 2 (high hERG activity) and class 3 (medium hERG activity). The dataset compounds were divided into the training set by random selection (80:20) or by diverse subset selection (80:20). The architecture of the CPG-NNs was designed with 3 output layers representing the classes 1-3. The CPG-NN models were trained with 16 different sets of descriptors. The CPG-NN model obtained using the set of 11 hERG relevant descriptors and the reference set of 20 diverse inhibitors showed the best performance, reaching a total accuracy of 0.73-0.74 for the test set and 0.92-0.93 for the training set. The best binary QSAR and CPG-NN models were validated using 1,806 compounds published in the PubChem compound library and 58 compounds collected from the literature. With the threshold value of 1 μM the binary QSAR model achieved the total accuracy of 0.93. The CPG-NN model correctly classified 68% of the compounds in class 1, 100% of the compounds in class2 and 75% of class 3.

Jia et al.[70] designed a classification model using an SVM and the atom type as molecular descriptors. The model yielded an overall accuracy of 99.59% for the 977 compounds of the training set and of 94% for the 66 compounds of the test set. The use of the atomic molecular descriptors makes the classification model easy to interpret. The most important atom-type descriptors were the N16 (nitrogen atom in an aliphatic ring), C17 (unsubstituted carbon atom next to the N16 in a ring), H4 (acidic hydrogen) and M12 (number of aromatic rings). The nitrogen of the N16 descriptor was usually protonated, and it was found in hERG blockers 308 times out of 322 times of the total occurrence of the descriptor. The C17 descriptor was found 1,635 times of which 1,318 times it was associated with hERG inhibitors. The H4 descriptor was found in 60 structures of which only 5 were hERG positives, in agreement with the fact that negatively charged compounds are normally non-blockers. The M12 descriptor occurred 2,581 times. Only 94 out of 535 molecules containing 3 aromatic rings were hERG inactive, while 14 out 19 compounds containing one aromatic ring were non-blockers.

New descriptors generated from the Shape Signature method were used by Chekmarev et al.[66] in combination with k nearest neighbors (k -NN), support vector machines (SVM) and Kohonen self-organizing maps (SOM) to create classification models. The models were built based on a dataset of 83 compounds divided in strong binders ($IC_{50} < 1 \mu\text{M}$) and weak binders ($IC_{50} > 10 \mu\text{M}$). Two different sets of molecular descriptors were calculated, one based only on molecular shape and the other one based on molecular shape and polarity. The SVM models showed a better performance than the k -NN with an overall accuracy of 69-74% and of 66-67%, respectively.

1.4.1 Decision trees

The decision tree approach was chosen by several groups in order to classify hERG blockers and non-blockers. The ClogP, MR and pKa were used by Buyck et al.[71] Compounds were classified as hERG inhibitors when the following three conditions are satisfied: $\text{clogP} \geq 3.7$, $-110 \leq \text{MR} < 176$ and $\text{pKa} \geq 7.3$.

A Neural Network model based on the E-state keys and Barnard 4096-bit fingerprints, and a Bayesian model based on FCFP_6, AlogP, Molecular Weight, and the counts of hydrogen bond acceptors and donors descriptors were used by O'Brien et al.[68] to generate a consensus model. The models were built using a dataset of 58,963 compounds randomly divided (80:20) to obtain a training and a test set of 46,967 and 11,996 compounds, respectively. The Neural Network model shows slightly better results than the Bayesian model with 85% vs 82% of compounds correctly classified. To improve the ability to correctly classify the compounds the Neural Network and the Bayesian models were combined. The "recover +ve" classifies the compound as hERG blockers if one of the models predicts it to be positive. The "recover -ve" classifies a compound as negative if one of the models predicts it to be negative. With this classification model the number of false positives increases. The "consensus model" classifies a compound as positive or negative if both models agree. The model correctly classifies 91% of hERG blockers and 87% of hERG non-blockers. With the "consensus model" the rates of false positive and false negative are reduced compared to the Neural Network and Bayesian models.

A total of 155 descriptors such as physicochemical, topological, SMARTS strings and SIMAST descriptors were computed on a dataset of 264 compounds by Gepp et al.[76] in order to generate two decision trees composed by a maximum of 8 branches. The first descriptor used in

the two partitioning models is the pharmacophoric string PHARM\$, which correctly classifies 71% of the compounds in both models and only 13 compounds were misclassified as false positives. The two models are identical in the two subsequent layers, which contain the descriptors HACSUR (ratio of surface of hydrogen-bond acceptor atoms to total surface), T1E (topological electronic index using the number of non-hydrogen atoms), HY (number of hydrogen atoms), DIPDENS (dipolar density) and T2E (topological electronic index using the number of bonds between non-hydrogen atoms). The differences between the two models start from the fourth layer. The first model contains 7 branches and the last fourth layers contain the descriptors HLSURF (ratio of surface on halogen atoms to total surface), MDE23 (molecular distance-edge vector λ_{23}), MR, CHBBA (covalent hydrogen-bond basicity), logP, QSUMN (sum of atomic charges on nitrogen atoms) and MGHBD (minimal geometric distance between two hydrogen-bond donor atoms). In the second partitioning model the SIMAST descriptor (fingerprint similarity compared to astemizole) is used in many branches instead of the MR and logP, and the QSUM- (sum of negative ESP charges) replaces the MR in the fifth layer. The first and the second partitioning models achieve overall accuracies for the training set of 91.7% and 93.2%, and of 76% and 80% for the test set.

Recursive partitioning models were developed by Dubus et al.[62] They used 203 molecules from the Aureus Pharma database, 32 P_VSA descriptors and 23 uncorrelated relevant descriptors selected from 184 2D-descriptors calculated with the Molecular Operating Environment (MOE) software. Model1 used an active/inactive boundary of 1 μM . Classification accuracy for the training set of 96% and 97.5%, and of 74% and 81% for the test set was achieved by the model using the relevant and P_VSA descriptors, respectively. In detail the models correctly classified 94% of the strong blockers, while the precision for classification of the weak blockers decreased to 63% for the relevant descriptors model and to 74% for the P_VSA model. Noteworthy, the misclassified molecules mainly showed an IC_{50} between 1 and 10 μM .

In a study performed by Ekins et al.[51] recursive partitioning, Sammon nonlinear mapping and Kohonen self-organizing maps were investigated with the aim to analyze the performance of these techniques individually or in a consensus approach. The recursive partitioning model was built using a training set of 99 compounds providing an r^2 of 0.90. Interestingly, the performance of the test set of 35 compounds was improved (from an r^2 value of 0.33 to 0.83) when the

Tanimoto index was introduced to filter the molecules according to their similarity to those used in the training set. The Sammon nonlinear mapping and Kohonen self-organizing maps models were generated using a dataset of 93 compounds and 8 descriptors selected with the PCA technique from more than 150 descriptors. The 8 descriptors selected are the Wiener index (a measure of molecular branching), the topological Balaban index (provides information on the connectivity and branching of the molecule and is related to the hydrophobic interaction of the molecule), number of H-bond donors, hydrophilicity index and electrotopological state indices (CH_2 , CH and $> \text{N}$, which provide information on the topology, polarity and hydrogen bonding capabilities of the compound). These descriptors suggest that the topology of the molecule plays an important role for hERG inhibition. The 93 molecules of the training set were divided in three classes based on their activity: class0 ($\text{IC}_{50} < 1 \mu\text{M}$), class1 ($1 < \text{IC}_{50} < 10 \mu\text{M}$) and class2 ($\text{IC}_{50} > 10 \mu\text{M}$). The analysis of the non linear map generated with the Sammon non linear mapping technique revealed that the compounds of the classes 0 and 2 were mapped in two different areas. The compounds of class 1 were mapped in a wide area of the map overlapping the areas occupied by class 0 and 2, resulting in a poor prediction ability. The model predicts correctly 86% and 100% of the compounds in the classes 0 and 2, giving an overall classification accuracy of 95%. As happened for the Sammon non linear mapping, also in the map generated with the Kohonen self-organizing map the molecules belonging to class 0 and 2 were mapped in distinct areas, while the area occupied by compounds belonging to class 1 overlapped the sites of the other two classes. The method correctly classified 86% and 79% of the compounds belonging to class 0 and 2, respectively. A consensus analysis performed using the three methods resulted in 86% of the compounds correctly classified in the classes 0 and 2. The consensus approach did not improve the results obtained with the individual methods.

Doddareddy et al.[67] designed 24 binary models by using Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM) to classify 2,644 compounds. 4 molecular fingerprint descriptors belonging to the extended connectivity fingerprints (ECFPs) and to the functional class fingerprints (FCFPs) were chosen. Four representative models out of 24 were selected for further validation. The 4 classification models yielded overall accuracies of 82-86% for the training set and 83-85% for the test set. The selected models were further tested by using the PubChem database. The models showed a poor predictivity in the classification of the test set compounds, with only 78 out of 193 compounds correctly classified. In addition, 50 compounds from the Cambridge database that were predicted as hERG blockers and 10 compounds that were

predicted as non-blockers by two or more models were selected for experimental validation. 18 of the 50 predicted hERG inhibitors showed more than 50% displacement of astemizole, while all the predicted non-blockers were found to be inactive.

The SOM was used by Hidaka et al.[77] to classify the compounds using structural information. First, 37 compounds were divided into three classes depending on whether the activity (pIC_{50}) was below 5, between 5 and 7, or higher than 7. The analysis of the map obtained reveals that the potent blockers and the inactive compounds occupy two different areas, while the compounds with the intermediate activity overlap the two areas. The same method was then applied to the public available dataset. They divided this database into active (“Hit”) and inactive compounds. The “Hit” molecules were subdivided into compounds that cause hERG blockade between 20-30%, between 30-50%, and more than 50%. Considering only the “Hit” compounds they established a line which divides the map into two parts: hERG positive and hERG negative areas. All compounds that block the hERG channel by more than 50% are mapped in the area of the hERG positive potential, with the exception of one false negative.

1.5 Matched molecular pairs

In a recent and interesting study Papadatos et al.[78] applied the matched molecular pairs technique to three large data sets: hERG (76,266 compounds), solubility (94,053 compounds) and lipophilicity (180,440 compounds) to find the most frequent modifications of the molecules. The matched molecular pair analysis of the hERG database identified 15 frequent transformations. These modifications are related to only one or two heavy atoms, except for the substitution of a hydrogen atom with a phenyl ring. In 9 transformations out of 15 the modifications have only a small or no effect for the hERG affinity of the compounds. According to previous results, they found that replacing a hydrogen atom with a hydroxy group is detrimental for the hERG affinity in 45% of the cases, while replacing a hydrogen atom with a phenyl ring increases the hERG inhibition in 65% of the cases. The use of context descriptors, such as reduced graphs, Murcko frameworks and Daylight fingerprints, as well as more local descriptors such as localized RG nodes and atom environments, highlights significant trends that are not evident when only the matched molecular pairs technique is used. For example, considering the substitution of a hydrogen atom with a methoxy group the global distribution indicates that the possibilities to reduce or to improve the hERG affinity are more or less

identical. However, if one takes into account the reduced graph node the scenario appears to be more complex. The possibility to reduce the hERG affinity increases if the transformation is adjacent to an aliphatic chain. The opposite effect is obtained if the transformation is next to a hydrogen-bond acceptor in an aromatic ring. On the contrary, if the transformation is adjacent to an aromatic ring the hERG inhibition is not affected. These results are related to the possibility to form hydrogen-bonds by the oxygen atom of the methoxy group. In the case of the substitution of a methyl group with a fluorine atom, the global distribution indicates that this might result in an increase or a decrease of the hERG affinity. The presence of the fragment A (fig. 3) is correlated with the increase of hERG potency. The analysis of the global distribution shows that replacing the cyclohexyl with the phenyl group can reduce or improve hERG potency. The cyclohexyl >> phenyl transformation in the fragment B (fig. 3) increases the probability to obtain a more potent hERG blocker. This result is due to the increased lipophilicity and hydrophobicity of the molecule, and to the reduced ability to form hydrogen-bonds of the adjacent amidic carbonyl. These three examples show the importance to include contextual information in the drug discovery process to develop compounds with a good toxicological profile.

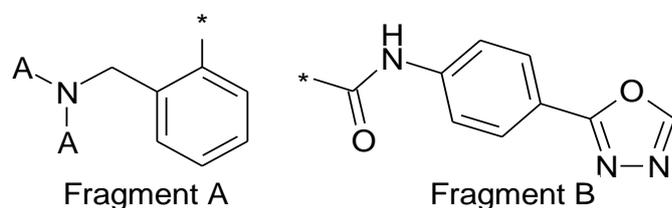


Fig. 3 Fragments found in hERG blockers. The asterisk indicates the attachment point.

1.6 Structure based approaches

1.6.1 Homology models of the hERG channel

Up to date there is no crystal structure of the hERG channel and most of our knowledge comes from studies on the Shaker channel (Kv1.1) and on mammalian channels (Kv1.2), KcsA, MthK and KvAP. Since the eukaryotic and prokaryotic pores are closely related, it appears reasonable to use the crystal structures of KcsA (pdb codes 1BL8,[79] 1K4C[80] and 1R3J[81]) and of KirBac1.1 (1P7B[82]) to build homology models of the hERG channel in the closed state, or

MthK (1LNQ[83]) and KvaP (1ORQ[84]) for the open state. It has to be noted that the degree of the pore opening varies in the crystal structures, from the closed state (KcsA) to the open state (KvaP), up to an even more open state in MthK. These different degrees in pore opening might represent different gating properties of the K^+ channel, or different snapshots in the gating trajectory. The first homology models of hERG channel were discussed also in several reviews.[26, 27, 85-87]

Many groups modeled the hERG channel in the closed state using as template the crystal structures of KcsA (1BL8 or 1K4C),[4-6, 11, 12, 88-94] as well as of MthK (1LNQ),[11, 21, 49, 88, 89, 94] KvaP (1ORQ, or 1R3J)[91, 94-99] and Kv1.2 (2A79)[100] were used as template to construct the homology model of hERG channel in the open state (table 1).

1.6.2 hERG inhibition and drug trapping

In 1969 Armstrong[101] demonstrated that the channel opening is a necessary condition to obtain the inhibition of the voltage gated K^+ channel by a small quaternary ammonium (QA) ion. Normally, the QA interferes with the channel gating upon repolarization causing a slow deactivation due to the impossibility to close the activation gate in a mechanism called “foot-in-the-door”. In a few words, the activation gate cannot close until the inhibitor dissociates from the hERG channel. Armstrong noted that the channel block due to small QA compounds, such as tetraethylammonium, did not show interference with the deactivation rate, suggesting that these compounds might be trapped in the hERG channel in the closed state. A molecule can only be trapped only if it is small enough to fit into the central cavity. If the molecule is charged, it cannot leave the central cavity through the hydrophobic environment of the inner pore or through the membrane, hence the block is irreversible until the channel reopens. Mitcheson et al.[3] used the potent hERG blocker MK-499 and the hERG mutant D540K to test the trapping hypothesis. This mutant of the hERG channel has the particular characteristic to open in response to hyperpolarization. They observed that the channel reopening during the hyperpolarization allowed the recovery from the block of MK499, as postulated by the trapping hypothesis. Moreover, the observation that molecules with large size such as MK499 ($7 \times 20 \text{ \AA}$) can be trapped, suggests that the hERG channel has a central cavity bigger than the one of the Shaker K^+ channel. The Shaker K^+ channel is blocked by tetraethylammonium (6.9 \AA of diameter), but it cannot trap MK499.

1.6.3 Case studies: propafenone derivatives trapping

Docking studies combined with alanine scanning of the amino acids facing the central cavity were performed to investigate the molecular determinants of hERG inhibition by propafenone, as well as the amino acids involved in the drug trapping.[11] The mutagenesis data showed that propafenone's inhibition of hERG was strongly dependent on the interactions between the compound and the amino acid Phe656, while it was not affected by mutations of Tyr652, Thr623, Ser624, Val625, Gly648 or Val659. The analysis of recovery from the propafenone block showed that the compound was not released faster from the mutant channels T652A, V625A and S624A than from the wild-type channel. Only in the case of F656A the recovery from the block is slightly faster. These results suggest that only the mutation F656A slightly reduces the interaction of propafenone with the hERG channel in the closed state. Their results also indicate that the mutagenesis data are better rationalized with the docking poses obtained with the hERG channel in the open state. Almost all of the top ranked poses form π -stacking interactions with two Phe656 of adjacent subunits. The model of the hERG channel in the open state suggests that the four Phe656 are highly accessible to the compound, whereas in the closed state model the space between the Phe656 units is reduced, making it impossible for propafenone to form π -stacking interactions with Phe656. This indicates that there are gating-induced changes in the position of Phe656 side chains.

In a recent study, Thai et al.[102] used 5 propafenone derivatives to perform a systematic analysis of use-dependency and recovery from the block of the hERG channel. The pose of propafenone docked into the hERG channel in the closed state predicts that the phenyl ring forms π -stacking interactions with Tyr652, that the Thr623 makes hydrogen-bonds with the hydroxy group and the protonated nitrogen atom, and that Ser624 forms hydrogen-bonds with the hydroxy moiety. The pose of propafenone docked into the homology model of the hERG channel in the open state shows π -stacking interactions between the phenyl ring and Phe656, and that the hydroxyl and the carbonyl groups as well as the ether oxygen form hydrogen-bonds with Thr623 and two Ser624 of adjacent subunits, respectively. The docking pose of propafenone into the hERG channel in the open state shows that the protonated nitrogen moves in the direction of the cytoplasm. This indicates that the propafenone moves upward when the hERG channel closes. The other trapped compound, the piperidine analogue, shows similar poses. The poses of non-trapped propafenone derivatives bearing a 2,3-dimethylphenyl-1-piperazinyl ring docked into the

hERG open state, predict that the carbonyl group and the ether oxygen form two hydrogen-bonds with Ser624 of different subunits. The conformation of the non-trapped compounds is similar to the pose of propafenone, with the substituent located into the inner pore. The non-trapped 1-[4-{3-[4-(2,3-dimethylphenyl)piperazin-1-yl]-2-hydroxypropoxy}-phenyl]ethanone makes a π -cation interactions with Phe656, and a hydrogen-bond with Ser624. Like the other non-trapped molecule, the 2,3-dimethylphenyl ring is placed in the inner pore. Even if the compound has a smaller volume than propafenone due to the lack of the second phenyl ring, the molecule is not trapped. These poses indicate that electrostatic interactions are important for drug trapping, because the protonated nitrogen is pushed up during the channel gating. Witchel et al.[11] indicated that the amino acid Phe656 plays an important role in the interaction with propafenone, while the mutations of other amino acids to alanine do not affect the hERG block. In this study, the results suggest that also the amino acids Thr623 and Ser624 might interact with propafenone. The main structural difference between trapped and non-trapped compounds is the bulkiness of the substituent attached to the protonated nitrogen atom, which is larger in the non-trapped ones. In light of this consideration, the authors suggest that the size of the substituent might be one of the contributing factors playing a role in the drug trapping/non-trapping. In detail, all these results indicate that the compounds that are small enough and that can alter the conformation can be trapped, in contrast to compounds with a bulky substituent that can prevent the closure of the activation gate with a “foot-in-the-door” like mechanism.

1.6.4 Amino acids involved in hERG inhibition

At the base of the hERG channel, two concentric rings formed by four Tyr652 and four Phe656 are located. The ring formed by the four Phe656 is placed near the cytoplasm, whereas the ring made by four Tyr652 faces the central cavity. These two rings can make hydrophobic, π -stacking and π -cation interactions, explaining the polyspecificity of the hERG channel.

Several authors studied the physicochemical properties responsible for interactions with Tyr652 and Phe656 with the ligands. Fernandez et al.[103] studied the nature of the interactions of MK-499, terfenadine and cisapride with the amino acids Tyr652 and Phe656. They observed that the mutation of Phe656 to Trp, Tyr, Met, Leu or Ile altered the block of MK-499, cisapride and terfenadine only slightly. In contrast, the mutation of Phe656 to the polar amino acids Arg, Glu, Ser and Thr, or to small amino acids such as Ala and Gly increased the IC₅₀. All together these

results indicate that the most important physicochemical feature of Phe656 is the hydrophobic volume and not the aromaticity. Mutation of Tyr652 to Phe or Trp did not influence significantly the sensitivity of hERG to MK-499, terfenadine and cisapride, indicating that the hydroxyl group of Tyr652 is not essential for the interaction with hERG blockers. The IC_{50} of MK-499, terfenadine and cisapride dramatically increased when Tyr652 was mutated to Ile, Val, Ala, Thr, Gln and Glu. All together these results highlight that the aromaticity at position Tyr652 is an important feature for hERG inhibition by these three potent blockers.

Mitcheson et al.[6] found that the hERG block mediated by MK-499 was affected by mutations of Thr623, Ser624, Val625, Gly648, Tyr652, Phe656 and Val659. The experiments performed on terfenadine and cisapride supported the interactions with Tyr652 and Phe656, but not the interaction with Val625. Same results were obtained by Karczewski et al.[104] for MK-499, its enantiomers, and its analogs obtained replacing the hydroxyl group with a hydrogen or an amine group. In contrast, the carbonyl analog was insensitive to the Phe656Ala mutation. This result indicates that the carbonyl oxygen interacts with the hERG channel.

The reduction of sensitivity due to the mutation of Thr623, Ser624 and Val625 to alanine was observed also for clofilium and ibutilide. Sanchez-Chapula et al.[5] determined that the binding site for the weak blocker chloroquine is formed by Tyr652 and Phe656. In contrast to previous results,[103] they found that the mutation Tyr652Phe was detrimental for the affinity, indicating that the hydroxyl group is essential for the inhibition of the hERG channel mediated by chloroquine.

Mutation of the Phe656 and Tyr652 to alanine reduce the binding affinity of the potent blockers MK-499, terfenadine, cisapride,[6, 9] clofilium, ibutilide,[12, 105] thioridazine,[106] and of the weak blockers chloroquine,[5] nifekalant, bepridil,[8] maprotiline[107] and mianserin[108] indicating the importance of the presence of aromatic amino acids in this position.

In contrast to these results, hERG channel block mediated by some blockers was almost insensitive to the mutations of Tyr652 and/or Phe656 to alanine. The inhibition of hERG channels by fluvoxamine was only partially attenuated by the Phe656Ala mutation.[109] Mutation of Phe656 to alanine did not affect the hERG block by dronedarone and amiodarone.[110] Tyr652Ala mutation showed a modest effect on the block mediated by dronedarone, while it had a more marked effect on the amiodarone binding. All these data suggest that neither Tyr652 nor Phe656 interact with dronedarone, while Tyr652 might be part of

the amiodarone binding site. Mutation of Tyr652 to alanine did not affect the hERG block by erythromycin, and mutation of Phe656 to alanine had only a weak effect on the inhibition.[88]

An interesting study of Shamovsky et al.[54] analyzed the contribution of the desolvation component of hERG potency and the nondesolvation-related interactions with the hERG channel. They used 25 3,9-diazaspiro-[5.5]undecane analogs and 8 hERG mutants. The PLS discriminant analysis (PLS-DA) revealed that the desolvation component of hERG potency increases the block of the mutants Tyr652Ala and Tyr652Phe, while the Thr623Ser and Ser624Thr mutants are less sensitive to inhibitors. These results suggest that the desolvation component of hERG potency is related to the lipophilicity of the compounds. Within the nondesolvation-related interactions with the hERG channel the Thr623Ser, Ser624Thr, Tyr652Ala, Phe656Met, Phe656Thr and Phe656Trp mutants increase hERG potency. This indicates that the lipophilic substituents cannot fit into the inner pore due to clashes with Phe656. This hypothesis is confirmed by docking studies on the homology model of hERG in the closed state, which predicts that the molecules cannot fit well at the level of the four Phe656. All together these results suggest that bulky and rigid substituents are detrimental for the hERG affinity.

1.6.5 Hydrogen bonds with Ser624, Thr623 or Val625?

The amino acids Thr623, Ser624 and Val625 lie at the top of the inner cavity. Dougherty et al.[111] mutated the amino acids facing the hERG pore to determine the binding efficacy of astemizol, dofetilide, haloperidol, risperidone, droperidol, pimozide, loxapine, amoxapine, imipramine, fluphenazine, triflupromazine, *cis*-flupenthixol, and amperozide. Astemizole showed an increased IC₅₀ value with the double flourinated Phe relative to the single flourinated Phe and the Tyr652Phe mutants, indicating that Tyr652 interacts with astemizole through π -stacking and/or π -cation interactions. The mutation of Phe656 into the two flourinated phenylalanines did not affect hERG channel block. This suggests that Phe656 may not be involved in the binding of astemizole, or that it forms hydrophobic interactions. Mutation of Thr623 to the non-natural amino acid Thr623-(OH) led only to slight increase of the binding affinity, while the Thr623Val mutant is less sensitive to astemizole. The Ser624Thr mutation slightly decreased the astemizole hERG block, while the affinity for the hERG channel was lost in the Ser624Ala mutation. These results show the importance of the hydroxy group of Ser624.

Perry et al.[12] used site-direct mutagenesis to investigate the interactions of clofilium and ibutilide with hERG channel. The authors performed an alanine scanning of the amino acids of the S6 domain facing the pore of the hERG channel. The results show that both blockers were affected by mutation of the amino acids Thr623, Ser624, Val625, Gly648, Tyr652, Phe656 and Val659 to alanine.

Using the same method, Kamiya et al.[8] determined that the binding site of the hERG blockers E-4031 and dofetilide consists of the amino acids Thr623, Ser624, Val625, Tyr652, Phe656 and Val659. The same effect was obtained for the hERG blocker nifekalant, except for the Val659Ala mutation that did not affect the hERG inhibition and for the Ile655Ala that reduced channel block. The Thr623, Ser624, Val625 and Phe656 mutations reduced the block of the hERG channel by bepridil.

Using alanine scanning, Kamiya et al.[9] showed that amino acids Thr623, Ser624, Tyr652 and Phe656 play an important role in hERG inhibition by terfenadine and cisapride.

Finally, Hosaka et al.[95] applied a site-direct mutagenesis to analyze the interactions between nifekalant and amino acids in the pore region of hERG. The mutation of Thr623, Val625, Gly648, Tyr652 and Phe656 to alanine abolished the block of the hERG channel. The mutant Val625Ala disrupted the K⁺ selectivity. This indicates that the side chain of Val625 might be essential for the stability of the selectivity filter structure, hence the mutant Val625Ala might perturb the surface of the inner cavity reducing indirectly the affinity for nifekalant. The homology model of the hERG channel in the open state showed that Gly648 was not part of the pore channel, so the reduction of the hERG block was due to an indirect action.

1.6.6 Role of Gly648

Siebrands et al.[112] created the hERG mutants T623A, S624A, V625A, G648A, Y652A, Y652T, F656A and F656T by site-direct mutagenesis to analyze the interactions between bupivacaine and the hERG channel. All the mutants abolished the channel block by bupivacaine. The mutation of Gly648 to alanine might lead to a reorientation of the amino acids in the S6 domain, thus indirectly reducing the affinity for the blocker.

1.6.7 Which subunits are involved in drug binding?

Recently it was investigated which subunits of the hERG channel are involved in the interaction

with the blockers.[113, 114] Myokai et al.[113] constructed seven tandem dimers with single or double mutations (Y652A and/or P656A) to test which mutations affect the binding of cisapride. The results brought to light that the binding site of cisapride consists of several subunits. Combining the voltage dependence of the cisapride block, the steady-state block, mutagenesis and kinetic data, it was suggested that cisapride binds at first to the low-affinity binding site formed by the two Phe656 of opposite subunits. Only when the voltage-dependent conformational changes reorient the residues in the pore, cisapride binds to the high-affinity binding site constituted by the two Tyr652 of adjacent subunits and Phe656.

Imai et al.[114] constructed the tandem dimers mutating the amino acids Ser624, Tyr652, and Phe656 of opposite subunits to investigate the binding site of cisapride, E-4031 and terfenadine. The inhibition curves of the mutant channels revealed that cisapride and E-4031 interact with Tyr652 and Phe656 of adjacent subunits and with Ser624, while terfenadine interacts with Tyr652 and Phe656 of opposite subunits but not with Ser624. Based on the results of the inhibition curves, the three compounds were docked into homology models of the hERG channel in the open state. The docking pose of cisapride shows π - π interaction between one aromatic ring of the molecule and Tyr652, and CH- π interaction between the Tyr652 of adjacent subunits and the methylene group near the protonated nitrogen in the piperidine ring. The second aromatic ring of the molecule forms π -stacking interactions with Phe656, and the carbonyl oxygen interacts via hydrogen bond with Ser624. The docking pose of E-4031 indicates that the aromatic rings form π - π interactions with herringbone geometry with two Tyr652 of adjacent subunits, that the pyridine ring makes a hydrogen-bond with Ser624, and that the methyl group in the methansulfonamide moiety forms a CH- π interaction with the amino acid Phe656. The pose of terfenadine shows that the two terminal aromatic rings and the second benzene ring of the benzhydryl moiety interact with Tyr652 and Phe656 through π - π interaction with herringbone geometry. Based on these interactions, the introduction of an electron withdrawing group, or the introduction of a bulky substituent, which disrupts the interactions with Tyr652 and Phe656, can lead to a terfenadine derivative with an attenuated hERG binding affinity. The docking poses of cisapride, E-4031 and terfenadine proposed in this study, in contrast to a previous study,[103] highlighted that π -cation interactions might not always play a dominant role, and that π - π interactions with herringbone geometry and CH- π interactions could be important in the hERG-blocker complex formation.

1.6.8 Influence of *para*-substituents on the phenyl ring

Mutagenesis data and docking studies suggest that the *para*-substituent of phenyl rings in inhibitors forms polar interactions with Thr623 and Ser624.[12] To test this hypothesis, a structure activity relationship of ibutilide derivatives was performed to analyze the influence on the IC₅₀ of different *para*-substituents.[105] The IC₅₀ of ibutilide analogs measured on wild-type hERG channels showed a rank order of nitro > chlorine > amine > amide. Similar results were obtained also with dofetilide derivatives. These results are in agreement with the 3D-CoMSiA and the 3D-pharmacophore models developed by Cavalli et al.,[30] which indicated the importance of a polar or a polarizable region close to an aromatic ring. All these information suggest that it is possible to reduce the hERG potency by modifying the nature of the substituent in the *para* position of phenyl rings and thereby develop compounds with a better toxicological profile.

