Computational Study of the α_1-Adrenoceptor Subtypes and their Ligands

By: Berry Matijssen
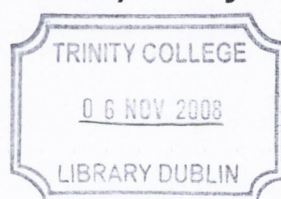
A thesis presented to Trinity College, University of Dublin, Dublin
for the degree of Doctor of Philosophy

July 2008
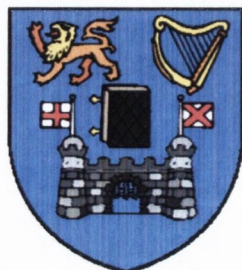
Declaration

I hereby declare that:

This thesis has not been submitted as an exercise for a degree at this or any other University.

The work contained herein is entirely my own, except where otherwise cited, referenced, acknowledged or accredited.

I agree that the Library of the University of Dublin, may at their discretion lend or copy the thesis upon request

C. Matijssen

....

Berry Matijssen

July 2008
University of Dublin, Trinity College

Thesis summary

Adrenoceptors (AR) belong to the G-protein coupled receptor (GPCR) family. These receptors play an important role in regulating many processes in the body related to the central nervous system, vascular system and many others. The importance of these receptors as regulators can be further stressed as they are considered promising drug targets. Currently 40% of all marketed drugs are targeted at GPCRs. The adrenoceptors can be subdivided into nine different classes of which the α_1-AR class consists of three subtypes (α_{1A}-AR, α_{1B}-AR and α_{1D}-AR). This α_1-AR class plays an important role in the condition known as benign prostatic hyperplasia (BPH). BPH affects 50% of men over 50 years old and with the aging of the population percentage, this percentage is expected to rise. BPH is manifested by the enlargement of prostate tissue that constrains the urethra Prostatic smooth muscle contraction occurs mainly via the α_{1A}-AR subtype. The effect of BPH is manifested through the impaired flow of urine through the urethra when is passes the prostate. Inhibition of the α_{1A}-AR by using an antagonist has shown to increase the urine flow and therefore, decrease the physiological aspects of BPH. The class α_{1D}-AR is located in the neck of the bladder where the urine leaves the bladder and it is believed that inhibiting this receptor could be beneficial in the treatment of BPH. The α_{1B}-AR does not play a role related to BPH, but is present in the brain. Therefore, it is believed that inhibition of this receptor could result in unwanted side-effects. Hence, antagonists which are selective for the α_{1A}-AR and α_{1D}-AR would be useful for treating BPH with reduced side effects. In our research we develop models for each of the three α_1-AR subtypes which can be used in determining ligand-specific interaction with each subtype.

Structure based drug design requires the knowledge of the structure of target receptor, which in our case are the three α_{1A}-AR, α_{1B}-AR and α_{1D}-AR subtypes. The 3-dimensional structures for these receptors are still unavailable and until recently, the only GPCR crystal structures that were available were that of bovine rhodopsin. The structure of bovine rhodopsin has a low similarity (<20%) to the adrenoceptors but shares a number of residues in the transmembrane helices which are conserved amongst all GPCRs making this structure potentially useful as a template for homology models and is the template we choose for our studies. During the writing of this thesis, the structure of the β_2-adrenoceptor has been discovered, which shares higher similarity (60%) with the other adrenoceptors, although this discovery was made towards the end of this study.

A number of ligands are known to selectively interact with the three α_1-AR subtypes and they provide a good starting point for modelling the interaction between ligand and protein. These ligands are described for the neutral state, but at physiological pH they should be protonated. In general, the ligand-receptor complex involves different interactions such as hydrogen bonds and ionic interaction and thus, an accurate description of the protonated state of the ligand is very important.

Using homology modelling we have developod models of the three adrenoceptor subtypes α_{1A}-AR, α_{1B}-AR and α_{1D}-AR based on the crystal structure of bovine rhodopsin. Several approaches have been taken and compared to obtain accurate 3D models. These models are tested for stability using molecular dynamics (MD) simulations. The adrenoceptors are normally found in a phospholipid bilayer, but such bilayers are difficult to model. An alternative approach to model this environment is to make use of solvent layers consisting of water/chloroform/water in which chloroform represents the hydrophobic phosplipid bilayer and water the intra- and extra-cellular space. A selection of models is made and can be further used for study of protein-ligand complexes.

A number of ligands selected from the literature were optimized for this protein-ligand interaction study. This optimization consisted of a conformational search to find a range of minimum energy conformations. Each conformation was then further optimized using hybrid functional theory (B3LYP) which provides accurate minimum energy conformations. The lowest energy conformation of all these calculated conformations is then selected which is regarded as the global minimum. This minimum energy conformation is then subjected to proton affinity calculations by protonation of each nitrogen independently and optimization of each structure using B3LYP. The conformation with the largest difference in energy between the neutral state and the most stabel protonated state is regarded as the site for protonation at physiological pH.

Using the homology models and the optimized ligands, the drug-receptor interaction can be investigated using molecular docking followed by MD simulations. The ligand has several possibilities of fitting into the receptor. Therefore, by means of molecular docking different ligand-protein complexes are generated and an optimal complex is selected. To refine this, further MD simulations are performed using the three-layer solvent model to model the membrane bilayer. From our ligand set a selection of 12 antagonists has been made which are docked individually into each of the three α_1-AR subtypes giving rise to

36 different antagonist-adrenoceptor complexes. The refinement of the interaction between ligand and receptor within the binding pocket of each subtype allows the determination of subtype specific interactions.

Analysis of the final complexes identifies a number of residues which are involved in binding. Most of the residues commonly considered to be involved in the binding to antagonists are conserved the three adrenoceptors and can therefore not be used to design subtype specific antagonists. They can be used to identify the binding pocket and used for identification if a ligand has bound correctly, as binding to these residues is expected. A considerable number of residues involved in binding can be considered as ligand-specific but our studies have not allowed the explaination of subtype specificity. Furthermore, it appears that some or our resulting ligand-protein complexes lack some strong interactions which were known to be important for strong binding affinity. This suggests that an optimal interaction has not been achieved. However, we believe that for four of the compounds studied an optimal binding has been achieved and for each adrenoceptor subtype establishing the reported needed interactions. Considering the diversity of these four structures and the different interactions established with each α_1-AR subtype, these compounds could be considered as good 'lead' comnpounds for developing subtype specific antagonists.

Acknowledgements

I would like to take this opportunity to acknowledge and thank a number of people who I have worked with and helped me out in many ways on several occasions over the past few years in Trinity College.

Firstly I would like to thank my two supervisors Graeme Watson and Isabel Rozas who gave me the opportunity of taking up a PhD studentship at Trinity College, Dublin and have been a great support for me during the last few years. They have helped me out on numerous occasions and provided me with a new set of insights and knowledge. One of the things which I have enjoyed while working with them, were the discussions we have had from different points of view. These have introduced new aspects of research for me as Graeme has a computational background, Isabel an organic medicinal chemistry background and myself a medicinal biochemistry background.

The second group of people I would like to thank is the people who I have worked the most with, my colleagues. I would like to start with Gemma Kinsella. I continued the adrenceptor project which she started which provided a good basis for my work. Many things came straight into my path which made it all somewhat easier for me. Although some of the people in our group have moved onto a different organisation, I would also like to thank Joanne Fearon, Aron Walsh, Michael Nolan, Ben Morgan, Oscar Rubinho, Justine Taylor, David Scanlon, Kate Godinho and Ní Chaoimh Dewdney for the enjoyable experience I had while working with them, but also for the social aspects for which I have a lot of good memories.

The other group of people I would like to thank is the research group which is supervised by Isabel. Although my attention sometimes slips during organic chemistry theory I learned a lot about the different issues that exist in chemistry outside my own field. I would like to thank Jonathan Corcoran, Sergio Lopez, Fernando Rodriguez, Padraic Nagle, Amila Kahvedzic and Aoife Flood for the many discussions on chemistry and other enjoyable moments.

Furthermore I would like to thank the people from HPC, Dermot Frost, Jimmy Tang, Jose Refojo, Geoff Bradley and Bob Crosbie. You guys have provided me with all of my working tools and have spent considerable time keeping them up to date and, more importantly,

fixed things as quickly as possible when they were broken. Also the advice on how to make my programmes run more efficiently was well appreciated.

The last people I would like to thank are my parents and my family for supporting me during the years and especially during my PhD study.

Table of contents

List of abbreviations

1D	One dimensional
3D	Three dimensional
3D-QSAR	Three Dimensional Quantitative Structure Activity Relationship
AM1	Austin Model 1
AMBER	Assisted Model Building with Energy Refinement
ARF	Alternative Reading Frame protein
B88	Becke functional
BPH	Benign Prostatic Hyperplasia
CHARMM	Chemistry at HARvard using Molecular Mechanics
CLR	Calcitonin Like Receptor
DAS	Dense Alignment Surface
DFT	Density Functional Theory
DHT	DiHydroxyTestosterone
DZ	Double Zeta
ECL	ExtraCellular Loop
ER	Endoplasmic Reticulum
FDA	Food and Drug Administration
FFT	Fast Fourier Transform
GDP	Guanine DiPhosphate
GRK	G-protein coupled Receptor Kinases
GROMACS	GROningen MAchine for Chemical Simulations
GTO	Gaussian Type Orbitals
GTP	Guanine TriPhosphate
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
HF	Hartree Fock
HP	Hartree Product
LCAO	Linear Combination of Atomic Orbitals
LDA	Local Density Approximation
LUTS	Lower Urinary Tract Symptoms
LYP	Lee-Yang-Par functional
MD	Molecular Dynamics
MNDO	Modified Neglect of diatomic Differential Overlap

MO	Molecular Orbital
NMR	Nuclear Magnetic Resonance
PA	Proton Affinity
PBC	Periodic Boundary Conditions
PCM	Polarizable Continuum Model
PDF	Probability Density Functional
PGTO	Primitive Gaussian Type Orbitals
PM3	Parametric Method number 3
PMF	Potential of Mean Force
PMES	Particle Mesh Ewald Sum
QM	Quantum Mechanics
QOL	Quality Of Life
QSAR	Quantitative Structure Activity Relationship
R_A	Active state of the receptor
RAMP	Receptor Activity Modifying Protein
RCSB	Research Collaboratory for Structural Bioinformatics
R_I	Inactive state of the receptor
RMSD	Root Mean Square Deviation
SBDD	Structure Based Drug Design
SCF	Self Consistent Field
SOMFA	Self Organizing Molecular Field Analysis
STO	Slater Type Orbitals
SV	Split Valence
TMHMM	Transmembrane Prediction with a Hidden Markov Model
TMpred	Transmembrane Prediction
VDW	Van der Waals
ZDOA	Zero Differential Overlap Approximation

Chapter I

Introduction: Benign Prostatic Hyperplasia, Adrenoceptors and GPCRs.

Heavier-than-air flying machines are impossible

Lord Kelvin (1824-1907)

1.1. Introduction

In this introduction several aspects of benign prostatic hyperplasia (BPH), G-protein coupled receptors (GPCRs) and the adrenoceptors will be discussed. First, an introduction to BPH which describes the prostate, the pathological processes involved, the prevalence and the current drug therapies for BPH will be presented. The receptors which are the object of this work are the adrenoceptors which belong to a class of proteins known as GPCRs. An overview of different structural and functional aspects of GPCRs will be given. This will be followed by a short overview of rhodopsin which was the only GPCR with a known crystal structure at the time of this study. Recently, the crystal structure of the β_2-AR has been discovered which shares more sequence similarity to the α_1-ARs. The next section will deal with the adrenoceptors. Several of their known structural features will be discussed and the different binding pockets are described. The last part of the introduction will consist on an overview of the research that has been performed on the adrenoceptors using pharmacophore analysis, quantitative structure activity relationships (QSAR), homology modelling, molecular docking and other computational methods.

