Exploring Ligand Binding in HIV-1 Protease and $K^+$ Channels Using Computational Methods

FREDRIK ÖSTERBERG
Dissertation presented at Uppsala University to be publicly examined in B42, BMC, Uppsala, Friday, December 2, 2005 at 14:00 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

Understanding protein-ligand interactions is highly important in drug development. In the present work objective is to comprehend the link between structure and function using molecular modelling. Specifically, this thesis has been focused on implementation of receptor flexibility in molecular docking and studying structure-activity relationships of potassium ion channels and their blockers.

In ligand docking simulations protein motion and heterogeneity of structural waters are approximated using an ensemble of protein structures. Four methods of combining multiple target structures within a single grid-based lookup table of interaction energies are tested. Two weighted average methods permit consistent and accurate ligand docking using a single grid representation of the target protein structures.

Quaternary ammonium ions (QAls) are well known K⁺ channel blockers. Conformations around C–N bonds at the quaternary centre in tetraalkylammonium ions in water solution are investigated using quantum mechanical methods. Relative solvation free energies of QAls are further estimated from molecular dynamics simulations. The torsion barrier for a two-step interconversion between the conformations $D_{3d}$ and $S_1$ is calculated to be 9.5 kcal mol⁻¹. Furthermore $D_{3d}$ is found to be more stable than the $S_1$ conformation which is in agreement with experimental studies. External QAI binding to the K⁺ channel KcSa is also studied. Computer simulations and relative binding free energies of the KcSa complexes with QAls are calculated. This is done with the molecular dynamics free energy perturbation approach together with automated ligand docking. In agreement with experiment, the Et,N⁺ blocker in $D_{3d}$ symmetry has better binding than the other QAls.

Binding of blockers to the human cardiac hERG potassium channel is studied using a combination of homology modelling, automated docking and molecular dynamics simulations. The calculations reproduce the relative binding affinities of a set of drug derivatives very well and indicate that both polar interactions near the intracellular opening of the selectivity filter as well as hydrophobic complementarity in the region around F656 are important for blocker binding. Hence, the derived model of hERG should be useful for further interpretations of structure-activity relationships.

Keywords: hERG, KcSa, AutoDock, LIE, molecular dynamics, ion channels

Fredrik Österberg, Department of Cell and Molecular Biology, Structural Biology, Box 596, Uppsala University, SE-75124 Uppsala, Sweden

© Fredrik Österberg 2005

ISSN 1651-6214
ISBN 91-554-6411-4
urn:nbn:se:uu:diva-6167 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-6167)
Ad astra per aspera
Publications


Reprints were made with permission from the copyright holders.

Related publications

## Contents

1 Introduction ........................................................................................................ 9  
1.1 Outline of the thesis .................................................................................. 10

2 Computational methods in molecular modelling ........................................... 11  
2.1 Molecular mechanics .............................................................................. 12
2.2 Docking ...................................................................................................... 12
2.3 Quantum chemical calculations ............................................................. 14
2.4 Molecular dynamics ................................................................................. 14
2.5 Free energy perturbation ........................................................................... 15
2.6 Linear interaction energy method ............................................................ 16

3 Exploring methods for modelling protein motion in docking simulations (I) .......................................................... 18  
3.1 Grid maps in AutoDock ........................................................................... 20
3.2 HIV-1 protease – a molecule with small molecular movements ............. 20
3.3 Docking techniques and new grid methods .......................................... 21
3.4 Conclusions and perspectives ................................................................. 24

4 Ion channels .................................................................................................... 25  
4.1 The KcsA potassium channel (II, III) ....................................................... 26  
4.1.1 Structure and function of KcsA ......................................................... 27
4.1.2 Quaternary ammonium ions – important channel blockers .......... 28
4.1.3 Conformations and solvation free energies of QAIs (II) ................. 28
4.1.4 Exploring blocking by QAIs to the outer cavity of the KcsA channel (III) ................................................................. 30
4.1.5 Conclusions and perspectives (II, III) .............................................. 32
4.2 Human potassium channel – hERG (IV) ................................................. 32
4.2.1 Structure and function of hERG ....................................................... 33
4.2.2 Modelling, docking and evaluation of hERG (IV) ......................... 34
4.2.3 Conclusions and perspectives .......................................................... 36

5 Summary in Swedish (2-5 sidor) .................................................................. 37

6 Acknowledgements ......................................................................................... 39

7 References ...................................................................................................... 41
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>hERG</td>
<td>Human ether a go-go related gene</td>
</tr>
<tr>
<td>HIV-1</td>
<td>Human immunodeficiency virus type 1</td>
</tr>
<tr>
<td>$I_{Kr}$</td>
<td>Delayed rectifier current</td>
</tr>
<tr>
<td>$I_{to}$</td>
<td>Transient outward current</td>
</tr>
<tr>
<td>KcsA</td>
<td>Streptomyces lividans K$^+$ channel</td>
</tr>
<tr>
<td>Kv</td>
<td>Voltage-gated K$^+$ channel</td>
</tr>
<tr>
<td>LQTS</td>
<td>Long QT syndrome</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>MM</td>
<td>Molecular mechanics</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>QAI</td>
<td>Quaternary ammonium ion</td>
</tr>
<tr>
<td>QM</td>
<td>Quantum mechanics</td>
</tr>
<tr>
<td>RESP</td>
<td>Restrained electrostatic potential</td>
</tr>
<tr>
<td>SAR</td>
<td>Structure-activity relationships</td>
</tr>
<tr>
<td>TdP</td>
<td>Torsades de Pointes</td>
</tr>
<tr>
<td>TEA</td>
<td>Tetraethylammonium ion</td>
</tr>
<tr>
<td>TMA</td>
<td>Tetramethylammonium ion</td>
</tr>
<tr>
<td>TPrA</td>
<td>Tetrapropylammonium ion</td>
</tr>
</tbody>
</table>
1 Introduction

Protein-ligand interactions are highly important in biochemistry, medicinal chemistry and drug development. In the latter one tries to comprehend the link between structure and function in order to be able to predict or design new drugs. This is done by using different techniques depending on prior information, such as 3D protein structure and binding data and, in cases of database scanning the amount of data to be processed.

The present work has been focused on potassium ion channels and studying structure-activity relationships (SAR) of different ion blockers using molecular modelling techniques. Another part of this thesis deals with molecular docking programs and implementation of ways to handle small receptor motion, i.e. allowing small flexibility in the target structure.

All cells have a membrane protecting them from the outer environment in the same manner that skin protects the human body. Although, neither cells nor humans can live in a closed environment, they still must have connections to their outer world. In cells this is mediated by integral membrane proteins. There is a very large variety of these proteins both in composition and function. Molecules that are allowed to pass membrane proteins can be of various types, e.g., water (through aquaporines) and ions (through ion channels), or more unspecific as other small solutes (through non-selective channels). The channel protein part of this work (section 4) deals with K⁺ channels.

Roderick MacKinnon was awarded the Noble Prize in Chemistry (2003) for his achievements in the field of channels in the cell membrane. The announcement read "for structural and mechanistic studies in ion channels" [1]. These discoveries have been of great importance in this field of research, especially the first X-ray crystallographic structure of an ion channel, KcsA by Doyle et al. [2]. From this X-ray structure hypotheses could be validated and important structure-activity properties examined. These membrane spanning proteins provide gates in the cells for ion transport in and out. Ion channels are found in different shapes and in different cell types and are of great importance, e.g., if ion channels in heart cells are blocked the result is arrhythmia. The theoretical research on ion channels needs some atomic information regarding the structure to be studied or information from a re-
lated structure. Therefore the X-ray structure of KcsA was of major importance.