The docking poses of clofilium and the nitro analog show that the *para*-substituent interacts with Thr623 and Ser624 and that the phenyl ring forms π -stacking interaction with Tyr652. The ethyl group attached to the protonated nitrogen, which is placed in the center of the central cavity, forms hydrophobic interactions with Tyr652. The second ethyl group present in some docked inhibitors may form hydrophobic interactions with a second Tyr652. The docking poses indicate that bulky substituents in *para* position prevent to adopt the conformations of clofilium and its nitro analog. The pose of clofilium predicts that the chlorine atom forms hydrogen-bond interactions with Ser624. In the proposed binding conformation of ibutilide, the *para*-substituent interacts with Thr623, Ser624, and the phenyl ring forms π -stacking interaction with Tyr652. The amide analog cannot adopt the binding mode of ibutilide and interacts with Thr623 and Ser624, probably due to the lack of the sulfonyl and of the hydroxyl groups. The interactions with Tyr652 are weak, and this explains the low affinity of the amide analog for the hERG channel. In all four compounds, the tail makes hydrophobic interactions with Phe656.

1.6.9 Two or three binding interactions?

Mutagenesis studies of the S6 domain identified Tyr652 and Phe656 as important sites of interaction. These results led to propose that the charged nitrogen atom may form π -cation interactions with Tyr652 or Phe656, and that the aromatic moieties of the blockers can make π -stacking interactions with the amino acids cited above. Choe et al.[97] suggested a model with

three key interactions: hydrogen-bond interaction between the protonated nitrogen atom and the carbonyl oxygen of Thr623; π -stacking interactions between an aromatic ring and Tyr652; hydrophobic interaction between an hydrophobic moiety of the blocker and Phe656. To test the three key interactions model, 69 known hERG blockers were divided into eight binding types and further subdivided into two groups based on the number of interactions with the hERG channel (two or three) predicted by the model. Consistent with the three key interactions model, the different distribution of pIC₅₀ values between the two groups reveals that the blockers can form three interactions, and that the compounds, which are predicted to form more interactions, show a higher mean pIC₅₀ value.

1.6.10 Docking studies and prediction of hERG binding affinity

Rajamani et al.[89] developed a two-state binding affinity model to predict the IC₅₀ values of potential hERG blockers. Homology models of the hERG channel in the closed and open states were constructed using as template the crystal structures of KcsA and MthK, respectively. Initially, the S6 helix of the reference closed state model was rotated to match MthK S6 helix, and subsequently the channel was closed rotating the S6 helix by 1°. The partial open state model (10° translation away from the reference model) and the fully open state (19° translation away from the reference model) were used to dock 32 hERG inhibitors. The best pose for each ligand was then minimized within the channel, and finally the minimized pose (bound state) was extracted and minimized in water (free state) to obtain the reference electrostatic and van der Waals energies. The difference in the computed energy values (Δ_{ele} and Δ_{vdw}) between the bound and the free states were used to derive linear regression fits to the experimental pIC₅₀. The estimated interaction energy was used to establish the preference of each ligand for one of the two states. The computed pIC₅₀ values of the 21 ligands that prefer to bind the open state model were predicted with an RMSD of 1.2 [pIC_{50Open}= -0.166(Δ_{vdw}) +0.002(Δ_{ele})]. For the 11 compounds that preferentially interact with the partially open state model, the pIC₅₀ values were calculated with an RMSD of 0.85 [pIC_{50Closed}= -0.155(Δ_{vdw}) +0.0004(Δ_{ele})]. Given the fact that both models had essentially the same coefficients for the (Δ_{vdw}) term, a single model was generated combining the energies for each ligand docked into its best fit state. Plotting the experimental versus computed pIC₅₀ values five outliers were identified. The model obtained [pIC_{50Combined}= -0.163(Δ_{vdw}) +0.0009(Δ_{ele})] omitting those 5 outliers showed an RMSD of 0.56 and an r² of 0.82. The equation reveals that the most important contribution to the hERG affinity

arises from Δv_{dw} , in agreement with the previous observations, which indicated that hydrophobicity and aromaticity were the most important physicochemical features of Phe656 and of Tyr652, respectively.[103]

Österberg et al.[98] used molecular dynamics (MD) simulations and the linear interaction energy (LIE) method to calculate the binding affinity of six sertindole analogues docked into the homology model of the hERG channel in the open state. For each ligand, the pose with the lowest energy was selected from the two or three best clusters, and was submitted to MD simulations. The K^+ ions can occupy the selectivity filter with the 1010 (K^+ -H₂O- K^+ -H₂O) or with the 0101 (H₂O- K^+ -H₂O- K^+) configurations. The binding free energies for hERG indicated that the 1010 is the most favorable conformation. The higher value of binding free energies obtained with the 0101 configuration is due to the electrostatic repulsion between the basic nitrogen of the blocker and the K^+ ion facing the central cavity. The plot of the calculated LIE free energies versus the experimental values shows a good correlation between these two terms.

Also, Farid et al.[99] obtained a good correlation between the predicted and the experimental ligand binding free energy of four sertindole analogs. The compounds were docked into the homology model of the hERG channel in the open state using the induce-fit protocol. The correlation between the Extra Precision (XP) scoring in Glide and the experimental binding affinity shows an r^2 of 0.95. The analysis of the terms of the Extra Precision (XP) scoring indicates that the Glide-XP lipophilic contact and the Glide-XP van der Waals energy terms are favorable for hERG inhibitors, while the Glide-XP penalty for buried polar groups term is unfavorable.

The GOLD docking software was used to dock 56 known blockers into a closed state model.[92] For each ligand, the best pose was selected and the docking GOLDScore fitness was used to derive a linear regression fit to the experimental pIC_{50} . The model achieved an r^2 of 0.60 and a q^2 of 0.56, demonstrating that it can be used to predict the binding affinity of hERG inhibitors.

1.6.11 Case studies: docking studies and improvement of the selectivity

A structure-activity relationship combined with docking studies was used by Price et al.[115] to reduce the hERG affinity in a series of CCR5 antagonists. Docking of the lead compound into a homology model of the hERG channel in the closed state revealed that the benzimidazole group fits perfectly to the lipophilic region described by the four Tyr652. This result suggested that it

was possible to reduce the affinity of the compounds for the hERG channel by replacing the benzimidazole moiety with other moieties. This led to the discovery of maraviroc, a potent CCR5 antagonist, which does not interact with the hERG channel.

Micheli et al.[100] performed an interesting study on the combination of docking experiments with structure-activity relationship (SAR) to improve the affinity of a series of 1,2,4-Triazol-3-yl-thiopropyl-tetrahydrobenzazepines with the dopaminergic receptor D₃ and to avoid the interaction with the hERG channel. The compounds were docked manually into homology models of the hERG channel in the closed and the open state. The docking experiments predict that the lead compound assumes a U-conformation, probably due to interactions with Tyr652, Phe656 and intramolecular π -stacking interaction. The pose shows that the charged nitrogen atom and the quinoxaline ring form hydrogen-bonds with serines in the pore of the hERG channel. Based on the docking results, two strategies were used to tackle hERG liability. In the first one the hydrophilicity of the compounds was increased. This strategy led to a compound highly selective for the D₃ receptor and with a reduced affinity for the hERG channel. The second strategy was to reduce the π interactions between the isoxazolyl group and the hERG channel breaking the coplanarity between the isoxazolyl and the benzazepine moieties. This strategy led to a reduction of hERG activity, without affecting the D₃ potency.

Also, Dinges et al.[116] used a combination of SAR and docking results to develop KDR kinase inhibitors with an optimized hERG profile. One kinase inhibitor was manually docked into a model of the hERG channel in the closed state. The best fit was obtained by orienting the compound parallel to the pore of the channel, with the acetylenic ether group oriented towards the cytoplasmatic side of the hERG channel. The pose predicts that the 1,4-dihydroindeno[1,2-*c*]pyrazole forms π -stacking interactions with Phe656, that the charged nitrogen atom on the *N*-methylpiperazine moiety can make π -cation interactions with Tyr652, and that the external nitrogen forms a hydrogen-bond with Ser624. Based on this pose three strategies were developed. The first strategy consisted on the modification of the basic side chain. This approach led to compounds with a reduced hERG affinity, but the antitumoral efficacy was compromised. In the second strategy the polarity of the acetylenic chain was increased. This led to molecules with an attenuated hERG activity, but also the KDR affinity was reduced, except for the compound bearing the glycol ether moiety that inhibited 76% tumor growth in the MX-1 tumor xenograft model. The third approach used the introduction of groups that disrupt electronically

or sterically the π -stacking interaction between the 1,4-dihydroindeno[1,2-*c*]pyrazole and Phe656. The substitution of the methylene bridge with a carbonyl group reduced the hERG affinity, without affecting the KDR inhibition.

In a recent study, Levoïn et al.[117] combined homology modeling and docking to refine a homology model of the H₃ receptor. The model of the H₃ receptor was optimized in three different ways: ten independent molecular dynamics simulations of the proteins embedded in the membrane; ten independent simulated annealing (SA) runs with the most active compound of ten clusters obtained from the training set of inhibitors of each protein; iterative simulated annealing (ISA) with a rigid potent compound. The ligands were docked in the refined models and subsequently the first ranked pose of each ligand was selected, and the affinity with the protein was assessed using the DOCK_SCORE, Ligscore1-2, PLP1-2, Jain, PMF, and Ludi_1-3 scoring functions. The performance of the prediction was then evaluated using the ROC curve and the area under the curve (AUC). The best performing model in the screening of the training set compounds was the one refined via the ISA method, which achieved an AUC value of 0.73. Similar results were obtained with the test set, where the model showed an AUC of 0.71. To predict also the interactions with antitarget proteins, the same procedure was applied to refine the homology models of hERG and CYP2D6. Also in these cases, the refinement process leads to a better performance of the homology models by producing an AUC > 0.69 for both antitarget proteins.

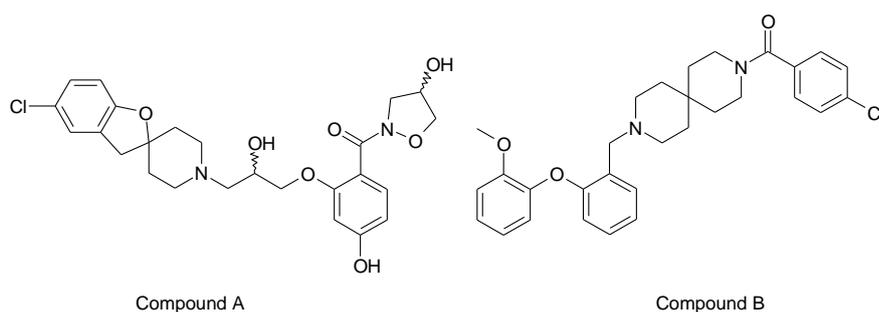


Fig. 4 2D structure of the two compounds docked into the hERG channel.

Recently, Shamovsky et al.[38] used a combination of pharmacophore modeling, QSAR and docking to develop a successful lead optimization strategy that overcomes the undesirable interactions with hERG. The aim of the docking experiments performed on a homology model of the hERG channel in the closed state was to explain the intrinsic hERG binding. To achieve this aim two compounds, which represent extreme cases of intrinsic hERG binding were docked into

the homology model of hERG. The compound A (fig. 4) makes cooperative hydrogen-bond interactions with Ser624, whereas the compound B forms cooperative π -stacking interactions with Tyr652. The docking poses suggest that also the amino acids Leu622, Ser649, and Phe656 are involved in the intrinsic interactions between the hERG channel and the blockers.

1.6.12 Orthogonal binding site?

In contrast to the accepted idea that the blockers bind the hERG channel longitudinally to the z axis, Zachariae et al.[118] proposed an orthogonal binding site. During MD simulations of a homology model of the hERG channel in the closed state the four Tyr652 adopted a “down conformation” with the aromatic ring plane pointing into the central cavity. The “down conformation”, which is stabilized by the interactions with Phe656, opens a way to the membrane. This can explain why the cavity size of hERG in the closed state is larger than the vestibule of KcsA and can accommodate bulky compounds, such as MK-499.[3] Docking studies performed on several compounds such as dofetilide, terfenadine, and cisapride, and previously published pharmacophore models[28-30, 33] corroborate this model. The pose of dofetilide suggests that the charged nitrogen atom is close to the K^+ binding site, and that the two aromatic rings form π -stacking interactions with two Tyr652 of opposite subunits. The binding position of terfenadine shows that the two neighboring aromatic rings of the molecule interact with two Tyr652 of adjacent subunits. In the case of cisapride, the two phenyl rings interact with two opposite Tyr652.

1.6.13 Docking results

The main goal of the docking technique is to model the molecules in the binding site and to find the bound conformation of the compounds as close as possible to the crystal poses, allowing the analysis of the interactions between the ligand and the target protein.

Mitcheson et al.[6] docked MK-499 into a closed state homology model of the hERG channel. The analysis of the poses shows that the p-CN phenyl ring and the benzopyran form π -stacking interactions with Phe656 and Tyr652, while the methanesulfonanilide moiety is placed into the pocket delimited by Gly648, Thr623 and Ser624.

Karczewski et al.[104] docked MK-499 into a homology model of the hERG channel in the closed state. The docking poses of MK-499 predict that the hydroxyl group forms a hydrogen-

bond with Ser624, but the contribution of this interaction to the complex stabilization is modest due to the distance between the two hydroxy oxygens (3.4 Å). In the docking pose of the carbonyl analog the distance between the oxygen of the side chain of Ser624 and the one of the substituent is 3.0 Å. This indicates that the hydrogen-bond between the carbonyl analog and Ser624 is stronger than the previously discussed one, which might explain the insensitivity to the Phe656Ala mutation.

Two possible binding modes of chloroquine were proposed by Sanchez-Chapula et al.[5] In the first binding mode, the molecule forms π -stacking interactions with Phe656 on three subunits, as well as a hydrogen-bond with Ser649, while the diethyl groups of the tail of the molecule form hydrophobic interactions with Tyr652. In the second binding pose, the quinoline group makes π -stacking interactions with Tyr652 and Phe656. The nitrogen atom attached to the quinoline forms a hydrogen-bond with Ser649 and the ethyl group attached to the charged nitrogen interacts with Tyr652 through hydrophobic interactions.

The aligned inhibitors used to generate the CoMSiA model of Pearlstein et al.[21] were docked into a homology model of the hERG channel in the open state. The proposed binding mode suggests that the Phe656 forms π -stacking interactions with an aromatic ring, that a second hydrophobic group can interact with another Phe656, and that the charged nitrogen atom makes π -cation interactions with Tyr652. The tail region of the blockers reaches into the pore region that extends from Tyr652 to the selectivity filter.

Moreno et al.[4] studied the effect of irbesartan on hERG, KvLQT1+minK, hKv1.5 and Kv4.3 channels using the patch clamp technique. The homology models of hERG and hKv1.5 channels in the closed state were generated using the KcsA crystal structure as template. The pose obtained from manual docking of irbesartan shows that the blocker forms a hydrogen-bond and π -stacking interactions with Tyr652, π -stacking interactions with Phe656 and van der Waals interactions with Thr623.

The pose of clofilium proposed by Perry et al.[12] shows π -stacking interactions between the Tyr652 and the phenyl ring of the molecule, in addition to the hydrophobic interactions of the aliphatic tail of clofilium with Phe656 and the interactions of the chlorine with Thr623, Ser624, and Val625.

Duncan et al.[88] docked erythromycin into a model of the hERG channel in the open state. All the poses with low energy interact with Phe656, while the interaction with Tyr652 is prevented

by the large size of the molecule that restricts the ability of erythromycin to move up into the inner cavity. The docking poses obtained with a model of the closed state shows high energy values due to steric clashes, indicating that the molecule cannot fit into the closed state channel. Visual inspection reveals that the molecule is too large for the central cavity of the hERG channel in the closed state.

Yoshida et al.[49] docked the hERG blocker pimoziide into an open state model to examine the correspondence between the molecular determinants derived from the 2D-QSAR model with the 3D structure of the channel. The docking pose suggests that the fluorine atoms and the carbonyl group of pimoziide form hydrogen bond interactions with Ser649 and Ser624, respectively.

Hosaka et al.[95] performed a flexible docking of nifekalant into an open state model. Consistent with the mutagenesis data, docking simulations suggest that the entire molecule is placed in the central cavity surrounded by Thr623, Ser624, Tyr652, and Phe656 of different subunits and located in close vicinity to Ser649.

Österberg et al.[98] docked a series of sertindole analogs into a homology model of the hERG channel in the open state. The docking pose of sertindole suggests that the carbonyl oxygen forms a hydrogen bond with the water molecule located in the selectivity filter. The imidazolidinone group makes good interactions with Thr623 and Ser624. Interestingly, the charged nitrogen is almost superposed to the crystallographic position of the K^+ ion in KvaP.

Farid et al.[99] used the induce-fit docking protocol to dock 12 known hERG blockers into hERG channel models in the open and closed state. S-terfenadine interacts with four Tyr652 and two Phe656. It forms T-shaped interactions with four of the six aromatic amino acids. The docking results show that the poses of R- and S-terfenadine are similar. S-terfenadine interacts simultaneously with four Tyr652 and two Phe656. The replacement of a T-shape interaction with a parallel one is the only important difference between the S- and R-terfenadine. S-terfenadine forms hydrogen-bonds with the backbone oxygen of the amino acid Leu622, and with the side chains of Ser624 and Ser649. R-terfenadine makes four hydrogen-bonds, three of them are with Ser624 of different subunits and the fourth one is with Tyr652. Also, the poses of (+)-cisapride and (-)-cisapride are similar. (+)-cisapride interacts with three Tyr652 and two Phe656, making two T-shaped interactions with the Tyr652 on opposite subunits, and a parallel interaction with one Phe656. The pose predicts also that the piperidine NH forms a hydrogen-bond with Ser624. The only difference between (+)-cisapride and (-)-cisapride is that the interaction with Phe656 is

replaced by an interaction with Tyr652. MK-499 interacts simultaneously with four aromatic amino acids (two Tyr652 and two Phe656). The molecule is predicted to be near Thr623, Ser624, Ser649, and Ala653, with which it can interact. The aliphatic chain of *S*-ibutilide interacts with three Tyr652 and one Phe656. The phenyl ring forms a T-shaped interaction with Tyr652. The methanesulfonamide group and the basic nitrogen are predicted to make hydrogen-bonds with two Ser624 of adjacent subunits. The pose of *R*-ibutilide is a mirror image of the one of *S*-ibutilide. Clofilium makes simultaneous interactions with four Tyr652 and three Phe656. The pose predicts one T-shaped interaction with Tyr652 and two with Phe656. Sertindole interacts with three Tyr652, two of which make T-shaped interactions. The docking pose shows also interactions with one Phe656. The pose of sertindole A5 predicts the interactions with two Tyr652 and one Phe656. It makes T-shaped interactions with one Tyr652 and one Phe656. The pose predicts also a hydrogen-bond between the backbone oxygen of Thr623 and the dimethylamine NH group. Sertindole A1-A4 interact with Tyr652 and Phe 656. Moreover, it makes hydrogen-bonds with Ser649 and with the backbone oxygen of Leu622.

Masetti et al.[91] constructed a homology model of the hERG channel in the open and in the closed state. Subsequently, the two models were embedded in a membrane bilayer, solvated, and the system was neutralized. Both models were subjected to molecular dynamics simulations of 5ns. The docking of astemizole into the hERG channel in the closed state could not provide any reasonable pose. Reliable binding poses were identified only when astemizole was docked into snapshots of MD simulations of the hERG channel in the open state. The top ranked pose shows that the benzimidazole ring forms π -stacking interactions with Tyr652 and parallel displaced π - π interactions with Phe656 of the same subunit. The *p*-fluorophenyl interacts with Tyr652 through parallel displaced π - π interactions. The pose predicts also a possible hydrogen-bond between the fluorine atom and Ser624, and the charged nitrogen atom of the piperidine ring forms a π -cation interaction with Phe656. The *p*-methoxybenzene ring is shown to be exposed to the cytoplasm.

The pose of terfenadine reported by Du et al.[92] shows that the *t*-butylphenyl moiety forms hydrophobic interactions with the amino acids Ser649, Tyr652, Ala653, and Phe656, whereas the diphenylmethanol group makes hydrophobic interactions with Ser624, Ser649 and Tyr652. The phenyl ring of the hERG blocker ibutilide is predicted to form hydrophobic interactions with Thr623, whereas the alkyl chain makes hydrophobic interactions with Thr623, Ser649, Tyr652, Ala653 and Phe656. Furthermore, the nitrogen atom of the methylsulfonamide moiety can form

a hydrogen-bond with Thr623.

Singleton et al.[93] docked a series of dofetilide derivatives bearing a fluorescent group into a hERG channel homology model of the closed state. The poses highlight that the compounds lie in the inner pore, where they interact with Tyr652 and Phe656. The polycyclic conjugated dyes occupy the central cavity.

Models of hERG channel in the open and closed state were generated by Stansfeld et al.[94] using as template the structure of MthK, KvaP, and KcsA. As the models showed only partial agreement with mutagenesis data, a series of KcsA-based intermediate models were generated rotating the four S6 domains. To obtain homology models where the Phe656 can interact with the blockers also a series of intermediate models, which simulate the channel opening, were created. The docking poses of compounds E-4031, dofetilide, ibutilide, and dronedarone show that the methanesulphonamide makes hydrogen-bonds with Thr623 and Ser624. A phenyl ring is predicted to form π -stacking interactions with Tyr652. The methanesulphonamide MK-499 does not interact with the amino acids Thr623 and Ser624. The pose of dronedarone predicts that the charged nitrogen atom is placed in the same position of the K^+ identified in the inner pore in the KvaP crystal structure. The docking pose of fluvoxamine suggests that the protonated nitrogen is placed below Phe656, whereas the trifluoromethyl group lies in the central cavity and shows hydrogen-bond interactions with Thr623 and Ser624. The poses of propafenone and vesnarione make π -stacking interactions with Tyr652. For propafenone, it is also predicted that the charged nitrogen atom is placed between the Phe656 residues, with which it can form π -cation interactions. The pose of terfenadine indicates that this compound interacts with Tyr652 and Phe656. In the case of clofilium and cisapride the molecules make π -stacking interactions with Tyr652, whereas the polar groups interact with the amino acids Thr623 and Ser624. In the docking poses of quinidine and chloroquine, the basic nitrogen is placed above Phe656 and the ring system forms π -stacking interactions with Tyr652.

In summary, numerous docking studies have been conducted and they support findings from QSAR and mutation studies. However, with respect to prediction of strong hERG binders, docking definitely suffers from the still unsolved issue of proper binding free energy calculations. In addition, the channel is quite flexible and compounds might bind to the closed, semiopen and/or open states. This renders hERG binding prediction solely based on docking quite risky.

1.7 Conclusions

In this chapter several methods to discover potential hERG blockers have been reviewed. The hERG channel opening and gating are multistep mechanisms and the blockers might inhibit the channel at any step of the processes. Even if the ligand-based methods presented highlight some structural trends for hERG blockers, and the docking results show that the interactions with some amino acids are crucial, the general limitations of *in silico* techniques render it difficult to perform virtual screening for discovery of potential hERG blockers. Thus, biological tests during the lead optimization process are still the first choice to develop compounds that don't interact with the hERG channel.

Table 1. Published homology models of the hERG potassium channel in the closed and open state.

Ref.	Template	Homology model / Docking
[6]	KcsA (1BL8)	Insight II / FLOG
[5]	KcsA (1BL8)	MOE / FLOG
[21]	MthK (1LNQ)	Modeller
[4]	KcsA (1BL8)	QUANTA/CHARMm / Manual docking
[119]	KcsA (1BL8)	SWISS-MODEL / Chimera
[12]	KcsA (1BL8)	Modeller v6.0 / GOLD
[11]	KcsA (1BL8), MthK (1LNQ)	Insight II / FlexiDock
[89]	KcsA (1K4C), MthK (1LNQ)	PRIME / Glide
[105]	KcsA (1BL8)	Modeller v6.0 / GOLD
[98]	KvaP (1ORQ)	SWISS-MODEL / AutoDock 3.0
[88]	KcsA (1BL8), MthK (1LNQ)	Insight II / FlexiDock
[97]	KvaP (1R3J)	Modeller v.8.0 / GOLD v.2.2
[49]	MthK (1LNQ)	MOE / Glide
[99]	KvaP (1ORQ)	PRIME / Glide 3.0
[120]	Kv1.2 (2A79)	SWISS-MODEL / AutoDock 3.0
[90]	KcsA (1BL8)	SWISS-MODEL / ZDOCK
[95]	KvaP (1ORQ)	Modeller v6.0 / SievGene
[96]	KvaP (1ORQ)	Modeller
[100]	KcsA (1BL8), Kv1.2 (2A79)	QUANTA / Manual docking
[39]	KcsA (1BL8)	Insight II / DOCK v.5.2
[121]	KcsA (1BL8)	Moloc / Manual docking
[91]	KcsA (1K4C), KvaP (1ORQ)	Modeller v.7.7 / GOLD v.3.0.1
[92]	KcsA (1BL8)	Modeller v.7.7/Insight II / GOLD v.3.1
[116]	KcsA (1BL8)	Insight II / Manual docking
[93]	KcsA (1BL8)	Modeler v6.0 / GOLD
[94]	MthK (1LNQ), KvaP (1ORQ), KcsA (1BL8)	Modeler v.6.2 and Modeler v.7.7 / GOLD v.3.0
[117]	MthK (1LNQ)	Modeler v.8.2
[38]	KcsA (1R3J)	Insight II / Glide
[104]	KcsA (1BL8)	Insight II / FLOG
[102]	KcsA (1K4C)	Modeller / FlexX
[118]	KcsA (2BOB, 2HjF and 2HVK)	Modeller / GOLD

2

THE TRAPPING PHENOMENA

Several studies demonstrate that drugs block the hERG channel when it is in the open or inactive state. During the closure of the activation gate some drugs dissociate from the hERG channel with a mechanism called “foot in the door”, or through rapid unbinding upon repolarization, while some drugs are trapped into the hERG channel in the closed state. In the “foot in the door” type mechanism the blocker do not allow the channel to close completely, leaving open a passage in the inner pore which is used by the drug to dissociated from the channel in the closed conformation.

The first evidence of the drug trapping phenomenon was provided by Armstrong et al.[101, 122, 123] Performing experiments on voltage-gated K^+ channels, they demonstrated that the QAs inhibit the channel only if they are present in the cytoplasm, and that the inhibitors can gain

access to the central cavity only if the channel is in the open state. They observed that several QAs prevented channel gating, with a mechanism which was then called “foot in the door”, while small QAs slowed the recovery, hypothesizing that the compounds were trapped into the closed channel conformation. Other studies demonstrated that when the Na^+ and K^+ ion channels are inhibited they can close without affection on the deactivation rate, and that after the drug washout the block persists. These results indicate that the drug can be trapped into ion channels in the closed state. Consistently with the trapping hypothesis, Carmeliet showed that the recovery from the almokalant block is negligible at -75 mV, when the channels are closed, and slow at -50 mV, when the channels can open, even if the drug is not present in the bathing medium.[124] These results indicate that the opening of the activation gate is necessary for the drug-ion channel dissociation. The confirmation of the drug trapping hypothesis's correctness came from Mitcheson et al.[3] through the mutant D540K. The hERG mutant D540K has the peculiarity to open at hyperpolarized potential. Hence, after the membrane hyperpolarization the mutant channel reopen facilitating the hERG mutant-drug dissociation and the recovery from the block. Noteworthy, the size of the ligand used (MK-499; 7×20 Å) indicates that the pore of the hERG cavity is bigger than other K^+ channels.

Stork et al.[10] analyzed the use-dependent block and the recovery of eight hERG blockers divided into two groups: group 1 (amiodarone, cisapride, droperidol and haloperidol), and group 2 (bepridil, domperidone, E-4031 and terfenadine). The drugs of the first group show a complete recovery from the block in contrast to the compounds of the second group, which are trapped into the closed state of the hERG channel. Interestingly they found that the molecules in the group 1 can dissociate from the closed hERG channel conformation. A possible explanation is that these drugs might not allow the channel to close completely through the “foot in the door” mechanism. They found that the drug trapping phenomenon was not correlated with the lipophilicity, volume, accessible surface area and molecular size of the ligands.

In summary, the drug trapping phenomenon is characterized by ultra-slow recovery/lack of recovery from the block, as demonstrated also by other groups such as Kikuchi et al.[125] and Kamiya et al.[8]

2.1 The trapping phenomenon in the propafenone and derivatives

As other hERG blockers, propafenone has to cross the cellular membrane and to be present in the cytoplasm to inhibit the hERG channel.[16] Witchel et al.[11] investigated the molecular determinants involved in the hERG inhibition and in the drug trapping phenomenon. In contrast to the results obtained by Arias et al.,[13] which suggest that propafenone blocks the open and inactive conformations of the hERG channel, they demonstrated that it binds to the hERG channel in the open state and that the inactivation is not mandatory for the channel block. As already found by Carmeliet et al.,[124] they demonstrated that propafenone can exit from the central cavity only if the hERG channel is in the open state. Alanine scanning showed that only Phe656 is important in the propafenone inhibition, while it was weakly or not affected by mutations of Thr623, Ser624, Val625, Gly648, Tyr652 and Val659. Mutagenesis data highlighted also that in the drug trapping only the amino acid Phe656 has a small influence. The four Phe656, which form an aromatic ring at the cytoplasmic side of the channel, have different accessibility when the channel is in the open or closed state. In the open hERG conformation the opening of the activation gate improves the distances between the four amino acids above-mentioned, making possible their interactions with the compounds, while in the closed state this space is too narrow, the molecules cannot interact with them and the trapped compounds cannot dissociate from the channel. The docking poses of propafenone docked into an open hERG model show multiple π -stacking interactions between the ligand and Phe656 of adjacent or opposite subunits. In contrast to the results obtained with the hERG model in the open state, in the closed conformation the low energy poses cannot completely rationalize the mutagenesis data, suggesting that the interactions between the molecule and the hERG channel are stronger when it is in the open state rather than in the closed one. The authors suggest also that in the closed hERG channel conformation the distance between the C α of Phe656 of opposite subunits might be bigger than the 10 Å suggested by their model. They also observed a direct influence on the propafenone binding from the electric field, forcing the molecule to dissociate from the central cavity when the membrane is hyperpolarized and the channel is in the open state.