1.1.1. Benign prostatic hyperplasia

The term BPH[1] is used to describe histopathologic hyperplastic changes in the prostate. Hyperplastic changes refer to an increase in the number of cells. Clinicians commonly use the term BPH to describe a clinical syndrome consisting of thee components: lower urinary tract symptoms (LUTS), benign prostatic enlargement and bladder outlet obstruction. A fourth term, benign prostatic obstruction is used when bladder outlet obstruction and benign prostatic enlargement occur. Several common physiological aspects of BPH such as flow rate, obstruction and the size of the prostate have been measured. However, there appear to be no or weak correlations between the occurring symptoms of BPH and these physiologic measures.[1,2]

1.1.2. Prostate

The main function of the prostate is to store and secrete a clear, slightly basic fluid that constitutes up to one-third of the volume of semen. The prostate also contains some smooth muscles that help expel semen during ejaculation.

A healthy human prostate is slightly larger than a walnut. It surrounds the urethra just below the urinary bladder (Fig. 1.1) and can be felt during a rectal exam. Within the

prostate, the urethra coming from the bladder is called the prostatic urethra and merges with two ejaculatory ducts.
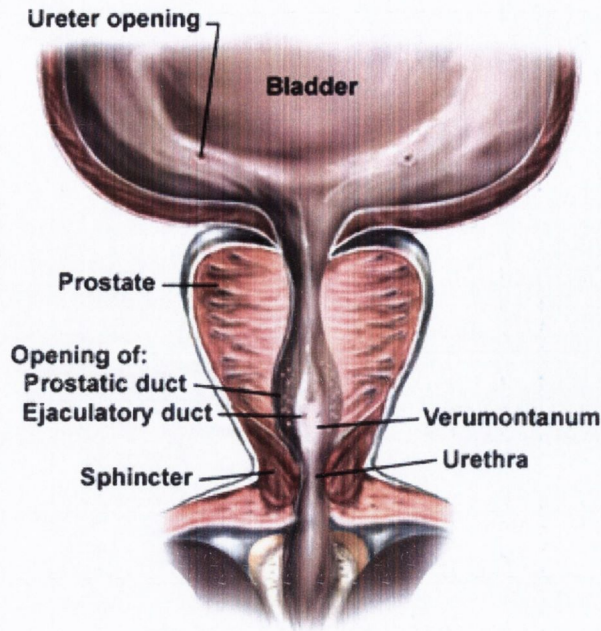
Figure 1.1: Anatomy of man showing the bladder and prostate (picture taken from indianhealthguru.wordpress.com)

The prostate consists of four different glandular regions[3] (Fig. 1.2). The peripheral zone is the bottom part of the outside of the prostate gland which surrounds the urethra. The transition zone surrounds the proximal urethra and is the region of the prostate gland which grows throughout life and is responsible for the benign prostatic enlargement. The anterior fibromuscular zone or preprostatic sphincter consists of muscle and fibrous tissue. The verumontanum is an elevation in the wall of the urethra where the seminal ducts enter it.
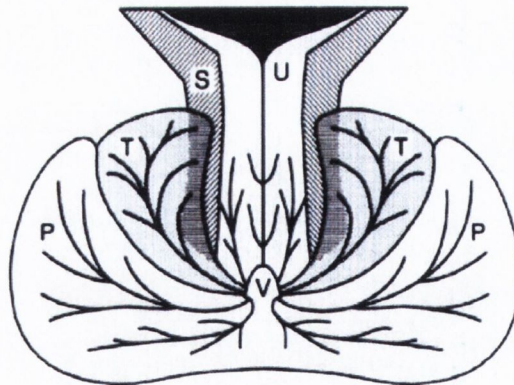
Figure 1.2: Overview of the prostate where P is the peripheral zone, T is the transition zone, S is the preprostatic sphincter, U is the urethra and V is the verumontanum (picture taken from Oesterling et al.[3])

1.1.3. Progression of BPH

Benign prostatic hyperplasia nodules originate within the transistion zone that lies proximal to the verumontanum and in proximity to the cylindrical urinary sphincter that arises form the bladder neck.[3] Inside the internal sphincter and immediately adjacent to the urethra is the periurethral tissue. BPH nodules can arise from different tissues. Those formed in the transition zone develop early and consist of large epithelium groups constituting the main mass of BPH tissue enlargements. Those nodules that orginate from the periurethral zone tend to have a later onset and to be smaller, fewer in number and composed primarily of fibroblastic tissue. The initial lesion of BPH does not arise from supportive tissue (stromal nodules), but from a formation of glandular budding and branching towards a central focus that occurs primarily in the transition zone.

In the early states of BPH, the transition zone undergoes a diffuse enlargement without any significant change in its architecture. The prostate shows microscopic enlargement, but no clinical symptoms occur. In the next stage of development, the initial microscopic nodules undergo further hyperplasic changes, resulting in a substantial increase in size. The process of budding and branching continues within the nodule and adjacent ducts become stimulated and are incorporated into the process as well. Microscopic nodules develop into macroscopic nodules and the original anatomy of the prostate gland becomes distorted.

The final stage in the histogenesis of BPH is the development of clinical symptoms as a result of the enlarged prostate gland. Prostatic enlargement by itself is not sufficient to produce clinically significant bladder outlet obstruction. Additional factors are necessary for the progression of pathological BPH into clinical BPH. An event such as prostatic infarction could be a cause for developing pathological BPH.

1.1.4. Prevalence of BPH

The estimated prevalence of BPH from community-based studies is 40% of men in their 70s, but histological evidence in autopsy studies has been found BPH in 60-80% of men aged 60-69 years.[1] Further research into histological development of BPH shows that initial development of BPH can begin as early as 25-30 years of age with a prevalence of approximately 10%.[3] With advancing age the occurence of BPH increases such, that by the age of 51-60 year the prevalence is more than 50%. At the age of 85 approximately 90% of men will have this condition. This means that BPH is a condition that starts early in life and if men live long enough, full BPH develops later.

1.1.5. Impact on quality of life

BPH has in general a great impact on the quality of life (QOL).[2] QOL is an important factor in an aging society. The two most troublesome symptoms are sleep interruption to empty the bladder (nocturia) and incomplete emptying of the bladder. The difference between the groups with moderate decrease in QOL and severe decrease of QOL was related to a significant difference in the symptoms that were experienced. Because LUTS and, and in particular BPH develop with age and result in a decrease of QOL, several treatments are focussed on increasing the QOL and not necessarily on the treatment of the causes of BPH itself.

1.1.4. Treatment of BPH

The treatment of BPH is dependent on the progression of the hyperplasia in the prostate and on the size of the prostate.[4] A medicinal approach can be taken for moderate and severe BPH accompanied with an enlarged prostate. Current therapies used in clinic are drugs that target 5-α-reductase which prevent further progression of BPH and target the α_1-adrenoceptors which limits the physiological effect of BPH.

5-α-reductase[5] is an enzyme that is involved in the progression of BPH. It has been discovered in men with a disease termed pseudovaginal perineoscrotal hypospadias. This condition is caused by the absence of 5-α-reductase. This disease is characterized with underdeveloped genital organs, but in late puberty a gender reassignment takes place in response to a significant rise in testosterone. It was also found that these men do not develop benignly enlarged prostates or prostate cancer. In addition they do not develop male pattern baldness. Testosteron can be degraded by 5-α-reductase to dihydroxytestosterone (DHT). This suggests that DHT instead of testosterone is needed to develop BPH. The drug Finasteride has been developed to inhibit 5-α-reductase. Finasteride competes with testosterone for the active side of the enzyme without affecting the interaction of testorone or DHT with the androgen receptor. Therefore finasteride reduces the serum DHT level by 70% and locally at the prostate by 90% which stops the progression of BHP.

The α_1-adrenoceptors are receptors embedded in the membrane of cells present amonst other tissues in the prostate. Several subtypes of the adrenoceptor are known. Of these

the α_{1A}-AR is the most prominent in the prostate and therefore this receptor is a drug target. A more detailed description of adrenoceptors can be found in chapter 2. It has been found that the alpha-blockers help to relax the smooth muscle in the prostate, therefore allowing a better urine flow. Drugs that target the α-AR adrenoceptor have been evaluated for over 30 years and have resulted in six drugs being marketed[6], (Fig. 1.3). First, the non-selective α-inhibitor phenoxybenzamine followed by selective α_1-AR antagonists. The first selective antagonist was prazosin, which was short acting, followed by the long-acting antagonists terazosin, doxazosin, tamsulosin and alfuzosin. The improvement of the long-acting antagonists arise from the reduction of side effects but these newer drugs have not shown an effective improvement in the treatment of BPH itself.

Phenoxybenzamine
(Rogitine, Novartis)

Prazosin
(Minipress, Pfizer)

Terazosin
(Hytrin, Abbott)

Doxazosin
(Cardura, Pfizer)

Tamsulosin
(Flomaxtra/Flomax, Astellas Pharma,
Boehringer-Ingelheim)

Alfuzosin
(Uroxatral, Sanofi-Aventis)
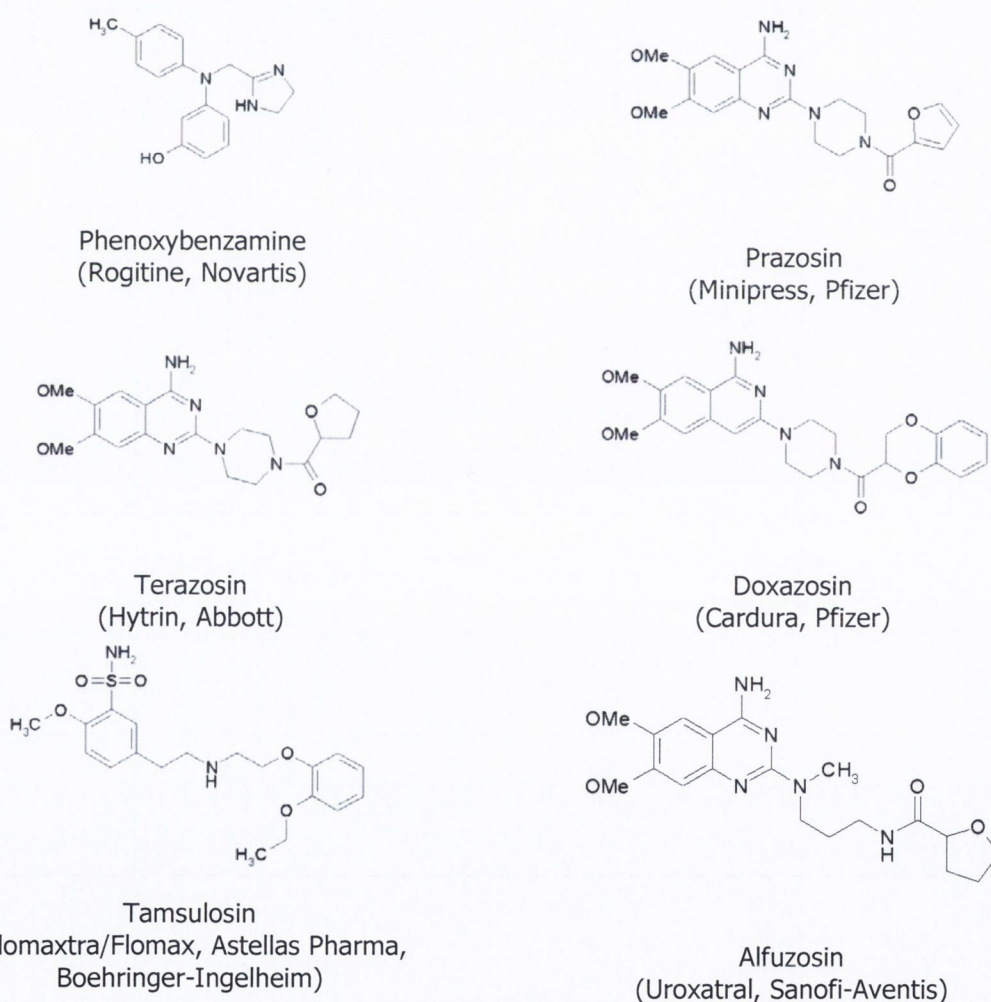
Figure 1.3: Overview of the structures of the different antagonists that target the α_1-AR.

Phenoxybenzamine was the first drug to be effective for the treatment of BPH. However, it came with several side-effects such as tiredness, dizziness, impaired ejaculation, nasal stuffiness, and hypertension.

Prazosin was the second drug and reduced the side effects of phenoxybenzamine significantly. The downside of prazosin was that it requires multiple daily dosing and has some adverse events related to its blood pressure-lowering effect.