In predicting 3D structures of ligand-receptor complexes it is useful to have an unbiased way of discerning how a ligand binds. The purpose of this is manifold, e.g., the active site may not be known or the binding conformation is known for one ligand but not for others. Knowledge of the binding conformation of a ligand does not necessarily guarantee that the conformation of other ligands can be predicted. Unbiased programs can be constructed to screen large databases which would be impossible to do manually.

In improving docking programs, flexibility to the ligand has been added but still very few treat the receptors flexible. The main obstacle in obtaining a flexible receptor is that straightforward methods are very time consuming in the evaluation of possible receptor conformations in addition to the ligand positions. However during the recent years considerable improvements have been achieved.

1.1 Outline of the thesis

The four papers included in this thesis will be referred to by their roman numerals in bold (I-IV) throughout the text.

The aim of section 2 is to give a short introduction to some of the different techniques in the molecular modelling field that have been used in paper I-IV. This part is not meant to be a complete description of each subject, but rather a short summary. In section 3 a brief description of the docking field and molecular flexibility is presented together with a summary of paper I. In this paper four new methods are presented for accommodating receptor flexibility without increasing computation time significantly.

Section 4 deals with ion channels and blockers. Here a synopsis of voltage-gated potassium channels is given, describing some important structures and functions. In section 4.1 the potassium channel from *Streptomyces lividans* (KcsA) is scrutinized to some extent together with the properties of the important ammonium ion blockers (summary of paper II) and the binding of the ammonium ion blockers to KcsA is addressed (summary of paper III).

The human potassium channel hERG is discussed in section 4.2. Here a short summary of the hERG structure, function and importance is given. This is followed by a description of the work done in paper IV where a homology model has been built and tested against known blockers.
2 Computational methods in molecular modelling

In monitoring chemical processes the concept of equilibrium is of fundamental importance. The state of equilibrium can be seen as a measure of how far a reaction is driven, e.g., how strongly a ligand binds to a receptor. This can be written as

$$A + B \xrightleftharpoons{K_{\text{bind}}} AB$$

(1)

where $A$ and $B$ are the reactants, $AB$ the complex and $K_{\text{bind}}$ the binding equilibrium constant. Such constants can be converted into binding free energies or relative binding free energies. In most chemical experiments the Gibbs free energy ($G$) typically is used as the pressure is most often kept constant (compared to the Helmholtz energy ($A$) which is used with constant volume). Since volume changes and the associated $p\Delta V$ term are small the difference between $\Delta G$ and $\Delta A$ is usually negligible in practice. The Gibbs free energy is defined as

$$G = H - TS$$

(2)

where $H$ is the enthalpy, $T$ the temperature and $S$ the entropy of the system.

In order to relate the experimentally determined equilibrium constants to the free energy, the standard binding free energy is written as follows

$$\Delta G_{\text{bind}}^\circ = -RT \ln \left( K_{\text{bind}}^\circ c^\circ \right)$$

(3)

where $R$ is the gas constant and $c^\circ$ is the standard state concentration (usually 1 M).
2.1 Molecular mechanics

In molecular mechanics (MM) atoms are treated as a set of soft spheres with point charges. This is done with force field methods and these can be used to describe the system's energy, which is a simplified description of reality but the accuracy of many force fields is quite good. A typical potential energy function contains a combination of terms for bonded and non-bonded interactions (Equation 4).

\[ V = \sum_{\text{bonds}} \frac{1}{2} k_b (b - b_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{\text{tortions}} k_\phi (1 + \cos(n\phi - \delta)) + \]

\[ \sum_{\text{improper tortions}} \frac{1}{2} k_\zeta (\zeta - \zeta_0)^2 + \sum_{\text{non-bonded pairs}} \varepsilon \left( \frac{r^*}{r} \right)^{12} - 2 \left( \frac{r^*}{r} \right)^6 + \sum_{\text{non-bonded pairs}} \frac{q_i q_j}{4\pi\varepsilon_0 r} \]  

In this function bond lengths and angles are evaluated in terms of deviations from their equilibrium values \((b_0, \theta_0, \zeta_0)\) or minimal value \((\cos(n\phi - \delta))\) in the first four bonded terms. For the non-bonded terms van der Waals and electrostatic interactions are evaluated for atoms that are separated by at least three bonds or are in different molecules. The van der Waals contribution is described by a Lennard-Jones 12-6 function composed of a repulsive (varies as \(r^{-12}\)) and an attractive component (varies as \(r^{-6}\)). The electrostatic contribution is obtained by applying Coulomb's law. The constants and the other non-variable parameters in Equation 4 are adjusted against either experimental and/or quantum mechanical (QM) studies to reproduce correct energies, frequencies and geometries.

Today there are several different force fields with their own parameterizations, e.g., GROMOS [3], Amber [4], CHARMM [5], OPLS-AA [6]. They differ in how they are developed and for which type of molecular systems they are designed.

2.2 Docking

When there is no prior knowledge about the binding pose of a protein blocker or if an unbiased starting position is wanted, docking programs [7-12] are very useful. With these programs an attempt is made to find the 'correct' structure of intermolecular complexes. With some of these programs it is possible to scan a large database for potent inhibitors when the structure of the target molecule is known. This is the main advantage of docking pro-
grams, their speed. Depending on the type of methods the docking program uses and the composition of the target molecule, regarding structure and residues, the best docking mode can be more or less hard to find. An easy way to see how well a program performs is to re-dock a complex with known structure.

Most docking programs are able to handle flexible ligands, i.e., rotatable bonds, but few have any degree of freedom in the target structure. This is of course a major drawback for most proteins since induced fit is common in proteins when ligands bind. In order to search conformational space and generate new ligand docking modes the programs use different methods, e.g., the Metropolis algorithm in Monte Carlo methods [13], genetic algorithms [8] and incremental construction algorithms [9], and these are described in section 3.

Scoring functions can be used for ranking a large number of possible docking poses in a fast manner and to evaluate if the ligand position is favorable. A good docking pose is, in this respect, a conformation with low energy. This type of functions attempts to estimate the binding free energy by using different contributions to the scoring function. These contributions can rely on knowledge-based potentials such as PMF score [14] and DrugScore [15], force field methods such as DOCK [16] and GOLD [10], or empirical free energy scoring functions such as ChemScore [17] and SCORE [18]. Scoring functions are calibrated against experimental data for optimal performance. The purpose of using scoring functions instead of more rigorous methods (see below) is their speed. This enables scanning large databases for putative substrate binders. The scoring function can look like a simple energy function:

\[ \Delta G_{\text{bind}} = \Delta G_{\text{vdw}} + \Delta G_{\text{hbond}} + \Delta G_{\text{el}} + \Delta G_{\text{tor}} + \Delta G_{\text{sol}} \]  

This specific function is implemented in the program AutoDock [8] used in paper I, II and IV. It is a semi-empirical scoring function which includes five energy terms, the first three are standard molecular mechanics terms. \( \Delta G_{\text{tor}} \) handles the internal torsion as well as the global rotation and translation contributions. \( \Delta G_{\text{sol}} \) handles the desolvation upon binding and the hydrophobic effect. Other scoring functions can include other contributions like lipophilic terms, ionic terms, aromatic terms etc.
2.3 Quantum chemical calculations

In QM the main objectives are to obtain energies, structural geometries and charge distributions of the molecules. This is done by solving the Schrödinger equation and obtaining the wavefunction that describe the electrons probability amplitude. Most often the time-independent Schrödinger equation is considered and this can only be solved analytically for a few simple cases.