In recent studies Thai et al.[102] and Windisch et al.[126] investigated the drug trapping phenomenon through the voltage clamp and docking techniques applied on a series of propafenone derivatives, characterized by a common scaffold and different substituents attached to the protonated nitrogen. In agreement with previous studies,[11] Thai et al.[102] observed that

from the closed and open state of the hERG channel there is a small movement of the amino acids at the base of the selectivity filter, while the amino acids located in the inner cavity show a significant movement during the channel gating, indicating that this portion is responsible of the channel closure. The volume of the channel pore decrease from 1642.2 Å³ (open state) to 1048.2 Å³ (closed state), with a reduction of more than 30%. Their data demonstrate that the non-trapped compounds can dissociate from the hERG channel in the closed state with the “foot in the door” mechanism. The docking studies performed into the homology models of hERG channel in the open and closed state, show that propafenone forms hydrogen bonds with Thr623 and with two Ser624 of adjacent subunits, and π -stacking interaction with Phe656.[102] During the channel gating the protonated nitrogen moves up into the central cavity where it makes hydrogen bonds with Thr623 and two Ser624, the hydroxyl group of the ligand conserves the hydrogen bond with the amino acid Thr623, and the phenyl ring forms π -stacking interaction with Tyr652. These poses suggest that during the channel closure the charged nitrogen of trapped compounds moves upward into the central cavity, probably due to the negative field in the channel pore. The docking poses show that the non-trapped propafenone derivatives place the bulky tail attached to the protonated nitrogen into the ring formed by the four Phe656 when the hERG channel is in the open state, with which the basic nitrogen can form π -cation interactions. The bulky and rigid moiety located at the center of this ring hinders the channel gating through the “foot in the door” mechanism. Taken together these results suggest that small and flexible molecules can be trapped into the closed state of hERG channel, while bulky and rigid compounds are forced to dissociate from the channel during the gating process. Windisch et al.[126] observed a correlation between the presence of a hydrogen bond acceptor in the tail attached to the protonated nitrogen, the pKa, the rigidity of the molecule, and the recovery from propafenone derivatives block, while the trapping phenomenon was not correlated with the molecular weight. The non-correlation between drug trapping and the molecular weight indicates that the size of the molecule is not sufficient to determine whether a molecule can be trapped or not. Their results suggest that the pKa of the compounds is an important feature in the drug trapping. In a few words, the trapped molecules have pKa values higher than the non-trapped propafenones, indicating that the protonation state of the molecule might hinder the drug dissociation during the channel closure. On the opposite, a bulky tail attached to the charged nitrogen of propafenone derivatives might hinder the channel gating. The experiments suggest that the molecules enter rapidly into the central cavity and block the hERG channel when it is in

the open conformation. The mutagenesis data indicate that Phe656 is essential in the hERG inhibition, while the Tyr652Ala mutant shows only a reduced sensitivity to the block mediated by propafenone and derivatives, as already found by Witchel et al.[11] In agreement with Carmeliet et al.[124] and with Witchel et al.[11] they found that the trapped compounds dissociate from the channel when it is in the open conformation, hence the propafenones enter and dissociate from the hERG channel in the open state.

Summarizing, almost all compounds which block the hERG have to be present in the cytoplasm, and they inhibit the channel efficiently when it is in the open or inactive conformation. During the channel gating many compounds dissociate through the “foot in the door” mechanism, or unbinding upon repolarization. Some compounds do not dissociate from it during the channel closure and are thus trapped into the closed channel state. The biological studies show that these trapped molecules can dissociate from the complex only when the channel reopens.

2.2 How is possible to investigate in silico the trapping phenomena?

Several techniques, such as classification methods, pharmacophore modeling, docking and molecular dynamics simulations can be applied to investigate the drug trapping phenomena. Through the classification methods it is possible to discriminate between trapped and non-trapped compounds, and investigate the chemical-physical properties of hERG blockers, but it does not provide information regarding the binding modes. Also, the 3D-pharmacophore modeling technique can be used to investigate the trapping/non-trapping phenomena creating two pharmacophore models, one for the hERG channel in the closed state, and one for the open state. The models provide important information, such as the spatial distribution of molecular properties when the ligands bind the hERG channel in the open or closed conformation, but it is not possible, as for the classification methods, to hypothesize a possible mechanism explaining the drug trapping phenomenon. The docking technique can be applied to investigate the binding modes of trapped compounds when they interact with the hERG channel in the closed state, and of blockers when they bind to the channel in the open conformation, and hypothesize possible mechanisms to explain drug trapping phenomena and drug dissociation. Finally, molecular dynamics is a powerful method that can be used to investigate drug trapping phenomena. With this technique it is possible to determine the stability of binding modes, investigate the protein plasticity, simulate the drug trapping mechanism and the channel gating, identify intermediate

states occurring during the channel closure, and study the drug-hERG dissociation.

3

COMPUTATIONAL BACKGROUND

The advent of computers deeply revolutionized every scientific field. The continuous increases of computing power allow scientists to study in detail complex systems, such as protein-ligand interactions, chemical reactions, etc..., and to develop complex models able to explain the observed phenomena. In particular, in drug discovery the computational techniques are used to classify compounds into active and inactive (classification techniques), to predict the activity (QSAR), to predict the protein structure (homology modeling), to study the ligand-protein interactions (docking), to study the protein dynamics (molecular dynamics), and to investigate the characteristics of a compound necessary to interact with the target (pharmacophore modeling).

In this thesis, the docking technique was used to investigate the interactions between

propafenone derivatives and homology models of the hERG channel in the open and in the closed state. In the following chapter I will illustrate the method in more detail.

3.1 Docking

The docking technique is a computational method used to predict ligand-protein or protein-protein interactions. In drug discovery, this method is used to predict the orientation of a molecule in the binding site of a target protein, with the aim to study the interactions, to predict the activity, and to investigate how we can increase or decrease the potency of the investigated drug.

A ligand-protein interaction is a system with a high number of degrees of freedom, which is difficult to simulate without introducing approximations. To solve the problem in reasonable calculation time it is necessary to decrease the degrees of freedom of the system by e.g. specifying the binding site, fixing bond lengths and allowing rotations only around single bonds. Docking techniques can be classified into three classes, which treat the ligand and the protein with different degrees of freedom:

1. *Rigid docking*: the ligand and the protein are rigid bodies and rotation around bonds is not allowed;
2. *Semiflexible docking*: the ligand is considered as a flexible body and rotation around bonds is allowed, while the protein is considered as a rigid body;
3. *Flexible docking*: ligand and protein are flexible.

Of course, the calculation time and the accuracy of results increase with the number of degrees of freedom considered. Nowadays almost all docking programs perform a semiflexible docking.

3.1.1 Search algorithms

The methods used to search the conformational space of the ligand can be deterministic or stochastic. A deterministic method gives always the same results. A stochastic method most often returns different solutions due to the random factors used in this approach.

The principal approaches are:

1. *Simulated annealing*: it takes the name from a metallurgical process to increase the size

of the crystals in which the temperature of the material is increased above the melting point and subsequently cooled until the crystallization temperature is reached. In this method, the cost function measure the internal energy of the molecule and a controlled parameter simulates the temperature. At every step, a Monte Carlo simulation is used to reach a thermal equilibrium dependent to the temperature of the system. At high temperatures, the energy barriers between conformations are overcome and the system can escape from a local minimum. As the system is cooled, only the low energy conformations are explored until the system is trapped in a local minimum.

2. *Tabù search*: a list of tabù conformations already explored is created. The new solutions cannot adopt one of the conformations on the list. This force the molecule to examine part of the conformational space not previously explored.
3. *Incremental search*: the ligand is divided into rigid fragments cutting all acyclic single bonds. Subsequently one or more base fragments are docked into the binding site and used as anchors to which the other fragments are attached.
4. *Genetic algorithms*: this method mimics the process of biological evolution. At the beginning, a population of conformations is randomly generated and a fitness function calculates the ligand-protein interactions to measure the quality of possible solutions. Each member of the population (chromosome) encodes for a linear string of bits describing the properties of the conformation. Subsequently a new generation of possible solutions is generated. The parents of the new generation are randomly selected with a bias for the best individuals to ensure that the population evolves towards the best solution. The reproduction of the population is subjected to crossover and mutation. With crossover two chromosomes are cut in a random point and exchanged to create two new conformations. With the operator mutation one random bit is changed (i.e. 1 to 0) to create a new chromosome.
5. *Glide*: it uses a series of hierarchical filters to find possible orientations of the ligand in the binding site. The shape and the properties of the binding site are represented on a grid precomputed upfront to perform the docking. The grid has different sets of fields that lead to an accurate scoring of the pose. A set of initial ligand conformations are generated performing an exhaustive conformational search, and the conformers generated are presented in a compact combinatorial form. Each cluster, which contains one conformation of the core and all possible conformations of the rotamer groups, is docked

as a single object. The search of the ligand orientation begins with an approximate positioning of the conformations in the binding site and a scoring phase, which reduce the number of poses considered in the further steps. The conformations selected are minimized on precomputed OPLS-AA Van der Waals and electrostatic grids. In the final step, the three-six lowest energy poses are subjected to a Monte Carlo simulation to examine the near torsional minima, and to correctly orient the peripheral groups. The final poses are then rescored.

6. *Triangle Matcher*: an ensemble of hydrophilic and hydrophobic alpha spheres defines the binding site. The compound is placed in the binding site aligning triplets of ligand atoms with triplets of alpha spheres.
7. *Alpha PMI*: the molecule is placed in the binding site aligning the principal moment of inertia of the conformations to alpha spheres randomly generated.

3.2 Scoring functions

Docking programs can generate a large amount of poses that need to be ranked to find the pose most similar to the crystallographic conformation. This operation is done by scoring functions, which attempt to estimate the binding free energy of the ligand-protein complex. The binding free energy is defined by the Gibbs-Helmholtz equation:

$$\Delta G = \Delta H - T\Delta S \quad 3$$

where ΔG is the binding free energy, ΔH is the enthalpy, T is the temperature, and ΔS is the entropy. The binding constant K_i can be calculated from the Gibbs-Helmholtz equation:

$$\Delta G = -RT \ln K_i \quad 4$$

where R is the gas constant.

There are three principle types of scoring functions:

1. *Empirical scoring functions*: several weighted terms are used to calculate the binding affinity:

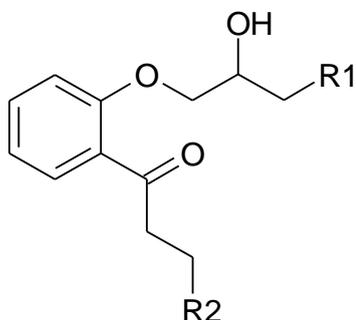
$$\Delta G = \sum_i W_i \times \Delta G_i \quad 5$$

where ΔG_i is determined by the contribution of different terms (hydrogen bonds,

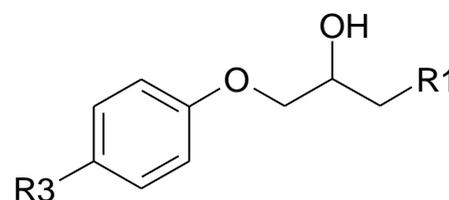
hydrophobic interactions, entropy, Van der Waals, electrostatic, desolvation, etc...). The weighting terms W_i are optimized using a dataset of affinity data of ligand-protein complexes of known 3D-structure through a multilinear regression analysis.

2. *Force field scoring functions*: are based on the non-bonded energy terms of molecular mechanics force field.
3. *Knowledge-based scoring functions*: are based on the statistical analysis of preferred ligand-protein interactions. They are based on the assumption that the more often a ligand atom x and a protein atom y are found at a certain distance d the more favorable is the interaction. To the interaction between the ligand atom x and the protein atom y at the distance d is assigned a free energy interaction depending on the frequency of the interaction. The sum of all interatomic interactions is the ligand-protein interaction free energy.

3.3 Compounds investigated

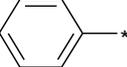
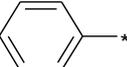
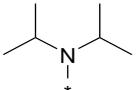
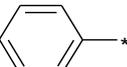
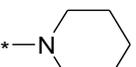
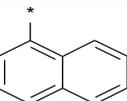
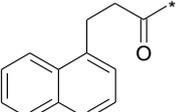
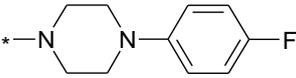
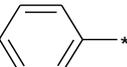
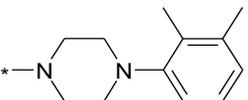
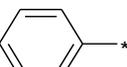
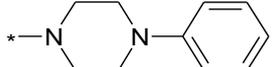
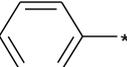
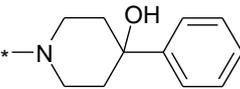
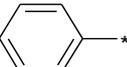
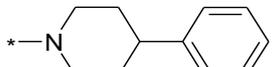
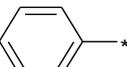


Scaffold 1



Scaffold 2

Table 2. Propafenone derivatives ^[a]Values are mean of at least three experiments and measured in heterologous hERG expressed in *X. laevis* oocytes.[126]

Compound	Scaffold	R1	R2	R3	IC ₅₀ μM ^[a]	Tapped
Propafenone	1				3.79 ± 0.24	Yes
GPV0005	1				1.34 ± 0.04	Yes
GPV0009	1				0.84 ± 0.08	Yes
GPV0180	1				1.78 ± 0.15	Yes
GPV0929	2				1.27 ± 0.05	Yes
GPV0031	1				1.46 ± 0.19	No
GPV0576	1				1.84 ± 0.24	No
GPV0019	1				1.42 ± 0.18	No
GPV0062	1				1.05 ± 0.05	Yes
SCT-AS03	1				0.77 ± 0.11	Yes

4

AIM OF THE STUDY

To predict correctly the adverse ADMET effects during the drug discovery process is a really important step, in which many different proteins play an important role. One of them definitely is the hERG potassium channel, which has been extensively studied due to its role in Torsade de Points, a lethal arrhythmia caused by drugs blocking the hERG channel. Although wealth of information has been obtained, there are still many open questions: are there other factors influencing the polispecificity? What is the conformation of the S5-P linker? What is the correct alignment of the S5? What are the molecular reasons of drug trapping?

To investigate the drug trapping phenomenon is important to understand the molecular mechanisms involved in the hERG channel gating and drug-hERG complex dissociation.

The aim of this thesis was to investigate trapping phenomenon using the docking technique. This

method was applied to analyze the interactions between propafenone and derivatives with the hERG channel in the open and the closed state, in order to explain the different trapping behavior of these compounds.

5

RESULTS

5.1 Docking into the homology model of hERG channel in the closed state

5.1.1 Material and methods

5.1.1.1 Docking protocol

The closed state of hERG channel homology model has been previously described by Thai et al.[102]

The trapped compounds propafenone, GPV0005, GPV0009, GPV0180 and GPV0929 were

docked into the homology model of the hERG channel in the closed state using the Dock docking tool of MOE.[127] Due to the fact that it was not possible to select one of the possible binding modes through consensus scoring and mutagenesis data, the Glide docking tool of Schrödinger, LLC, New York, NY, 2008[128] was used with the aim to prioritize one of the selected clusters.

MOE: to the pdb file of a homology model of the hERG channel in the closed state the hydrogen atoms were added and then minimized in the Amber99 force field using the MOE2007.09 tool Energy Minimize. During the minimization the partial charges of the model were recalculated. The ligands were drawn in the R configuration using the Molecule Builder tool of MOE2007.09,[127] and subsequently minimized, the partial charges calculated and the lone pairs atoms adjusted according to the default parameters with the program Energy Minimize using the MMFF94x force field. The nitrogen was maintained in the protonated state.[102] For every ligand a systematic conformational search was performed. The 100 most different conformations concerning the potential energy and the RMSD values were selected with Diverse Subset tool of MOE2007.09 and the partial charges recalculated according to the MMFF94(modified) force field. Subsequently, the 100 most diverse conformations were docked into the homology model of the hERG channel in the closed state. All atoms of the amino acids with an atom within 6.5 Å from a pose obtained in a previous study[102] were considered as part of the binding site and selected during the docking runs. For every run, only 20 poses per conformation were retained and minimized by the MMFF94x force field. Other parameters were kept as the default setting.

Glide: The ligands were prepared in the R configuration with the Maestro tool LigPrep, version 2.2, Schrödinger, LLC, New York, NY, 2005.[128] The nitrogen was maintained in the protonated state. The protein was prepared with the Protein Preparation Wizard included in Maestro, version 8.5, Schrödinger, LLC, New York, NY, 2008. Glide version 5.0, Schrödinger, LLC, New York, NY, 2008 was used to dock the compounds. We chose to carry out 100 poses for each ligand. Other parameters were kept as the default setting. The binding site was defined by the amino acids Ser624, Ser649, Tyr652 and Phe656 of each subunit.

A second docking run using Glide was performed to investigate the distributions of similar poses in the four possible directions that they can have in the central cavity. With the aim to limit the number of duplicate poses deleted by the docking program, the compounds were docked using

the value of RMSD of 0.5 Å and the maximum atomic displacement lower than 0.001 Å.

5.1.1.2 The symmetry problem in the hERG channel

The hERG channel has homotetrameric architecture. The four monomers contain six transmembrane domains (S1-S6). Generally only the S5-P-S6 segment of one subunit is modeled while the other three subunits are just copies,[129] hence only the parts that define the binding site, the selectivity filter, the central cavity and the inner pore, are included in the model. This operation leads to a perfect symmetry, so when a docking is run in a symmetric binding site, it is possible to have the same pose in four different directions. The docking programs cannot detect these poses as duplicates, because, due to the placement with different directions, they have a high RMSD values even if they have the same interactions pattern. Hence, poses that would be detected as duplicates whether they have the same directions are classified as non-duplicates due to their different orientation. Thus we had to develop a method which takes this into account and harmonises equivalent poses.

The symmetry problem can be faced in two ways. The first possible solution, as proposed by Farid et al.,[99] consists on a rigid body rotation of the complexes around to the identified symmetry axis. It is possible to apply this strategy also in non symmetric hERG channels, for example after molecular dynamics simulations, once the rotational axis is identified. We developed a new method in which only the ligand is rotated and the coordinates of the channel are kept fixed. This method can be used only if the protein, in this case the hERG channel, is symmetric. Our approach is computationally less expensive than the first method mentioned, because in our case only the ligand is rotated and not the entire complex. These operations are useful to facilitate the visual inspection of the pose and the identification of possible binding modes.

5.1.1.3 Alignment

When a docking study is performed in symmetrical binding sites, such as in the hERG channel, the alignment of the poses facilitates the visual comparison of the conformations and, through the generation of an RMSD matrix and the agglomerative hierarchical clustering, makes it possible to identify possible binding modes without considering the original direction of the poses. In this subchapter, I will explain how the symmetry problem can be faced.

In order to align the poses it is necessary to define a rotational axis in the center of the channel, and a reference pose to which the other poses can be compared. For this purpose, it is needed to define four amino acids, one per subunit, on the top of the hERG channel, and four on the bottom. The centers on the top and on the bottom of the hERG channel are defined as the points that lie in the middle distance between the C α of an amino acid of two opposite subunits. The rotational axis is the line that links these two points. The rotational angle used to rotate the ligand inside the binding site, in order not to change the distances and the interactions between the

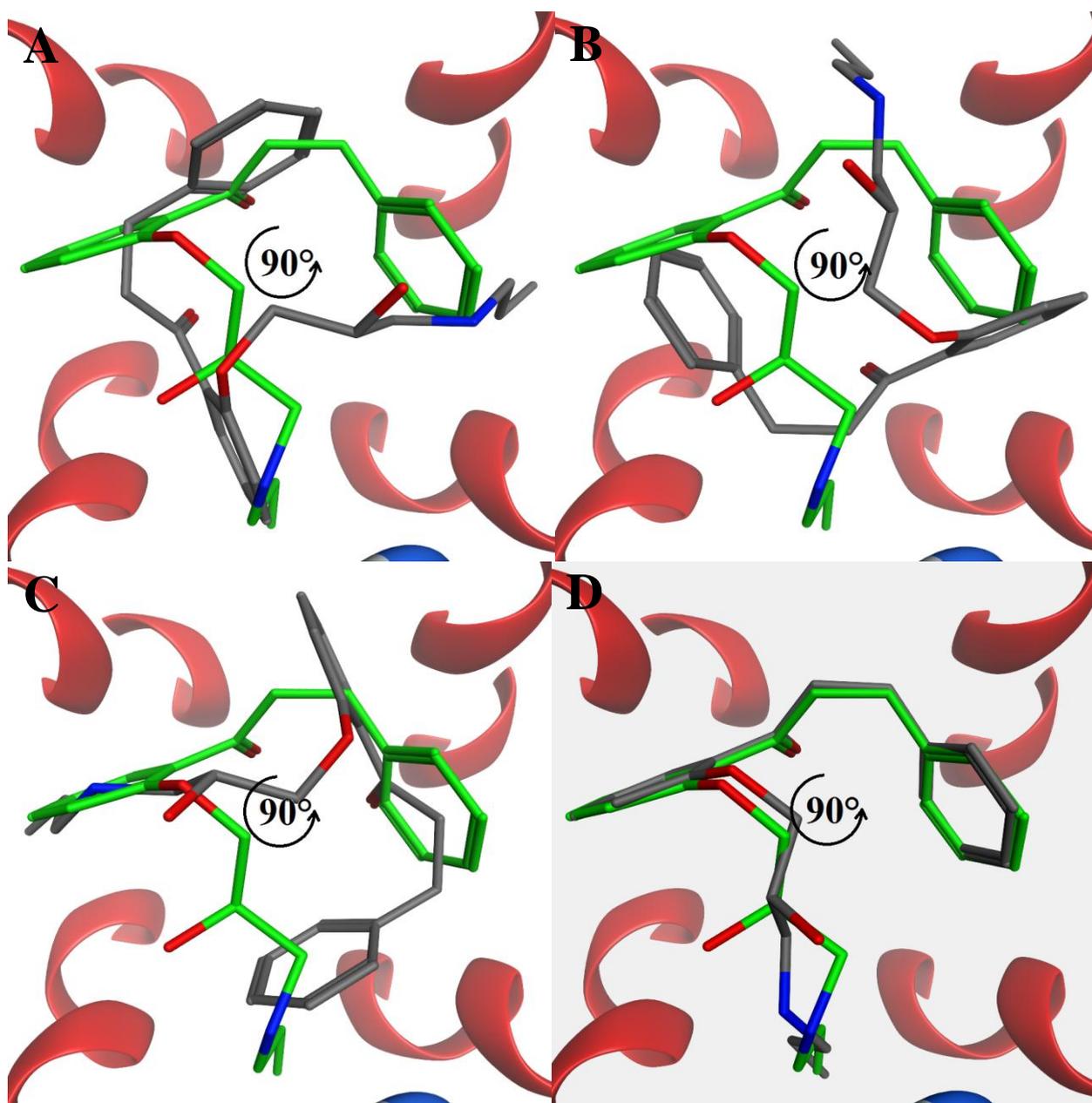


Fig. 5 Pose alignment. The green carbon colored is the reference pose. (A) a second pose is loaded in the original orientation and the RMSD0° is calculated; (B) the pose is rotated of 90° and the RMSD90° is calculated; (C) the pose is rotated again of 90° and the RMSD180° is calculated; (D) the pose is rotated for the for the last time and the RMSD270° is calculated.

compound and the amino acids, must be 90°. The reference pose to which all the other conformations are compared and aligned is the first one of the database. The second pose is loaded in the channel and the RMSD0° is calculated using the first pose as reference. Subsequently the second pose is rotated three times and for each rotation the RMSD value is calculated (RMSD90°, RMSD180° and RMSD270°) (fig. 5). The “rotational pose”, which corresponds to the lowest value of RMSD, is the one that is aligned or “adjusted” with the reference pose. To perform the alignment was written an svl script (MOE) called ROTALI.

5.1.1.4 RMSD matrix calculation

The ROTALI svl script was used also to generate the RMSD matrices used to delete and to identify the binding modes. As for the pose alignment, the ROTALI script compares the reference pose with all the others in the database. The first reference pose is the first one of the database. The value of RMSD0° of the second pose is calculated, then the pose is rotated three times and for every rotation the value of RMSD (RMSD90°, RMSD180° and RMSD270°) of the “rotated pose” is calculated. The lowest value of RMSD is written in a RMSD matrix, because it corresponds to the value of the “rotational pose” that is aligned or at least is “adjusted” with the reference pose. When all the poses are compared to the first one, the reference pose changes to the second one and all the operations are repeated again. The program ends when all the poses of the database are taken as reference.

5.1.1.5 Duplicate poses deletion

Through the RMSD matrix it is possible to detect and delete the duplicate poses.

MOE defines two poses as duplicates if they have the same hydrogen bond and hydrophobic interaction pattern.

Glide defines duplicate poses as the poses with an RMSD value lower than 0.5 Å and a maximum atomic displacement lower than 1.3 Å.

In order to use a definition of duplicate poses similar to the one of MOE applicable to the RMSD matrix, they were defined as the poses with an RMSD value lower than 0.1 Å. The poses obtained from the Glide docking program were deleted using their definition of duplicate poses.

5.1.1.6 Identification of the common binding modes

The poses obtained with the MOE docking program were analyzed in two manners and the results were then compared with the results obtained with Glide. For the poses generated with MOE, we decided to analyze those within 10 kcal from the pose with the best London dG scoring function value (personal communication from Dr. Wolfram Altenhofen, Chemical Computing Group), from now it will be called database of poses A. With the aim to see if some common binding modes are lost considering only the poses within the 10 kcal from the best ranked pose according to the London dG scoring function, a second analysis was performed using a number of poses slightly higher than the maximum number of poses within the 10 kcal (195 in the compound propafenone). In order not to be too far from the ideal selection of poses suggested by Dr. Wolfram Altenhofen, we decided to consider only fifteen more poses than the maximum number previously mentioned. If some possible binding modes are missed considering only the poses within 10 kcal, this would be immediately clear even considering only a few more poses in the analysis. Hence, for every ligand the first 210 poses ranked according to the London dG scoring function (from now database of poses B) were selected. All molecules docked are characterized by a common scaffold (fig. 7). The common scaffold was extracted using the function `abstract_commonscaffold` of the previously loaded script `FORMELPOOL.SVL`, in which the string of the scaffold was specified. All the ligands docked are characterized by the presence of two aromatic rings connected by a ketone linker, and by a protonated nitrogen linked to one of the two phenyl rings by an ether linker with a hydroxyl group in position 3 (fig. 7). The rationale for using the common scaffold to analyze the poses, is due to the fact that it is reasonable to assume that compounds with the same scaffold have a common binding mode in the binding site. In both analyses, the non-duplicate poses were merged in a database and an RMSD matrix calculation based on the common scaffold (fig. 7) was performed using the `ROTALI svl` script. A hierarchical clustering of the RMSD matrix, using the Ward's agglomerative method, was performed with the `XLSTAT`[130] added in MS Excel. The Ward's method uses the variance to measure the distance between two objects. At the beginning the Ward's method considers each object (in this case the poses) as single cluster. In each step the two clusters whose merging increases least the intracluster variance are merged. The agglomerative cluster analysis was performed until all the poses of the clusters were within a RMSD value of 2.5 Å from the centroid to obtain a good intracluster homogeneity.

With the aim to identify the most energetically favourable clusters, we performed a consensus scoring on the two databases of MOE using the scoring functions ASE, Affinity dG, Alpha HB and London dG developed in MOE. The analysis of the ligand interactions with the hERG channel in the closed state was performed using the centroid of those clusters containing all the ligands and whose poses were ranked within the first 10 positions in at least three scoring functions out of four. The clusters were also validated using the mutagenesis data. Witchel et al.[11] demonstrated that only the Phe656 have a small influence on the trapping of propafenone, while mutation of the other residues in the central cavity did not show any influence on the recovery from block. The fact that the influence of Phe656 on drug trapping is small might indicate that the interaction between the channel in the closed state and the molecule is weak.

5.1.1.7 Analysis of ligand interactions

The interactions between the ligands and the hERG channel were analyzed with the MOE2007.09 software package.

5.1.2 Results

5.1.2.1 Alignment and deletion of duplicate poses

In the hERG channel, each side of the binding site is identical to both adjacent sides, and this renders it difficult to analyze the poses generated. This problem can be solved if we fix 5 points, one in the center of the channel and one in every subunit, so we can define four directions from the center. With the purpose to define the direction of the pose, we can select a point in the ligand, and the direction of the line that links the point in the center of the channel with the point on the molecule can be assumed as the direction of the pose. In this work, the point selected in the ligand is the charged nitrogen.

The docking programs produce for every ligand several poses which are spread in all four directions (fig. 6 A and B). In another docking study in the hERG channel, in order to facilitate the visualization of the complex, the poses were aligned by a rigid body rotation of the ligand-hERG channel complex.[99] In ROTALI the coordinates of the hERG channel are fixed and the alignment is performed through the rotation of the poses and the comparison with a reference pose. The result of our script is that the poses that are similar to the reference pose are aligned

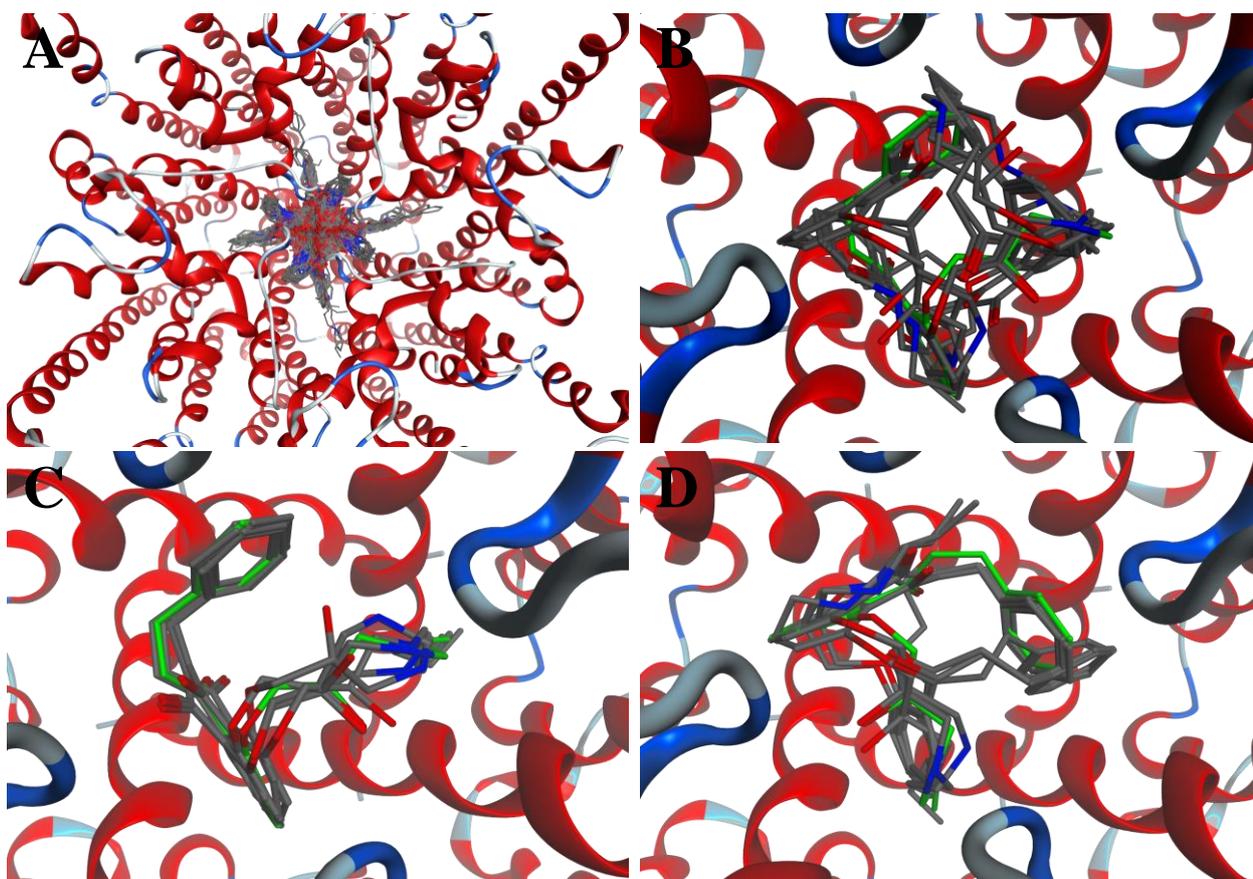


Fig. 6 (A) Docking results of propafenone docked with MOE Dock program; (B) 10 poses of the compound propafenone before the alignment; (C) 10 poses aligned with the reference pose (green); (D) poses adjusted with the reference pose.

with it (fig. 6C), while the poses that are not similar to the reference one are “adjusted” with it (fig. 6D). It means that the poses aligned have the same direction of the reference one, while the poses adjusted can have also different directions.