Terazosin was the first long-acting drug which in contrast to Prazosin only needs daily dosing. Doxazosin was the second long-acting drug for treatment of BPH. The advantage of doxazosin was its longer half-live tolerability. However, in practice this did not give it a therapeutical benefit. Tamsulosin was the third long-acting drug and was marketed as an antagonist for treatment of BPH. However, it did not show clinically sufficient improvements. The advantage of tamsulosin is that it does not need a dose titration and it has a minimal effect on blood pressure. One of the side effects of tamsulosin is that it causes anejaculation. Alfuzosin is the fourth drug that was approved by the FDA for treatment of BPH. It shares the same advantages as tamsulosin by not needing a buildup of the dose (dose titration), but has the benefit that is does not cause ejaculatory dysfunction. Alfuzosin has also the most minimal effects on dizziness, asthenia and ejaculatory dysfunction relative to all the other α-blockers.

Both terazosin and doxazosin have observed effects of lowering blood pressure in men who are hypersensitive at the baseline. This was desirable because it suggested that two common conditions, hypertension and hyperplasia, could be treated with a single drug. However, it was found that drugs that have both effects, are significantly less effective on releaving hypertension than the available antihypertensive drugs.

With the introduction of selective α_1-AR blockers the focus on improving drugs for the treatment of BPH targeted at reducing side-effects, but there has been little progress made in the treatment of BPH itself. It has also been found by a number of pharmaceutical companies that antagonists that increase the urine flow do not necessarily relieve the symptoms of LUTS. This has moved the focus of research for BPH away from selective targeting of α_1-AR to other pathophysiological processes surrounding LUTS.

1.2. Adrenoceptor

Adrenoceptors are metabotropic receptors (GPCRs) and the target of the neutrotransmitter/hormone noradrenaline and adrenaline.

1.2.1. Subtypes

The adrenoceptors (AR) can be subdivided into α-AR and β-AR[7] (Fig 1.9). The α-AR mediate most excitatory functions such as vasoconstriction, uterine musculature contraction, urethra contraction and pupil dilation and one important inhibitory function, intestinal relaxation. β-AR mediate most inhibitory functions such as vasodilation, uterine musculature relaxation and bronchodilation and one excitatory cardiac function. The α-AR can be further subdivided into α_1-AR and α_2-AR while the β-AR can be subdivided into β_1-AR, β_2-AR and β_3-AR. Due to our interest in the α_1-AR, the α_2-AR and β-AR will not be discussed further. The current classification of α_1-AR consists of α_{1A}-AR, α_{1B}-AR and α_{1D}-AR.
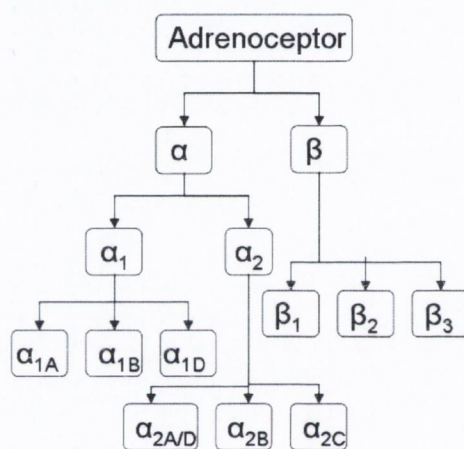
Figure 1.3: Classification of the adrenoceptors

The α_1-AR usually mediates excitatory responses in the effector organ. The α_2-AR is located presynaptically and regulates the release of the neurotransmitter, but is also present in postsynaptical locations. Both α-ARs are important for the control of vascular tone. There is however a difference in the distribution of the different α-AR subtypes in different organs. This difference in distribution can be used to develop treatments based on selective ligands.[8]

1.2.2. Treatment of BPH by subtype selective antagonists.

The prostate contains a high concentration of α_{1A}-AR, but not α_{1B}-AR and α_{1D}-AR. It is also found that the stimulation of the adrenoceptor leads to a less than favourable flow of urine through the prostate. Therefore the α_{1A}-AR is an ideal target for suppressing symptoms of BPH. Besides the prostate, adrenoceptors are also present in the bladder

mouth. But these are of type α_{1D}-AR and stimulation of these receptors leads to symptoms such as incomplete emptying of the bladder. This receptor can be targeted as well to treat BPH. Thus targeting the receptors α_{1A}-AR and α_{1D}-AR an effective suppression of BPH symptoms is obtained. However, most drugs are not selective for those two receptors alone. If a ligand has an affinity for the α_{1B}-AR which is present in the brain, this could result in unwanted side-effects. Therefore an ideal drug should be designed to selectively inhibit the α_{1A}-AR and α_{1D}-AR, but not interact with the α_{1B}-AR.

1.2.3. Structural features of the adrenoceptors.

There are several known structural features of adrenoceptors which have been primarily derived from mutation studies. A characteristic structural feature is the presence of two cysteine residues which are able to form a disulphide bond and are conserved in most class A GPCRs. Based on the arrangement of the ligand in the receptor this leads to a reaction mechanism for the cleavage of the disulfide bond catalysed by the ligands.[9] The interaction between the carboxylic acid side chain of aspartic acid and the disulfide bond leads to the polarization and withdrawal of a proton from the protonated nitrogen of the ligand to one of the sulphur atoms. The accompanying cleavage of the disulfide bond may produce a conformational change in the extracellular loops, opening the pore formed by the seven-helix bundle. This means that not only can the ligands move into the receptor, but water and ion flux can also occur through the receptor from the extracellular to the intracellular regions of the cell. Water or ion penetration through the pore of the helices is probably an additional trigger leading to the activation, by means of conformational changes in the receptor, of the G-proteins and followed biochemical responses.

Another characteristic feature is that the conformational change associated with the activation of GPCRs (section 1.3), involves the disruption of a salt bridge constraint between Glu[113] on TM-III and a Lys[196] in TM-VII.[10] When there is a *cis-trans* conversion this salt bridge is broken and the receptor is activated.

The residues Asp[125] and Lys[331] in the α_1-AR have the potential of forming a constraining salt bridge holding the receptor to an inactive protein configuration.[11] This salt bridge constraint is then released upon binding by the receptor agonist. Substitution of the lysine with an alanine and, thereby eliminating the positive charge at position 331, resulted in a 6-fold increment of adrenaline binding affinity with no alterations of affinity values for selective adrenoceptor antagonists. These mutations suggest a molecular mechanism by which the positively charged Lys[331] stabilize the negatively charges Asp[125] via a salt bridge

constraint until the receptor is bound by an agonist. The salt bridge and ionic lock are thought to be common in most GPCRs.

Another important structural feature is the second extracellular loop (ECL-II) which has been studied extensively for the dopamine D_2 receptor[12]. It is proposed that this loop can enter the binding site crevice of aminergic and other rhodopsin-like GPCRs that have small molecule ligands. Experimental findings suggest that the residues in the ECL-II loop that contribute to ligand binding are not freely accessible.[12] Thus, ECL-II is likely to be folded into the binding-site crevice a significant fraction of the time. The experiments also suggest that some amino acids of the ECL-II loop could have their residues in close proximity to the binding pocket and be involved in binding or activation.

In the case of the third extracellular loop (ECL-III), it is not a simple peptide that links TM-VI to TM-VII.[13,14] This loop occupies a significant position in the GPCR architecture and it is likely that it is involved in providing important epitopes for normal receptor functioning. It has been found that the replacement of the ECL-III does give a different affinity for antagonists if the ECL-III is replaced by the ECL-III of another subtype of the adrenoceptor. It is unlikely that one or more key amino acids are responsible for this, but that for the different GPCRs there are different amino acids located at different positions on the ECL-III loop.

A very important feature of GPCRs is the highly conserved Glu/Asp-Arg-Tyr motif (E/DRY) which plays an important role in the activation process.[15,16] It has been found that the arginine of the DRY-motif can drive the isomerization of the α_{1B}-AR into different states, therefore having a crucial role in the activation process of the receptor.[17] Due to the sequence similarity with other other GPCRs this is likely to occur in several GPCRs.
It has been demonstrated that different mutations of Phe^{143} in the α_{1B}-adrenoceptor result in diverse functional effects. The hypothesis is that Arg^{143} in the α_{1B}-adrenoceptor plays a role in receptor isomerization towards different states. The mutation of Arg^{143} to lysine seems to trigger receptor isomerization to an 'active' state. In contrast, it is found that the mutation of Arg^{143} to glutamic acid seems to induce the isomerization of the receptor to an 'active-like' state that shares some properties with the 'active' forms, including high affinity for agonists, enhanced basal phosphorylation, and the opening of a cytosolic crevice between the ICL-II and ICL-III, but is severely impaired in its ability to mediate a response.

The apparent polymorphism of the phenotype, which can be associated with mutations of the arginine of E/DRY motif in different GPCRs, does not invalidate the hypothesis that this residue might play an important role in the activation process of various receptors. However, the true mechanism of the arginine in this E/DRY motif remains unclear. There are several papers reporting the protonation of the Asp in the strongly conserved E/DRY motif as a requirement for activation.[11,18-20]

1.2.4. Mutation data

It has been found that the residue Phe^{303} is highly conserved in GPCRs and is located in TM-VI[21] of the α_{1B}-AR. Phe^{303} is a key residue in coupling TM helical movements to G-protein-activation.

In the α_{1A}-AR Phe^{308} and Phe^{312} are involved in the binding of imidazoline-type agonists such as oxymetazoline, cirazoline, and clonidine.[22] This in contrast to phenethylamine-type agonists suggesting a different binding mode for different types of ligands.

Studies indicate that the Phe^{310} side chain of the α_{1A}-AR is well positioned to interact with the catechol-ring of an agonist. Substituted cysteine accessibility method studies reveal that the side chain of the Phe^{310} residue is solvent accessible and directed into the agonist-binding pocket.[23] They conclude that for catecholamine agonists the interaction with Phe^{310} is essential for the transformation from the ground state to the fully activated state.

Another mutation study found that neither the mutation of Phe^{163} to Gln^{163} nor Phe^{187} to Ala^{187} had any effect on the affinity of the α_1-antagonists[24]. However, the affinity of the endogenous agonist adrenaline was reduced 12.5 and 8-fold by the $Phe^{163}Gln$ and $Phe^{187}Ala$ respectively. An additive loss in affinity (150-fold) was observed for adrenaline for both mutations. It was proposed that both Phe^{163} and Phe^{187} are involved in independent aromatic interactions with the cathechol ring of agonists, but are not involved in the activation of the receptor.

1.2.5. Binding pocket

Several residues have been found to be involved in binding of agonists and antagonist. These residues have been found using experimental methods or computational methods. Most of the residues involved in binding have been determined for the α_{1A}-AR and several for the α_{1B}-AR. However, due to the similarity in amino acid sequence and structure it is likely that residues which are aligned similarly are involved in binding in the α_{1A}-AR, α_{1B}-AR and α_{1D}-AR. Therefore, the aligned residues can be found in all of the adrenoceptor subtypes, leading to the following residues being possibly involved in binding of agonists (Table 1.2) or antagonists (Table 1.3).