The Schrödinger equation can only be solved exactly for a few simple problems, such as the hydrogen, so for larger systems approximations has to be used. One technique used is the Hartree-Fock method where an iterative process produces better and better wavefunctions [19]. In the Hartree-Fock approach each electron interacts with the mean field of the other electrons. A higher level of theory is the density functional theory where one method is B3LYP. Paper II uses both Hartree-Fock and B3LYP methods whereas paper III and IV only use Hartree-Fock.

The mentioned approximations for solving the wavefunction uses the approximation of Gaussian functions, as an approximation of atomic orbitals, called basis sets, e.g., 6-31G*. Depending on the number of functions included in the basis set, a higher accuracy can be achieved.

2.4 Molecular dynamics

By using the molecular dynamics (MD) method it is possible to simulate the motion of a system of particles as a function of time. This is done by solving Newton's law of motion and the result is a trajectory with successive positions and velocities of the simulated system that change with time. Since the phase space is large not all points can be visited, but by running long trajectories good estimates of thermodynamic quantities can be calculated. [19]

In order to run an MD simulation of a molecular system, starting coordinates and initial velocities are required. Positions can for example be taken from a structure obtained from X-ray crystallography or a docked complex. The initial velocities can be random numbers that obey Maxwell's probability distribution at the specific temperature. This is then combined with Equation 5 from MM and Equation 6 where the derivative of the potential energy is taken with respect to the atom positions to give the force acting on each atom.
From Newton's second law of motion in Equation 6 the atoms' accelerations can be calculated. With these properties new positions and velocities can be calculated at \( t + \delta t \) by using finite difference methods. One of the most frequently used is the leap-frog algorithm (Equation 7) [20], which is a variant of the Verlet algorithm [21].

\[
\frac{-\partial V}{\partial r_i} = F_i = m a_i \tag{6}
\]

\[
v_i \left( t + \frac{1}{2} \delta t \right) = v_i \left( t - \frac{1}{2} \delta t \right) + a_i (t) \delta t
\]

\[
r_i \left( t + \delta t \right) = r_i (t) + v_i \left( t + \frac{1}{2} \delta t \right) \delta t
\]

2.5 Free energy perturbation

The free energy perturbation (FEP) method is an exact technique derived from statistical mechanics and it is very useful for investigating relative free energies. By using Equation 8 [22] it is possible to calculate the free energy difference between two different states, \( i \) and \( j \).

\[
\Delta G_{i \rightarrow j} = -RT \ln \left\{ \exp \left( -\frac{1}{RT} (V_j - V_i) \right) \right\}_i \tag{8}
\]

The potential energies \( V_i \) and \( V_j \) are derived for the two different states from Equation 4. The ensemble average on potential \( V_i \) denoted \( \langle \cdot \rangle_i \), can be calculated by using e.g., MD. An important remark is that the phase space of \( i \) and \( j \) must overlap, otherwise the phase space of \( j \) will not be sampled well enough when running a trajectory on \( V_i \). A result of this is bad accuracy of the calculated free energy.

In situations when larger perturbations are done one has to use a coupling parameter \( \lambda \) to gradually transform the potential from state A to state B (Figure 1). This is done in small steps with an effective potential \( V(\lambda) = (1-\lambda)V_A + \lambda V_B \) where \( \lambda \) goes from 0 to 1 in small steps \((\Delta G_{\lambda(n) \rightarrow \lambda(n+1)} \sim kT)\). The consecutive potential pairs are then used in Equation 8 and all the free energies are summed up to a resulting total free energy difference.
The FEP method is often used together with thermodynamic cycles [23,24]. This is useful in e.g., comparing the binding free energy of two different ligands (Figure 1). Since the free energy is a state function the relative binding free energy can be calculated as \( \Delta \Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \). It is difficult to directly calculate the binding free energies \( \Delta G_3 \) and \( \Delta G_4 \) (due to molecular rearrangement) so here the easier legs in the thermodynamic cycle can be chosen, \( \Delta G_1 \) and \( \Delta G_2 \). These are plain mutations of A to B in the free and the bound state.

\[
\begin{align*}
A+C &\rightarrow A-C \\
B+C &\rightarrow B-C \\
\Delta G_1 &
\end{align*}
\]

\( \Delta G_3 \) and \( \Delta G_4 \) describe the binding free energy of A to C and B to C respectively. This approach was used in paper II and III for analysis of the affinity of different quaternary ammonium ions to the ion channel KcsA.

2.6 Linear interaction energy method

A less computationally demanding method for calculating free energy differences is the linear interaction energy (LIE) method [25]. It is a semi empirical method that only relies on simulations of two corners of the thermodynamic cycle described in Figure 1, and not the hybrid states in between.

According to LIE theory (Equation 9), the free energy is divided into two different contributions, a nonpolar part (\( v_{\text{dw}} \)) and a polar part (\( e_l \)). The nonpolar contribution is described by an empirical expression and the polar contribution is derived from the electrostatic linear response approximation.

\[
\Delta G = \alpha \left( \langle V_{l-s}^{\text{dw}} \rangle_p - \langle V_{l-s}^{\text{dw}} \rangle_w \right) + \beta \left( \langle V_{l-s}^{\text{el}} \rangle_p - \langle V_{l-s}^{\text{el}} \rangle_w \right) + \gamma \quad (9)
\]

The arrow brackets indicate ensemble averages and the subscripts \( p \) and \( w \) stand for protein (ligand in bound state) and water (ligand in free state) re-
spectively. The subscript $l-s$ stands for ligand-surrounding interactions. The constant $\alpha$ is an empirically derived parameter. This scaling factor has been calibrated against experimental binding data [25,26] and is relatively robust for different molecular systems [27-30]. Calibrated into this constant are energy contributions from hydrophobic and entropic effects. The second part of Equation 9 is scaled by $\beta$. This constant describes the electrostatic contribution and is a result from the electrostatic linear response approximation which indicates that the induced polarization in some dielectric medium, due to an external electric field, is proportional to the applied field. The $\gamma$ constant is in some cases needed to reproduce the absolute experimental binding free energies [30,31].

It is necessary that the ligand has a more or less 'correct' starting position in the protein since the MD simulations can not explore the entire phase space due to energy barriers in the potential surface and relatively short time frame. This can be solved by using docking techniques to get starting conformations, described in section 2.2, together with the LIE method as was done in paper IV.
3 Exploring methods for modelling protein motion in docking simulations (I)

When receptors bind ligands they usually undergo some degree of conformational change, induced fit. It can be a small rotation of a side chain or something larger as a loop movement. With this in mind it is obvious that an optimal docking program should include some means of describing plasticity of the receptor. Additionally it should also, as mentioned in section 2.2, be able to generate and evaluate molecular complexes at high speed.

Docking programs that first came in use had almost no flexibility of either ligand or receptor. They treated the molecules as rigid bodies with only translation and rotation, as in the DOCK program [16]. Since then there has been major improvements. Today there are numerous docking programs with different ways of searching docked positions. Most of these programs can model full flexibility of the ligand but only a fraction of them have any flexibility of the receptor and then only to a limited extent [32-34].