The fourfold symmetry of the hERG channel allows the placement of the same pose in every side of the channel (fig. 6 A and B). If we consider this in terms of duplicate poses, it would mean that almost 75% of them could be duplicates. This can happen only if there is a homogeneous distribution of the poses in the four directions of the hERG channel, and almost all the poses should have more than one similar pose. In order to investigate the similar poses, their distribution in the hERG channel was analyzed. The similar poses were defined as the poses that have a value in the RMSD matrix lower than 0.5 Å.

In MOE (table 3) similar poses almost never lay in all four directions. Indeed the highest number of similar poses with four directions is obtained for compound GPV0180, and they are only 10% of the total number of conformations. For propafenone 53% of the total number of poses is

Table 3. Distribution of similar poses in MOE in the first 100 poses selected according to the scoring function. Dir.= number of directions in which the similar poses can be founded; N. poses = number of poses; % of poses = percentage of poses regard to the total number of the poses in the database.

Dir.	Propafenone		GPV0005		GPV0009		GPV0180		GPV0929	
	N. poses	% of poses	N. poses	% of poses	N. poses	% of poses	N. poses	% of poses	N. poses	% of poses
1	1	1.00%	7	7.00%	2	2.00%	3	3.00%	1	1.00%
2	24	24.00%	27	27.00%	25	25.00%	18	18.00%	30	30.00%
3	28	28.00%	16	16.00%	3	3.00%	9	9.00%	13	13.00%
4	0	0.00%	7	7.00%	0	0.00%	10	10.00%	8	8.00%
Total	53	53.00%	57	57.00%	30	30.00%	40	40.00%	52	52.00%

Table 4. Distribution of similar poses in Glide. Dir. = number of directions in which the similar poses can be founded; N. poses = number of poses; % of poses= percentage of poses regard to the total number of the poses in the database; N.D = no docking pose obtained.

Dir.	Propafenone		GPV0005		GPV0009		GPV0180		GPV0929	
	N. poses	% of poses	N. poses	% of poses	N. poses	% of poses	N. poses	% of poses	N. poses	% of poses
1	3	3.03%	0	0.00%	1	12.50%	0	0.00%	N.D	N.D
2	25	25.25%	11	39.29%	2	25.00%	3	17.65%	N.D	N.D
3	14	14.14%	0	0.00%	0	0.00%	3	17.65%	N.D	N.D
4	2	2.02%	0	0.00%	0	0.00%	0	0.00%	N.D	N.D
Total	44	44.44%	11	39.29%	3	37.50%	6	35.29%	N.D	N.D

similar and 52 of them are present in two or three directions, while for other compounds this percentage varies between 27% for GPV0180 and 43% for GPV0929 and GPV0005. In the case of GPV0005 57% of the poses have one or more similar poses and only 7 of them lie in all four directions.

The compounds were docked also with Gide. The poses were considered duplicates if the RMSD value was lower than 0.5 Å and the maximum atomic displacement was lower than 0.001 Å. The results (table 4) show that only for propafenone similar poses lie in the four directions and it happens only two times, while in the GPV0180 they are at most in three directions, and in the compounds GPV0005 and GPV0009 they have only one or two directions. Also in the poses from the Glide docking program, as in MOE, the majority of the similar poses have two or three

Table 5. Number of similar poses in MOE in the first 100 poses selected according to the scoring function. N. poses = Number of similar poses; N. simil.= Number of poses similar to the reference pose; % of poses = percentage of N. simil poses regard to the total number of the poses in the database.

N. poses	Propafenone		GPV0005		GPV0009		GPV0180		GPV0929	
	N. simil.	% of poses	N. simil.	% of poses	N. simil.	% of poses	N. simil.	% of poses	N. simil.	% of poses
1 pose	24	24.00%	30	30.00%	20	20.00%	18	18.00%	27	27.00%
2 poses	20	20.00%	15	15.00%	10	10.00%	12	12.00%	16	16.00%
3 poses	4	4.00%	12	12.00%	0	0.00%	0	0.00%	5	5.00%
4 poses	3	3.00%	0	0.00%	0	0.00%	0	0.00%	4	4.00%
> 4 poses	2	2.00%	0	0.00%	0	0.00%	10	10.0%	0	0.00%

Table 6. Number of similar poses in Glide. N. poses = Number of similar poses; N.simil.= Number of poses similar to the reference pose; % of poses = percentage of N.simil poses regard to the total number of the poses in the database; N.D = no docking pose obtained.

N. poses	Propafenone		GPV0005		GPV0009		GPV0180		GPV0929	
	N. simil.	% of poses	N. simil.	% of poses	N. simil.	% of poses	N. simil.	% of poses	N. simil.	% of poses
1 pose	29	29.29%	8	28.57%	2	25.00%	2	11.76%	N.D	N.D
2 poses	5	5.05%	3	10.71%	1	12.50%	2	11.76%	N.D	N.D
3 poses	5	5.05%	0	0.00%	0	0.00%	2	11.76%	N.D	N.D
4 poses	3	3.03%	0	0.00%	0	0.00%	0	0.00%	N.D	N.D
> 4 poses	2	2.02%	0	0.00%	0	0.00%	0	0.00%	N.D	N.D

directions.

The number of similar poses for every pose was investigated with the intent to understand the reason why so few of them lie in all four directions (table 5 and table 6). Finding a pose in all four directions can happen only if the pose has three or more similar poses. In MOE the percentage of poses which have at least 3 similar conformations varies between 0% for GPV0009 and 12% for GPV0005. In Glide this percentage varies between 0% for GPV0005 and GPV0009, and 11.76% for GPV0180. These low percentages explain the low number of similar poses in all four directions.

These results might indicate that both docking programs do not explore efficiently the conformational space of the ligands in all four directions of the symmetric binding site, otherwise the number of conformations with 3 or more similar poses and the number of similar

Table 7. Deletion of the poses from MOE docking in the first 100 poses selected according to the London dg scoring function.

Ligand	Poses start	Poses end	Poses deleted	% deletion
Propafenone	100	79	21	21%
GPV0005	100	77	23	23%
GPV0009	100	88	12	12%
GPV0180	100	84	16	16%
GPV0929	100	78	22	22%

Table 8. Deletion of the poses from Glide docking.

Ligand	Poses start	Poses end	Poses deleted	% deletion
Propafenone	99	81	18	18.18%
GPV0005	27	22	5	18.52%
GPV0009	8	8	0	0%
GPV0180	16	14	2	12.5%
GPV0929	No poses	No poses	No poses	No poses

poses with 3 and 4 directions would be higher. However, the fact that we were able to identify similar poses with different directions means that the ROTALI svl script works, and that it is a useful tool to facilitate visual inspection of the poses.

As mentioned before MOE defines the duplicate poses based on the hydrophobic and hydrogen bonds interaction pattern. It means that the RMSD value for the duplicate poses has to be low; hence we decided to consider as duplicates the poses that have a value of RMSD lower than 0.1 Å. The highest number of duplicate poses deleted is obtained with the ligand GPV0005, where 23 poses out of 100 were duplicates. Table 7 clearly shows that the method to recognise the duplicate poses based on the interaction pattern fails in the symmetric binding site of hERG if the duplicate poses have different directions, because we detected a percentage of duplicates between 12-23%. On the other hand these percentages show that our approach to detect the duplicate poses works even if they are not originally aligned.

We docked the trapped compounds with Glide using the default parameter, in order to investigate the ability of the program to detect duplicate poses in the hERG binding site and to identify possible binding modes. We were not able to dock the compound GPV0929. Also the

protocol of Glide to detect duplicate poses based on the RMSD value and the maximum atomic displacement fails, as demonstrated by the detection of duplicates with our approach. This was foreseeable, because the poses that have different directions have a high value of RMSD, even if they are duplicates after the alignment (table 8).

Table 7 and table 8 clearly show that the inability to recognise duplicates in a symmetric binding site is independent on the method, while they demonstrate that the alignment of the poses performed through the ROTALI svl script and our approach to detect duplicate poses work for both docking programs even if the poses have different directions.

5.1.2.2 Pose analysis

The binding site of the hERG channel blockers is placed in the inner cavity.[6, 9, 12, 22, 131] At the base of the hERG channel there are two concentric rings formed by the symmetric arrangement of the four Phe656 and the four Tyr652. The ring composed by the four Phe656 is located at the cytoplasmatic side of the channel. The second ring is formed by four Tyr652. Phe656 and Tyr652 are facing the inner cavity, but they have different orientations. Furthermore, Phe656 and Tyr652 are perpendicular to the channel axis, but in the case of Tyr652 they are slightly bent outwards. The distance between the C α of Phe656 of opposite subunits in the closed channel is 9.4 Å and this spatial constriction of the ring of the four Phe656 leads to the impossibility for the ligand to interact through π -stacking interactions with the Phe656, which is important to block the channel when it is in the open state.[11] The hERG blockers bind the hERG channel in the open conformation from the intracellular side. In the open state the volume of the binding site of the hERG channel is 1642.2 Å³. [102] In the closed conformation the volume of the binding site is reduced more than 30% to 1048.2 Å³, and it implies that only small and flexible compounds can be trapped in the closed state.[102]

The compounds propafenone, GPV0005, GPV0009 and GPV0180 were docked into the closed state homology model of the hERG channel. The compound GPV0929 was not analyzed, because the different scaffold and the visual inspection of the first ranked poses suggest that it might have a different binding mode respect to the four ligands above cited, and because it was not possible to dock it in Glide. Due to the distribution of the poses in all four directions, the symmetry operation with ROTALI was applied to delete the duplicate poses. The RMSD matrix obtained from the ROTALI script, calculated based on the common scaffold of the non-duplicate

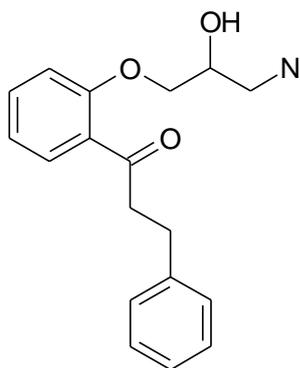


Fig. 7 Common scaffold used to calculate the RMSD matrix.

poses (fig. 7), was used to identify the common binding modes through agglomerative hierarchical cluster analysis. In the cluster analysis of the RMSD matrix of the database of poses A, 33 clusters were obtained with a cut off of the RMSD value of 2.5 Å from the centroid. In six of these clusters at least one pose of all the ligands was present. In the agglomerative cluster analysis of the RMSD matrix of the database of poses B, the clusters containing all the trapped ligands were 11 (out of a total number of 45). A further selection of the clusters was performed based on consensus scoring, in which the scoring functions ASE, Affinity dG, Alpha HB and London dG of MOE were used. Considering the consensus scoring for the database of poses A and B, it was possible to select 4 clusters in both cases. The interaction analysis of the centroids shows that some poses with a similar interaction pattern have the scaffold oriented with opposite directions (fig. 8 A, C and D). If the orientation of the scaffold is considered, we can distinguish four different binding modes, which can be reduced to two considering the interaction pattern.

In the first binding mode (fig. 9) the two phenyl rings forms parallel-displaced π - π and/or T-stacking interactions with Tyr652 of opposite subunits, and hydrophobic interactions with Phe656. The hydroxyl group and the charged nitrogen don't interact with the hERG channel. The substituent R1 is placed near the the Tyr652 with which it forms hydrophobic interactions.

In the second binding mode (fig. 10) the two phenyl rings of the ligand interact with two Tyr652 of adjacent subunits through parallel-displaced π - π and/or T-stacking interactions, and with Phe656 through hydrophobic interactions. The hydroxyl group and the charged nitrogen don't interact with the amino acids in the central cavity of the channel. As found in the first binding mode, the substituent R1 makes hydrophobic interactions with Tyr652.

The alanine scanning performed by Witchel et al.[11] shows that only Phe656 has a small influence on the propafenone recovery, while Ser624, Val625 and Tyr652 don't have any effect in that way. This suggests that propafenone interacts weakly with the Phe656, making T-stacking, parallel-displaced π - π , or hydrophobic interactions when the hERG channel is in the closed conformation. In agreement with these results, our poses cannot make π -stacking interaction with Phe656, due to the spatial size reduction of the ring of the four Phe656.

The compounds were also docked with Glide. Of 13 clusters obtained with the agglomerative

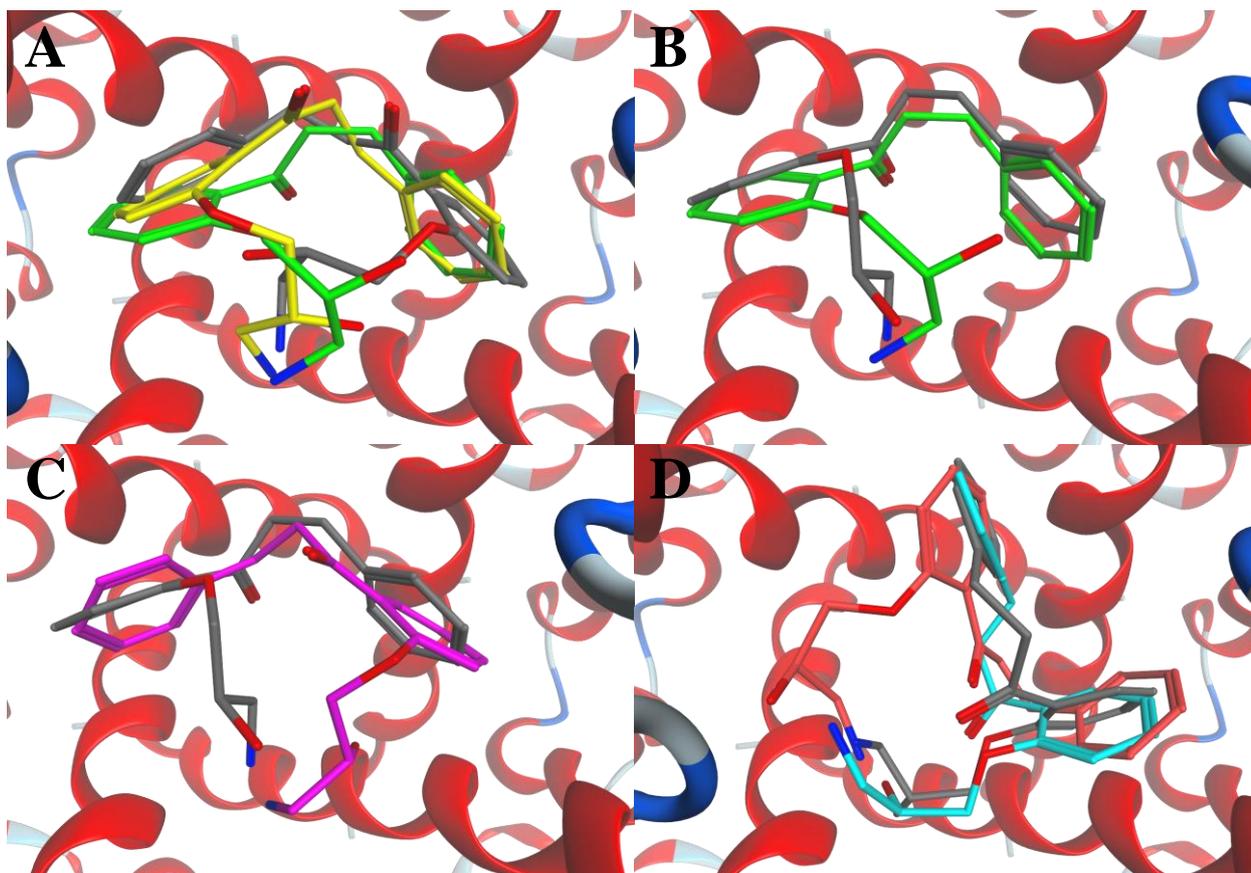


Fig. 8 Centroids of the four selected clusters. (A) Centroids of database of poses A. The centroid green colored and the yellow colored have the phenyl groups and the charged nitrogen placed in the same position, while in the centroid element colored the position of the two phenyl rings is inverted and the charged nitrogen is in the same position of the first two centroids. (B) The centroid of the database of poses A (green colored) is superposed with the centroid of database of poses B. (C) Centroids of the database of poses B with the scaffold inverted. (D) Superposition of two selected centroids of the database of poses B (pink colored and element colored) with a centroid of the database of poses A (azure colored). The pink colored centroid has the direction of the scaffold inverted with respect to the other two centroids. Only the scaffolds of the centroids are visualized.

hierarchical cluster analysis of the RMSD matrix, only 1 cluster contains poses of all trapped ligands. As shown in fig. 9D, the phenyl rings and the charged nitrogen of the first binding mode obtained in MOE and the centroid of Glide are placed in the same positions, while the position of the hydroxyl group is slightly different and in the centroid of Glide forms a hydrogen-bond with the Ser624 (fig. 11). The fact that similar poses were found with MOE and with Glide might indicate that this binding mode could be the most reasonable to describe the interactions of trapped ligands with the hERG channel in the closed state.

As found by Witchel et al.[11] and Thai et al.[102] in the closed state conformation of hERG channel the molecules docked interact also with the amino acid Tyr652 and not exclusively with Phe656. In agreement with the mutagenesis data, the selected poses do not form strong interactions with Phe656.

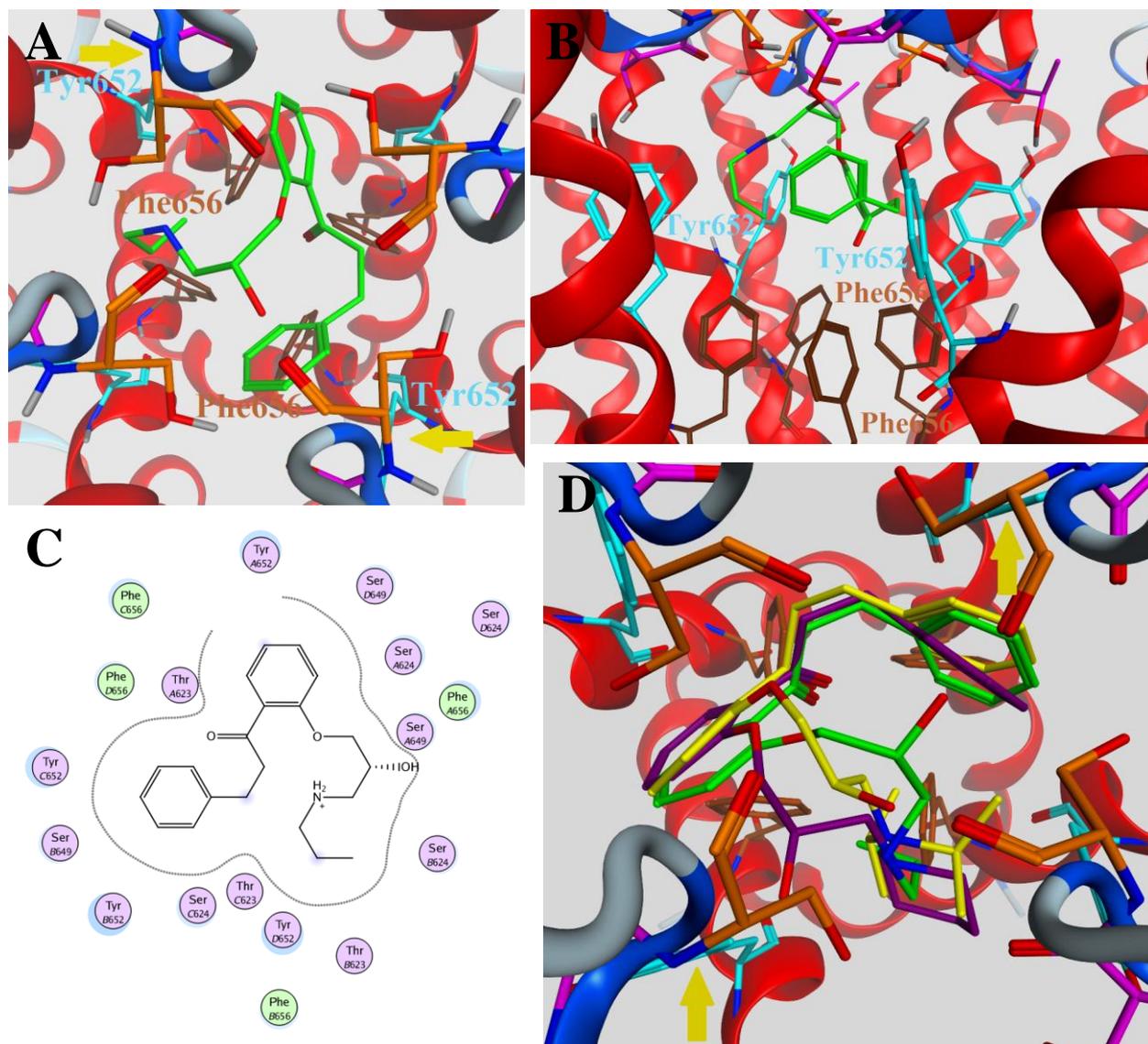


Fig. 9 First binding mode. Top view (A); lateral view (B); interaction pattern of the first binding mode (C); (D) superposition of the centroid of the validated clusters of database of poses A (green), database of poses B (yellow), and Glide (dark violet). Only the amino acids Thr623 (violet), Ser624 (orange), Tyr652 (cyan) and Phe656 (brown) of the four subunits are visualized. The yellow arrows indicates the two Tyr652 interacting with the trapped compounds through parallel π - π displaced and/or T-stacking interactions.

In the first binding mode (fig. 9) the two phenyl groups of the scaffold are placed near the Tyr652 of opposite subunits with which they can interact through parallel-displaced π - π and/or T-stacking interactions. The hydroxyl groups are placed between Ser624 of the adjacent subunits and the charged nitrogens lie between two Ser624 of adjacent subunits.

In the second binding mode (fig. 10) the two phenyl rings interact with Tyr652 of adjacent subunits through parallel-displaced π - π and/or T-stacking interactions. In two of the three poses of the second binding mode, the hydroxyl groups are downward and placed under the Ser624, while in the third pose the hydroxyl group is upward and placed near the hydroxyl group of the

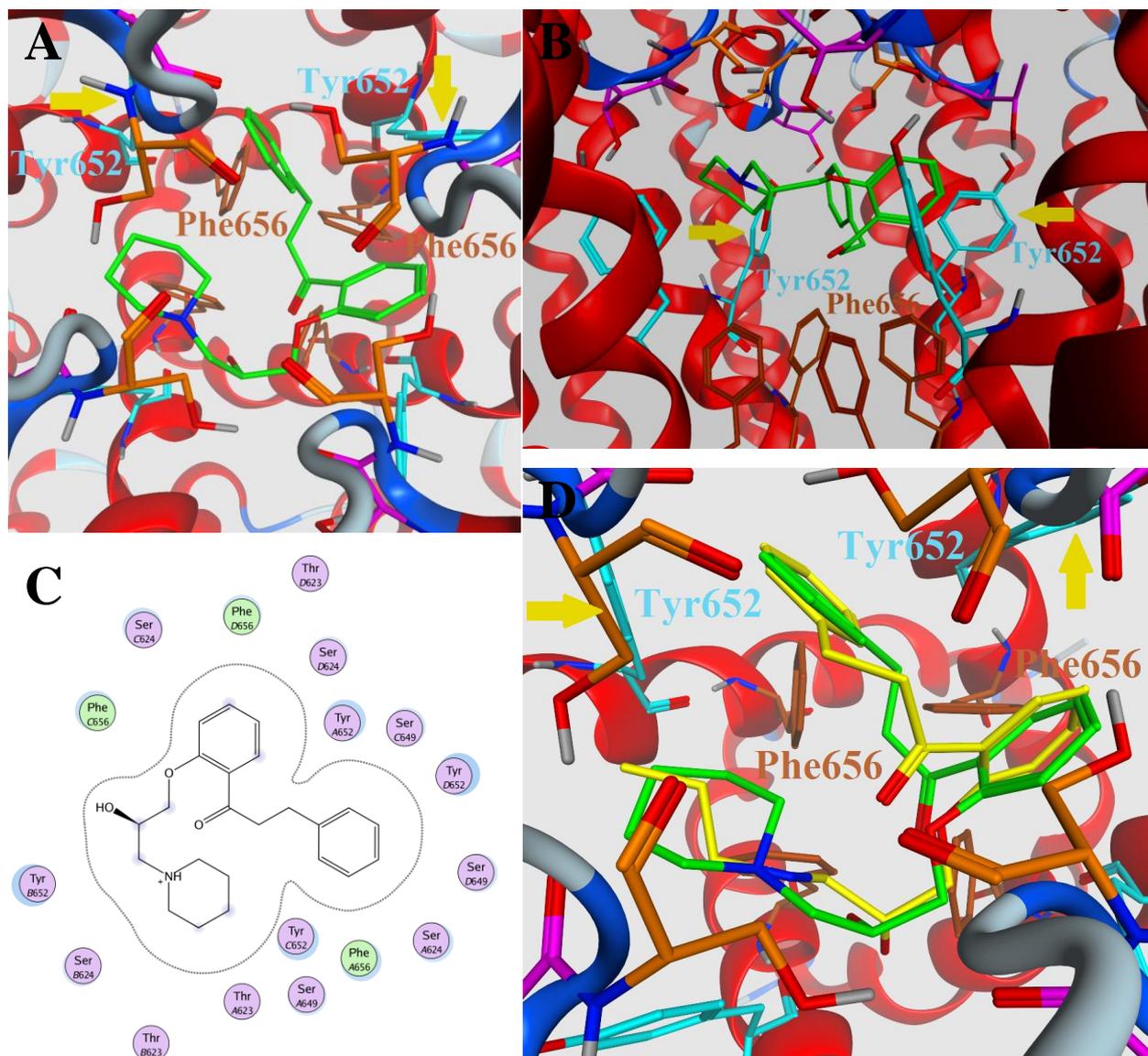


Fig. 10 Second binding mode. (A)Top view, (B) lateral view, and (C) interaction pattern of the second binding mode. (D) Superposition of the centroid of the validated clusters of database of poses A (green), and database of poses B (yellow). Only the amino acids Thr623 (violet), Ser624 (orange), Tyr652 (cyan) and Phe656 (brown) of the four subunits are visualized. The yellow arrows indicate the Tyr652 of adjacent subunits interacting with the trapped compounds through parallel-displaced π - π and/or T-stacking interactions.

previously published pose.[102] In both binding modes, the charged nitrogens are grouped under the Ser624, as found by Thai et al.[102]

In the pose published by our group[102] the phenyl ring of propafenone is positioned in an optimal position to make π -stacking interaction with Tyr652, the hydroxyl group is placed under Thr623 with which it forms a hydrogen bond and the charged nitrogen, located close to the position of the first binding mode, makes hydrogen bonds with Thr623 and Ser624. In contrast to this result, in the poses obtained in this study the two aromatic rings of the selected binding

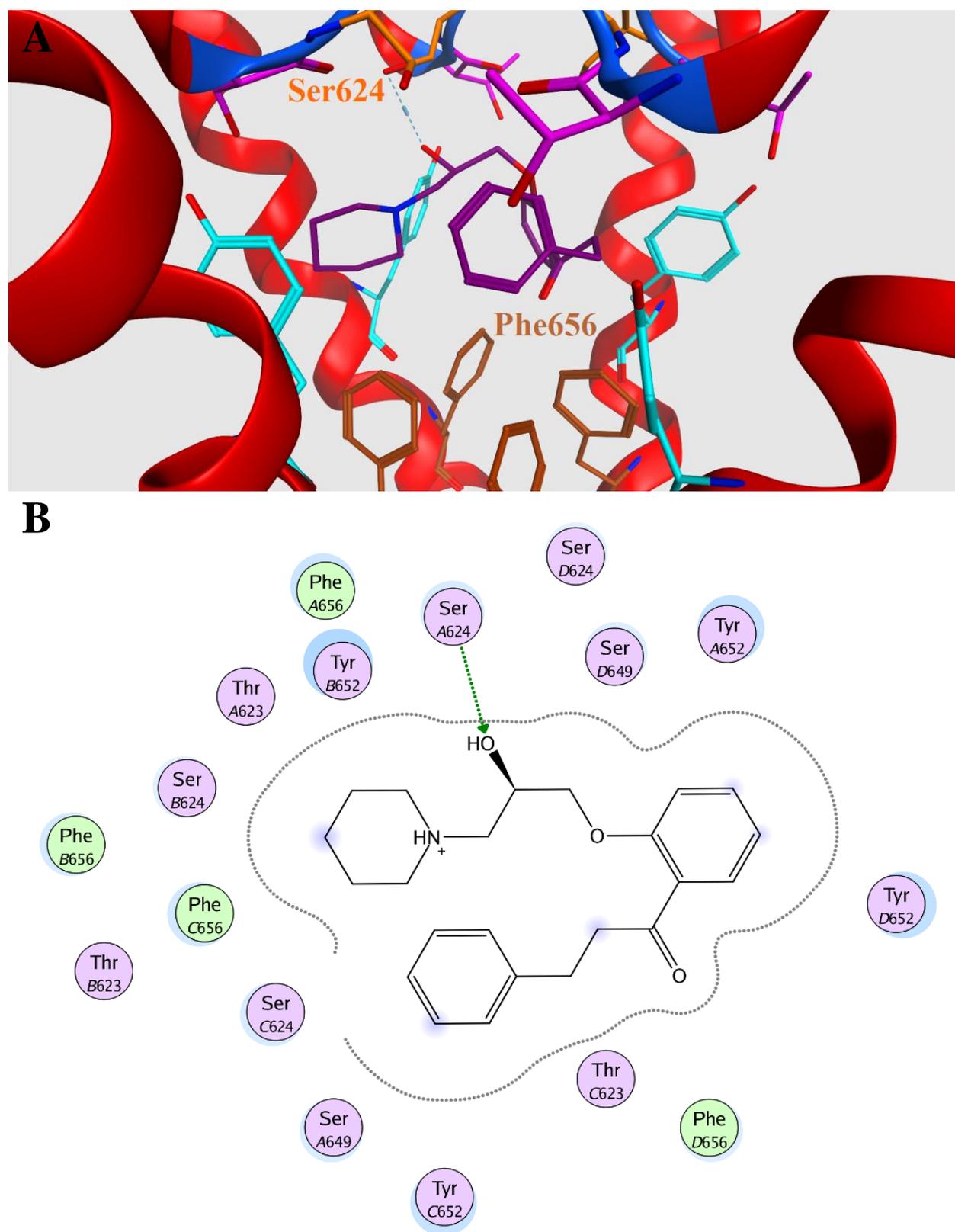


Fig. 11 Glide result. (A) Propafenone binding mode docked with Glide. (B) Propafenone interaction pattern. Only the amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented.