Table 1.2: Residues of the α_1-AR subtypes involved in binding of an agonist

Residue in α_{1A}-AR	Residue in α_{1B}-AR	Residue in α_{1D}-AR	Reference
Asp[106] *	Asp[125] *	Asp[176] *	25
Ser[158]	Ser[177]	Ser[228]	26
Phe[163]	Leu[182]	Leu[233]	27
Gln[177] *	Gly[196] *	Gly[247] *	25
Val[185]	Ala[204]	Ala[255]	25
Phe[187]	Phe[206]	Phe[257]	25,27
Ser[188]	Ser[207]	Ser[258]	25
Ser[192]	Ser[212]	Ser[262]	27
Phe[193] *	Phe[213] *	Phe[263] *	27
Phe[281]	Phe[303]	Phe[357]	29
Trp[285]	Trp[307]	Trp[361]	26
Phe[288] *	Phe[310] *	Phe[365] *	18,25
Met[292]	Leu[314]	Leu[368]	25

* residue is involved in binding of both agonist and antagonist

Table 1.3: Residues of the α_1-AR subtypes involved in binding of an antagonist

Residue in α_{1A}-AR	Residue in α_{1B}-AR	Residue in α_{1D}-AR	Reference
Asp^{106} *	Asp^{125} *	Asp^{176} *	25
Cys^{110}	Cys^{129}	Cys^{180}	26
Gln^{177} *	Gly^{196} *	Gly^{247} *	25
Ile^{178}	Val^{197}	Ile^{248}	25
Asn^{179}	Thr^{198}	Thr^{249}	25
Phe^{193} *	Phe^{213} *	Phe^{263} *	27
Phe^{288} *	Phe^{310} *	Phe^{365} *	17,25
Phe^{308}	Phe^{330}	Phe^{384}	25,29
Phe^{312}	Phe^{334}	Phe^{388}	25,29
Tyr^{316}	Tyr^{338}	Tyr^{392}	26

* residue is involved in binding of both agonist and antagonist

1.3. G-protein coupled receptors

G-protein coupled receptors (GPCRs) are the target for a large amount of neurotransmitters. GPCRs are characterized by seven membrane-spanning domains (α-helices) with an extracellular N terminus and a cytoplasmic C terminus (Fig 1.4). GPCRs can be divided into three major subfamilies[30] of which the rhodopsin-related GPCRs is the largest family and also the most investigated:

- Class A: receptors related to rhodopsin
- Class B: receptors related to the calcitonin receptor
- Class C: receptors related to the metabotropic glutamate receptors
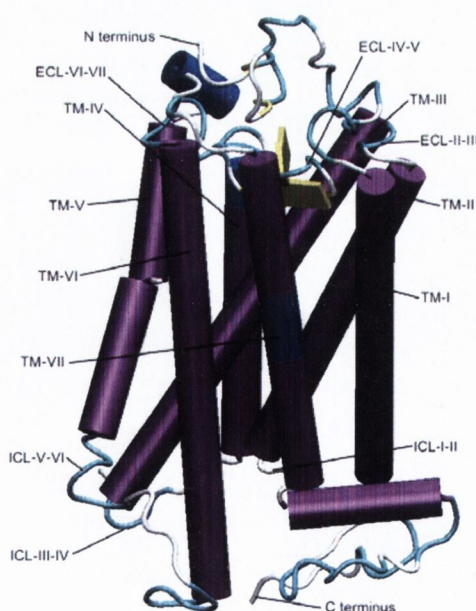
Figure 1.4: General structure of a GPCR

Activation of a GPCR consists of a ligand entering the protein setting off a cascade of reactions, leading to a desired biological effect. The first models that describe activation for GPCRs are described as 'lock-and-key' models (Section 1.3.1). However, further insight into GPCRs shows that there are more possible mechanisms involved in activation as explained by the ternary complex model (Section 1.3.2). Besides the activation by the ligand there appear to be other proteins such as receptor acitivity modifying proteins (Section 1.3.3.) and Arrestins (Section 1.2.4.) involved in activation and deactivation of GPCRs. Finally there is also evidence that GPCRs can form oligomers (Section 1.2.5.) and this formation is needed to perform their function. This has been demonstrated only for some of the GPCRs, leading us to believe that these mechanisms exist for several GPCRs, but not for all. Still, these potential mechanisms should be considered when analysing modelling results. Due to the computational costs and lack of data on which influences to model for each GCPR, these factors are usually not considered in computational models.

1.3.1. Activation of GPCRs

The activation of GPCRs has been studied extensively, but is not completely understood. However, several basic steps have been found to occur in all of the GPCRs leading to a basic model for GPCR activation[31]. The basic model of activation involves the following steps (Fig 1.5):

1. The receptor is in an unbound and inactive state.
2. The ligand binds to the GPCR.
3. The GPCR undergoes a conformational change.
4. The GDP that is attached to the G-protein is replaced by a GTP.
5. The G-protein is split up into two subunits: the α-subunit (carrying GTP) and the βγ-subunit. The α-part binds to the effector mechanism (i.e. adenyl cyclase, ion channels) which is then activated.
6. The GTP of the α-subunit is dephosforylated by GTP-ase to GDP and the α-part and the βγ-part can merge back together.
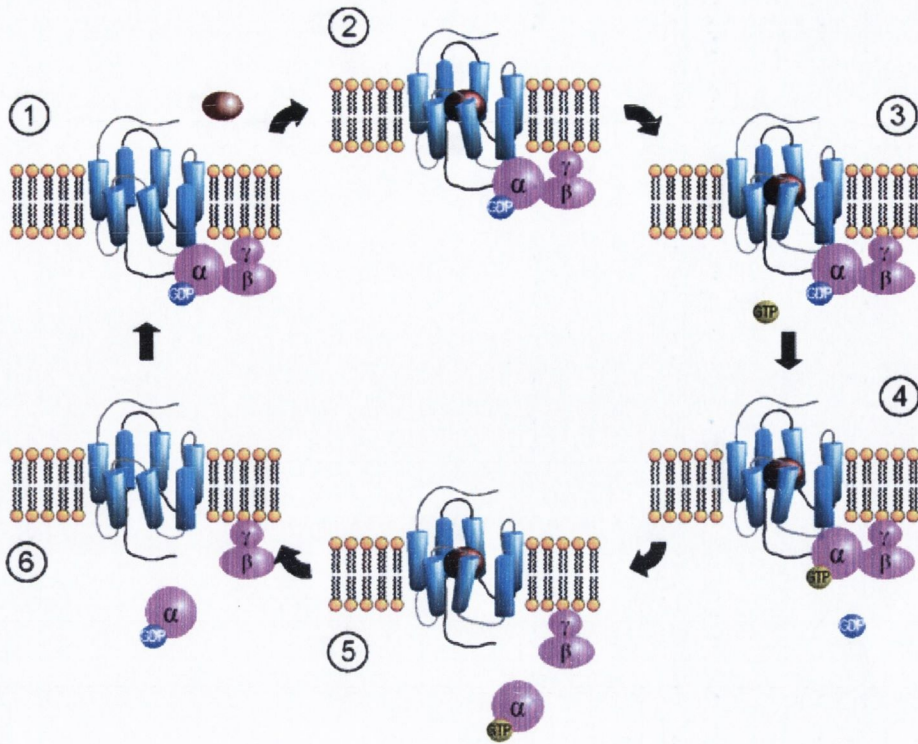
Figure 1.5: Activation of a GPCR (taken from wikipedia.de)

The inactive structure of the GPCR rhodopsin has been determined by crystallisation. However, the structural changes that need to occur to obtain an active conformation and the active conformation itself are unknown. There is an indication that after the ligand binds to the receptor, the protein undergoes a linear rearrangement, adopting a β-strand conformation[32].

The structure of rhodopsin does not include the attachment of the G-protein. It is suggested that the G-protein is attached to the highly conserved DRY motif in helix three and the highly conserved NP—Y—ERD motif in helix eight.[33,34]

1.3.2. Ternary complex model

The GPCR can exist in an inactive state (R_I) and an active state (R_A). The G proteins which mediate the physiological response, can only bind to the active-state R_A. This means that the receptor can be in three different states, the inactive state, the active state and the active state with the g-protein being bound[35]. A ligand can bind to all of these three forms and form three corresponding ligand bound species with different preferences to form the active state and different affinities for G proteins. This is known as the ternary complex model and a general scheme is shown in Fig. 1.6.

$$AR_I \longleftrightarrow AR_A \longleftrightarrow AR_A G$$
$$\updownarrow \qquad\qquad \updownarrow \qquad\qquad \updownarrow$$
$$R_I \longleftrightarrow R_A \longleftrightarrow R_A G$$
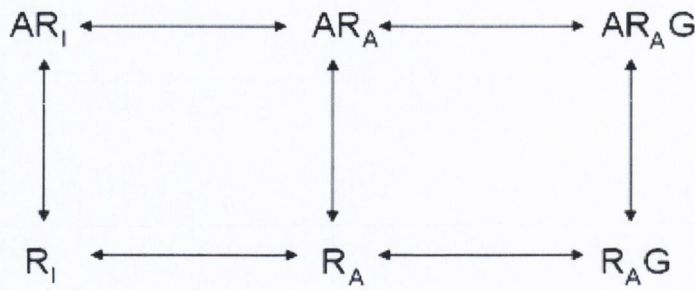
Figure 1.6: The extended ternary complex model showing the inactive state (R_I), the active state (R_A), the ligand (A) and the G-protein (G).

A thermodynamically more complete but more complex model is known as the cubic ternary complex model[36] (Fig 1.7). In this model the inactive state of the receptor can also form a complex with the G protein that does not signal. Still, this model can exchange GDP for GTP and activate the physiological response without a ligand being bound, an occurrence known as constitutive activity.
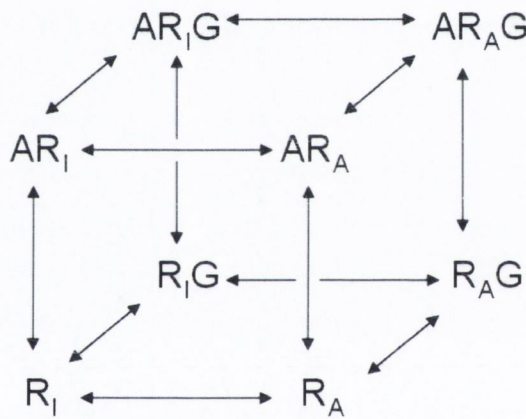
Figure 1.7: the cubic ternary complex model showing the inactive state (R_I), the active state (R_A), the ligand (A) and the G-protein (G).

1.3.3. Receptor activity modifying proteins

The receptor activity modifying proteins (RAMPs) are single transmembrane proteins that heterodimerize with GPCRs[37]. RAMPs have an influence in obtaining the appropriate cellular localization and can influence the function of a GPCR. For most GPCRs, the function of RAMPs in not known, but it has been studied extensively for the calcitonin-like receptor (CLR). There have been three different RAMPs found for the CLR. The interaction of these three different RAMPs with CLR gives rise to receptors for different ligands. This creation of diversity in receptor interaction has been confirmed by the finding that four ligands could interact with two GPCRs and three RAMPs differentially to yield seven different receptor phenotypes.[38] RAMPs interact with CLR at the endoplasmic reticulum

(ER) of Golgi and then stay together during the life cycle. This has lead to the belief that RAMPs can be therapeutically targeted to influence the behaviour of GPCRs.

Due to the clear influence of RAMP for the CLR, it is likely that RAMPs interact with a whole series of GPCRs and are a factor that should be taken into account. However, it is not known which GPRCs have an interaction with a RAMP and how strong the influence of the RAMP on the function of the GPCR is.

1.3.4. Desensitization

Desensitization is the decrease in the signal of a GPCR after a long exposure to an agonist. GPCRs are regulated by a common desensitizing mechanism. This underlying mechanism involves two families of proteins, the GPCR kinases (GRKs) and the arrestins.[39] GRKs bind specifically to GPCRs and phosphorylate agonist-occupied receptor systems. Arrestins bind to phosphorylated receptors and prevent their coupling to G-proteins.

GRKs have been linked to rhodopsin and the β_2-adrenergic receptor, leading to the belief that there must be more widely distributed members of the GRK and arrestin families which can be tissue or receptor specific.

After the activation of the GPCR, GRKs are recruited to the ligand-bound GPCRs and phosphorylate residues within both the intracellular loops and cytoplasmic tail of the GPCR. Ligand activated phosphorylated receptors are targeted to clathrin-coated pits using β-arrestins. By binding to the small G protein β-arrestins accelerate the exchange of GDP for GTP. The receptors are internalized when these pits are 'pinched off' from the plasma membrane and placed into coated vesicles.

1.3.5. Oligomerization

Recently, it has become evident that GPCRs can exist as either dimers or oligomers and this could be related to their function.[40] The addition of either an agonist or antagonist can not alter these interactions. It has also been shown that the α_{1A}-AR can form intracellular dimers/oligomers suggesting that this receptor is a constitutive dimer/oligomer at different stages in its life history.[41] The interaction points to form a

dimer are not known. Some points of interaction have been proposed, but they are widespread, and provide little understanding of how dimers are formed.

1.3.6. Rhodopsin

Before the structure of rhodopsin became available, there were no crystallographic 3D structures of any GPCR. The first structure of a seven transmembrane helix structure that was resolved was that of bacteriorhodopsin[42-46] (pdb: 1AP9). This protein shares the similarity of all GPCRs that it has seven transmembrane helices, but lacks the attachment of the G-protein and most of the residues that are highly conserved among GPCRs. However, due to the structure and positioning of the transmembrane helices it is useful in homology modelling.