Different methods are used today in docking programs to explore the potential surface of the complex and find the optimal docking pose of a ligand. In combination with a rigid receptor there are for example:

- **Incremental construction**, e.g., [7,9]. A base fragment is first identified in the ligand. The conformational space is then searched for this often significant part of the molecule. After identifying positions for this first fragment, these will be starting points for the following conformational analysis and adding additional fragment to the base fragment. This method is similar to the depth-first systematic conformational search using a search tree.[19]

- **Monte Carlo methods**, e.g., [13]. The ligand conformation is changed by small random displacement, producing a random walk in the space around the starting position. The new conformation is accepted or rejected using the Metropolis algorithm, \( P_{\text{accept}}(\Delta E) = \min(1, \exp(-\Delta E/RT)) \). \( P_{\text{accept}}(\Delta E) \) is the probability for accepting the new conformation, \( \Delta E \) is the energy difference, \( k \) is Boltzmann's constant and \( T \) the temperature.
- Genetic algorithms, e.g., [8,10,35]. The Lamarckian genetic algorithm uses terminology from genetic evolutionary biology. The first step is creation of a population of ligands that are randomly placed in the docking area (Figure 2). These parameters are state variables and are describing the translation, orientation and conformation of the ligand with respect of the protein. State variables correspond to a gene (sometimes defined as chromosome) in the genetic algorithm. As in nature, these genes (with all information stored into a linear bit string with 0s and 1s), describing the ligands state, can undergo mutations and crossovers in random positions in the bit string, thereby creating new conformations. From these a choice will be made which individuals that shall survive to the next generation.

![Flow chart of a docking round using the genetic algorithm for searching the potential surface of the docked complex.](image)

Figure 2. Flow chart of a docking round using the genetic algorithm for searching the potential surface of the docked complex.

Regarding the receptor there are obviously different ways to get some flexibility. Paper I deals with this problem by using four new methods in treatment of small molecular movements in receptors.
3.1 Grid maps in AutoDock

The docking program that has been used in paper I is AutoDock [8]. This automated and unbiased docking program uses a grid based method for energy evaluation of the flexible ligand in complex with a rigid protein. Points on a 3D grid, placed to cover the entire inner cavity of the receptor, are probed with the atoms that constitute the ligand, e.g., C, O, N, H (Figure 3). These points make up lookup tables with interaction energies between the atoms in the ligand and the receptor. AutoDock can in this way vary the atomic positions of the ligand and rapidly calculate the energies used in the docking process.

![Figure 3. The picture shows the grid points that the probe atom visits for calculations of the maps over the active site. The white sphere indicates the probe atom. This is repeated with different probe atoms reflecting all the atoms in the ligand. (Picture adopted from the AutoDock 3.05 manual.)](image)

3.2 HIV-1 protease – a molecule with small molecular movements

The work here is on a simplified/empirical way of adding receptor flexibility to AutoDock. The goal was to make grid tables describing an 'average protein' from a large set of X-ray structures, and in this way catches small movements in the protein when it binds different ligands. For this purpose the human immunodeficiency virus type 1 (HIV-1) protease is a good target, because there is numerous crystallographic structures and binding studies reported on this system. This protease has a well defined active site shaped like a tunnel right through the protein. The enzyme is opened and closed by two large flaps and under these two flaps a structural water is often found important in ligand binding (Figure 4).
A set of 21 HIV-1 protease complexes with different peptidomimetic inhibitors was selected. The experimental binding free energy ranged from −14 to −8 kcal mol$^{-1}$ and in all of these, except the complex with PDB code 1hvr, the structural water molecule under the flaps is present. The peptidomimetic inhibitors bind similarly in an outstretched way right through the binding site. All structures have been solved to a resolution of 2.8 Å or lower and R-values below 0.2.

### 3.3 Docking techniques and new grid methods

As a first test of AutoDock's performance, all inhibitors were separated from respective protease and re-docked. This served as a check of the ability of the docking program to reproduce the binding mode and binding free energy. The result of this turned out good for all complexes except two. Although, these two ligands from the complexes in 1hte and 1htf show high RMSD to the crystal structure, it can be seen that if 1htf is rotated 180 degrees around the central axis of enzyme it fits the crystal structure with an RMSD of 1.25 Å. For 1hte the small size is a problem in the docking. This inhibitor is small and only covers half of the active site. The two top predicted conformations have an RMSD >6 Å although the remaining conformations are in good agreement with the crystal structure.
The next step was to dock all the 21 complexes against each, i.e. a complete cross-docking of the complexes, other to see how well these different inhibitors would bind without any induced fit. The result was reasonable but clearly a few inhibitors were unsuccessfully docked (Figure 5). The inhibitors in the protease complexes with PDB codes 1hvi, 1hvj, 1hvk, 1hvl, and 1hvs show a similar pattern and this is correlated with their length. These need more space in the active site and thereby push Arg8 (Figure 4) outward about 2 Å and is a behavior that the four new methods are trying to accommodate. The inhibitors in 1hte and 1htf showed the same behavior as when re-docked.

![Inhibitor Structures](image)

Figure 5. These inhibitors have some difficulties to find their crystallographic positions in cross-docking with AutoDock due to their size, either longer or shorter than the rest of the test set of inhibitors. a) Inhibitor 1hvi, 1hvj, 1hvk, 1hvl, 1hvr and 1hvs are shown schematically, R1 = H or OH. b) Inhibitor 1hte and 1htf are shown schematically. R2 = OH or C9NOH11.

The large number of complex-specific grid map tables cover points of interaction energies for all the 21 proteases combined with all the 21 inhibitors. These tables were finally combined by four different methods to see if any of these new sets of maps could be used as an 'average' protease. The following methods were used:

- **Mean grid.** Point-by-point averages are taken over all corresponding grid values.
- **Minimum grid.** The minimum values of all corresponding grid points are taken.
- **Clamped grid.** A weight \( w \) is used to scale the energies before summing them, \( w=1 \) if the interaction energy is lower than 0 kcal mol\(^{-1}\) and \( w=0.0001 \) otherwise. The weight factor is then normalized. Table 1 describes the use of this energy weighted method. This method is similar to a method used by Knettel et al. [36].
- **Energy-weighted grid.** A weight \( w \) is also used here in scaling the energies. It uses a Boltzmann assumption based on the interaction energy, \( w=\exp(-\Delta G/RT) \). Table 2 describes the use of this energy weighted method.
Table 1. Energy calculation of an average grid point from 3 hypothetical map tables, using the method clamped grid. For values above the selected threshold of 0 kcal mol\(^{-1}\) w=0.0001 otherwise 1.

<table>
<thead>
<tr>
<th>Interaction energy†</th>
<th>Weight</th>
<th>Normalized weight‡</th>
<th>Clamped. interaction energy†,‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.0001</td>
<td>0.00005</td>
<td>0.05</td>
</tr>
<tr>
<td>−3.0</td>
<td>1</td>
<td>0.5</td>
<td>−1.5</td>
</tr>
<tr>
<td>−2.0</td>
<td>1</td>
<td>0.5</td>
<td>−1.0</td>
</tr>
<tr>
<td>Sum: 2.0001</td>
<td>1</td>
<td></td>
<td>−2.55</td>
</tr>
</tbody>
</table>

†[kcal mol\(^{-1}\)]
‡These numbers are rounded for clarity.