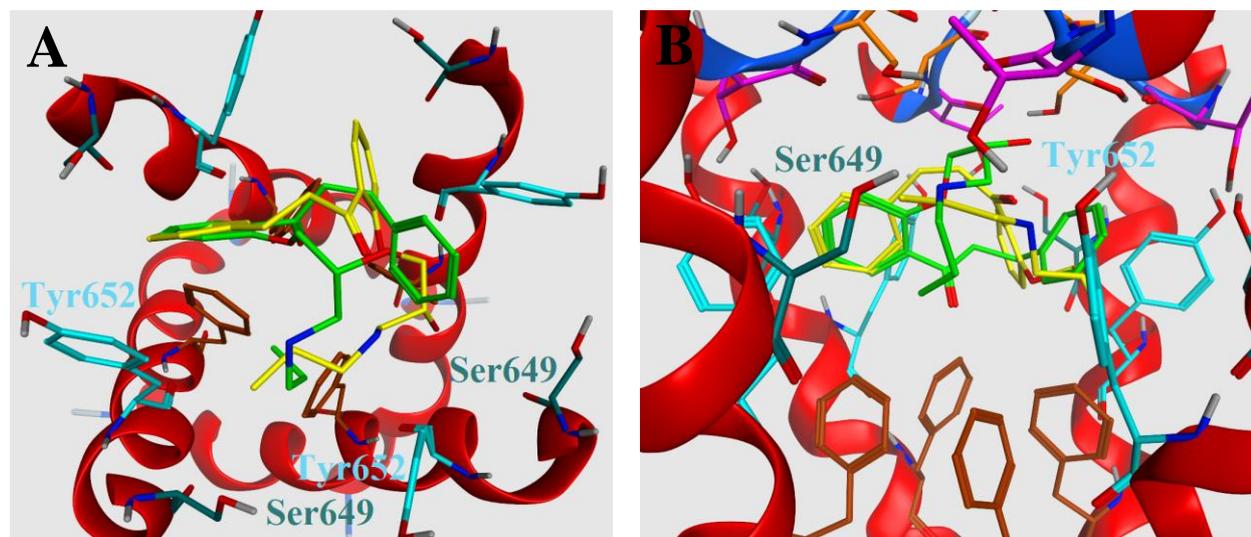


Fig. 12 First and second binding modes. Top view (A) and lateral view (B) of the first (green) and the second (yellow) binding mode. Only the amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented. Due to the lack of space between the amino acids Ser649 and Tyr652, and between them and the protonated nitrogen, it is not possible for a trapped compound to have a bulky and rigid substituent attached to the basic nitrogen.

modes form parallel-displaced π - π and/or T-stacking interactions with Tyr652 of opposite or adjacent subunits, while the protonated nitrogen doesn't make any hydrogen bond interaction. In addition, Glide suggests a possible hydrogen bond in the first binding mode between the hydroxyl group and Ser624.

As indicated in our previous work,[102] the size of the substituent attached to the charged nitrogen might be the first hint to discriminate between trapped/non-trapped compounds. A recent study[126] shows that the rigidity of the substituent R1 (table 2), the pKa of the basic nitrogen and the presence of a hydrogen bond acceptor in the substituent attached to it might play a role in the drug trapping. In this study, the poses show that the charged nitrogen is placed near the Tyr652 and Ser649, in a zone where it is not possible to place a large and rigid substituent due to the lack of space between the two amino acids above-cited (fig. 12), and between the two amino acids and the protonated nitrogen. A bulky and rigid substituent attached to the charged nitrogen would push back the entire molecule, forcing it to assume an unfavourable high-energy conformation, while the hydrogen bond acceptor might form a hydrogen bond with the hERG during the channel closure, hindering the gating mechanism and forcing the dissociation. Small and flexible substituents can bend in the small space between the basic nitrogen and the amino acids Ser649 and Tyr652, without destabilizing the conformation of trapped propafenones in the closed hERG channel. The drug trapping phenomena is a multistep mechanism that starts in the open state and finish in the closed state, in which many intermediate

conformations of the channel might play an important role. Probably the dissociation of non-trapped compounds occurs in one of the intermediate conformations that might be formed during the hERG closure.

5.1.3 Conclusions

A series of trapped propafenone derivatives with a different size of the substituent attached to the charged nitrogen were docked into a homology model of the hERG channel in the closed state. The fourfold symmetry of the hERG channel renders it possible to find similar poses with four different directions, hence it is also possible to find duplicate poses with four different directions. In a symmetric protein like the hERG channel it is thus necessary to align and delete the duplicates that have originally different directions. This can be done easily through the RMSD matrix calculated with the svl script ROTALI. Through cluster analysis of the RMSD matrix based on the common scaffold, consensus scoring and mutagenesis data evaluation two possible binding modes were selected, which might explain the trapping/non trapping behaviour. One of these two possible binding modes was also found with Glide and it might indicate that this pose could better explain the trapping behaviour. Interestingly, in both binding hypothesis the R1 group is located in a zone where it is not possible to place bulky and rigid substituents.

5.2 Docking studies into a homology model of the hERG channel in the open state

5.2.1 Material and methods

5.2.1.1 Docking protocol

The homology model of the hERG channel in the open state was obtained using the crystal structure of KvAP as previously described by Tseng et al.[96] The hydrogen atoms were added to the pdb file kindly provided by Prof. Dr. R. H. Guy and minimized using the Amber99 force field with the MOE2008.10 tool Energy Minimize. The partial charges of the amino acids were recalculated during the minimization. The molecules propafenone, GPV0005, GPV0009, GPV0031, GPV0180 and GPV0576 were built and the most 100 diverse conformations were selected as described in the subchapter 5.1.1.1 using the programs implemented in MOE2008.10. The aniline nitrogen of the compounds GPV0031 and GPV0576 was assigned as sp^2 hybridization state. In the systematic search, for the compound GPV0576, due to the huge number of conformations generated using the default conditions (10.480.616 conformations) which made the size of the database too big to calculate E and RMSD, the step of the phenyl-carbonyl carbon bond rotation was modified to 180° in order to reduce the number of conformations generated. The MOE2008.10 docking tool Dock was used to dock the 100 most diverse conformations into the hERG channel in the open state, retaining 20 poses per run. The binding site was defined by the atoms of the amino acids Ser624, Ser649, Tyr652 and Phe656 of each subunit. We decided to use the Alpha PMI placement method and the minimization of the poses according to the MMFF94x force field to obtain a better performance of the docking program (personal communication from Dr. Wolfram Altenhofen, Chemical Computing Group). All the other parameters were left in their default values. In contrast to the protocol applied in the closed state, the Glide docking program was not used due to the fact that MOE showed a better performance to explore the conformational space of the trapped compounds in all 4 directions of the central cavity, and that the pose obtained with Glide was similar to the first binding mode hypothesized with MOE, indicating that Glide finds the same pose of MOE, and that with the latter more possible binding modes are taken into account. Moreover, it was possible to prioritize one of the 3 possible binding modes through consensus scoring and mutagenesis data.

5.2.1.2 Alignment

The methodology used to align the poses has been described in detail in subchapters 5.1.1.2 and 5.1.1.3.

5.2.1.3 RMSD matrix calculation and duplicate poses deletion

These processes are described in the subchapters 5.1.1.3, 5.1.1.4 and 5.1.1.5.

5.2.1.4 Identification of the common binding modes

For each docked ligand were selected the first 210 poses ranked according to the London dG scoring function. It was not possible to use the poses within 10 kcal due to the fact that only a maximum of 8 poses per ligand were considered and it was not possible to select a common binding mode through the agglomerative hierarchical clustering. Moreover, the docking study in the closed state channel demonstrates that we achieve the same results using the poses within the 10 kcal from the best pose or the first 210 poses ranked according to the London dG scoring function, and that with Glide one of the possible binding modes selected with MOE is reproduced.

A RMSD matrix based on the common scaffold was generated using the non-duplicate poses (see the subchapter 5.1.1.6). The MS Excel Add-In XLSTAT[130] was used for hierarchical clustering of the RMSD matrix above-mentioned using the Single Linkage method. Bottegoni et al.[132] demonstrated that the Single Linkage rule is more aggregative than the Ward's method, hence the number of clusters obtained in the first case is lower. The use of the Single Linkage instead of the Ward's method, used to analyze the poses of trapped propafenones docked into the closed conformation of the hERG channel, comes from the idea that a lower number of clusters facilitates the analysis of the docking poses, and that similar poses are less dispersed in different clusters. In the Single Linkage clustering method the similarity between two clusters is determined by the distance between the two closest pair of objects belonging to the two clusters considered. In the first step, all objects are considered as a single cluster. At each step, the distance between the clusters is computed. To calculate the distances between the clusters all objects of all clusters are considered. The distance between two clusters is the shortest distance from any object of one cluster with any object in the second one. The clusters with the shortest distance are merged into a new cluster. With the aim to obtain a good intracluster homogeneity,

only when all the poses in all clusters were within 2.5 Å from the centroid the hierarchical clustering procedure was completed. Subsequently, the poses of the clusters containing at least one pose of all compounds and the amino acids within 5 Å from the ligand were minimized using the MMFF94x force field. After the minimization, the conformations with distorted aromatic groups were deleted after a visual comparison with some distorted crystallized aromatic rings obtained from the Cambridge Crystallographic Data Centre.[133] To analyze the remaining poses, the RMSD_MATRIX.SVL was written to calculate a second RMSD matrix based on the common scaffold. The Single Linkage agglomerative hierarchical clustering method was then applied to identify the common binding modes. The clusters containing one or more poses of all ligands were selected. The RMSD_MATRIX.SVL generates an RMSD matrix without performing the rotation of the poses as in ROTALI. Only the heavy atoms of the scaffold are considered to calculate the RMSD values. At the beginning it takes the first pose of the database as the reference pose against which the RMSD value of the other poses are calculated, then the second pose is taken as the reference one. The operations are repeated until all the poses of the database are taken as reference. To refine the selection of the binding modes, an appropriate svl script called RECEPTOR_RMSD.SVL was written to calculate the RMSD matrix of the complexes generated merging the RMSD matrices of the minimized models and poses of each selected clusters, taking into account the common scaffold of the ligands (fig. 7) and the amino acids that have at least one atom within 4.5 Å in 90% of the poses. The RECEPTOR_RMSD.SVL calculates the RMSD matrix of the common scaffold (through the function RMSD_MATRIX.SVL) and of the amino acids selected in each subunits. To find the amino acids within 4.5 Å from the ligand the script COUNT_SIDECHAIN.SVL was written. The script detects which amino acids are within the cut off distance from the ligand, and it creates columns headed with the subunit number and the name of the amino acids detected. Every time that an amino acid of a certain subunit is within the distance of 4.5 Å from the pose, the script writes the number 1 under the respective column. Subsequently, manually ranking the columns it is possible to determine whether the amino acid considered is within 4.5 Å in 90% of the poses or not. On the RMSD matrix of the complexes was performed a cluster analysis using the Single Linkage clustering method to identify the common binding modes. The clusters containing poses of all ligands were validated through consensus scoring analysis using the London dG, ASE, Affinity dG and Alpha HB scoring functions as described in the subchapter 5.1.1.6, and through mutagenesis data.[11, 126] Both Witchel et al.[11] and Windisch et al.[126]

demonstrated toward the mutagenesis data that only the amino acid Phe656 is essential in the hERG inhibition mediated by propafenone and derivatives, while it is less sensible or insensible to the mutations of the amino acids Thr623, Ser624, Val625, Ser631, Gly648, Tyr652 and Val659.

In order to discriminate between the trapped and non-trapped compounds, potential energies considering the heavy atoms of the scaffold, poses, and substituents R1 (table 2) were calculated with the software package MOE2008.10 using MMFF94x force field, and finally these values were ranked in ascending order. In order to get a size independent parameter potential energy values were divided by the total number of heavy atoms. To extract the heavy atoms of the common scaffold, poses and R1 substituents, the smile string of the function `abstract_commonscaffold` in the script `FORMELPOOL.SVL` (subchapter 5.1.1.6) was from time to time modified specifying the smile strings regarding the molecules and each moieties of the compounds considered in the calculation of the potential energy. For each ligand a database was created considering only the heavy atoms of the poses, common scaffold and R1 substituents. Subsequently the databases were merged and the potential energy values calculated.

5.2.1.5 Analysis of ligand interactions

The interactions between the ligands and the hERG channel were analyzed with the MOE2008.10 software package.

5.2.2 Results

5.2.2.1 Poses analysis

Many studies demonstrate that the hERG blockers inhibit the channel in the open state. Several lines of evidences indicate that the binding site is located in the central cavity.[6, 9, 12, 22, 131]. At the level of the inner pore there are two concentric rings. The ring formed by four Tyr652 faces to the inner cavity. During the transition from the closed state to the open state the distance between the C α of Tyr652 of opposite subunits increases from 14.3 Å to 19.3 Å. Four Phe656 form the ring at the cytoplasmic side of the hERG channel. When the hERG channel is in the closed conformation, the space between the C α of Phe656 of opposite subunits is too tight to allow the molecules to interact with Phe656.[102] During the channel opening the distance

between the C α increases from 9.4 Å to 15.6 Å, allowing the compounds to enter into the central cavity and to block the channel.[102] In the open conformation of the hERG channel, the four Phe656 can form hydrophobic, π -cation or π -stacking interactions with hERG blockers, making the channel polyspecific in ligand recognition. The increase in the size of the two concentric rings from the closed to the open state corresponds also to an increase of the volume from 1048.2 Å³ to 1642.2 Å³. The reduction of more than 30% of the volume during the channel gating might explain why only small and flexible compounds are trapped into the closed state.[102]

In this study, we docked the compounds propafenone, GPV0005, GPV0009, GPV0180, GPV0031 and GPV0576 (table 2) into the homology model of hERG channel in the open state, in order to investigate the interactions that determine the inhibition of the channel and to shed light on the drug trapping phenomena. Due to the symmetry of the homology model of the hERG channel in the open state, the poses were distributed in all four directions (see also the subchapter 5.1.1.2). The symmetry operation was applied to align the poses and to delete the duplicates. The non-duplicate poses were used to generate a RMSD matrix based on the common scaffold. Through hierarchical cluster analysis, we identified 3 clusters containing one or more poses of all compounds. A visual inspection of the clusters highlighted that some poses had a distorted aromatic group. For this reason, the poses of the selected clusters and the amino acids that were within 5 Å from the ligand were minimized. After the minimization, the poses that still had a distorted aromatic group were deleted. The remaining poses were analyzed through the Single Linkage clustering method performed on the RMSD matrix based on the common scaffold (fig. 7), and the clusters containing poses of all ligands were selected. The poses of the selected clusters and the amino acids defined as part of the binding site were used to calculate the RMSD matrix of the complexes. Through the Single Linkage clustering method of the RMSD matrix 3 clusters were selected.

In the first cluster, the analysis of the centroid interactions shows that it forms three hydrogen-bonds with two amino acids of the same subunit: two hydrogen-bonds are between the hydroxyl group of the propafenone scaffold and the amino acids Thr623 and Ser624, and one involves the charged nitrogen atom and the Ser624 (fig. 13). The two aromatic groups make T-stacking and hydrophobic interactions with three Phe656 of the subunits B-D. The poses in the first cluster show that the two aromatic groups of the ligands form π -stacking, and/or T-stacking, and/or parallel-displaced π - π interactions with two or three Phe656 of the subunits B-D. Due to the small distance between the four subunits near the selectivity filter, the hydroxyl group of the

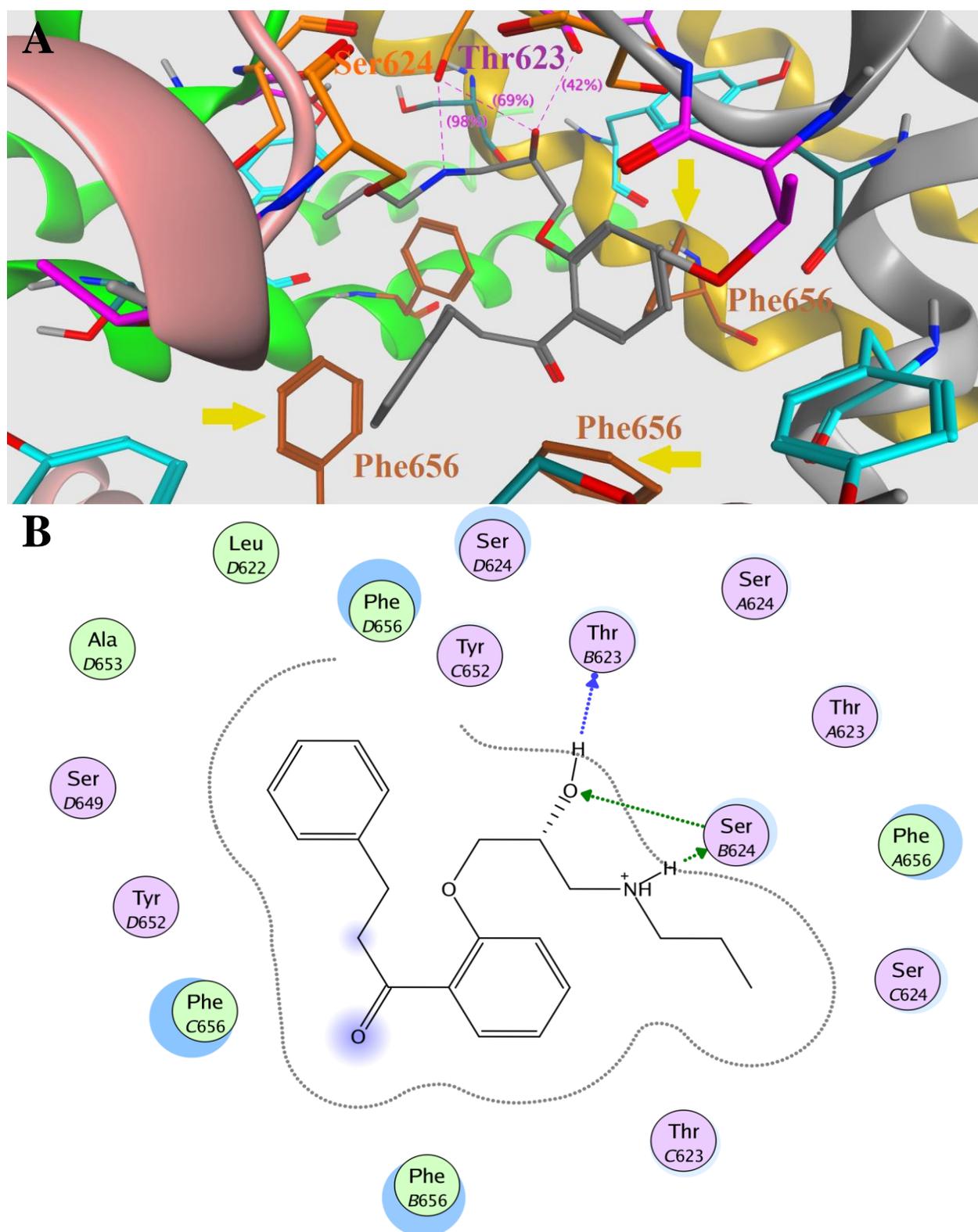


Fig. 13 Centroid of the first cluster. 3D (A) and 2D (B) interaction pattern of the centroid of the first cluster. Only the amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented. The subunits A, B, C and D are colored in green, yellow, grey and pink respectively. The yellow arrows indicate the amino acids Phe656 interacting with the centroid through T-stacking and hydrophobic interactions.

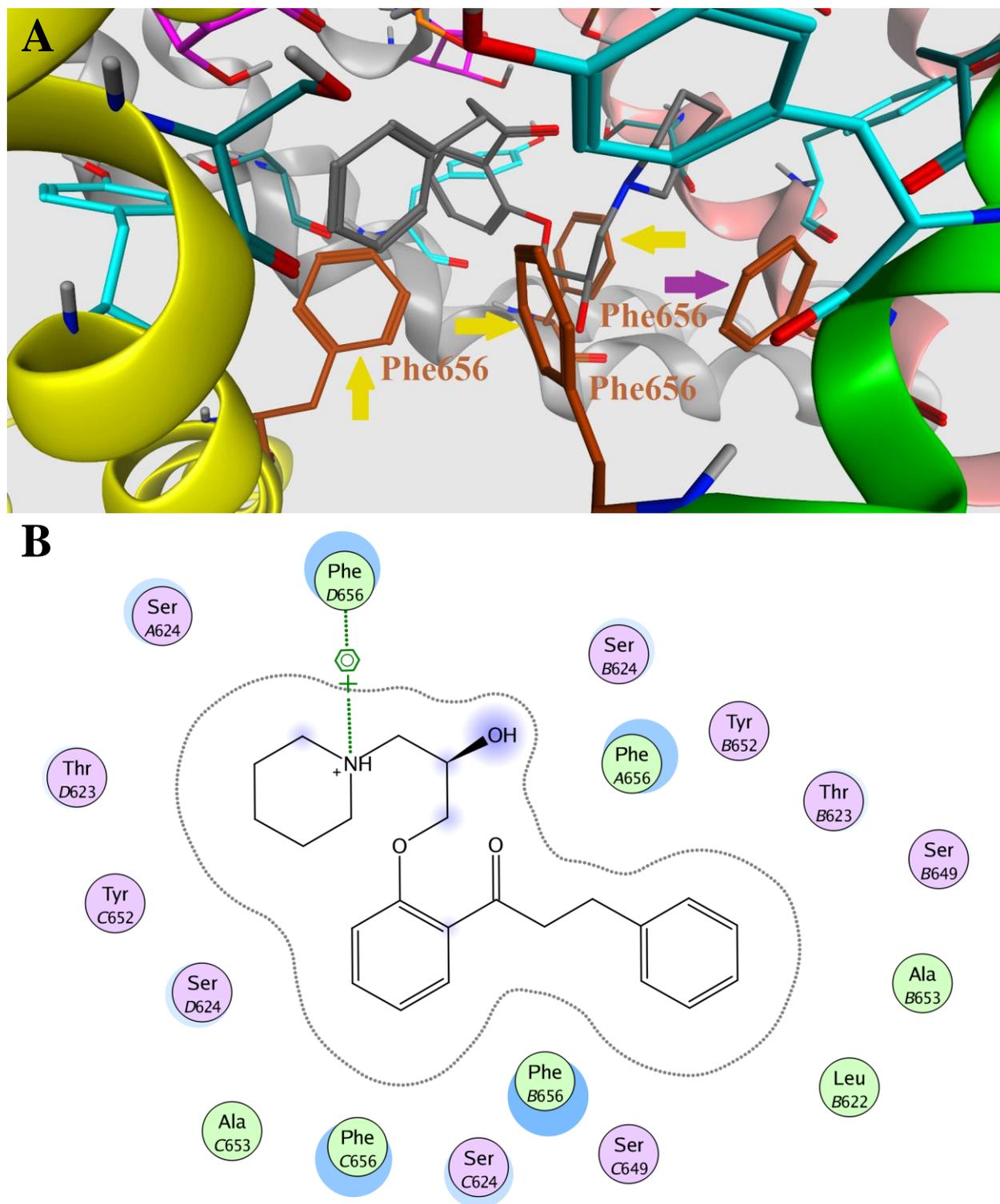


Fig. 14 Centroid of the second cluster. 3D (A) and 2D (B) interaction pattern of the centroid of the second cluster. The amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented. The subunits A, B, C and D are colored in green, yellow, grey and pink respectively. The violet and yellow arrows indicate respectively the Phe656 forming π -cation and T-stacking/parallel displaced π - π interactions with the centroid.

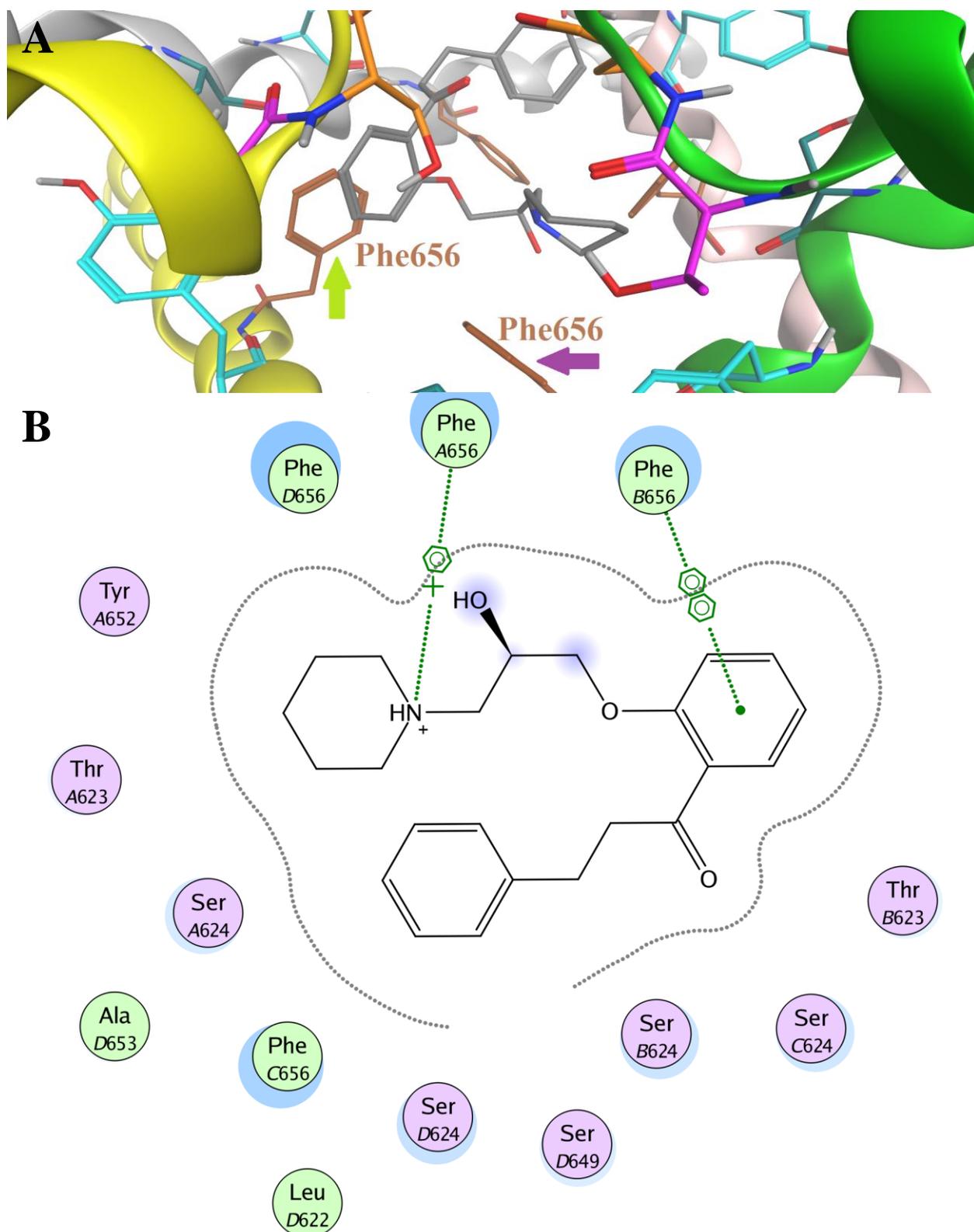


Fig. 15 Centroid of the third cluster. 3D (A) and 2D (B) interaction pattern of the centroid. Only the amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented. The subunits A, B, C and D are green, yellow, grey and pink colored respectively. The green and the violet arrows indicate respectively the Phe656 making π - π and π -cation interactions with the centroid.

ligands can form hydrogen bonds with the amino acid Ser624 of the subunits A-B and D, and/or with the Thr623 of the subunit B. The charged nitrogen can make hydrogen-bond interactions with Ser624 and/or with Thr623 of the subunits A-C and A-B, respectively.

In the second cluster, the charged nitrogen atom of the centroid forms π -cation interaction with Phe656, while the phenyl rings make T-stacking and parallel-displaced π - π interactions with the other three Phe656 (fig. 14). The analysis of the poses in the second cluster reveals that the two aromatic rings of the ligands interact with two or three Phe656 of subunits A-C through π -stacking, and/or T-stacking, and/or parallel-displaced π - π interactions, while the charged nitrogen atom forms π -cation interactions with one or two Phe656 of subunits A and D. In this cluster there is at least one pose per ligand that makes π -cation, and/or π -stacking, and/or parallel-displaced π - π , and/or T-stacking interactions with Phe656, but not with other amino acids, in agreement with the mutagenesis data.[11, 126]

In the third cluster, the centroid interacts with the hERG channel through π -stacking and π -cation interactions with two Phe656 of adjacent subunits (fig. 15). The two aromatic rings of the poses in the third cluster interact with the hERG channel through π -stacking, and/or T-stacking, and/or parallel displaced interactions of one or two Phe656 of the subunits B-D. The protonated nitrogen forms π -cation interactions with one or two Phe656 of the subunits A and D. In these cluster not all ligands show π -cation, and/or π -stacking, and/or T-stacking, and/or parallel-displaced π - π interactions with the amino acid Phe656 in contrast to the mutagenesis data.