Until very recently, the only known crystallographic structure that have been determined for this class of proteins wat that of bovine rhodopsin (Fig 1.8). Several crystal structures have been published with different resolutions.

- 1F88, 2.8 Å, June 2000[47]
- 1L9H, 2.60 Å, March 2002[48]
- 1GZM, 2,65 Å, May 2002[33]
- 1U19, 2.2 Å, July 2004[49]
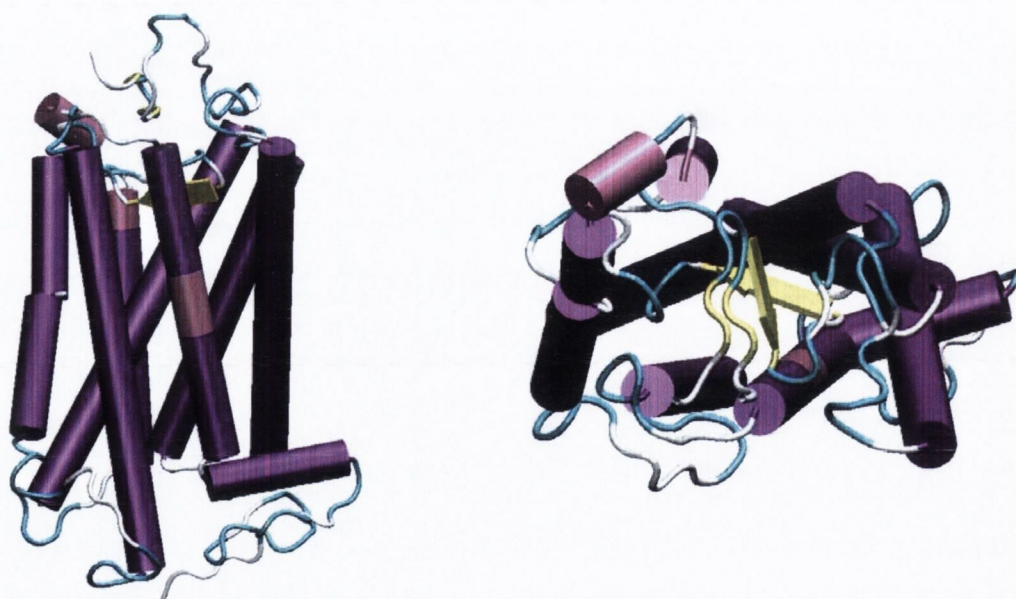
Figure 1.8: Crystal structure of rhodopsin viewed from the side (left) and the top (right). In the side view, the bottom part of the model reaches into the intracellular part and the top part into the extracellular compartments.

The structure of rhodopsin consists of a bundle of seven transmembrane helices that surround the photoreactive chromophore, 11-cis retinal. Retinal is made in the retina from Vitamin A. Isomerization of 11-*cis*-retinal into all-*trans*-retinal by light induces a conformational change in the opsin that activates the associated G protein and triggers a second messenger cascade.

One difference between rhodopsin and other GPCRs is that rhodopsin acts via a photo-induced conformational change, while other GPCRs act via binding of an agonist. However, it is believed that this activation mechanism involves similar principles.[50] Another issue to consider is the fact that despite a ligand being bound, the rhodopsin receptor is still in an inactive state, whereas the active state is more interesting for research purposes.

While this thesis was being written the crystal structure of the β_2-AR has been published[51] which will affect the research in the area of GPCRs in the future.

1.4. Current AR computational models

Several papers have been published that model adrenoceptors or ligands binding to these receptors. These models have been generated using a variety of techniques. One possible approache is the development of a pharmacophore model. A pharmacophore is the minimal structural requirement for activity. As pharmacophores do not require 3D models of the receptor, but are based on a set of ligands, they can be easily generated because there are known ligands with binding affinities and selectivity for each of the adrenoceptor subtypes. The development of QSAR models is a different approach and do not necessarily need a 3D model either. However in the case of 3D-QSAR a 3D model of the adrenoceptor can be used. Another approach is the development of 3D structural models of the receptors. Usually these consist of homology models with a ligand data set used in docking experiments. The resulting ligand-protein complexes can then be used for further study of protein-ligand interactions using MD simulations for the understanding of the activity of know drugs or the design of new ligands.

1.4.1. Pharmacophore models

A pharmacophore is an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response. Usually they describe part of the molecule as

hydrophobic or aromatic and identify hydrogen bond donors or acceptors and ionisable groups. Pharmacophores can be used to select compounds from a database for further screening. As pharmacophores attempt to describe interactions that are needed for a ligand to interact with a receptor, not all compounds that fit this pharmacophore actually bind to the receptor. Although pharmacophore models can be validated by comparing them to activity, they usually do not correlate very well. Still, they can be used to describe essential features of ligands, to show activity and are a usefull tool to search databases. There have been several pharmacophore models published for the α_1-adrenoceptor, but few of them are selective for the α_{1A}, α_{1B} and α_{1D} subtypes. This is due to the few compounds that are known to be subtype specific. Thus, these pharmacophore models are based on sets of ligands with limited structural variety limiting the quality of the generated pharmacophores and therefore limiting the utility of these pharmacophores for the development of structurally similar ligands.

1.4.1.1. General α_1-AR pharmacophore models.

A pharmacophore model for α_1-AR is given by Barbaro[52] et al based on pyrizadinone inhibitors. This model however is not α_1-AR subtype specific (Fig. 1.10, Table 1.4). The model they derive using the program Catalyst makes use of a positive ionisable nitrogen (PI), hydrophobic regions (HY) and hydrogen bond acceptors (HBA). Despite the lack of subtype specificity there are some requirements derived from this pharmacophore model that should be considered as essential for good α_1-AR antagonists:

1. A basic, positive ionisable nitrogen (PI), Fig. 1.10) accessible to the receptor and easily protonable at physiological pH is required. Although not part of the pharmacophore, it is thought that this nitrogen interacts with TM-III.
2. The phenyl ring occupies both the hydrophobic regions HY1 and HY2 features (Fig. 1.10) of the pharmacophore. This suggests that both HY1 and HY2 might constitute a unique and large hydrophobic pocket. HY3 is considered an additional pharmacophore feature, but not a necessity.
3. A polar group at the other end of the molecule relative to the arylpiperazine moiety corresponds to a hydrogen bond acceptor (HBA, Fig. 1.10).
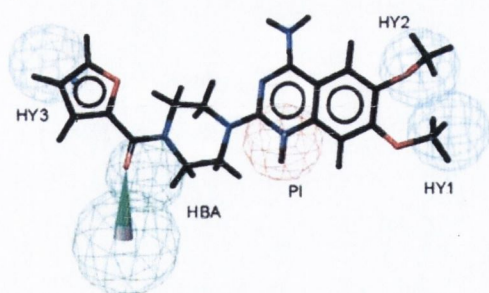
Figure 1.10: Pharmacophore model of the α_1-AR.[52] HBA is the hydrogen bond acceptor (green), HY is a hydrophobic region (blue) and PI is a positive ionisable feature (red).

Table 1.4: Pharmacophore model for the α_1-AR.

Feature	Distance (Å)
PI-HY1	6.69
PI-HY2	6.17
PI-HBA	5.62
PI-HY3	9.78

1.4.1.2. α_{1A}-AR pharmacophore

The specific α_{1A} pharmacophore hypothesis of Bremner et al[53] using Catalyst for antagonists consists of three features including a positive charge in the middle of the system (Fig 1.11, Table 1.5), a hydrogen bond acceptor group and an aromatic ring system at opposite ends of the molecules. This pharmacophore model is not very good for compounds with weak α_{1A} affinity. To design this pharmacophore model, four selective α_{1A}-AR antagonists have been used. The correlation coefficient obtained was 0.35 which is described as very poor.
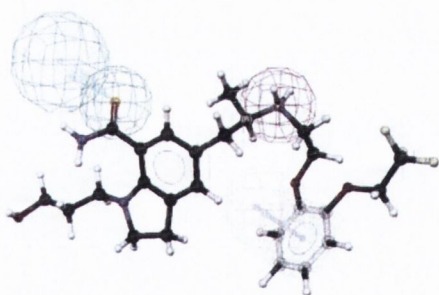
Figure 1.11: Pharmacophore by Bremner *et al*[53] of α_{1A}-AR with KMD-3213 mapped onto it (C=black, H=white, O=red, N=blue, F=green, red=positive ion (P), green=hydrogen bond acceptor (HBA), light brown=aromatic ring system (A))

Table 1.5: Features of the α_{1A} pharmacophore by Bremner et al. (P=positive charge, HBA=hydrogen bond acceptor, A=Aromatic ring centre)

Feature	Angle
P-A	5.5°
P-HBA	7.1°
A-P-HBA	100°

Another α_{1A} pharmacophore hypothesis has been published by MacDougall et al[54] which is very limited as it only states properties that have a high significance for the ligand-protein interaction (Fig. 1.12). To design this pharmacophore model, a training set of 27 α_1-AR antagonists has been used and four different features have been derived (Table 1.6) using Catalyst. Four features were derived being a hydrogen bond acceptor (HBA), hydrophobic group (Hal), Aromatic hydrophobic group (Har) and a positive ionisable

group (PI). Unfortunately, no distances between the groups are published. The correlation coefficient that was obtained with the training set was 0.95 which is very good.
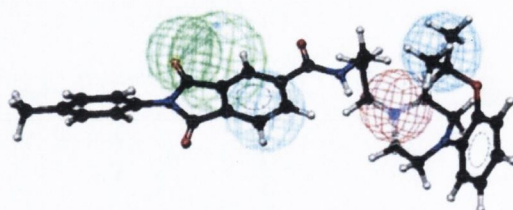
Figure 1.12: Pharmacophore by MacDougall *et al*[54] for the α_{1A}-AR. (red=positive ion, blue is hydrophobic part, green=hydrogen bond acceptor)

Another α_{1A}-AR pharmacophore was developed by Li et al[55] using the program Catalyst. It is based on 30 compounds in the training set and 15 compounds in the validation set (Fig. 1.13, Table 1.6). The resulting pharmacophore consists of a hydrogen bond acceptor, a hydrogen bond donor, a hydrophobic group, a positive charged group and an aromatic ring centre. They found a correlation of 0.94 with the training set and a correlation of 0.84 with the validation set which can be described as good.

a)

b)
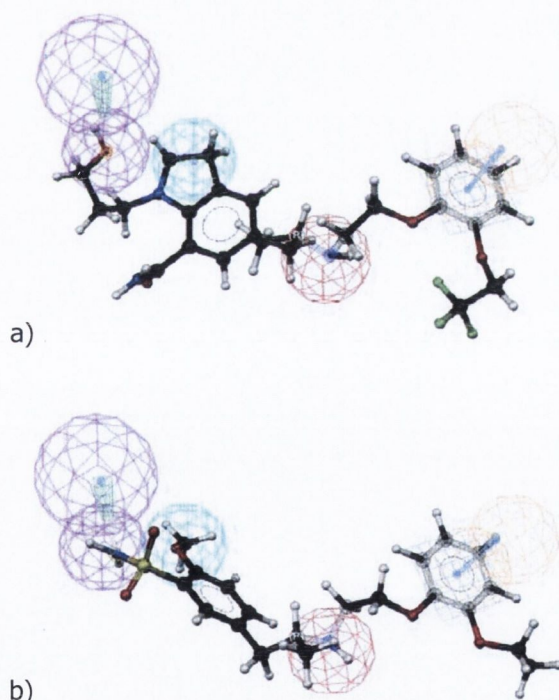
Figure 1.13: The pharmacophore of Li et al with Silodosin (a) and Tamsulosin (b) aligned.[55] Red is a positive ionisable group (PI), orange is an aromatic ring (R), blue is a hydrophobic part (H) and violet is a hydrogen bond donor (HBD).

Table 1.6: Features of the α_{1A}-AR pharmacophore by Li et al.[55] HBA is a hydrogen-bond acceptor, HBD is a hydrogen bond donor, H is a hydrophobic group, PI is a positive charge and R is an aromatic ring center.