Table 2. Energy calculation of an average grid point from 3 hypothetical map tables, using the method energy-weighted grid.

<table>
<thead>
<tr>
<th>Interaction energy†</th>
<th>Weight</th>
<th>Normalized weight‡</th>
<th>Energy-weighted interaction energy†,‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0*</td>
<td>0*</td>
<td>0*</td>
</tr>
<tr>
<td>−3.0</td>
<td>160</td>
<td>0.8</td>
<td>−2.5</td>
</tr>
<tr>
<td>−2.0</td>
<td>30</td>
<td>0.2</td>
<td>−0.3</td>
</tr>
<tr>
<td>Sum: 190</td>
<td>1.0</td>
<td></td>
<td>−2.8</td>
</tr>
</tbody>
</table>

†[kcal mol\(^{-1}\)]
‡These numbers are rounded for clarity.
* Very small positive number.

The result of the mean grid method was not particularly good. Only 62% of the 21 ligands found the proper conformation with an RMSD <2 Å. This was expected since the truly bad interaction energies (high positive values) will dominate points in the grid maps over the relatively lower, good binding energies. This creates large repulsive areas in the mean grid.

In the minimum method the results were the opposite of the mean method. This was also anticipated since all high positive values are removed and replaced by the lowest values. This results in grid maps with large favorable areas for docking, giving the inhibitor too large space for finding its best docking pose. The ligands find their 'correct' interaction conformations in all of the 21 cases. The drawback is that the estimated binding free energies are largely overestimated.

Using the clamped method in constructing a combined representation of all the proteases gave good results in finding the crystallographic binding poses. Furthermore also the estimated binding free energies show good correlation with the experimental values. Here it demonstrates clearly how nice the repulsive part around Arg8 is softened but it also displays how the attractive energy part is somewhat too generous.
In the last method, the energy-weighted maps, the same result as in the clamped method are obtained but with one improvement. The areas with attractive interaction energies are reduced and still the 'correct' binding poses are found as well as the correct estimated binding free energies. This method also has a sounder justification, than the clamped method, as based on the Boltzmann assumption in assigning the weight factor instead of an empirical threshold value that can work for some molecular complexes but maybe not another.

3.4 Conclusions and perspectives

It is clear from paper I that for limited protein motions the energy-weighted map method can be a good alternative. This method only consumes some extra calculation time in combining a set of new grid maps and the actual docking time is the same as with the original grid maps. In paper I crystallographic structures have been used as templates for a small molecular movement in accommodating different inhibitors. Another way to get sets of receptor coordinates could be from MD and MC methods or by rotating a selected number of residue torsions in an ordered way. The common dilemma with these methods is the time scale, since a docking program should be able to screen large databases promptly.

In striving for methods to handle induced fit in drug design, the temperature factors (B-values) can also be of interest when the starting coordinates come from a crystallographic structure. An automated implementation could incorporate B-values of the active site residues and from this calculate the distance that each atom can be displaced. For those above a certain minimum distance threshold, torsion angles are set to be rotatable and thereby generate new possibilities for induced fit.
4 Ion channels

Ion channels are major pharmaceutical targets as they regulate several important functions in humans, e.g., nerve signaling and heart rhythm. A large problem in studying channels is the difficulties in getting crystals for X-ray diffraction [37]. Despite these problems, crystallographers have been able to crystallize a few very interesting and important structures [2,38-41].

Ion channels can be divided into a wide range of different families. For example, they can be sorted under how they are activated, e.g., by voltage or ligands, or which ions they regulate, e.g., K⁺, Na⁺, H⁺, Ca²⁺, Mg²⁺ and Cl⁻. In the activation process, so called gating, the channel is opened and can thereby let ions flow in or out of the cell. This ion flux is often very fast and can support 10⁶-10⁸ ions per second [42]. The actual way how opening and closing of the channels is undertaken is still under debate. There have been several proposals for how potassium channels do this, e.g., helical screw, helical twist and paddle model, and what parts are involved. [39,43-48]

Two major classes of K⁺ channels activated by outer stimuli are (1) the voltage gated ion channels, such as the channel from thermophilic archaeabacteria \textit{Aeropyrum pernix} (KvAP) [44], and (2) the ligand gated ion channels, to which the Ca²⁺-activated \textit{Methanobacterium thermoautotrophicum} (MthK) [38] channel accounts. There are also channels regulated by other means e.g., pH as in KcsA [49]. These types of channels have subunits with membrane spanning helices, that form the pore when they are assembled (Figure 6). In the narrow pore with an inside aligned by backbone carbonyl groups, ions and water can pass through. This part is called the selectivity filter and it has the conserved sequence of TVGYG (e.g., in the channels KcsA, MthK, KvAP, Kv1.3, Kv1.5), in the hERG channel the sequence is SVGFG. Another conserved sequence is the hinge region which has a sequence of PXP [50,51] where X can be any residue (found in e.g., the Kv1 family of voltage-gated human channels) or a specific G [52] (found in e.g., in the channels hERG and KvAP). These sequences are found in the helix forming the inner cavity, M2 or S6. This helix has hydrophilic residues pointing inwards to the large water cavity while interactions with the helix next to it are hydrophobic as well as the residues in this neighboring helix.
When ions enter the selectivity filter, the coordination of eight carbonyl oxygens compensates for the dehydration cost and also the size is optimal for the specific ion. This has been shown by several groups using different techniques [53-56]. It has been elucidated from MD simulations [24] and later verified by a high resolution X-ray structure [57] that the low energy configurations in the selectivity filter have the order K⁺-water-K⁺-water or water-K⁺-water-K⁺.

![Figure 6. A schematic picture of an ion channel and its selectivity filter. The circles in the selectivity filter mark the four binding site with eight oxygen interactions per binding site. The narrow selectivity filter, 12 Å long, only allows ions to move in a concerted linear way. There has also been observed a position in the outer cavity, 0, and an 5th position in the inner cavity by X-ray crystallography [55,58].](image)

In paper II-IV K⁺ ion channel interactions and blockers are examined. Paper II deals with solvation free energies and conformations of quaternary ammonium ions, paper III focuses on the same blocker but in the KcsA channel, and paper IV examines the hERG channel with sertindole blockers. These channels are all highly selective towards potassium ions.

4.1 The KcsA potassium channel (II, III)

The ion channel KcsA comes from the bacterium *Streptomyces lividans* and was first crystallized by Doyle et al. in 1998 and solved to a resolution of 3.2 Å [2]. The structure has been of extreme importance and has provided new knowledge around several interesting properties. The sequence of the KcsA channel is similar to other potassium channels in higher organisms. This makes it interesting to examine both general properties and functions. By having the atomic coordinates it was possible to use molecular modelling
techniques to validate old data, e.g., a molecular model [59], and explore new properties.

4.1.1 Structure and function of KcsA

The KcsA channel is a tetramer with two membrane spanning helices and a pore helix in each subunit (Figure 7). The selectivity filter has a conserved sequence of residues TVGYG and is 12 Å long. This sequence comes right after the pore helix (P) and controls the potassium selectivity. The four inner helices (M2) of KcsA [2] form an inverted tepee with a large water cavity near the selectivity filter and a closed part towards the cell interior. An overall dipole moment from the P-helices is assumed to stabilize the 5th binding site inside the water cavity, below the selectivity filter. This site has been shown to bind a quaternary ammonium ion [27,60].