With the aim to further select the clusters, we performed a consensus scoring based on the first ten poses ranked according to the Affinity dG, Alpha HB, ASE and Londong dG scoring functions of MOE. The first and the second clusters have one or more poses ranked within the first ten in all four scoring functions, while in the third cluster the first ten poses of the London dG scoring function are missing. In an interesting study Witchel et al.[11] performed a detailed alanine scanning to determine the binding site of propafenone. Noteworthy they found that only the Phe656Ala mutation affects the block of the hERG channel by propafenone, while it is insensitive or less sensitive to the mutations of Thr623, Ser624, Val625, Gly628, Ser631, Gly648, Tyr652 and Val659. Similar results were obtained by Windisch et al.[126] Considering the consensus scoring and the mutagenesis data we can prioritize the second cluster, because it has one or more conformations within the top ten in all four scoring functions, and only this cluster has at least one pose per ligand in agreement with the mutagenesis data.

Even if it is possible to prioritize the second binding mode, to not neglect any possible trapping

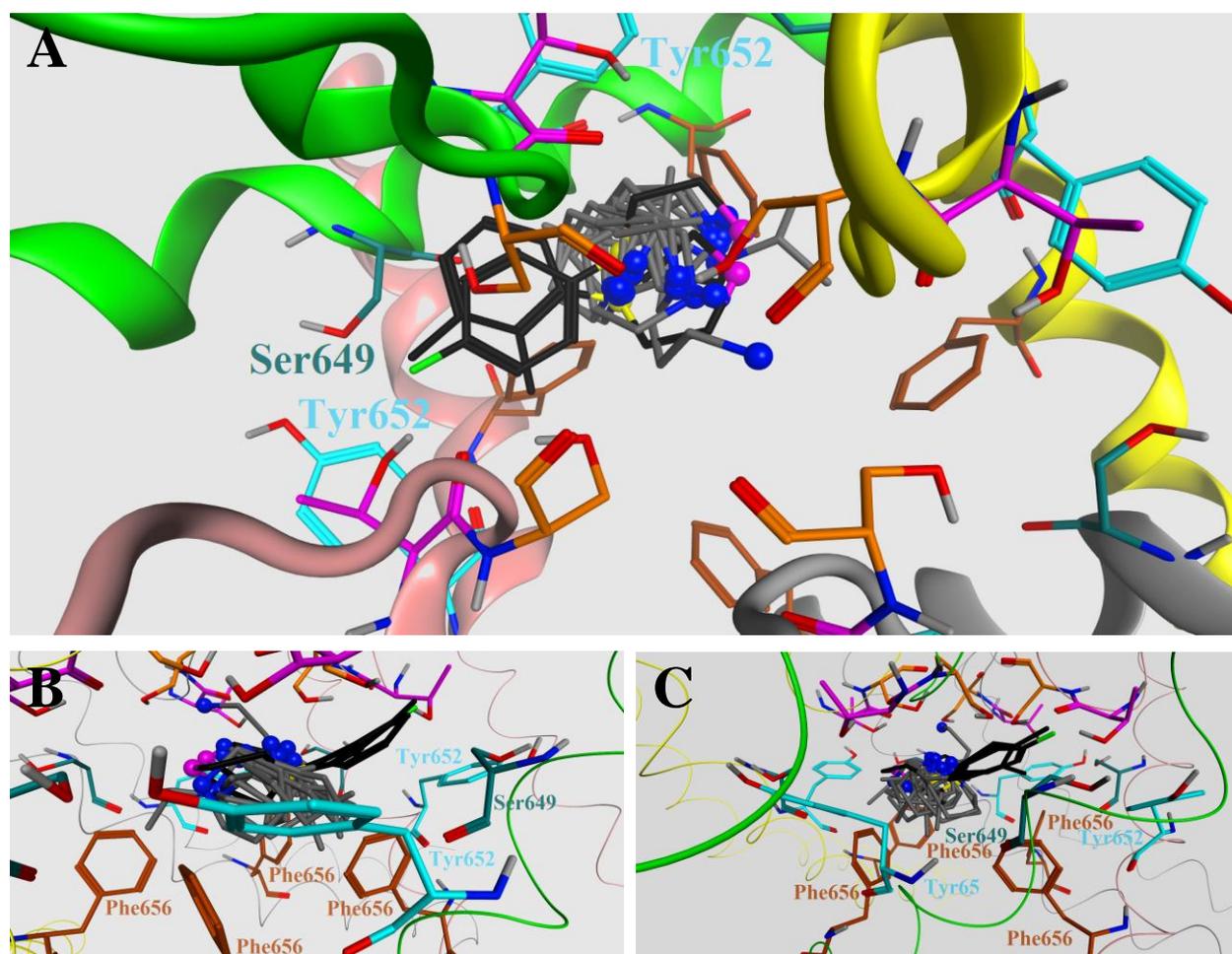


Fig. 16 Position of R1 in the poses of cluster1. Upper (A), lateral (B) and frontal view (C) of the substituent R1 in poses of cluster1. Only the amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented. The subunits A, B, C and D are colored in green, yellow, grey and pink respectively. The carbon atoms of the substituent R1 of the non-trapped compounds are colored in black. The charged nitrogen is colored in blue and in violet in the trapped and non-trapped molecules respectively. The non-protonated nitrogen of the non-trapped compounds is colored in yellow.

mechanisms, the positions of the substituent attached to protonated nitrogen in all three selected clusters will be discussed.

The poses of the first cluster show that the substituent attached to the charged nitrogen is placed near the amino acids Ser649 and Tyr652 of the same subunit (fig. 16). The presence of the backbone (fig. 16C) renders it impossible for the non-trapped propafenones to place the protonated nitrogen near the two amino acids above-mentioned due to the bulkiness and the rigidity of the R1 substituents, in contrast with the position of the charged nitrogen of trapped compounds. This leads to a push back of the charged nitrogen in the non-trapped compounds (fig. 16A), and forces their R1 substituent to bend upward (fig. 16B and 16C). In contrast to the results of Windisch et al.,[126] through these poses it is not possible to explain the “foot in the door” mechanism. This suggests that these conformations might be not as reliable as the poses in

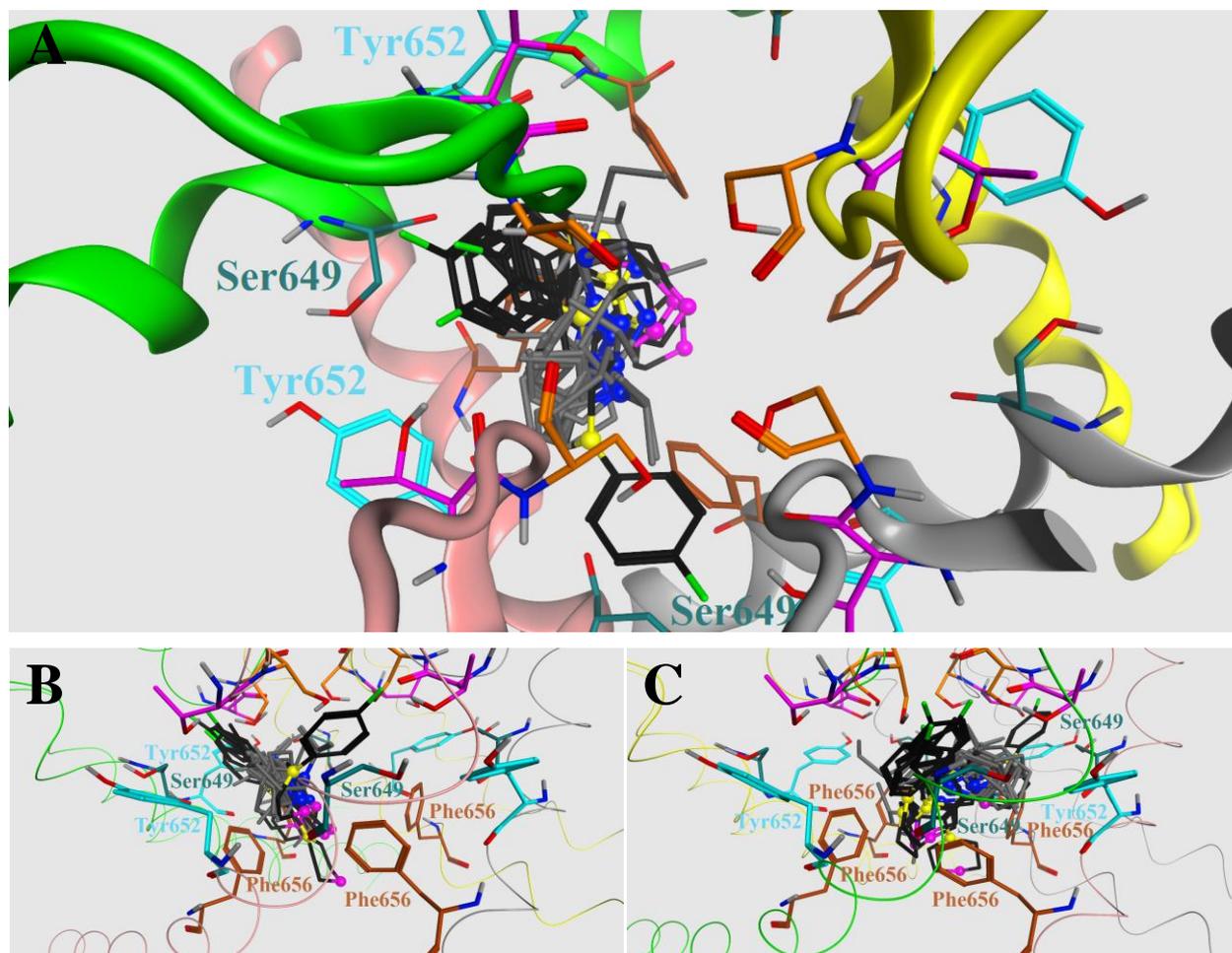


Fig. 17 Position of R1 substituents in the poses of cluster2. Upper (A), lateral (B) and frontal view (C) of the substituent R1 in poses of cluster2. Only the amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented. The subunits A, B, C and D are colored in green, yellow, grey and pink respectively. The carbon atoms of the substituent R1 of the non-trapped compounds are colored in black. The charged nitrogen is colored in blue and in violet for the trapped and non-trapped molecules respectively. The noncharged nitrogen of the substituent of the non-trapped compounds is colored in yellow.

the second and third clusters.

The poses of the second cluster place the R1 substituent in a zone of the channel delimited by the amino acids Ser649-Tyr652 of the same subunit and Tyr652 of an adjacent subunit (fig. 17A and B). As in the first cluster the protonated nitrogen of non-trapped propafenones is pushed away by the R1 substituent from the three amino acids above-mentioned. In contrast to the first cluster, the charged nitrogen of the poses in the second cluster is located lower, near the ring formed by the four Phe656. This forces the piperidin group present in some R1 substituents to bend upward to avoid clashes with the Phe656. The space between the position of the piperidin group in the trapped compounds and the top of the hERG channel is not big enough to accommodate the

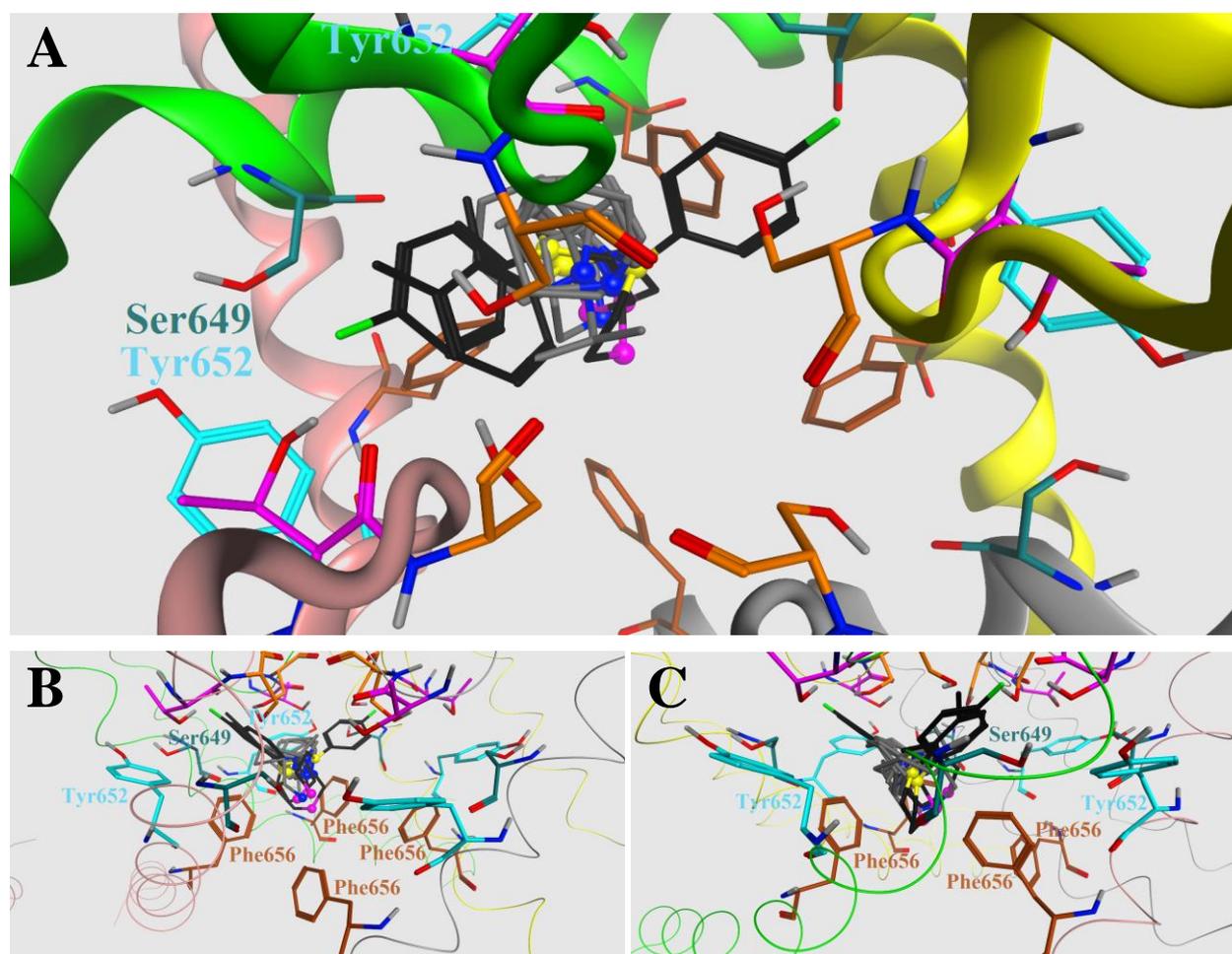


Fig. 18 Position of R1 in the poses of cluster3. Upper (A), lateral (B) and frontal view (C) of the substituent R1 in poses of cluster3. Only the amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented. The subunits A, B, C and D are colored in green, yellow, grey and pink respectively. The carbon atoms of the substituent R1 of the non-trapped compounds are colored in black. The charged nitrogen is colored in blue and in violet for the trapped and non-trapped molecules respectively. The noncharged nitrogen of the substituent of the non-trapped compounds is colored in yellow.

substituted phenyl rings in the R1 substituents of non-trapped propafenones. This determines the placement of the charged nitrogen in the piperazine group of non-trapped molecules within the four Phe656, while in the trapped compounds the protonated nitrogen is above these four amino acids (fig. 17B and C). During the channel gating the protonated nitrogen of the trapped propafenones moves near the amino acids Ser649 and Tyr652, as seen in the subchapter of closed hERG channel conformation. In non-trapped propafenones the rigidity and the bulkiness of the substituents R1 prevent this movement. Hence, due to the position of the charged nitrogen within the ring formed by the four Phe656, the substituents R1 of the non-trapped compounds hinder the channel gating with the “foot in the door” mechanism.

In the poses of the third cluster the substituent R1 is placed in the zone delimited by two Tyr652 and one Ser649 (fig. 18A, B and C). The charged nitrogen atoms of the poses in the second and

third clusters are located at the same level. Generally in the non-trapped propafenones the bulky and rigid substituent push the protonated nitrogen lower than to its position in the trapped compounds. The piperidine ring present in some R1 substituents is bent upward to avoid possible clashes with the four Phe656. Due to the lack of space between the piperidine ring of the compounds GPV0005 and GPV0180 and the top of the central cavity, the phenyl rings present in the R1 substituents of the non-trapped molecules force the piperazine moieties to be placed in a lower position. Hence, the protonated nitrogen of the non-trapped compounds is within the ring formed by the four Phe656, while in the trapped propafenones it is above it (fig. 18B and C). As previously described, in the non-trapped molecules the position of the charged nitrogen within this ring hinders the hERG channel gating with the “foot in the door” mechanism due to the rigidity and the bulkyness of the R1 substituent.

The poses of cluster1 are similar to the conformations of propafenone, GPV0005 and GPV0576 published by Thai et al.[102] (RMSD value from the centroid lower than 2.5 Å). The two phenyl rings of the centroid and of the published poses are near each other, with the aromatic rings of the previously published poses placed orthogonal to the channel axis, while in the centroid of cluster1 they are bent to form T-stacking and hydrophobic interactions with three Phe656 of adjacent subunits. The chain that connects the two rings is upward in the published poses, and downward in the centroid, explaining why it loses the hydrogen bond between the carbonyl group and the Ser624 predicted by Thai et al.[102] Another difference regards the position of the protonated nitrogen and the substituents attached. In the centroid of cluster1 the charged nitrogen forms a hydrogen bond with Ser624, placing the substituent in a zone delimited by Ser649 and Tyr652 of the same subunit. In the previously published poses, the protonated nitrogen is placed near the ring formed by the four Phe656, at the same level where the second and third clusters place the non-protonated nitrogen of the non-trapped compounds. Despite the fact that the centroid is similar to the poses published by Thai et al.,[102] the mutagenesis data, the consensus scoring analysis and the agreement with the “foot in the door” mechanism suggest that it is reasonable to prioritize the second binding mode.

The position of the protonated nitrogen of the non-trapped compounds within the ring on the cytoplasmic side of the hERG, as it happens in the second and third clusters, might hinder the closure of the activation gate. This is in agreement with the “foot in the door” mechanism,[101, 134] and explains why the molecules with a R1 substituent bulky and rigid are not trapped in the

Table 9. Poses of cluster1. E scaffold: potential energy calculated considering the heavy atoms of the scaffold (table 2). E pose: potential energy calculated considering only the heavy atoms of the pose. N1: number of heavy atoms of the pose. E R1: potential energy calculated considering only the heavy atoms of the substituent R1. N2: number of heavy atoms of the substituent R1. The values of E are in kcal/mol.

Poses	E scaffold	E pose	N1	E pose / N1	E R1	N2	E R1 / N2	Trapped?
Propafenone	60.58	88.07	25	3.52	0.72	4	0.18	Yes
Propafenone	77.94	111.62	25	4.46	0.80	4	0.20	Yes
Propafenone	76.79	100.86	25	4.03	0.80	4	0.20	Yes
Propafenone	71.19	90.32	25	3.61	0.83	4	0.21	Yes
Propafenone	73.70	114.09	25	4.56	0.90	4	0.22	Yes
Propafenone	74.69	93.40	25	3.74	2.45	4	0.61	Yes
Propafenone	83.22	114.61	25	4.58	2.79	4	0.70	Yes
Propafenone	73.88	118.95	25	4.76	4.48	4	1.12	Yes
GPV0009	73.39	102.49	28	3.66	10.66	7	1.52	Yes
GPV0180	66.24	109.49	31	3.53	12.38	6	2.06	Yes
GPV0005	70.65	90.18	27	3.34	13.10	6	2.18	Yes
Propafenone	64.45	99.49	25	3.98	9.03	4	2.26	Yes
GPV0005	63.96	86.46	27	3.20	13.62	6	2.27	Yes
GPV0005	70.82	95.07	27	3.52	13.76	6	2.29	Yes
GPV0005	75.54	96.77	27	3.58	13.76	6	2.29	Yes
GPV0180	67.75	112.49	31	3.63	16.21	6	2.70	Yes
GPV0005	65.19	91.42	27	3.38	17.39	6	2.90	Yes
GPV0031	77.48	170.26	34	5.01	77.35	13	5.95	No
GPV0576	61.29	160.88	35	4.60	88.66	14	6.33	No

hERG channel. It is interesting to note that the propafenones docked into the hERG channel in the closed state as well as in the open state place the R1 substituent near the amino acids Ser649 and Tyr652 of the same and/or adjacent subunits.

To discriminate between trapped and non-trapped propafenone derivatives the potential energy values of the poses, scaffold and R1 substituents (table 2), were calculated. Obviously in all three clusters it is not possible to separate the poses of the trapped compounds from the non-trapped ones considering the potential energy values of the scaffold (table 9, 10 and 11). In all three clusters the potential energy values of the non-trapped poses, and of their substituent R1, are always higher than the values obtained for the trapped compounds. This is due to the greater number of atoms in the non-trapped compounds, and to the less favorable conformation with which the non-trapped propafenone derivatives interact with the hERG channel in the open state. In the first and in the second cluster, ranking the compounds according to the values obtained dividing the potential energy by the total number of heavy atoms, it is not possible to discriminate between the trapped and non-trapped compounds, due to the presence of one outlier (one pose of propafenone and GPV0005 for cluster1 and cluster2 respectively) in the poses of

Table 10. Poses of cluster2. E scaffold: potential energy calculated considering the heavy atoms of the scaffold (table 2). E pose: potential energy calculated considering only the heavy atoms of the pose. N1: number of heavy atoms of the pose. E R1: potential energy calculated considering only the heavy atoms of the substituent R1. N2: number of heavy atoms of the substituent R1. The values of E are in kcal/mol.

Poses	E scaffold	E pose	N1	E pose / N1	E R1	N2	E R1 / N2	Trapped?
Propafenone	84.91	113.63	25	4.54	1.01	4	0.25	Yes
Propafenone	80.37	100.56	25	4.02	2.49	4	0.62	Yes
GPV0009	75.29	100.24	28	3.58	8.61	7	1.23	Yes
GPV0009	80.98	101.94	28	3.64	8.85	7	1.26	Yes
GPV0009	81.32	102.48	28	3.66	9.12	7	1.30	Yes
GPV0009	66.24	91.45	28	3.27	9.20	7	1.31	Yes
GPV0009	65.86	90.80	28	3.24	9.91	7	1.42	Yes
GPV0009	64.50	90.03	28	3.21	10.28	7	1.47	Yes
GPV0009	71.53	97.32	28	3.48	11.12	7	1.59	Yes
GPV0009	68.11	93.03	28	3.32	11.12	7	1.59	Yes
GPV0009	66.02	94.24	28	3.37	11.83	7	1.69	Yes
GPV0180	62.29	105.30	31	3.40	12.68	6	2.11	Yes
GPV0005	89.73	128.59	27	4.76	13.25	6	2.21	Yes
GPV0005	80.12	105.34	27	3.90	13.57	6	2.26	Yes
GPV0180	64.41	120.40	31	3.88	13.58	6	2.26	Yes
GPV0180	64.47	118.68	31	3.83	13.66	6	2.28	Yes
GPV0005	62.00	101.57	27	3.76	13.85	6	2.31	Yes
GPV0005	63.05	103.27	27	3.82	13.96	6	2.33	Yes
GPV0180	68.75	118.72	31	3.83	16.31	6	2.72	Yes
GPV0576	77.59	170.33	35	4.87	76.16	14	5.44	No
GPV0576	74.81	169.48	35	4.84	77.10	14	5.51	No
GPV0031	76.44	165.55	34	4.87	71.67	13	5.51	No
GPV0031	75.08	160.75	34	4.73	71.71	13	5.52	No
GPV0031	76.43	164.80	34	4.85	71.83	13	5.52	No
GPV0031	77.59	169.55	34	4.99	76.21	13	5.86	No
GPV0031	73.94	178.14	34	5.24	90.02	13	6.92	No

non-trapped propafenone derivatives. For the first cluster, it is not clear why this pose is an outlier (table 9). Regarding the second cluster (table 10), the ether linkage between the phenyl ring and the piperidine group of the outlier goes deep into the ring formed by the four Phe656 of the hERG channel, in the same position found in non-trapped propafenones, assuming an unfavourable conformation that explains the high value of potential energy. In contrast to the first two clusters, in the third one it is possible to discriminate between the trapped and non-trapped propafenone derivatives ranking the compounds according to the values obtained dividing the potential energy of the pose by the total number of heavy atoms (table 11). The high values of potential energy of non-trapped molecules are due to the phenyl-R1 linker that is placed deeply into the inner pore, assuming a U conformation, and to the rigidity and bulkiness of the R1 substituents. Only one pose of the trapped compounds have the phenyl-R1 linker that

Table 11. Poses of cluster3. E scaffold: potential energy calculated considering the heavy atoms of the scaffold (table 2). E pose: potential energy calculated considering only the heavy atoms of the pose. N1: number of heavy atoms of the pose. E R1: potential energy calculated considering only the heavy atoms of the substituent R1. N2: number of heavy atoms of the substituent R1. The values of E are in kcal/mol.

Poses	E scaffold	E pose	N1	E pose / N1	E R1	N2	E R1/ N2	Trapped?
GPV0009	65,67	89,13	28	3,18	8,87	7	1,27	Yes
Propafenone	56,58	90,04	25	3,60	7,61	4	1,90	Yes
GPV0180	79,49	124,37	31	4,01	12,13	6	2,02	Yes
GPV0005	71,04	101,04	27	3,74	12,90	6	2,15	Yes
GPV0005	75,82	100,13	27	3,71	15,91	6	2,65	Yes
GPV0005	55,82	87,74	27	3,25	16,68	6	2,78	Yes
GPV0005	74,33	109,08	27	4,04	17,32	6	2,89	Yes
GPV0005	71,04	106,82	27	3,96	17,53	6	2,92	Yes
GPV0005	60,63	89,65	27	3,32	17,75	6	2,96	Yes
GPV0576	78,17	178,60	35	5,10	88,14	14	6,30	No
GPV0031	58,85	152,94	34	4,50	85,21	13	6,55	No
GPV0031	82,91	189,38	34	5,57	90,94	13	6,99	No

goes deeply into the inner pore as in the non-trapped compounds, but it has a lower value of potential energy probably due to the smaller size and to the flexibility of the molecule.

In all three clusters the values of potential energy of the R1 substituents is much higher in the non-trapped propafenones than in the trapped ones. The discrimination between trapped and non-trapped compounds is still clear even when the value of potential energy is divided by the total number of heavy atoms of R1. This is due to the size and to the rigidity of the substituent in the non-trapped propafenone derivatives. This result allow to hypothesize that during the channel closure, the rigidity and the bulkiness of the substituent R1 of non-trapped compounds destabilize the ligand-hERG complex, forcing the dissociation through the “foot in the door” mechanism.

5.2.3 Conclusions

In this work, we docked a series of six propafenone derivatives into the homology model of the hERG channel in the open state. Hierarchical cluster analysis of the RMSD matrix based on the common scaffold and considering the amino acids interacting with the poses reveals three possible binding modes. The comparison of the selected clusters with the results obtained in the closed state shows that the substituent attached to the charged nitrogen is situated near the zone Tyr652-Ser649-Tyr652 in both cases. In the second and in the third cluster the bulky and rigid R1 substituents of non-trapped compounds force the protonated nitrogen to be placed within the

ring formed by the four Phe656. This might hinder the closure of the channel activation gate, leading to the dissociation of the ligand with the “foot in the door” mechanism. Hence, these clusters can explain why the propafenone derivatives with a rigid and bulky R1 substituent are not trapped. Interestingly, the values of potential energy of R1, and this value divided by the number of heavy atoms, allow us to discriminate between trapped and non-trapped propafenone and derivatives. This might indicate that the bulkiness and the rigidity of the R1 substituent play an important role in the propafenone derivatives-hERG complex dissociation during the channel gating.

In this study only a small set of propafenones were used, hence these results cannot be generalized for other classes of hERG blockers. It is also important to note that drug trapping is a multistep mechanism in which different intermediate states of the hERG channel gating might have a strong influence.

5.3 Docking studies into models of hERG channel in the open state obtained from molecular dynamics simulations

5.3.1 Material and methods

5.3.1.1 Docking protocol

With the aim to take into account the flexibility of the amino acid sidechains facing the central cavity of the hERG channel, nine propafenone derivatives were docked into several models of hERG channel in the open state obtained from snapshots of molecular dynamics simulations kindly provided by Anna Weinzinger.[135] To the pdb file of the eleven models the hydrogen atoms were added. The models were minimized by soft minimization operation performed with the MOE2008.10 tool Energy Minimize[127] in the Amber 99 force field, constraining the chirality, applying a root mean square gradient value of 0.5 and a value of 10 to the Thether Atoms option for heavy atoms. The partial charges of the atoms were recalculated during the minimization. The models in which the amino acids of the central cavity or of the inner pore violated the Ramachandran Plot were discarded. The remaining eight models were then superposed according to the C α of the amino acids with the MOE2008.10 tool Protein Superpose using the default conditions.

Recently Windisch et al.[126] synthesized and tested the new propafenone derivatives GPV0019, GPV0062 and STC-AS03. These molecules have only small differences in the R1 substituents, but they show a different recovery from the hERG block. In the non-trapped compound GPV0019 the substituent at the nitrogen atom is a piperazine group, while in the trapped molecules GPV0062 and STC-AS03 this moiety is replaced with a piperidine. To investigate in more detail the trapping phenomena these compounds were included in this study. The ligands GPV0019, GPV0062 and SCT-AS03 were built and minimized as described in the subchapter 5.1.1.1 using the tools implemented in MOE2008.10. The sp² hybridization was assigned to the aniline nitrogen of the compound GPV0019. The 100 most diverse conformations of the compounds GPV0019, GPV0062 and SCT-AS03 were selected through a systematic search and the tool Diverse Subset implemented in MOE2008.10 applying the protocol described in the subchapter 5.1.1.1. The selection of the 100 most diverse conformations of propafenone, GPV0005, GPV0009, GPV0031, GPV0180 and GPV0576 was already done in the previous

study (see the subchapter 5.2.1.1). The databases of the 100 most diverse conformations were then docked into the eight selected models. The binding site was defined by the atoms of the amino acids within a sphere of 6.5 Å from the amino acids Thr623, Ser624, Tyr652 and Phe656 of each subunit. This allows the selection of all the amino acids of the central cavity. In the previous study (chapter 5.2) this was not possible, due to the high number of poses docked outside the hERG channel applying this definition of the binding site. The placement method, the minimization and the other conditions used to dock the compounds are described in the subchapter 5.2.1.1.