Property	Distance or angle
PI-R	5.82 Å
P-HBD	9.08 Å
PI-H	7.53 Å
R-PI-HBD	125°
HBD-R	13.27 Å
HBD-H	3.95 Å
H-R	10.87 Å

Three different pharmacophore models have been described here for the α_{1A}-AR which all use similar features such as hydrogen bond donor/acceptor, positive charged ion and an

aromatic ring centre. The correlation of these models with their training set was usually very good, but correlation with the validation set ranges from poor to good. This suggests that the design of a pharmacophore model based on a set of known compounds is feasible, but applying this model to a larger database is not necessary valid. Therefore, pharmacophore models based on larger dataset should be used and care should be taken when applying a model based on a limited data set.

1.4.1.3. α_{1B}-AR pharmacophore

The α_{1B} pharmacophore hypothesis of Bremner et al[53] consists of four features including a hydrogen bond donor group (HBD) and a hydrophobic group (H) as well as an aromatic ring system (A) and a positive charge (P) (Fig. 1.14, Table 1.7). To design the pharmacophore model three selective α_{1B}-AR antagonists have been used. The obtained correlation coefficient was 1.00 but this number is due to the number of compounds.

Table 1.7: Features of the α_{1B}-AR pharmacophore by Bremner et al. (P=positive charge, A=Aromatic ring centre, H=Hydrophobic group, HBD=Hydrogen bond) donor)

Feature	Angle
P-A	6.2°
P-H	7.8°
P-HBD	4.9°
H-P-HBD	57°
A-P-HBD	52°
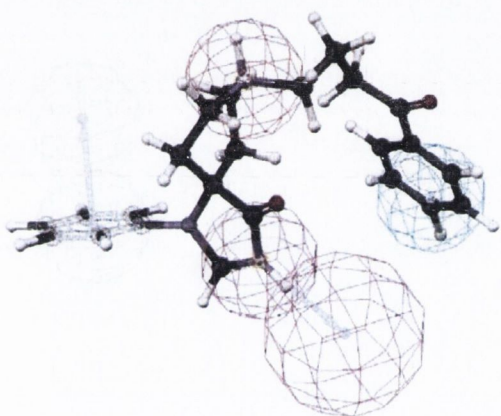
Figure 1.14: Pharmacophore by Bremner *et al*[53] of the α_{1B}-AR with KMD-3213 mapped onto it (C=black, H=white, O=red, N=blue, F=green, purple = hydrogen bond donor (HBD), blue = hydrophobic group (H), red=positive ion (P), light brown=aromatic ring system (A))

Another α_{1B} pharmacophore hypothesis has been published by MacDougall et al[54] which is very basic as it only states properties that have a high significance for the ligand-protein interaction (Fig. 1.15). To design the pharmacophore model 21 α_1-AR antagonists have been used and the pharmacophore was derived using Catalyst. The correlation coefficient obtained, is 0.95 within the training set.
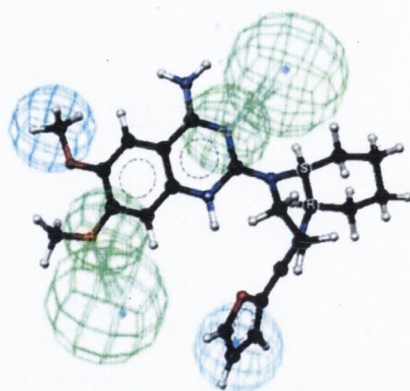
Figure 1.15: Pharmacophore by MacDougall et al[54] for the α_{1B}-AR.
(blue is hydrophobic part, green=hydrogen bond acceptor)

The pharmachophore model described first, is not good because too few ligands have been used for designing this model. The model described second has used more compounds but its weakness is in the number of descriptors that are used which are only a hydrogen bond acceptor and two hydrophobic regions. There is no mention of positive ionisable groups which are known to play a role in binding to adrenoceptors.

1.4.1.4. α_{1D}-AR pharmacophore

The α_{1D} pharmacophore hypothesis of Bremner et al[53] consists of a positive ion feature and an aromatic ring. It also included a hydrophobic group and a hydrogen bond acceptor. However, they only used three compounds for making their pharmacophore model and produced little predictive values for the ligands in their validation set (Fig. 1.16, Table 1.8). To design the pharmacophore model, only four selective α_{1A}-AR antagonists have been used. The correlation coefficient is 0.14 which is described as no correlation and limits the applicability of this pharmacophore model.
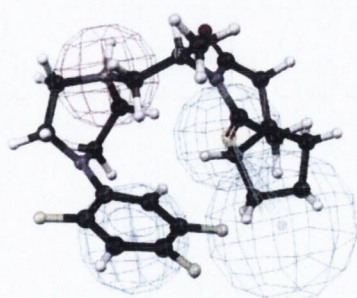
Figure 1.16: Pharmacophore by Bremner et al[53] of α_{1D}-AR with KMD-3213 mapped onto it (C=black, H=white, O=red, N=blue, F=green red=positive ion (P), green=hydrogen bond acceptor (HBA), blue = hydrophobic group (H)).

Table 1.8: Features of the α_{1D} pharmacophore by Bremner et al. (P=positive charge, HBA=hydrogen bond acceptor, H=Hydrophobic group)

Feature	Angle
P-H	5.4°
P-HBA	4.5°
H-P-HBA	47°

Another α_{1D}-AR pharmacophore hypothesis has been published by MacDougall et al[54] which is limited as it only states properties that have a high significance for the ligand-protein interaction (Fig. 1.17). To design the pharmacophore model, twenty α_{1D}-AR antagonists have been used and the pharmacophore was derived using Catalyst. The correlation coefficient is 0.91 with the training set which is described as very good.
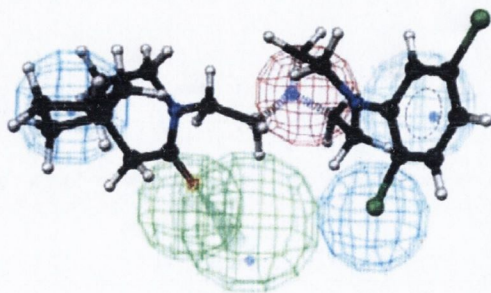
Figure 1.17: Pharmacophore by MacDougall[54] et al for the α_{1B}-AR. (red=positive ion, blue is hydrophobic part, green=hydrogen bond acceptor)

Another α_{1D} pharmacophore hypothesis has been published by Romeo *et al*.[56] This model gives an overview of the location of the centres of the different features of the pharmacophore and the positions of these features relative to each other (Fig. 1.18, Table 1.9). To design the pharmacophore model sixteen α_{1D}-AR antagonists have been used. The correlation coefficient is 0.91 within the training set and is described as very good.
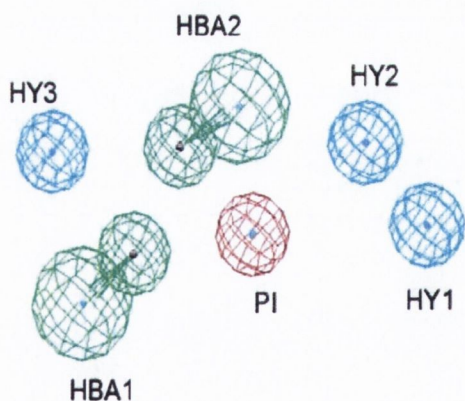
Figure 1.18: Pharmacophore by Romeo *et al*[56] of the α_{1D}-AR. HY is a hydrophobic group (blue), HBA is a hydrogen bond acceptor (green) and PI is a positive ionisable group (blue).

Table 1.9: Features of the α_{1D} pharmacophore by Romeo *et al*. (HY = hydrophobic group, PI = positibe ionisable group, HBA = hydrogen bond acceptor).

Feature	Distance/angle
HY1-HY2	4.1 Å
HY1-PI	7.5 Å
HY1-HBA2	11.30 Å
HY1-HBA1	12.6 Å
HY1-HY3	16.5 Å
HY2-PI	5.9 Å
HY2-HBA2	8.0 Å
HY2-HBA1	10.8 Å
HY2-HY3	13.7 Å
PI-HBA2	4.6 Å
PI-HBA1	5.2 Å
PI-HY3	9.4 Å
HBA2-HBA1	4.5 Å
HBA2-HY3	5.7 Å
HBA1-HY3	5.5 Å
HY1-HY2-PI	94.6°
HY2-PI-HBA2	97.8°
PI-HBA2-HBA1	69.8°
HBA2-HBA1-HY3	62.3°
HY2-PI-HY3	124.4°
HY2-HBA2-HBA1	117.4°

Three α_{1D}-AR pharmacophore models have been presented of which the first model by Bremner is probably the worst because of the use of only four compounds for designing the model. The other two use more compounds and show good correlation with the training although no validation set was used. However, the two models do show similarity with the same features being used and a similar positioning of the positive ionisable group, the hydrogen bond acceptor and two hydrophobic groups. The model by Romeo *et al*, mentions an extra hydrogen bond acceptor which is probably due to the selection of compounds used to generate the pharmacophore model.

1.4.2. Quantitative structure activity relationship models

Several (3D) quantitative structure activity relationships (QSAR) studies have been performed on adrenoceptor ligands. QSAR studies allow the prediction of an activity based on a set of properties of the ligands. 3D-QSAR extends these properties in space, such as hydrophobicity and electrostatic fields. These field can then be used to select favourable and non-favourable fields for functional groups. Most of the studies focus on one type of compounds using the same scaffold structure and adding different functional groups to this scaffold. Several papers describe 3D-QSARs for the α_1-AR[57,58], but there are limited papers published which have developed α_1-AR subtype selective QSAR models[59-61]. Due to our interest in subtype selectivity only the papers which have developed an α_1-AR subtype selective QSAR model are commented upon.

1.4.2.1. 3D-QSAR on sertindole-like antagonists

A 3D-QSAR study has been performed on sertindole (Fig. 1.19) and sertindole-like antagonists by Balle *et al*[59].
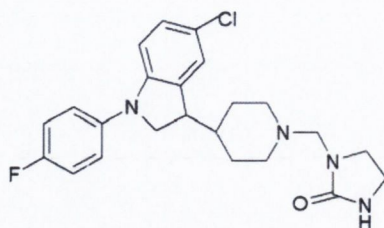
Figure 1.19: Structure of Sertindole.

The 3D-QSAR models have been designed with the use of 2 probes that interact with the ligand on each point of a 3D-grid. The probes that are used are methyl to describe the steric fields and water to show the electrostatic interaction. The resulting interaction fields are shown in the form of 3D contour maps. The C3 contour (Fig. 1.20) maps shows that

the binding pocket of α_{1A}-AR is able to accommodate larger substituents in the area corresponding to the indole 6-position compared to the other two receptor subtypes. An unfavourable interaction (positive interaction energy, i.e., steric repulsion) between a substituent in the molecule and the probe in areas with negative coefficients is predicted to lead to increased affinity (increased pK_i value). The areas of pK_i positive steric interactions occupy similar positions in space, and the size of these areas increases from α_{1A}-AR over α_{1D}-AR to α_{1B}-AR.

a) α_{1A}-AR contour map

b) α_{1B}-AR contour map

c) α_{1D}-AR contour map

d) α_1-AR contour map
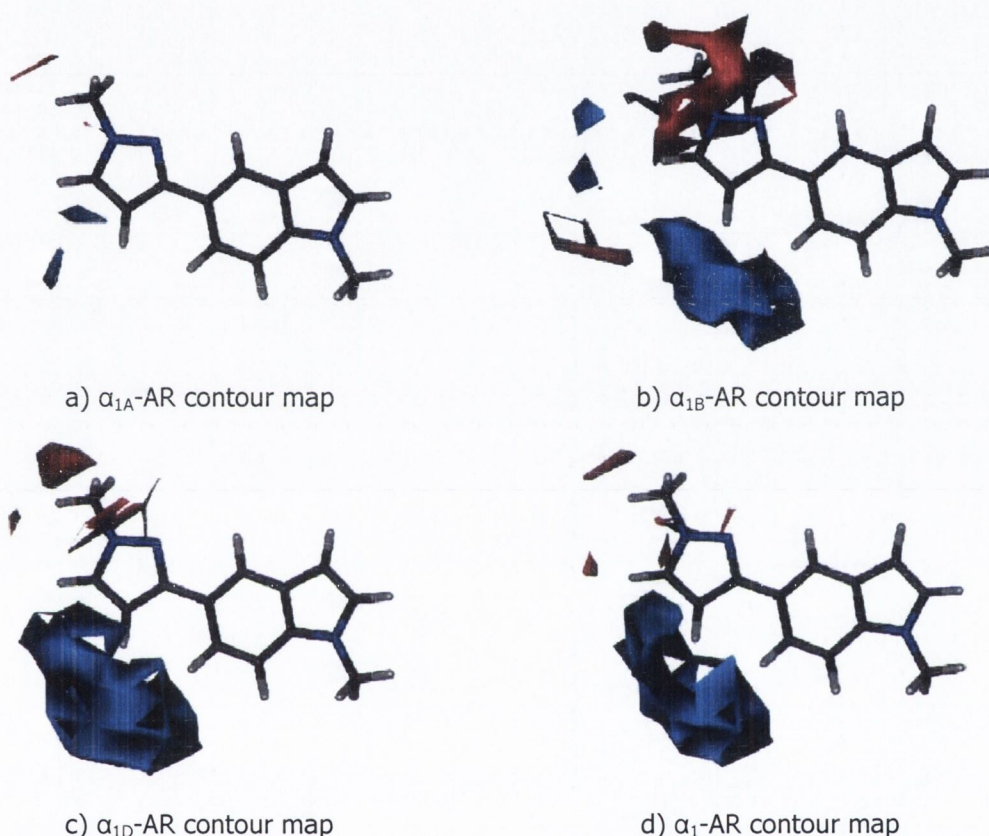
Figure. 1.20: Contour maps around the scaffold structure of sertindole-like compounds for the different α_1-AR subtypes[59] showing positive coefficients(red) and negative coefficients (blue) for the steric field.