![Figure 7. A schematic picture of one of the four subunits in the two transmembrane helix (2TM) channel KcsA. Helix M1 is the outer helix in contact with the membrane and the P and M2 helices are forming the pore and the inner cavity. The membrane is approximately 34 Å thick.](image)

In lowering the pH KcsA is found to open [49,61]. This is proposed to be caused by a tilt and screw in the conserved hinge region Gly99 in the M2 helices [62]. It has been shown that removal of this residue destroys the gating function [52]. Other potentially important residues are Glu71 and Asp80, which belong to the pore helix and the loop between P and M2, respectively. These residues have their carboxylic groups rather close to the filter. In the original structure of KcsA the resolution was too low to elucidate their protonation state but a correct solution was proposed by using MD/FEP methods [63] and later verified in the high resolution structure of KcsA [58], Glu71 is protonated and Asp80 is not protonated at normal pH. Finally, in the outer cavity four Tyr82 form a cage like binding site. This is where the
blockers supposedly bind at the outside of KcsA and it has also been shown that mutating tyrosine 82 to valine, reduces the binding affinity by several kcal mol$^{-1}$ [27].

4.1.2 Quaternary ammonium ions – important channel blockers

The quaternary ammonium ions (QAI) are a well studied set of molecules in ion channel blocking [64-66]. Considerable efforts have been made in examining their structure, thermodynamics and solution properties [67,68]. The conformations of these alkyl-types of ions are largely determined by the steric intramolecular interactions between the carbon in the second and more distant carbon positioned from the central nitrogen. Two main low energy conformations dominate the symmetry of QAI, quasi-planar conformation with $D_{2d}$ symmetry and quasi-pyramidal conformation with $S_4$ symmetry (Figure 8). The total charge of these molecules is +1 while at the same time the nonpolar hydrocarbon groups make it hydrophobic at the surface. This makes QAI's better solvated the shorter the hydrocarbon chains are.

It has been shown by MD simulations that Et$_4$N$^+$ binds to the 0th and 5th position in KcsA in both of its low energy conformations [27]. In order to more closely examine the difference in symmetry of QAI's, conformations in vacuum and water were examined by computational methods and NMR spectroscopy (II).

4.1.3 Conformations and solvation free energies of QAI (II)

QM calculations of the equilibrium geometric parameters of the QAI's show little dependency on theory level but for the electron charge distribution the differences are substantial, such as the net charge on N that can differ between $-1.064$ for HF/6-31G(d) to $-0.432$ for HF/6-311+G(2d,p) using Mulliken population analysis (MPA). The most stable values are found using electrostatic potential (ESP) derived partial atomic charges. Unfortunately these charges have a fairly large conformational dependency but using the restrained electrostatic potential (RESP) fitting procedure this can be removed. In the gas phase the $D_{2d}$ conformation is predicted to be the more stable and $S_4$ is better solvated than $D_{2d}$. $D_{2d}$ is shown to convert to $S_4$ via a two step mechanism with a barrier of 9.5 kcal mol$^{-1}$.

By using rigorous FEP/MD simulations in the thermodynamic cycle in Figure 8 the relative solvation free energies of the QAI's can be calculated. The FEP transformation was done in one direction, shrinking the molecule, in order to maintain the symmetry of the QAI's. In each case the C-atom is
mutated into a H-atom, and the H-atoms connected to the carbon are mutated into dummy atoms (d) with all non-bonded parameters set to zero. The bond length between H-d is shortened to avoid sampling problems at the end points of transformations and each simulation is done using the GROMOS [3] and Amber [4] force fields.

The transformations that were done are those presented in Figure 8, \( n\)-Pr\(_4\)N\(^+\)→Et\(_4\)N\(^+\) and Et\(_4\)N\(^+\)→Me\(_4\)N\(^+\). These were done in a standard FEP/MD manner and the atomic partial charges from RESP HF/6-31G(d) were used. The bond length between the N and the first C was also adjusted from Amber original parameters to the ones obtained by HF/6-31(d). The MD simulations were performed with and without the SHAKE procedure applied on their bonds [69].

![Diagram of thermodynamic cycle](image)

*Figure 8. Thermodynamic cycle of Me\(_4\)N\(^+\), Et\(_4\)N\(^+\) and Pr\(_4\)N\(^+\) in vacuum and water. The direction of the FEP mutations were done against the arrows, i.e., shrinking the molecules.*

The SHAKE constraints reduce fluctuations in the MD simulation but the procedure adds a contribution to the energies which has to be corrected for, through a potential of mean force (PMF) correction. In the FEP/MD calculations transformation Et\(_4\)N\(^+\)→Me\(_4\)N\(^+\) results in energies from –5 to –4 kcal mol\(^{-1}\) (depending on symmetry) and after adding the PMF contribution values are close to experimental –7.5 kcal mol\(^{-1}\) [68,70]. The same applies to
the transformation of $n$-Pr$_4$N$^+$→Et$_4$N$^+$ which has an uncorrected $\Delta G_{solv}$ of −0.1 - 0.7 kcal mol$^{-1}$.

The PMF contribution can be canceled out by looking at the relative solvation free energies. From the FEP/MD calculations it can be seen in the relative solvation free energies that the $S_4$ conformation is slightly better solvated than $D_{2d}$, in Et$_4$N$^+$ by up to −0.4 kcal mol$^{-1}$ and the difference between $S_4$ and $D_{2d}$ in $n$-Pr$_4$N$^+$ is 0.0 to 0.8 kcal mol$^{-1}$. The result from the NMR experiments predict a symmetric structure of the TEA in water, however the $D_{2d}$ and $S_4$ symmetries are not discernible.

4.1.4 Exploring blocking by QAIs to the outer cavity of the KcsA channel (III)

Although KcsA [2] is a bacterial channel it is still of considerable use since understanding it can provide valuable information of human channels, e.g., how to construct new pharmaceuticals. This channel has large structural similarities with human voltage-gated ion channel but is opened by pH shift [61] (though it has some minor voltage sensitivity [49]). In paper III the goal was to determine the favored binding mode of QAIs to KcsA, $n$-Pr$_4$N$^+$ (TPrA) Et$_4$N$^+$ (TEA) Me$_4$N$^+$ (TMA), and also elucidate what contributes to the preferred $D_{2d}$ symmetry over $S_4$. This is done with the low energy loading state of the selectivity filter, water-K$^+$-water-K$^+$ ordered from the outer cavity [24].

The outer cavity of KcsA has a relatively well defined binding site with Y82 forming a hydrophobic ring at the top of the bowl-formed cage and carbonyl oxygens of Y78 and G79 at the entrance of the selectivity filter (Figure 9). Although the binding position is quite clear, AutoDock was used to get unbiased starting positions of the three QAIs in the outer cavity.
Figure 9. Close-up of KcsA with T74, T75, V76, G77, Y78, G79 and Y82 shown in sticks together with a bound TPrA. The loading state of the selectivity filter is water-K⁺-water-K⁺. Two out of four subunits are shown for clarity.

The outcome of the docking shows that all QAIs fit the outer cage, even the larger TPrA after some accommodations (rotated 0.5 Å out from the binding site). Experimentally the QAIs are ordered as TMA < TPrA < TEA, with TEA being the best binder [71]. The estimated binding free energies from AutoDock ranked the QAIs as TMA < TEA < TPrA with an energy difference between TPrA and TEA of approximately 0.5 kcal mol⁻¹. It should be noted that AutoDock’s standard error is of 2 kcal mol⁻¹. Here the most important are the actual positions since the free energies are to be calculated in a much more rigorous way by FEP/MD.