5.3.1.2 Identification of the common binding modes

Due to the asymmetric binding site it was not necessary to apply the alignment of the poses and delete the duplicates with the processes described in subchapters 5.1.1.2, 5.1.1.3, 5.1.1.4 and 5.1.1.5.

For each database of poses, the first 210 poses ranked according to the London dG scoring function were selected. To reduce the amount of poses considered, and to take into account only the most different ones to cover as much conformational space as possible, of the 210 poses only the 20 most diverse conformations based on the RMSD and E were selected with the Diverse Subset tool of MOE2008.10.

The common binding modes were selected and validated, and the potential energy values of the poses, common scaffold and R1 substituents calculated as described in the subchapter 5.2.1.4.

5.3.1.3 Analysis of ligand interactions

See the subchapter 5.2.1.5.

5.3.2 Results

5.3.2.1 Pose analysis

A series of nine propafenone derivatives were docked into eight homology models of hERG channel in the open state derived from molecular dynamics simulations. Table 12 clearly shows that the cytoplasmic side of the inner cavity exhibits significant movements, while the

movements of the four amino acids Ser624 are smaller. This is in agreement with the fact that the movement of the amino acids in the inner pore determines the closure of the hERG channel. The volume of the hERG channel varies from a minimum of 1322.0 Å³ to a maximum of 2246.1 Å³.

Table 12. Structural information on the eight hERG models. Distance 1: distance between the C α of Ser624 of opposite subunits. Distance 2: distance between the C α of Phe656 of opposite subunits.

Models	Distance 1 (Å)	Distance 2 (Å)	Volume (Å ³)
Model 1	8.28 - 8.97	18.30 – 18.63	1670.8
Model 2	8.32 – 8.80	17.19 – 19.44	1681.5
Model 3	9.17 – 9.36	16.59 – 16.92	2097.3
Model 5	9.52 – 9.77	16.83 – 17.28	1647.1
Model 6	8.42 – 10.07	17.27 – 19.72	2246.1
Model 8	9.82 – 10.00	15.30 – 16.07	1623.8
Model 9	10.04 – 11.09	13.68 – 16.39	1370.3
Model 11	9.79 – 10.13	13.86 – 17.97	1322.0

In this study, we docked propafenone, GPV0005, GPV0009, GPV0019, GPV0031, GPV0062, GPV0180, GPV0576 and SCT-AS03 (table 2) into eight models of the hERG channel in the open state. Two clusters, from now on called cluster four and five, were selected through Single Linkage cluster analysis of the RMSD matrix based on the common scaffold and the successive refinements.

In cluster four, one phenyl ring of the centroid forms a parallel-displaced π - π interaction with Phe656 (fig. 19). The second phenyl ring is located under the four Phe656. The analysis of the poses of this cluster show that the aromatic ring form π -stacking, and/or parallel-displaced π - π interactions with Phe656 of the subunit D, π -cation interaction with Phe656 of subunit B, and parallel-displaced π - π interaction with the Tyr652 of subunit C.

The centroid of cluster five makes π -stacking interaction with the Phe656 of the subunit D (fig. 20). The protonated nitrogen is placed under the ring formed by the four Phe656 directly in contact with the cytoplasm (fig. 20A). In this position, the charged nitrogen cannot interact with the hERG channel. One phenyl ring of the scaffold is placed between the two concentric rings of the inner pore, while the second one is placed under the selectivity filter. The poses of the fifth cluster show that the phenyl ring form π -stacking or parallel-displaced π - π interactions with the Phe656 of the subunit D, and T-stacking, or parallel-displaced π - π interactions with Tyr652 of the subunit C. This binding mode seems to be not reliable, because it cannot explain the role of the protonated nitrogen in the hERG inhibition.

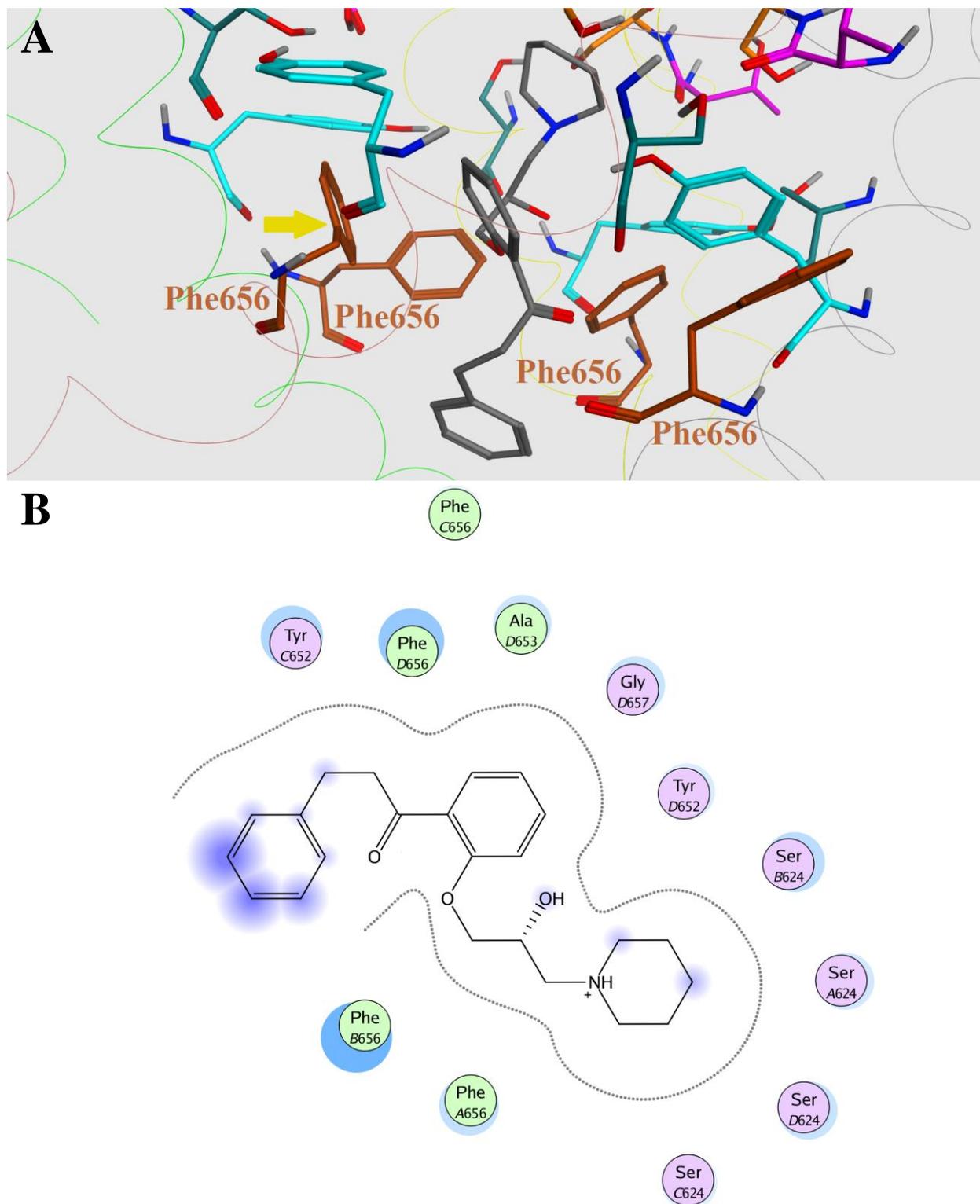


Fig. 19 Centroid of cluster four. 3D (A) and 2D (B) interaction pattern of the centroid of cluster four. Only the amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented. The subunits A, B, C and D are colored in green, yellow, grey and pink respectively. The yellow arrow indicates the Phe656 that interacts with the centroid through parallel-displaced π - π interaction.

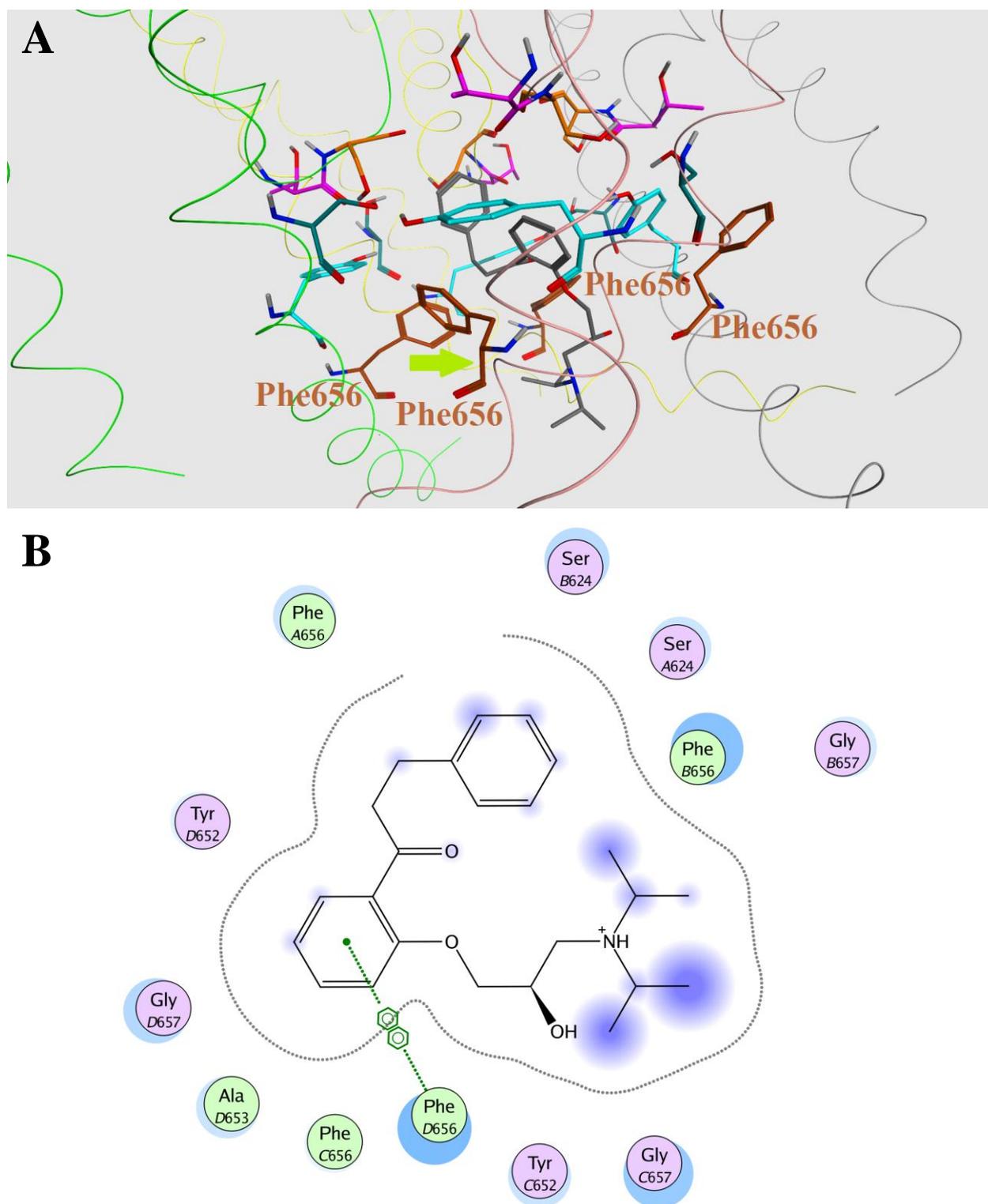


Fig. 20 Centroid of cluster five. 3D (A) and 2D (B) interaction pattern of the centroid of cluster five. Only the amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented. The subunits A, B, C and D are colored in green, yellow, grey and pink respectively. The green arrow indicates the Phe656 that makes π -stacking interaction with the phenyl ring of the centroid.

Neither through consensus scoring nor through mutagenesis data it was possible to prioritize one of the two clusters. In both clusters one or more poses are present which are ranked within the first ten positions in the four scoring functions used in the consensus scoring, and not all ligands have at least one pose in agreement with mutagenesis data,[11, 126] due to the lack of π -stacking, or π -cation, or T-stacking, or parallel displaced π - π interactions with the amino acid Phe656.

Both centroids show a different binding mode regarding the interactions and the scaffold orientations. In contrast to the three clusters obtained previously (subchapter 5.2), in cluster four the placement of the substituent R1 is “messy” (fig. 21A). This is probably due to the fact that in the first case the docking was performed in a symmetric model of the hERG channel, while the binding site of the eight selected models are snapshots of molecular dynamics simulations, hence the hERG channel is asymmetric. In cluster four some poses place the R1 substituent between the Ser649 of subunit A and the Tyr652 of subunit B (fig. 21B), while in other conformations the substituent is located between the Ser649 of subunit A and Ty652 of subunit D (fig. 21C), and in two poses of trapped compounds it is oriented towards the selectivity filter (fig. 21D). In contrast to clusters one, two and three, there are no clear differences in the charged nitrogen position of trapped and non-trapped propafenone derivatives (fig. 21E).

In the fifth cluster, the substituents R1 are situated under the ring formed by the four Phe656. As in cluster four, the substituents have different directions and it is not possible to distinguish between trapped and non-trapped compounds based on the protonated nitrogen position (fig. 22).

Both clusters are in agreement with the “foot in the door” mechanism. In the fourth cluster the second phenyl ring, as previously mentioned, lies under the ring formed by the four Phe656. During the channel gating the aromatic group has to move into the central cavity, forcing the protonated nitrogen to get closer to the amino acids Ser649 and Tyr652, as seen in the subchapter 5.1. In the non-trapped compounds this movement is prevented by the rigid and bulky R1 substituents. Hence, the tail of the scaffold placed in the middle of the ring formed by the four Phe656 hinders the hERG channel closure facilitating the drug dissociation. In cluster five, the charged nitrogen is located under the ring formed by the four Phe656. During hERG channel gating the trapped compounds move into the central cavity. In the non-trapped propafenones this movement is impeded by the bulkiness and rigidity of R1 substituents, which in turn hinders the channel gating with the “foot in the door” mechanism and thereby facilitates the ligand

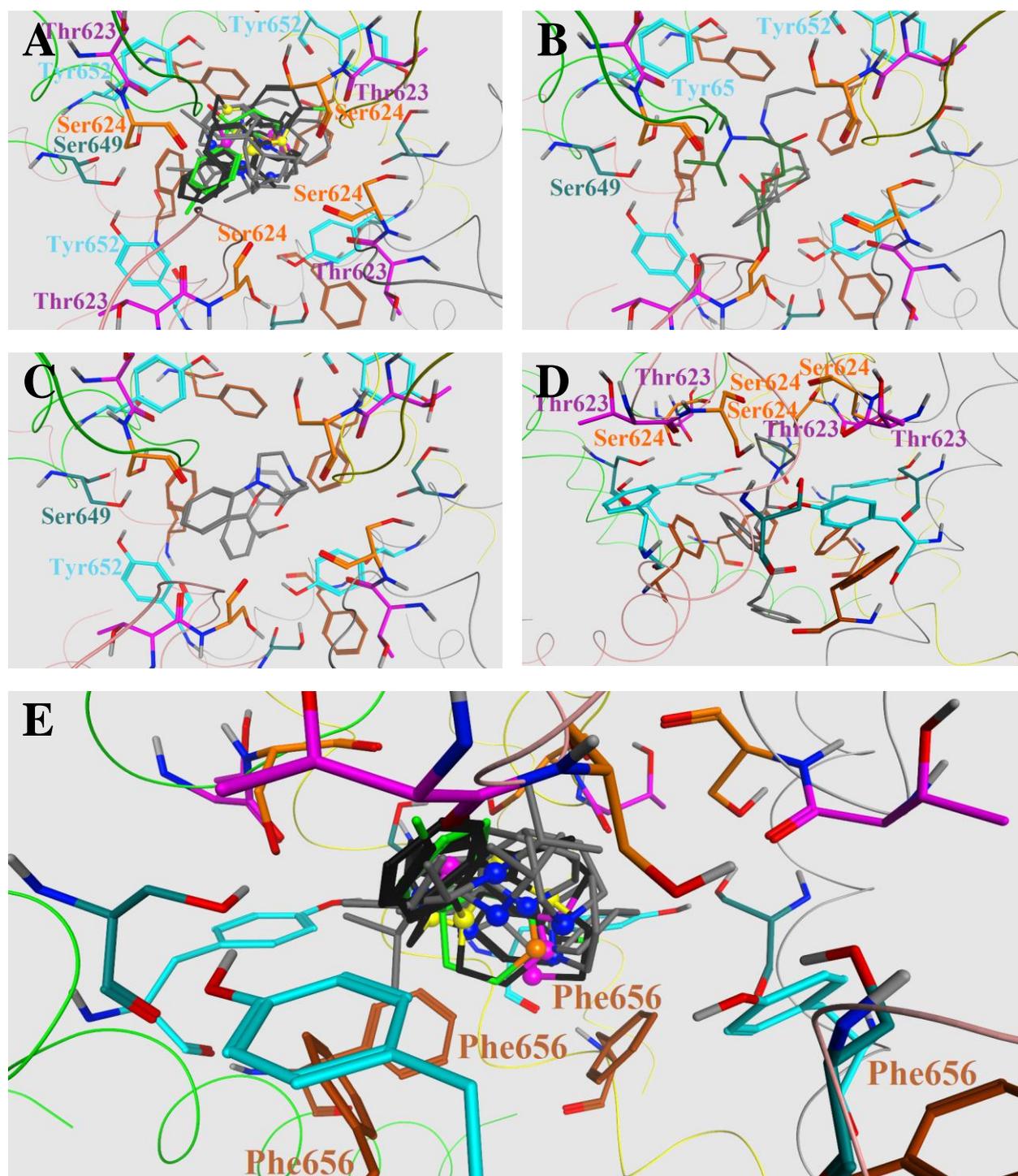


Fig. 21 Position of the substituents in the poses of cluster four. (A) top view of the substituent R1. (B) Propafenone (element) and GPV0009 (dark green) placing the substituent between Ser649 and Tyr652 of the subunits A and B. (C) Pose placing the substituent between Ser649 and Tyr652 of the subunits A and D. (D) Poses with the substituent R1 oriented toward the selectivity filter. (E) Lateral view of substituents R1. The protonated nitrogen of trapped, non-trapped ligands and GPV0062 is colored in blue, violet and orange respectively. The aniline nitrogen of the non-trapped compounds is colored in yellow. The carbon atoms of the non-trapped ligands and of the GPV0062 are colored in black and green respectively. The amino acids Thr623 (violet), Ser624 (orange), Ser649 (dark cyan), Tyr652 (cyan) and Phe656 (brown) are represented. The subunits A, B, C and D are colored in green, yellow, grey and pink respectively.

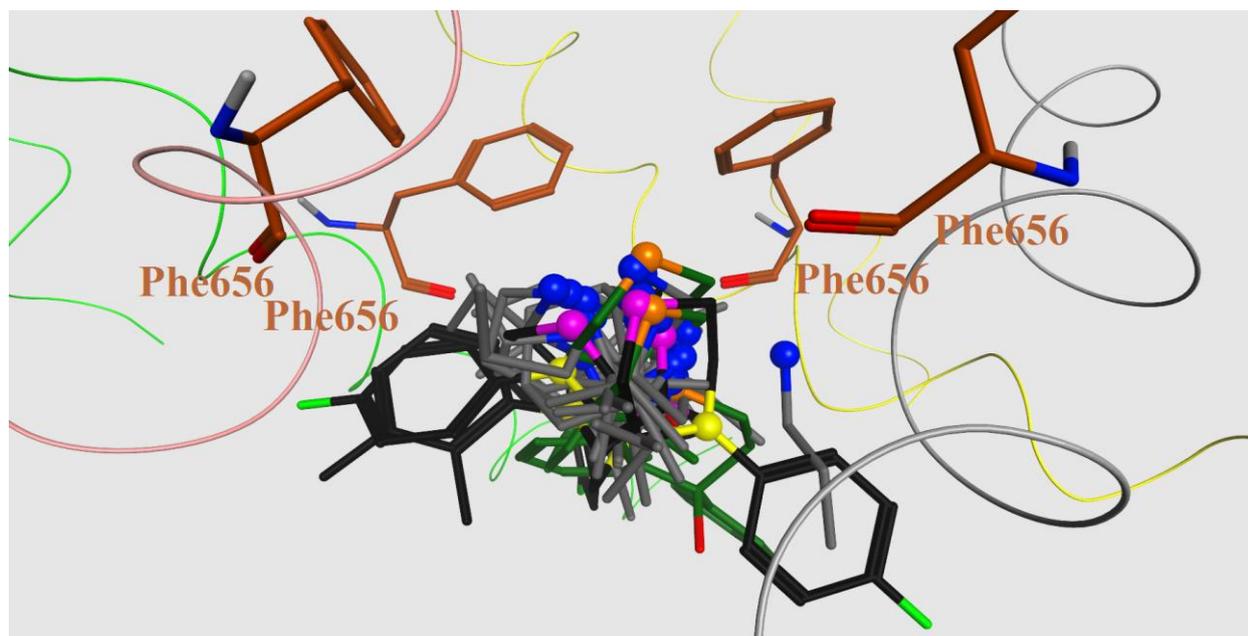


Fig. 22 Lateral view of the substituents R1 position in cluster five. The protonated nitrogen of trapped, non-trapped and GPV0062 compounds is colored in blue, violet and orange respectively. The non-protonated nitrogen of non-trapped molecules is colored in yellow. The carbon atoms of the non-trapped compounds and of the GPV0062 are colored in black and green respectively. The subunits A, B, C and D are colored in green, yellow, grey and pink respectively.

dissociation.

As already seen in the previous subchapter, the potential energy of the scaffold, pose and substituent R1 (table2) considering only the heavy atoms was calculated in order to discriminate between trapped and non-trapped propafenones. Taking into account the potential energy values of the scaffold, obviously, in both clusters it is not possible to distinguish between the poses of trapped and non-trapped compounds. As in clusters one, two and three, the conformations of trapped molecules have a lower value of potential energy than the one of non-trapped ligands. In cluster four, ranking the conformations according to the value obtained dividing the potential energy values of the poses by the total number of heavy atoms, it is still possible to discriminate between trapped and non-trapped molecules (table 13). The propafenone derivative GPV0062, which does not show a clear behavior with a recovery value between the trapped and non-trapped molecules, is ranked as a trapped compound.[126] The difference in the potential energy values of trapped and non-trapped compounds is due to the rigidity and bulkiness of R1 substituents. In contrast to cluster four, in the fifth one it is not possible to discern between trapped and non-trapped propafenones by ranking the values calculated dividing the potential energy of the pose by the number of heavy atoms, due to the incorrect classification of two propafenone poses (table 14). The reasons why these two conformations are outliers are not

Table 13. Poses of cluster4. E scaffold: potential energy calculated considering the heavy atoms of the scaffold (table 2). E pose: potential energy calculated considering only the heavy atoms of the pose. N1: number of heavy atoms of the pose. E R1: potential energy calculated considering only the heavy atoms of the substituent R1. N2: number of heavy atoms of the substituent R1. The values of E are in kcal/mol.

Poses	E scaffold	E pose	N1	E pose / N1	E R1	N2	E R1/ N2	Trapped?
GPV0009	68.26	97.64	28	3.49	10.17	7	1.45	Yes
GPV0009	70.43	95.10	28	3.40	11.06	7	1.58	Yes
GPV0009	81.11	108.68	28	3.88	12.30	7	1.76	Yes
Propafenone	66.79	104.50	25	4.18	7.35	4	1.84	Yes
GPV0180	71.59	112.88	31	3.64	13.10	6	2.18	Yes
GPV0062	60.53	107.47	34	3.16	31.72	13	2.44	Yes
GPV0005	77.28	109.81	27	4.07	15.08	6	2.51	Yes
SCT-AS03	75.15	129.23	33	3.92	36.00	12	3.00	Yes
GPV0005	71.82	103.52	27	3.83	20.12	6	3.35	Yes
GPV0031	61.74	156.56	34	4.60	83.54	13	6.42	No
GPV0019	84.21	196.07	33	5.94	81.41	12	6.78	No
GPV0031	91.20	185.00	34	5.44	90.09	13	6.93	No
GPV0576	89.82	193.77	35	5.54	97.88	14	6.99	No

clear.

As seen for the first, second and third clusters, also for the fourth and fifth ones it is possible to discriminate between trapped and non-trapped compounds ranking the ligands according to the potential energy values of the R1 substituents. The difference in the potential energy values is clear also when it is divided by the number of heavy atoms. Interestingly, the only difference between the compounds SCT-AS03 and GPV0019 is the substitution of the aniline-nitrogen with a carbon atom. The aniline nitrogen hybridized sp^2 renders the phenylpiperazine ring less flexible than the phenylpiperidine. Considering cluster four, during the channel gating the rigidity of the phenylpiperazine moiety prevents the protonated nitrogen of GPV0019 to get closer to the amino acids Ser649-Tyr652 and to adopt one of the two conformations hypothesized for the closed state. In cluster five, the rigidity of the phenylpiperazine ring and the bulkiness of the phenyl groups attached might impede to the protonated nitrogen to enter into the central cavity.

Interestingly, as previously mentioned, in all five clusters the substituent R1 in the non-trapped compounds have always a value much higher than in the trapped propafenones. Furthermore, in cluster five the substituent is placed under the ring formed by the four Phe656, hence the values of potential energy calculated do not depend on possible conformations influenced by the binding site. This result seems to indicate that for propafenones the trapping phenomena depends on intrinsic properties of the R1 substituents, such as the bulkiness and the rigidity, rather than the binding mode in the open or closed conformations of the hERG channel. This means that for

Table 14. Poses of cluster5. E scaffold: potential energy calculated considering the heavy atoms of the scaffold (table 2). E pose: potential energy calculated considering only the heavy atoms of the pose. N1: number of heavy atoms of the pose. E R1: potential energy calculated considering only the heavy atoms of the substituent R1. N2: number of heavy atoms of the substituent R1. The values of E are in kcal/mol.

Poses	E scaffold	E pose	N1	E pose / N1	E R1	N2	E R1/ N2	Trapped?
Propafenone	68.89	96.42	25	3.86	0.33	4	0.08	Yes
Propafenone	75.84	101.80	25	4.07	0.55	4	0.14	Yes
Propafenone	64.05	87.78	25	3.51	0.85	4	0.21	Yes
Propafenone	88.00	116.36	25	4.65	1.55	4	0.39	Yes
Propafenone	86.77	124.98	25	5.00	2.21	4	0.55	Yes
GPV0009	76.76	102.41	28	3.66	8.65	7	1.23	Yes
GPV0009	63.27	93.81	28	3.35	8.93	7	1.28	Yes
GPV0009	68.90	99.39	28	3.55	8.95	7	1.28	Yes
GPV0009	75.35	97.37	28	3.48	9.55	7	1.36	Yes
GPV0009	80.94	106.34	28	3.80	9.64	7	1.38	Yes
GPV0009	86.08	118.17	28	4.22	10.11	7	1.44	Yes
GPV0009	85.00	107.47	28	3.84	10.52	7	1.50	Yes
GPV0009	80.53	104.43	28	3.73	10.79	7	1.54	Yes
GPV0009	84.17	108.59	28	3.88	11.17	7	1.60	Yes
SCT-AS03	69.49	108.12	33	3.28	22.82	12	1.90	Yes
SCT-AS03	78.81	125.25	33	3.80	23.08	12	1.92	Yes
GPV0062	91.16	141.46	34	4.16	25.56	13	1.97	Yes
GPV0062	89.71	138.69	34	4.08	25.61	13	1.97	Yes
GPV0062	84.42	129.58	34	3.81	26.16	13	2.01	Yes
GPV0005	77.04	113.99	27	4.22	14.98	6	2.50	Yes
GPV0180	65.70	117.50	31	3.79	17.59	6	2.93	Yes
GPV0031	90.67	172.46	34	5.07	67.16	13	5.17	No
GPV0031	63.88	147.47	34	4.34	70.94	13	5.46	No
GPV0031	62.71	146.94	34	4.32	73.62	13	5.66	No
GPV0576	64.92	163.19	35	4.66	86.71	14	6.19	No
GPV0019	85.27	189.06	33	5.73	80.46	12	6.70	No

propafenone and derivatives it might be not necessary to perform a docking study to discriminate between the trapped and non-trapped compounds, but it might be enough to calculate the potential energies of the R1 substituents after a simple systematic search. However, in the last case it would be impossible to investigate and understand the drug trapping and drug dissociation mechanisms. Windisch et al.[126] hypothesized that the pKa of the molecule, the possibility to form hydrogen-bonds, the bulkiness and the rigidity of the compound, play a role in the drug trapping. Hence, by modulating these properties of the substituent it might be possible to avoid drug trapping. It has to be noted that we docked only propafenones and derivatives, hence this hypothesis cannot be generalized to other types of molecules.

5.3.3 Conclusions

In this study, we docked nine propafenone and derivatives into eight models of the hERG channel in the open state. Two binding modes were selected through Single Linkage cluster analysis of the RMSD matrix of the ligand-hERG complexes. In cluster four, the charged nitrogen is located near the amino acids Ser649 and Tyr652, as previously found in the poses obtained docking the molecules into the open and closed hERG channel conformations, while one of the two aromatic rings is placed under the four Phe656. During the channel gating the phenyl ring of the trapped propafenones has to move into the central cavity forcing the protonated nitrogen to be close to the amino acids Ser649 and Tyr652. For the non-trapped propafenones these movements are prevented by the R1 substituent: the phenyl ring placed under the ring formed by the four Phe656 cannot move into the central cavity hindering the channel gating with the “foot in the door” mechanism. In cluster five the protonated nitrogen is placed under the four Phe656. This renders the conformations of cluster five not as reliable as the poses in cluster four, because it cannot explain why the charged nitrogen is an important feature for hERG inhibition. The R1 substituents of the poses in cluster five during the channel gating have to move into the central cavity. In non-trapped propafenones this movement is prevented by bulky and rigid substituents, hindering the gating through the “foot in the door” mechanism.

In both clusters it is possible to discern between the trapped and non-trapped molecules ranking them according to the potential energy values of the R1 substituents, and according to this value divided by the total number of heavy atoms. The fact that also in cluster five the potential energy values of R1 substituents of the trapped compounds is lower than in the non-trapped ones, it might indicate that the trapping phenomena of propafenone derivatives is determined by intrinsic properties of R1 moieties.

In this study, we docked only propafenone derivatives, hence these considerations cannot be generalized to non-propafenone type molecules.