The electrostatic contour maps (Fig. 1.21) shows that the distinct areas of negative coefficients (blue) must represent specific electrostatic interactions with the hydrogen bond donating component of the probe. A favourable interaction (negative interaction energy, i.e., electrostatic attraction) between a possible substituent in the molecule and the probe in these areas is predicted to lead to increased affinity (increased pK_i value). The 'ortho' nitrogens of the pyrazoles are involved in binding to α_{1A}-AR and α_{1D}-AR receptors whereas nitrogens in the 'meta' position are involved in binding to α_{1B}-AR receptor.

a) α_{1A}-AR contour map   b) α_{1B}-AR contour map

c) α_{1D}-AR contour map   d) α_1-AR contour map
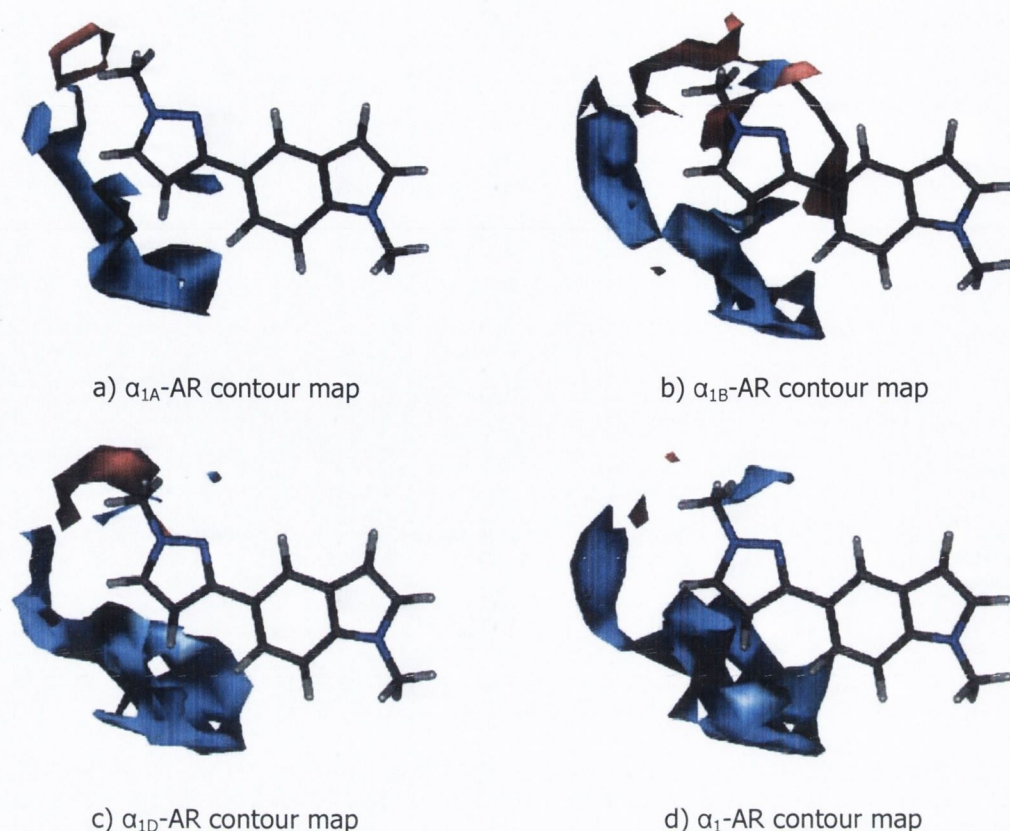
Figure 1.21: Contour maps around the scaffold structure of sertindole-like compounds for the different α_1-AR subtypes[59] showing positive coefficients(red) and negative coefficients (blue) for the electrostatic field.

1.4.2.2. 3D-QSAR on dihydropyridine antagonists.

A 3D-QSAR study has been performed on dihydropyridine derived antagonists for the α_1-AR by Li *et al*[60] using the program SOMFA. They found that S-(+)-Niguldipine (Fig. 1.22) has a higher affinity for the α_{1A}-AR than the other α_1-AR subtypes. Therefore, this class of dihydropyridine derivates (Fig. 1.22) could be used as a lead to find new antagonists which are selective for α_{1A}-AR antagonists.
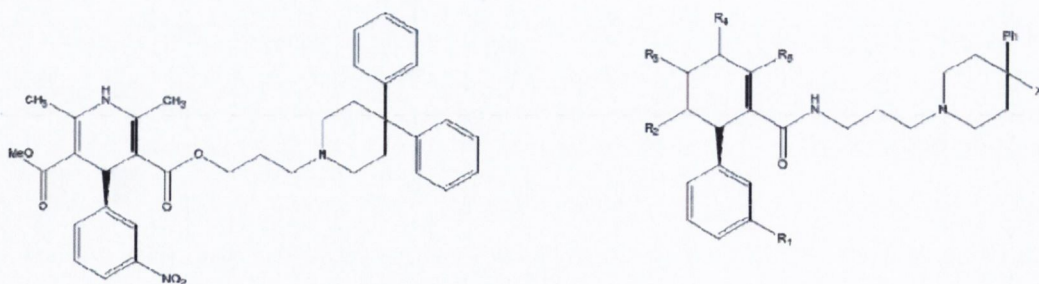
Figure. 1.22: Structure of S-(+)-Niguldipine (left) and the scaffold (right) used for 3D-QSAR compound comparison

The 3D-QSAR models were produced by making use of two descriptors, electrostatic potential and shape, which leads to two different plots. Their results indicated that the electrostatic contribution is of high importance (Fig. 1.23). A preference for electropositive groups was observed around the substituent R3 and R4 at the dihydopyridine ring and around the *ortho* and *meta* position at the phenyl ring. A preference for electronegative groups was observed around the position at the phenyl ring, and around the substituent R2 at the dihydropyridine ring.

Figure. 1.23: The electrostatic potential master grid around S-(+)-Niguldipine from two different viewpoints. Red represents a favourable positive potential of a unfavourable negative charge. Blue represents a favourable negative potential or an unfavourable positive charge.[60]

In the map corresponding to steric factors (Fig. 1.24), they found a preference for a favourable steric interaction around the substituent R2 at the dihydropyridine ring, and around the postion at the phenyl ring. Simultaneously, an unfavourable steric interaction was found around the substituent R3 and R4 at the dihydropyridine ring, and around the ortho and the meta position at the phenyl ring. In the area of substituent X neighbouring to a phenyl ring, steric interactions may be expected to enhance activities.
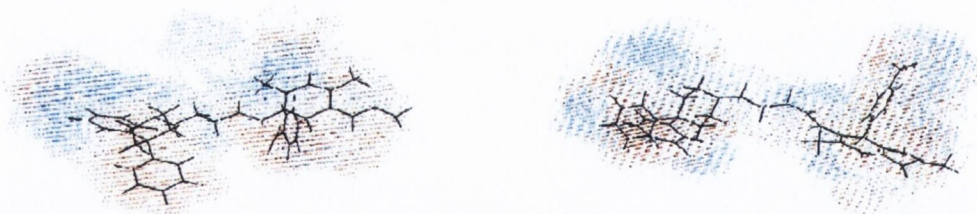
Figure. 1.24: The shape master grid around S-(+)-Niguldipine from two different viewpoints. Red represents areas of favourable steric interaction, blue represents areas of unfavourable steric interaction[60].

1.4.3. Docking/MD models

Several models have been developed for the α_1-AR subtypes that make use of homology modelling, docking, scoring and MD simulations. Some of these models were developed before the crystal structure of rhodopsin was determined and the crystal structure of bacteriodopsin was used. Bacteriodopsin shares the same 7 transmembrane helices as in GPCRs, but does not belong to the family of GPCRs. We would expect the models based on rhodopsin to be more accurate. An overview of the models that have been generated for the different α_1-AR subtypes is given next.

1.4.3.1. Alpha 1A models

Pedretti *et al*[62] designed a homology model for the α_{1A}-AR based on bovine rhodopsin and docked noradrenaline and WB-4101 into this receptor. Analysis of the noradrenaline-AR complex showed that Asp^{106} has a strong electrostatic interaction with the ammonium group of noradrenaline. Ser^{188} and Ser^{192} form a H-bond with both hydroxyl groups of the cathecol ring in which also Ser^{158} takes part. Ser^{192} interacts with the *para*-hydroxyl group, while Ser^{188} interacts with the *meta*-hydroxyl group. Gln^{177} forms H-bonds with the α-hydroxyl group. In this model noradrenaline is located in a site lined by two hydrophobic regions being Phe^{187} and Phe^{193} interacting with the cathecol ring and by aromatic residues in which Trp^{185} and Phe^{288} play key roles (Fig. 1.25).

Analysis of WB-4101 (Fig. 1.25) shows that Asp^{106} plays a key role forming strong electrostatic interaction with the ammonium group. The benzodoxane system forms a π-π interaction with Phe^{193}, while the phenyloxy ring is inserted in an aromatic pocked lined by Phe^{288}, phe^{289}, Phe^{308}, Phe^{312} and Tyr^{316}. Cys^{110} forms a hydrogen bond with the oxygen atom at position 1 of the benzodioxane moiety and Gln^{177} establishes a H-bond with a methoxyl group of the phenyloxy ring. An overview of the interactions of noradrenaline and WB-4101 is given in Fig. 1.26.
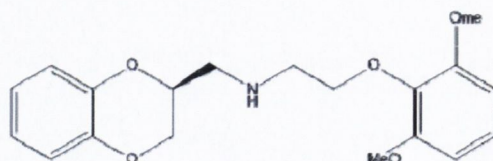
Figure 1.25: Structure of WB-4101
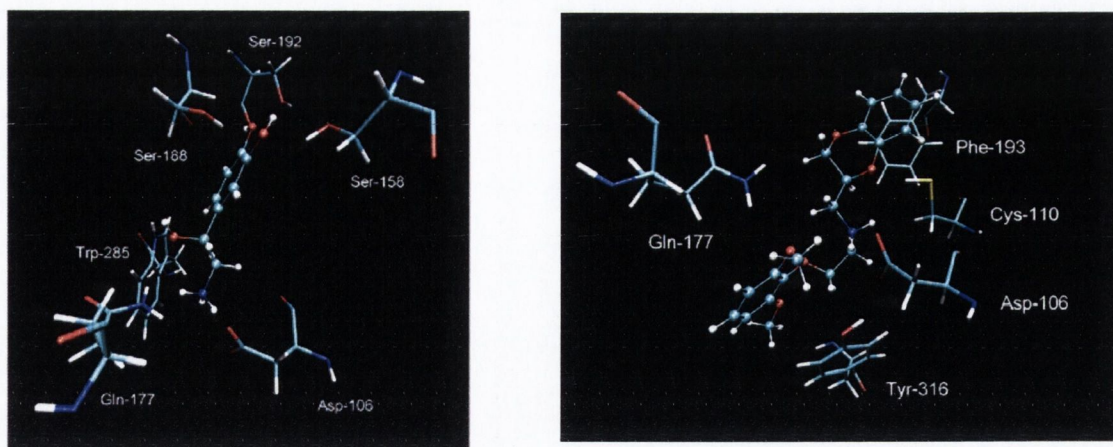
Figure 1.26: Interaction of the the α_{1A}-AR with norepinphrene (left) and WB-4101 (right) by pedretti *et al.*[62]

Leonardi *et al*[63] studied the positioning of BMY 7378 in the α_{1A}-AR and found that the Gln^{177} of the extracellular loop connecting TM-I and TM-II (ECL-II), which is directed towards the putative ligand binding pocket, exerts a steric hindrance and prevents most of the ligands from interacting with Asp^{106}. As a consequence, the protonated nitrogen atom of the two fragments has a preference for interaction with either Glu^{173} or Glu^{305}. A low interaction energy is characterized by the formation of a salt bridge between the positively charged nitrogen atom of the ligand and the Glu^{305} of the receptor.