With the blockers docked, FEP/MD simulations were done using the Amber force field [4] together with RESP charges calculated previously (II). The Amber force field was chosen for two reasons, it gave good results with QAI in paper II and it complies well with AutoDock’s Amber like force field. The SHAKE procedure was applied. The bond PMF contribution was explicitly taken care of by considering uncoupled mutations of the force field parameters and the coordinates in each step of the mapping procedure of the FEP.

$$\Delta G_{sp}(\tilde{\lambda}_m \rightarrow \tilde{\lambda}_{m+1}) = \Delta G(\varepsilon_m, \varepsilon_{m+1})$$

The FEP/MD calculations of the extracellular block of KcsA by $D_{2d}$ conformations showed that TEA binds with the best affinity, while TPrA and TMA have 1.6 kcal mol⁻¹ and 2.4 kcal mol⁻¹ less affinity respectively. This corresponds well to experimental data [71]. For the $S_4$ conformation good experimental agreement is obtained in mutating TPrA to TEA but not in the TEA to TMA simulation. The main contribution to this appears to come from non-
polar interactions and PMF contributions. Table 2 in paper III shows a preference of binding the $D_{2d}$ symmetry over $S_4$ for both TEA and TPrA, by 3.8 kcal mol$^{-1}$ and 2.9 kcal mol$^{-1}$, respectively, with TMA as a reference.

In these FEP/MD calculations the importance of hydrophobic interactions between the QAIs and KcsA is identified. This has also been reported in mutation studies both using computer simulations and experimentally, especially regarding the importance of blockers interactions with Y82 [27,49].

4.1.5 Conclusions and perspectives (II, III)

The actual structure of the hydrocarbon chain conformations is strongly correlated to the local environment. This is may be not so surprising since the energy difference is quite low between the two symmetries $D_{2d}$ and $S_4$ in water. A conclusion that can be drawn from paper II is that QAIs can easily adopt either $D_{2d}$ or $S_4$ symmetry in molecular complexes.

TEA in the $D_{2d}$ symmetry has a stronger interaction to KcsA than $S_4$ and the $D_{2d}$ TEA has better binding affinity than TMA and TPrA in the same conformation. This seems to originate from better van der Waals and steric interactions with the Tyr82 cage. In the absolute binding free energy the major interactions comes from the polar interactions, the non-polar contributions are also important for stabilization although cation-π interactions do not seem to play a significant roll. The amphiphilic nature of the outer cavity enables different blockers to bind which of course makes the selectivity problem harder for this type of channels due to the broader spectrum of possible blocker configurations.

4.2 Human potassium channel – hERG (IV)

Heart cells contain several different voltage-dependent ion channels, regulating the flux of Na$^+$, Ca$^{2+}$ and K$^+$. These channels work in concert and their ion-flow give rise to the measurable cardiac action potential and the electrocardiogram (ECG), as shown in (Figure 10). The main reason for studying the human ether a go-go related gene (hERG) channel is that serious arrhythmia and even sudden cardiac death can be traced back to be a result of blocking this specific channel. This can be seen as a prolongation of the QT wave in the ECG, reflecting a longer duration of the electrical activity that controls contraction of the heart muscle, long QT syndrome (LQTS). The aim is to provide a structural basis for interpreting functional properties of hERG.
Specific drugs that block the efflux of $K^+$ and therefore cause the prolongation of the QT wave are called class III antiarrhythmic drugs, e.g., dofetilide, clofilium and sotalol. These and other drugs that block the channel have been studied heavily in the experimental field to gain insight about the channel. Due to the problems in crystallizing membrane proteins there is no crystallographic structure of hERG although substantial insights have been attained after its discovery in 1994 by Warmke and Ganetzky [72].

With no X-ray structure of hERG a major theoretical task is to make models from which it is possible to draw conclusions about structure and activity, and in this respect homology modelling is the most useful approach.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{cardiac_cell_potential.png}
\caption{The picture shows an example of a cardiac cell potential (---) and the recorded electrocardiogram (----) corresponding to it versus time. The initial phase of the depolarization starts with activation of Na$^+$ channels and an influx of Na$^+$ into the cell ($I_{Na}$). The potential then reaches a plateau at the influx of Ca$^{2+}$ ($I_{Ca}$). This corresponds to the first sharp peak in the ECG. The remaining of the action potential corresponds to K$^+$ efflux ($I_{to}$ – Kv4.3, $I_{Kr}$ – hERG, $I_{Ks}$ – KvLQT1, $I_{Kl}$ – Kir2) from different channels during repolarization. The QT wave stretches from the start of depolarization and to the end of repolarization.}
\end{figure}

4.2.1 Structure and function of hERG

When the hERG channel was first isolated [72] it was shown that it is a six transmembrane (6TM) protein with voltage-gating properties (Figure 11). The bacterial channels with solved 3D structure, e.g., KcsA [2], KvAP [39]
and MthK [38], have similar structure around its selectivity filter, although the conserved sequence is SVGFG in hERG and TVGYG in the other three mentioned. Coordination of the potassium ions in the selectivity filter are regulated in the same manner and the low energy loading state are probably the same. In the first four helices lies the voltage-sensitivity, more specifically in the S4 helices. Here there are several charged Arg residues and these helices are considered to be largely involved in the gating mechanism of hERG.

**Figure 11.** Schematic picture over one of four subunits of a voltage-gated potassium channel topology. The S4 helix carries a number of positively charged Arg that contribute to the voltage sensor part. P-helices and S6-helices shape the interior of the channel, containing the conserved sequence SVGFG and the gating G. Between S5 and P the hERG channel has an longer loop than e.g., KvAP.

The movement of S4 leads to an activation of the channel and it opens although the direct mechanism remains unclear. A conserved residue in the S6 helix, Gly648, is assumed to be engaged as a hinge. In the open state blockers get access to the inner cavity of the channel (see Figure 6) and can bind. S6 also contain two other very important residues Tyr652 and Phe656, which are crucial for several blockers in obtaining strong binding affinity. This has been seen in experimental mutation studies where these residues have been mutated and thereby losing the channel's ligand specificity [73].

### 4.2.2 Modelling, docking and evaluation of hERG (IV)

In paper IV the aim was to construct a homology model of hERG and explore blocker binding to it. The sequence from S5 to S6 of KvAP was chosen as template structure in the homology modelling. The reason for this choice
is that it shows several similarities with hERG, e.g., voltage-gated, conserved Gly in S6 (reflecting its open nature needed for binding [74-77]) and an acceptable sequence homology especially between its S6 helices. The reason for just modelling S5-S6 is because the site of interest in binding is situated around the inner cavity. Parts of the structure at long distance from the cavity will not interact directly with binding compounds and can therefore be omitted from the homology modelling.

The two sequences were aligned with the program T-Coffee [78] and special attention was paid to the conserved sequences and the long turret between S5 and P. The outcome of the alignment showed good consistence between the conserved sequences and the long turret where treated with a large gap. This alignment was then used in the homology modelling program SWISS-MODEL [79] for producing one subunit of the hERG channel. This was then assembled to a tetramer, equilibrated with MD, and checked with respect to torsional angles and steric clashes. The model displayed features that were anticipated such as side chains from the residues Tyr652 and Phe656 pointing in to the cavity and the selectivity filter lined with carbonyl oxygens.