REFERENCES

1. *Gene Connection for the Heart*. <http://www.fsm.it/cardmoc/>.
2. Witchel, H.J. and J.C. Hancox, *Familial and acquired long qt syndrome and the cardiac rapid delayed rectifier potassium current*. Clin Exp Pharmacol Physiol, 2000. **27**(10): p. 753-66.
3. Mitcheson, J.S., J. Chen, and M.C. Sanguinetti, *Trapping of a methanesulfonanilide by closure of the HERG potassium channel activation gate*. J Gen Physiol, 2000. **115**(3): p. 229-40.
4. Moreno, I., et al., *Effects of irbesartan on cloned potassium channels involved in human cardiac repolarization*. J Pharmacol Exp Ther, 2003. **304**(2): p. 862-73.
5. Sanchez-Chapula, J.A., et al., *Molecular determinants of voltage-dependent human ether-a-go-go related gene (HERG) K⁺ channel block*. J Biol Chem, 2002. **277**(26): p. 23587-95.
6. Mitcheson, J.S., et al., *A structural basis for drug-induced long QT syndrome*. Proc Natl Acad Sci U S A, 2000. **97**(22): p. 12329-33.
7. Mitcheson, J.S., *hERG potassium channels and the structural basis of drug-induced arrhythmias*. Chem Res Toxicol, 2008. **21**(5): p. 1005-10.
8. Kamiya, K., et al., *Molecular determinants of HERG channel block*. Mol Pharmacol, 2006. **69**(5): p. 1709-16.
9. Kamiya, K., et al., *Molecular determinants of hERG channel block by terfenadine and cisapride*. J Pharmacol Sci, 2008. **108**(3): p. 301-7.
10. Stork, D., et al., *State dependent dissociation of HERG channel inhibitors*. Br J Pharmacol, 2007. **151**(8): p. 1368-76.
11. Witchel, H.J., et al., *The low-potency, voltage-dependent HERG blocker propafenone--molecular determinants and drug trapping*. Mol Pharmacol, 2004. **66**(5): p. 1201-12.
12. Perry, M., et al., *Structural determinants of HERG channel block by clofilium and ibutilide*. Mol Pharmacol, 2004. **66**(2): p. 240-9.
13. Arias, C., et al., *Effects of propafenone and its main metabolite, 5-hydroxypropafenone, on HERG channels*. Cardiovasc Res, 2003. **57**(3): p. 660-9.

14. Katchman, A.N., et al., *Comparative evaluation of HERG currents and QT intervals following challenge with suspected torsadogenic and nontorsadogenic drugs*. *J Pharmacol Exp Ther*, 2006. **316**(3): p. 1098-106.
15. Kiehn, J., et al., *Inhibitory effects of the class III antiarrhythmic drug amiodarone on cloned HERG potassium channels*. *Naunyn Schmiedebergs Arch Pharmacol*, 1999. **359**(3): p. 212-9.
16. Mergenthaler, J., et al., *Blocking effects of the antiarrhythmic drug propafenone on the HERG potassium channel*. *Naunyn Schmiedebergs Arch Pharmacol*, 2001. **363**(4): p. 472-80.
17. Paul, A.A., H.J. Witchel, and J.C. Hancox, *Inhibition of the current of heterologously expressed HERG potassium channels by flecainide and comparison with quinidine, propafenone and lignocaine*. *Br J Pharmacol*, 2002. **136**(5): p. 717-29.
18. Sanchez-Chapula, J.A., et al., *Voltage-dependent profile of human ether-a-go-go-related gene channel block is influenced by a single residue in the S6 transmembrane domain*. *Mol Pharmacol*, 2003. **63**(5): p. 1051-8.
19. Schwoerer, A.P., et al., *Molecular interaction of droperidol with human ether-a-go-go-related gene channels: prolongation of action potential duration without inducing early afterdepolarization*. *Anesthesiology*, 2007. **106**(5): p. 967-76.
20. Clancy, C.E., et al., *K⁺ channel structure-activity relationships and mechanisms of drug-induced QT prolongation*. *Annu Rev Pharmacol Toxicol*, 2003. **43**: p. 441-61.
21. Pearlstein, R.A., et al., *Characterization of HERG potassium channel inhibition using CoMSiA 3D QSAR and homology modeling approaches*. *Bioorg Med Chem Lett*, 2003. **13**(10): p. 1829-35.
22. Sanguinetti, M.C. and M. Tristani-Firouzi, *hERG potassium channels and cardiac arrhythmia*. *Nature*, 2006. **440**(7083): p. 463-9.
23. Morais Cabral, J.H., et al., *Crystal structure and functional analysis of the HERG potassium channel N terminus: a eukaryotic PAS domain*. *Cell*, 1998. **95**(5): p. 649-55.
24. Chen, J., et al., *The S4-S5 linker couples voltage sensing and activation of pacemaker channels*. *Proc Natl Acad Sci U S A*, 2001. **98**(20): p. 11277-82.
25. Cui, J., et al., *Cyclic AMP regulates the HERG K(+) channel by dual pathways*. *Curr Biol*, 2000. **10**(11): p. 671-4.
26. Thai, K.M. and G.F. Ecker, *Predictive models for HERG channel blockers: ligand-based and structure-based approaches*. *Curr Med Chem*, 2007. **14**(28): p. 3003-26.
27. Aronov, A.M., *Tuning out of hERG*. *Curr Opin Drug Discov Devel*, 2008. **11**(1): p. 128-40.

28. Morgan, T.K., Jr. and M.E. Sullivan, *An overview of class III electrophysiological agents: a new generation of antiarrhythmic therapy*. Prog Med Chem, 1992. **29**: p. 65-108.
29. Ekins, S., et al., *Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go-related gene potassium channel*. J Pharmacol Exp Ther, 2002. **301**(2): p. 427-34.
30. Cavalli, A., et al., *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**(18): p. 3844-53.
31. Aronov, A.M. and B.B. Goldman, *A model for identifying HERG K+ channel blockers*. Bioorg Med Chem, 2004. **12**(9): p. 2307-15.
32. Testai, L., et al., *Torsadogenic cardiotoxicity of antipsychotic drugs: a structural feature, potentially involved in the interaction with cardiac HERG potassium channels*. Curr Med Chem, 2004. **11**(20): p. 2691-706.
33. Aronov, A.M., *Common pharmacophores for uncharged human ether-a-go-go-related gene (hERG) blockers*. J Med Chem, 2006. **49**(23): p. 6917-21.
34. Crumb, W.J., Jr., et al., *Effects of antipsychotic drugs on I(to), I(Na), I(sus), I(K1), and hERG: QT prolongation, structure activity relationship, and network analysis*. Pharm Res, 2006. **23**(6): p. 1133-43.
35. Johnson, S.R., et al., *Estimation of hERG inhibition of drug candidates using multivariate property and pharmacophore SAR*. Bioorg Med Chem, 2007. **15**(18): p. 6182-92.
36. Kramer, C., et al., *A composite model for HERG blockade*. ChemMedChem, 2008. **3**(2): p. 254-65.
37. Garg, D., T. Gandhi, and C. Gopi Mohan, *Exploring QSTR and toxicophore of hERG K+ channel blockers using GFA and HypoGen techniques*. J Mol Graph Model, 2008. **26**(6): p. 966-76.
38. Shamovsky, I., et al., *Overcoming undesirable HERG potency of chemokine receptor antagonists using baseline lipophilicity relationships*. J Med Chem, 2008. **51**(5): p. 1162-78.
39. Coi, A., et al., *Identification of "toxicophoric" features for predicting drug-induced QT interval prolongation*. Eur J Med Chem, 2008. **43**(11): p. 2479-88.
40. Cianchetta, G., et al., *Predictive models for hERG potassium channel blockers*. Bioorg Med Chem Lett, 2005. **15**(15): p. 3637-42.
41. Li, Q., et al., *hERG classification model based on a combination of support vector machine method and GRIND descriptors*. Mol Pharm, 2008. **5**(1): p. 117-27.

42. Ermondi, G., S. Visentin, and G. Caron, *GRIND-based 3D-QSAR and CoMFA to investigate topics dominated by hydrophobic interactions: the case of hERG K⁺ channel blockers*. Eur J Med Chem, 2009. **44**(5): p. 1926-32.
43. Aptula, A.O. and M.T. Cronin, *Prediction of hERG K⁺ blocking potency: application of structural knowledge*. SAR QSAR Environ Res, 2004. **15**(5-6): p. 399-411.
44. Keseru, G.M., *Prediction of hERG potassium channel affinity by traditional and hologram qSAR methods*. Bioorg Med Chem Lett, 2003. **13**(16): p. 2773-5.
45. Fioravanzo, E., et al., *General and Independent Approaches to Predict HERG Affinity Values*, in *Internet Electronic Journal of Molecular Design*. 2005. p. 625–646.
46. Song, M. and M. Clark, *Development and evaluation of an in silico model for hERG binding*. J Chem Inf Model, 2006. **46**(1): p. 392-400.
47. Zolotoy, A.B., et al., *Physicochemical determinants for drug induced blockade of HERG potassium channels: effect of charge and charge shielding*. Curr Med Chem Cardiovasc Hematol Agents, 2003. **1**(3): p. 225-41.
48. Coi, A., et al., *Prediction of hERG potassium channel affinity by the CODESSA approach*. Bioorg Med Chem, 2006. **14**(9): p. 3153-9.
49. Yoshida, K. and T. Niwa, *Quantitative structure-activity relationship studies on inhibition of HERG potassium channels*. J Chem Inf Model, 2006. **46**(3): p. 1371-8.
50. Seierstad, M. and D.K. Agrafiotis, *A QSAR model of HERG binding using a large, diverse, and internally consistent training set*. Chem Biol Drug Des, 2006. **67**(4): p. 284-96.
51. Ekins, S., et al., *Insights for human ether-a-go-go-related gene potassium channel inhibition using recursive partitioning and Kohonen and Sammon mapping techniques*. J Med Chem, 2006. **49**(17): p. 5059-71.
52. Leong, M.K., *A novel approach using pharmacophore ensemble/support vector machine (PhE/SVM) for prediction of hERG liability*. Chem Res Toxicol, 2007. **20**(2): p. 217-26.
53. Gavaghan, C.L., et al., *Development, interpretation and temporal evaluation of a global QSAR of hERG electrophysiology screening data*. J Comput Aided Mol Des, 2007. **21**(4): p. 189-206.
54. Shamovsky, I., et al., *Increasing selectivity of CC chemokine receptor 8 antagonists by engineering nondesolvation related interactions with the intended and off-target binding sites*. J Med Chem, 2009. **52**(23): p. 7706-23.
55. Roy, K. and G. Ghosh, *QSTR with extended topochemical atom (ETA) indices. 13. Modelling of hERG K⁺ channel blocking activity of diverse functional drugs using different chemometric tools*. Molecular Simulation, 2009. **35**: p. 1256-68.

56. Diller, D.J., et al., *W. O. Patent WO2006055918*. 2006.
57. Diller, D.J., et al., *U. S. Patent US20060206269*. 2006.
58. Diller, D.J. and D.W. Hobbs, *Understanding hERG inhibition with QSAR models based on a one-dimensional molecular representation*. *J Comput Aided Mol Des*, 2007. **21**(7): p. 379-93.
59. Bains, W., A. Basman, and C. White, *HERG binding specificity and binding site structure: evidence from a fragment-based evolutionary computing SAR study*. *Prog Biophys Mol Biol*, 2004. **86**(2): p. 205-33.
60. Tobita, M., T. Nishikawa, and R. Nagashima, *A discriminant model constructed by the support vector machine method for HERG potassium channel inhibitors*. *Bioorg Med Chem Lett*, 2005. **15**(11): p. 2886-90.
61. Roche, O., et al., *A virtual screening method for prediction of the HERG potassium channel liability of compound libraries*. *Chembiochem*, 2002. **3**(5): p. 455-9.
62. Dubus, E., et al., *In silico classification of HERG channel blockers: a knowledge-based strategy*. *ChemMedChem*, 2006. **1**(6): p. 622-30.
63. Thai, K.M. and G.F. Ecker, *Similarity-based SIBAR descriptors for classification of chemically diverse hERG blockers*. *Mol Divers*, 2009. **13**(3): p. 321-36.
64. Thai, K.M. and G.F. Ecker, *Classification models for HERG inhibitors by counter-propagation neural networks*. *Chem Biol Drug Des*, 2008. **72**(4): p. 279-89.
65. Thai, K.M. and G.F. Ecker, *A binary QSAR model for classification of hERG potassium channel blockers*. *Bioorg Med Chem*, 2008. **16**(7): p. 4107-19.
66. Chekmarev, D.S., et al., *Shape signatures: new descriptors for predicting cardiotoxicity in silico*. *Chem Res Toxicol*, 2008. **21**(6): p. 1304-14.
67. Doddareddy, M.R., et al., *Prospective validation of a comprehensive in silico hERG model and its applications to commercial compound and drug databases*. *ChemMedChem*, 2010. **5**(5): p. 716-29.
68. O'Brien, S.E. and M.J. de Groot, *Greater than the sum of its parts: combining models for useful ADMET prediction*. *J Med Chem*, 2005. **48**(4): p. 1287-91.
69. Sun, H., *An accurate and interpretable bayesian classification model for prediction of HERG liability*. *ChemMedChem*, 2006. **1**(3): p. 315-22.
70. Jia, L. and H. Sun, *Support vector machines classification of hERG liabilities based on atom types*. *Bioorg Med Chem*, 2008. **16**(11): p. 6252-60.
71. Buyck, C., et al., *The 14th Euro. Symp. on QSARs*. Bournemouth, UK., 8–13 Sep. 2002.

72. Catana, C., *Simple idea to generate fragment and pharmacophore descriptors and their implications in chemical informatics*. J Chem Inf Model, 2009. **49**(3): p. 543-8.
73. Redfern, W.S., et al., *Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development*. Cardiovasc Res, 2003. **58**(1): p. 32-45.
74. Yap, C.W., et al., *Prediction of torsade-causing potential of drugs by support vector machine approach*. Toxicol Sci, 2004. **79**(1): p. 170-7.
75. Skold, C., et al., *Presentation of a structurally diverse and commercially available drug data set for correlation and benchmarking studies*. J Med Chem, 2006. **49**(23): p. 6660-71.
76. Gepp, M.M. and M.C. Hutter, *Determination of hERG channel blockers using a decision tree*. Bioorg Med Chem, 2006. **14**(15): p. 5325-32.
77. Hidaka, S., et al., *Nonlinear classification of hERG channel inhibitory activity by unsupervised classification method*. J Toxicol Sci, 2010. **35**(3): p. 393-9.
78. Papadatos, G., et al., *Lead optimization using matched molecular pairs: inclusion of contextual information for enhanced prediction of HERG inhibition, solubility, and lipophilicity*. J Chem Inf Model, 2010. **50**(10): p. 1872-86.
79. Doyle, D.A., et al., *The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity*. Science, 1998. **280**(5360): p. 69-77.
80. Zhou, Y., et al., *Chemistry of ion coordination and hydration revealed by a K⁺ channel-Fab complex at 2.0 Å resolution*. Nature, 2001. **414**(6859): p. 43-8.
81. Zhou, Y. and R. MacKinnon, *The occupancy of ions in the K⁺ selectivity filter: charge balance and coupling of ion binding to a protein conformational change underlie high conduction rates*. J Mol Biol, 2003. **333**(5): p. 965-75.
82. Kuo, A., et al., *Crystal structure of the potassium channel KirBac1.1 in the closed state*. Science, 2003. **300**(5627): p. 1922-6.
83. Jiang, Y., et al., *Crystal structure and mechanism of a calcium-gated potassium channel*. Nature, 2002. **417**(6888): p. 515-22.
84. Jiang, Y., et al., *X-ray structure of a voltage-dependent K⁺ channel*. Nature, 2003. **423**(6935): p. 33-41.
85. Jalaie, M. and D.D. Holsworth, *QT interval prolongation: and the beat goes on*. Mini Rev Med Chem, 2005. **5**(12): p. 1083-91.

86. Recanatini, M., et al., *QT prolongation through hERG K(+) channel blockade: current knowledge and strategies for the early prediction during drug development*. *Med Res Rev*, 2005. **25**(2): p. 133-66.
87. Aronov, A.M., *Predictive in silico modeling for hERG channel blockers*. *Drug Discov Today*, 2005. **10**(2): p. 149-55.
88. Duncan, R.S., et al., *Erythromycin block of the HERG K+ channel: accessibility to F656 and Y652*. *Biochem Biophys Res Commun*, 2006. **341**(2): p. 500-6.
89. Rajamani, R., et al., *A two-state homology model of the hERG K+ channel: application to ligand binding*. *Bioorg Med Chem Lett*, 2005. **15**(6): p. 1737-41.
90. Yi, H., et al., *Interaction simulation of hERG K+ channel with its specific BeKm-1 peptide: insights into the selectivity of molecular recognition*. *J Proteome Res*, 2007. **6**(2): p. 611-20.
91. Masetti, M., A. Cavalli, and M. Recanatini, *Modeling the hERG potassium channel in a phospholipid bilayer: Molecular dynamics and drug docking studies*. *J Comput Chem*, 2008. **29**(5): p. 795-808.
92. Du, L., et al., *A novel structure-based virtual screening model for the hERG channel blockers*. *Biochem Biophys Res Commun*, 2007. **355**(4): p. 889-94.
93. Singleton, D.H., et al., *Fluorescently labeled analogues of dofetilide as high-affinity fluorescence polarization ligands for the human ether-a-go-go-related gene (hERG) channel*. *J Med Chem*, 2007. **50**(13): p. 2931-41.
94. Stansfeld, P.J., et al., *Drug block of the hERG potassium channel: insight from modeling*. *Proteins*, 2007. **68**(2): p. 568-80.
95. Hosaka, Y., et al., *Mutational analysis of block and facilitation of HERG current by a class III anti-arrhythmic agent, nifekalant*. *Channels (Austin)*, 2007. **1**(3): p. 198-208.
96. Tseng, G.N., et al., *Probing the outer mouth structure of the HERG channel with peptide toxin footprinting and molecular modeling*. *Biophys J*, 2007. **92**(10): p. 3524-40.
97. Choe, H., et al., *A novel hypothesis for the binding mode of HERG channel blockers*. *Biochem Biophys Res Commun*, 2006. **344**(1): p. 72-8.
98. Osterberg, F. and J. Aqvist, *Exploring blocker binding to a homology model of the open hERG K+ channel using docking and molecular dynamics methods*. *FEBS Lett*, 2005. **579**(13): p. 2939-44.
99. Farid, R., et al., *New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies*. *Bioorg Med Chem*, 2006. **14**(9): p. 3160-73.

100. Micheli, F., et al., *1,2,4-triazol-3-yl-thiopropyl-tetrahydrobenzazepines: a series of potent and selective dopamine D(3) receptor antagonists*. J Med Chem, 2007. **50**(21): p. 5076-89.
101. Armstrong, C.M., *Inactivation of the potassium conductance and related phenomena caused by quaternary ammonium ion injection in squid axons*. J Gen Physiol, 1969. **54**(5): p. 553-75.
102. Thai, K.M., et al., *The hERG potassium channel and drug trapping: insights from docking studies with propafenone derivatives*. ChemMedChem Under submission, 2009.
103. Fernandez, D., et al., *Physicochemical features of the HERG channel drug binding site*. J Biol Chem, 2004. **279**(11): p. 10120-7.
104. Karczewski, J., et al., *Analogs of MK-499 are differentially affected by a mutation in the S6 domain of the hERG K⁺ channel*. Biochem Pharmacol, 2009. **77**(10): p. 1602-11.
105. Perry, M., et al., *Drug binding interactions in the inner cavity of HERG channels: molecular insights from structure-activity relationships of clofilium and ibutilide analogs*. Mol Pharmacol, 2006. **69**(2): p. 509-19.
106. Milnes, J.T., et al., *hERG K⁺ channel blockade by the antipsychotic drug thioridazine: An obligatory role for the S6 helix residue F656*. Biochem Biophys Res Commun, 2006. **351**(1): p. 273-80.
107. Kiesecker, C., et al., *Atypical tetracyclic antidepressant maprotiline is an antagonist at cardiac hERG potassium channels*. Naunyn Schmiedebergs Arch Pharmacol, 2006. **373**(3): p. 212-20.
108. Scherer, D., et al., *Inhibition of cardiac hERG potassium channels by tetracyclic antidepressant mianserin*. Naunyn Schmiedebergs Arch Pharmacol, 2008. **378**(1): p. 73-83.
109. Milnes, J.T., et al., *Blockade of HERG potassium currents by fluvoxamine: incomplete attenuation by S6 mutations at F656 or Y652*. Br J Pharmacol, 2003. **139**(5): p. 887-98.
110. Ridley, J.M., et al., *High affinity HERG K(+) channel blockade by the antiarrhythmic agent dronedarone: resistance to mutations of the S6 residues Y652 and F656*. Biochem Biophys Res Commun, 2004. **325**(3): p. 883-91.
111. Dougherty, D.A., H.A. Lester, and M.W. Nowak, *U. S. Patent US20060084102*. 2006.
112. Siebrands, C.C. and P. Friederich, *Structural requirements of human ether-a-go-go-related gene channels for block by bupivacaine*. Anesthesiology, 2007. **106**(3): p. 523-31.
113. Myokai, T., et al., *Topological mapping of the asymmetric drug binding to the human ether-a-go-go-related gene product (HERG) potassium channel by use of tandem dimers*. Mol Pharmacol, 2008. **73**(6): p. 1643-51.

114. Imai, Y.N., S. Ryu, and S. Oiki, *Docking model of drug binding to the human ether-a-go-go potassium channel guided by tandem dimer mutant patch-clamp data: a synergic approach*. J Med Chem, 2009. **52**(6): p. 1630-8.
115. Price, D.A., et al., *Overcoming HERG affinity in the discovery of the CCR5 antagonist maraviroc*. Bioorg Med Chem Lett, 2006. **16**(17): p. 4633-7.
116. Dinges, J., et al., *1,4-Dihydroindeno[1,2-c]pyrazoles with acetylenic side chains as novel and potent multitargeted receptor tyrosine kinase inhibitors with low affinity for the hERG ion channel*. J Med Chem, 2007. **50**(9): p. 2011-29.
117. Levoine, N., et al., *Refined docking as a valuable tool for lead optimization: application to histamine H3 receptor antagonists*. Arch Pharm (Weinheim), 2008. **341**(10): p. 610-23.
118. Zachariae, U., F. Giordanetto, and A.G. Leach, *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**(14): p. 4266-76.
119. Frenal, K., et al., *Exploring structural features of the interaction between the scorpion toxin CnErg1 and ERG K⁺ channels*. Proteins, 2004. **56**(2): p. 367-75.
120. Gomez-Varela, D., et al., *Different relevance of inactivation and F468 residue in the mechanisms of hEag1 channel blockage by astemizole, imipramine and dofetilide*. FEBS Lett, 2006. **580**(21): p. 5059-66.
121. Waldhauser, K.M., et al., *Interaction with the hERG channel and cytotoxicity of amiodarone and amiodarone analogues*. Br J Pharmacol, 2008. **155**(4): p. 585-95.
122. Armstrong, C.M., *Induced inactivation of the potassium permeability of squid axon membranes*. Nature, 1968. **219**(5160): p. 1262-3.
123. Armstrong, C.M., *Interaction of tetraethylammonium ion derivatives with the potassium channels of giant axons*. J Gen Physiol, 1971. **58**(4): p. 413-37.
124. Carmeliet, E., *Use-dependent block and use-dependent unblock of the delayed rectifier K⁺ current by almokalant in rabbit ventricular myocytes*. Circ Res, 1993. **73**(5): p. 857-68.
125. Kikuchi, K., et al., *Blockade of HERG cardiac K⁺ current by antifungal drug miconazole*. Br J Pharmacol, 2005. **144**(6): p. 840-8.
126. Windisch, A., et al., *Trapping and dissociation of propafenone derivatives in HERG channels*. Br J Pharmacol, 2011.
127. MOE, Chemical Computing Group Inc. Montreal.
128. Schrödinger, LLC, New York, NY.

129. Recanatini, M., A. Cavalli, and M. Masetti, *Modeling HERG and its interactions with drugs: recent advances in light of current potassium channel simulations*. ChemMedChem, 2008. **3**(4): p. 523-35.
130. XLSTAT, Addinsoft. www.xlstat.com.
131. Sanguinetti, M.C. and J.S. Mitcheson, *Predicting drug-hERG channel interactions that cause acquired long QT syndrome*. Trends Pharmacol Sci, 2005. **26**(3): p. 119-24.
132. Bottegoni, G., A. Cavalli, and M. Recanatini, *A comparative study on the application of hierarchical-agglomerative clustering approaches to organize outputs of reiterated docking runs*. J Chem Inf Model, 2006. **46**(2): p. 852-62.
133. *Cambridge Structural Database*.
134. Yeh, J.Z. and C.M. Armstrong, *Immobilisation of gating charge by a substance that simulates inactivation*. Nature, 1978. **273**(5661): p. 387-9.
135. Stary, A., et al., *Toward a consensus model of the HERG potassium channel*. ChemMedChem, 2009. **5**(3): p. 455-67.

CURRICULUM VITAE

PERSONAL INFORMATION

Name SCHIESARO ANDREA
Address 4/10, Neumayrgasse, 1160, Vienna, Austria
Telephone +4368183139226, +393479038935
Fax
E-mail andrea2179@libero.it

Nationality Italian
Date of birth 26/05/1979

EDUCATION AND TRAINING

- *Dates (from – to)* 2007 to now PhD under the supervision of Prof. Gerhard F. Ecker
 Thesis title: “In silico screening on the HERG potassium channel ”
 May 2010 to August 2010 visiting student under the supervision of Prof. Gabriele Costantino
 October 1999-March 2007 Degree in Chimica e Tecnologia Farmaceutiche (Pharmaceutical Chemistry and Technology), Department of Pharmacy, University of Padova
 Thesis title: “Progettazione di nuovi antagonisti dei recettori adenosinici A2A ed A3: uno studio computazionale” (“Design of new antagonists of adenosine receptors A2A and A3: a computational study”)
 October 1998-September 1999 Study of Biology, University of Padova
 September 1993-June 1998 High school certificate
- *Name and type of organisation providing education and training* 2007 to now University of Vienna-Department of Medicinal Chemistry
 May 2010 to now University of Parma-Dipartimento Farmaceutico (Department of Medicinal Chemistry)
 October 1999-March 2007 University of Padova-Facoltà di Farmacia (Faculty of Pharmacy)
 October 1998-September 1999 University of Padova-Facoltà di Biologia (Faculty of Biology)
 September 1993-June 1998 Istituto Biologico Sanitario (Biological Sanitary Institute) “Santa Bertilla Boscardin”
- *Principal subjects/occupational skills covered* Drug Design, Molecular modeling, Computational Chemistry, Medicinal chemistry
- *Title of qualification awarded* Maturità scientifica- High school diploma
 Laurea Specialistica in Chimica e Tecnologia Farmaceutiche-Master degree in Pharmaceutical Chemistry and Technology

**Personal skills
and competences**
*Acquired in the course of
life and career but not
necessarily covered by
formal certificates and
diplomas.*

Mother tongue Italian

Other languages

English

- Reading skills Good
- Writing skills Good
- Verbal skills Good

German

- Reading skills Basic
- Writing skills Basic
- Verbal skills Basic

*Technical skills and
competences*

Operating systems: MS Windows 98, XP, Vista and 7
MS Office: Power Point, Word, Excel
Molecular modeling: MOE, Glide, Gold, Amber, Namd, Vmd, Isis
draw, Charmm-GUI

Teaching Computational Life Sciences (10 hours)
Laboratory Course – Drug Analysis and – development (*in silico*
methods)

Congress, Symposia 2007 Summer School on Drug Design, Vienna
2008 XXth International Symposium on Medicinal Chemistry,
Vienna; Poster presentation
2009 21st Scientific Congress of the Austrian Pharmaceutical
Society (ÖPhG), Vienna; Poster presentation
2009 Joint Meeting on Medicinal Chemistry 2009, Budapest; Poster
presentation
2010 240th ACS National Meeting & Exposition , Boston; Poster
presentation

Additional information

References:

Gerhard Ecker Univ.-Prof. Mag. Dr.
Pharmacoinformatics Research Group
Department of Medicine / Pharmaceutical Chemistry
University of Vienna
Althanstraße 14
1090 Vienna, Austria
Tel: +43 1 4277 551 10

Fax: +43 1 4277 95 51
e-mail: gerhard.f.ecker@univie.ac.at
Webpage: <http://pharminfo.univie.ac.at>
Stefano Moro Ph.D Associate professor
Molecular Modeling Section
Department of Pharmaceutical Sciences
University of Padova
Via Marzolo 5
35131 Padova, Italy
Tel: +39 049 827 5704
Fax: +39 049 827 5366
e-mail: stefano.moro@unipd.it
Webpage URL <http://mms.dsfarm.unipd.it>

Gabriele Costantino Prof.
MedChemGroup
Pharmacy Department
University of Parma
Via GP Usberti 27/A-Campus Universitario
43100 Parma, Italy
Tel: +39 0521 905054
Fax: +39 0521 905006
e-mail: gabriele.costantino@unipr.it
Webpage: <http://www.unipr.it/arpa/dipfarm/medchem/people.htm>

Annexes

List of publications:

1. Thai KM, Windisch A, Stork D, Weinzinger A, **Schiesaro A**, Guy HR, Timin EN, Hering S, Ecker GF. "The hERG Potassium Channel and Drug Trapping: Insight from Docking Studies with Propafenone Derivatives" *ChemMedChem*. 5, 436-442, (2009)
2. Michielan L, Bacilieri M, **Schiesaro A**, Bolcato C, Pastorin G, Spalluto G, Cacciari B, Klotz KN, Kaseda C, Moro S "Linear and Nonlinear 3D-QSAR Approaches in Tandem with Ligand-Based Homology Modeling as a Computational Strategy To Depict the Pyrazolo Triazolo-Pyrimidine Antagonists Binding Site of the Human Adenosine A(2A) Receptor." *J. Chem. Inf. Model.* 48, 350-363 (2008)
3. Colotta V, Catarzi D, Varano F, Capelli F, Lenzi O, Filacchioni G, Martini C, Trincavelli L, Ciampi O, Pugliese AM, Pedata F, **Schiesaro A**, Morizzo E, Moro S "New 2 Arylpyrazolo[3,4-c]quinoline Derivatives as Potent and Selective Human A₃ Adenosine Receptor Antagonists. Synthesis, Pharmacological Evaluation and Ligand-Receptor Modeling" *J. Med. Chem.* 50, 4161-4074 (2007)

List of book chapters

Schiesaro A, Ecker GF. "Prediction of hERG channel inhibition using *in silico* techniques" in *HERG potassium channels in drug discovery and development*, Springer (2010) – in press