We have used the Amber95 force field [4] in the MD program Q [25] for two main reasons. First, because the charges that AutoDock uses for its macromolecule (in our case the hERG channel) are of the Amber type [8,80]. Secondly, because we could easily obtain the atomic charges for the set of ligands in the same way as were done for the amino acids in the Amber95 force field using the RESP procedure.

A test set of sertindole and some molecular analogues where selected for exploring the inner cavity binding site of hERG. Sertindole has been shown to be able to cause LQTS [81]. It has been predicted that sertindole binds in an extended way in the inner cavity interacting both with the lower part of the selectivity filter and the residues Tyr652 and Phe656 [82]. The difference in binding free energy between these blockers is around 6 kcal mol$^{-1}$ [82]. All these blockers have a positively charged nitrogen in the central part of the molecule.

Before running the MD simulation unbiased starting positions for the blockers were needed and for this purpose AutoDock was used. The parameter setup was chosen in a standard manner using settings from Hetenyi et al. [83]. Running the blocker with full flexibility against the hERG model and selecting top scores in each docking cluster resulted in ligand binding in the predicted orientation and conformation [82]. In the docking procedure the used loading state for the selectivity filter is only water. This state is just for simplicity in the docking setup and the resulting poses do not change de-
pending on the loading state (the other loading state was tested to check this).

MD simulations were done with the docked structures using the program Q [25]. Two simulations were performed for each ligand, one MD in water and one in solvated hERG model. The energy averages from these were then applied in the LIE equation (Equation 10) and the relative binding free energy was obtained for each complex.

The relative binding free energies finally were calculated and compared with the experimental relative free energies calculated from IC\textsubscript{50} values [82]. The ranking of the blockers was perfect and the correlation between the experimental values and the calculated was in excellent agreement.

The position of the blocker's positive nitrogen seem to be affected by the P-helices in the same way as the crystallographic cavity ion and the cyclic structures in the blockers make hydrophobic contacts with Phe656 as proposed by Mitcheson et al. [73] and Pearlstein et al. [82].

4.2.3 Conclusions and perspectives

The free energy calculations on the modeled hERG channel correctly ranked the sertindole analogs against the experimental IC\textsubscript{50} values. This result provides support for the model. Another observation is the position of the basic nitrogen which is positioned for the blocker with best affinity at the position of the crystallographic cavity ion and for blockers with lower affinity the nitrogen position tends to drift from this position. The low dielectric of the membrane has been shown to result in an enhanced effect on dipole moment of the pore helices [84] thereby helping the positioning of the charged nitrogen.

Work is now being done in improving the homology model. This can be performed in a number of ways, one example is to get a more stable structure by having a more optimal interactions between S5 and S6. A homology model that can show correct structure-activity relationships is extremely important for further research on pharmaceutical drugs and understanding the properties of existing drugs that interact with hERG.
5 Summary in Swedish (2-5 sidor)

Kartläggning av ligandbindning till HIV-1 proteas och K⁺-kanaler genom att använda datorberäkningsmetoder

Förståelsen av protein-ligand interaktioner är av största betydelse inom läkemedelsutvecklingen. I detta arbete är målet att förstå sambandet mellan struktur och funktion genom att användanda molekylmodellering. Denna avhandling har sitt huvudfokus på implementering av receptor flexibilitet i molekyldockning samt i studerandet av struktur-aktivitets samband mellan kaliumjonkanaler och deras blockerande molekyler.

I dockningssimuleringar av ligander till protein har rörelser i proteinet samt viktiga strukturella vatten approximerats till en medelstruktur genom att använda flera snarlika strukturer av samma protein. Fyra metoder för att kombinera ett mål protein (medelstruktur) utifrån multipla strukturkomplex har utvecklats och testats. Två metoder, användande en viktfunktion har visats ge konsistenta resultat i molekyldockning användandes en medelmatris representation av medelproteinet.

Kvartära ammoniumjoner är välkända K⁺ kanalblockerare. Konformationen vid C-N bindningen vid det kvartära centret i tetraalkylammoniumjoner i vattenlösning undersöktes med hjälp av kvantmekaniska metoder. Den relativasolvatiseringsframställningen av kvartära ammoniumjoner undersöktes vidare med hjälp av molekyldynamiksimuleringar. Torsionsbarriären för en tvåstegs mekanism, bytet av $D_{2d}$ symmetri till $S_4$, beräknades till 9,5 kcal mol⁻¹. Vidare framkom det att $D_{2d}$ symmetrin är mer stabil än $S_4$ vilket också bekräftas av experimentella studier. Bindning av kvartära ammoniumjoner till den yttre vestibulen av kaliumjonkanalen KcsA studerades också. Datorsimuleringar och beräkningar av relativa bindningsfriaenergier av KcsA i komplex med kvartära ammoniumjoner utfördes. Detta gjordes med molekyldynamik och störningarsteori i kombination med automatisk molekyldock-
ning. Resultaten ligger i linje med experimentella data, Et₄N⁺ i $D_{2d}$ symmetri binder bäst till KcsA.

Bindning av blockerare till den humana hjärtcellens kaliumjonkanal hERG har studerats genom att använda en kombination av homologi-modellering, automatisk molekyldockning och molekyldynamiks-simuleringar. Beräkningarna på en uppsättning läkemedelsderivat till jonkanalsblockerare reproducerade den relativa bindningsenergin mycket bra. Indikationer på att både polära interaktioner nära den intracellulära öppningen av selektivitetsfiltret, såväl som hydrofoba interaktioner i regionen runt F656, är mycket viktiga för bindning av denna jonkanalens blockerare. Den framtagna modellen av hERG borde vara användbar för vidare förståelse och tolkning av denna mycket viktiga humana jonkanal.
6 Acknowledgements

AstraZeneca for their financial support.

My supervisor Johan Åqvist – I am very grateful to you for accepting me to do my Ph.D. in your group. Your knowledge in the scientific fields has learned me much.

Anders Karlén – I think it was you Anders who got me in to this field from the beginning. For that I am very glad, one should never underestimate the roll of having summer students.

Prof. Arthur Olson and Sophie Coon, David Goodsell, Ruth Huey, Garrett Morris and Michel Sanner – Accepting me as a scholar in your group Art will probably have a positive effect for the rest of my life. The 5 month stay in San Diego was really a fantastic experience, both academically and outside the lab.

Victor Luzhkov – Your knowledge in ion channels gave me a nice start of my time as a Ph.D. student. I did learn me much in the two papers we collaborated in.

Karin Kolmodin and Isabella Feierberg – You both also gave me a good start in the MD field as well as being good colleagues.

Martin Almlöf, Martin Ander, Sinisa Bjelic, Björn Brandsdal, Jens Carlsson, Martin Nervall, Hugo Gutierrez de Teran and Stefan Trobro – What have my time at the lab been without you guys? Getting to know you and spending time together both in the lab and outside have been extremely nice! The daily contact with all of you!

DrLoco_2.0, Kournikova, SVEN, cartman and Bugs – We'll meet in DM-Antalus with the redeemer in the hands of foe!

alla@xray – Thanks for nice relaxing coffee breaks!
BK Smittan – A solid support in the cyberspace! We have had plenty of bullshit mixed with discussions on a very intellectually high level. Just keep it going!

My lovely wife Anna and son Erik – You two have been the best support one can think of during these last weeks. You are my everything!
7 References

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 124

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-6167