Evers *et al*[64] performed successful docking of a ligand set of 23.000 compounds into the α_{1A}-AR. To generate the homology model based on bovine rhodopsin, they designed 100 different models with compound 1 docked into it.

compound	K$_i$[nM]	Structures
Table 1.10: Two of the ligands that Evers *et al* used for docking as shown in Fig. 1.26		
1		
2	3.6	

By using mutation data and QSAR studies they identified which residue of the binding pocket in each of the homology models has the best contact with the compound. These positions of the residues from the different homology models were then combined into a final model. This model was then further used via a high-throughput screen. They established that the residues that are involved in binding are mainly located in TM-II, TM-III, TM-V and TM-VI, TM-VII (Fig. 1.27) and their results showed that a high throughput screen with a homology model based on rhodopsin is possible. The docking of some of the different compounds (Table 1.10) is shown in Fig. 1.28.
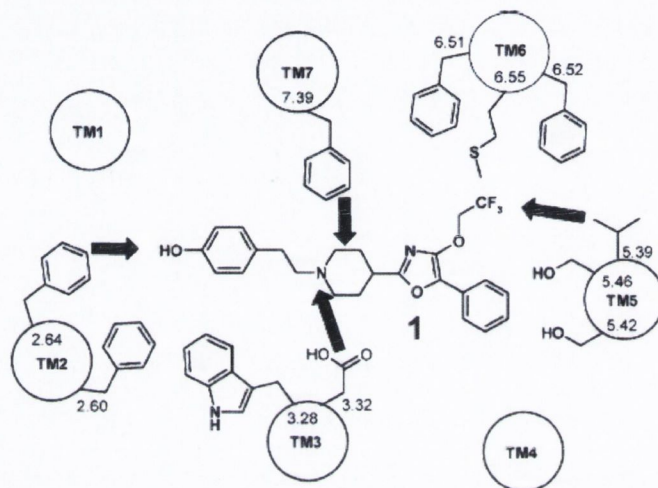
Figure. 1.27: Proposed binding pocket of the α_{1A}-AR with compound 1 docked by Evers et al[64]

a)   b)
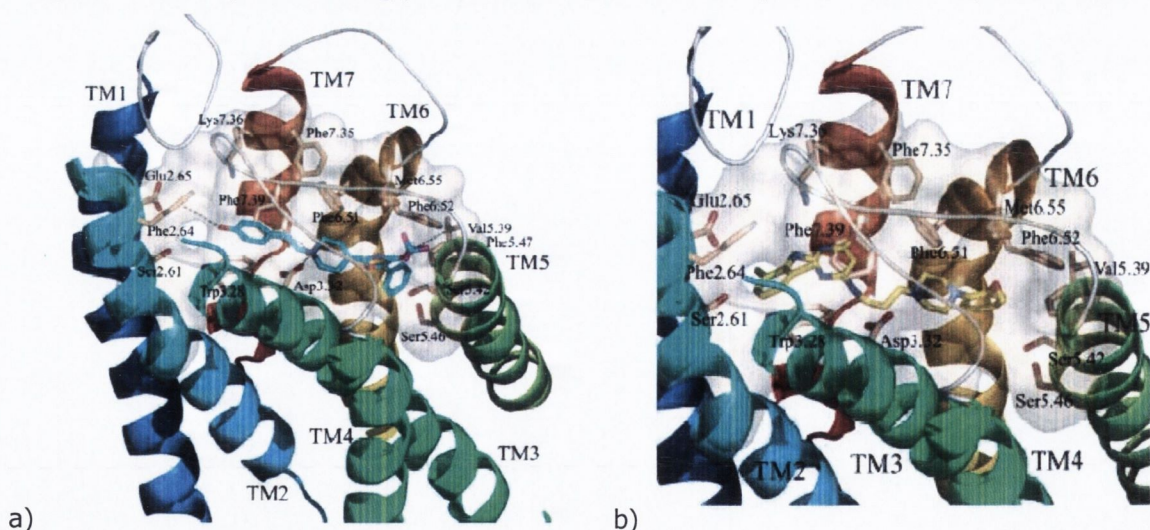
Figure. 1.28: Docking of different compounds into the α_{1A}-AR showing the residues that interact with compound 1 (a) and compound 2 (b) by Evers et al[64].

Our group[65] has previously studied the activation of the α_{1A}-AR by docking adrenaline and noradrenalin in the adrenoceptor followed by MD simulation. The interaction between the compound and the adrenoceptor should lead to an agonist-activated conformation. It was found that there may be be two new residues involved in binding of agonists being Thr^{174} and Cys^{176} which have not been described earlier in literature. A structural change was also observed, in which TM-V moves away form TM-III and increased flexibility about a Pro kink, which facilitates a movement of TM-VI. A change in the interactions of Asp^{124} of the conserved DRY motif could cause Arg^{124} to move out of the TM helical bundle and change the orientation of residues in ICL-I-II and IC-III-IV allowing for increased affinity of coupling to the G-protein. The position of adrenaline and noradrenaline in the α_{1A}-AR is given in Fig 1.29.
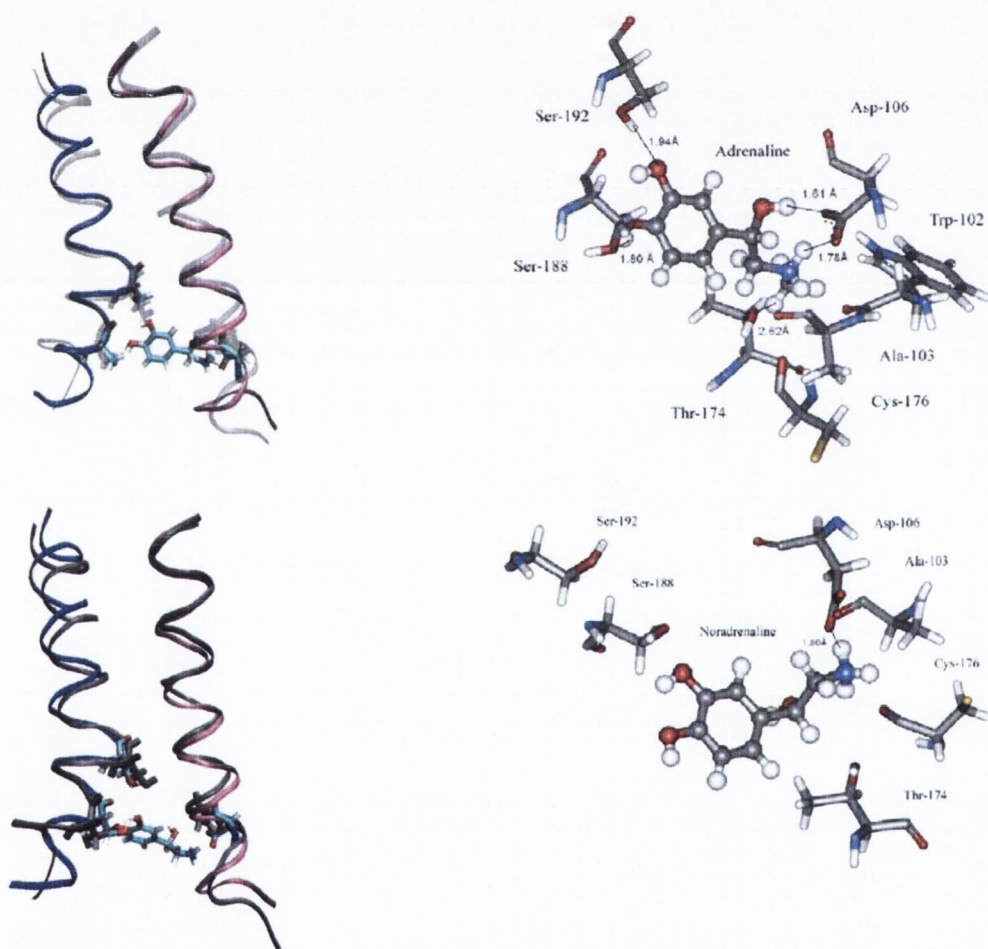
Figure. 1.29: Docking of adrenaline (top) and noradrenalin (bottom) in the α_{1A}-AR as predicted by Kinsella *et al.*[65]

Our group has earlier also studied the effect of docking antagonists in different positions in the α_{1A}-adrenoceptor. [66,67] It was found that different families of ligands induce different conformations of the adrenoceptors due to the size and interactions of the protein with the ligand. For this study, three antagonists were used, doxazosin,

tamsulosin and compound 3 (Fig. 1.30). These antagonists were docked into the α_{1A}-AR followed by a MD simulation of 1 ns. This gave rise to three different conformations of the adrenoceptor and can subsequently be used to explore three possible binding modes of each family of antagonists. These conformations were used to dock the ligands from our dataset, containing compound 4, 5, 6 (Fig. 1.30) into the adrenoceptor followed by a MD simulation of 0.5 ns to optimise the ligand-protein complex.

Dox 4

Tam 5

3 6
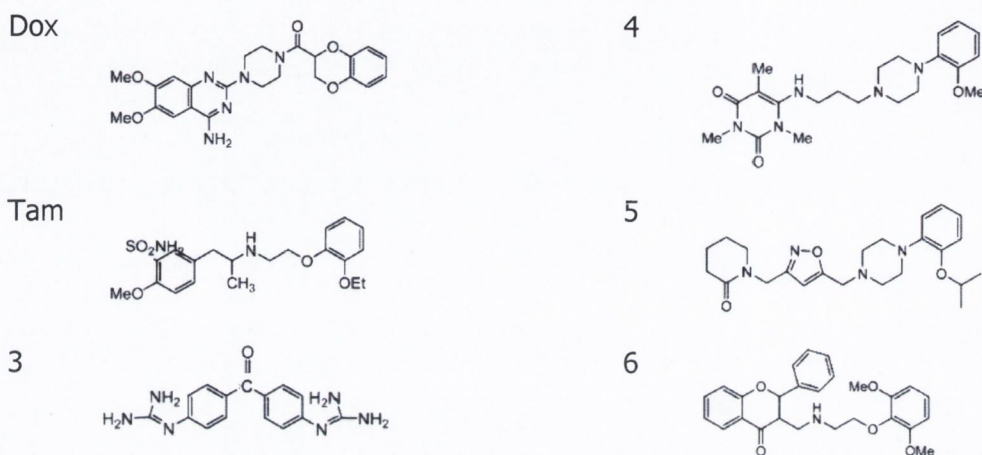
Figure 1.30: Ligands used by Kinsella et al[67] for doxasosin (dox), tamsulozin (tam) and compounds 3, 4, 5 and 6.

It was found that a number of ionic interactions with Ile[178], Asn[179], Glu[180] and Ser[188] occurred in TM-V. However, the binding site analysis showed that one class of antagonists is not directly suitable for screening of another structural class. To obtain novel antagonists there has to be a degree of flexibility in the receptor. An overview of the three ligand-protein complexes generated for docking of their ligand data set is given in Fig. 1.31. These complexes were subjected to a MD simulation giving rise to new conformations of the receptor with the binding pocket in three different positions being upper (doxazosin), middle (tamsulosin) and lower (compound 6). Each of the three conformations was used for docking of compound 4, 5 and 6. The interaction of three ligands with each of these conformations is shown in Fig. 1.32.

a) Doxazosin

b) Doxazosin

c) Tamsulosin

d) Tamsulosin

e) compound 3
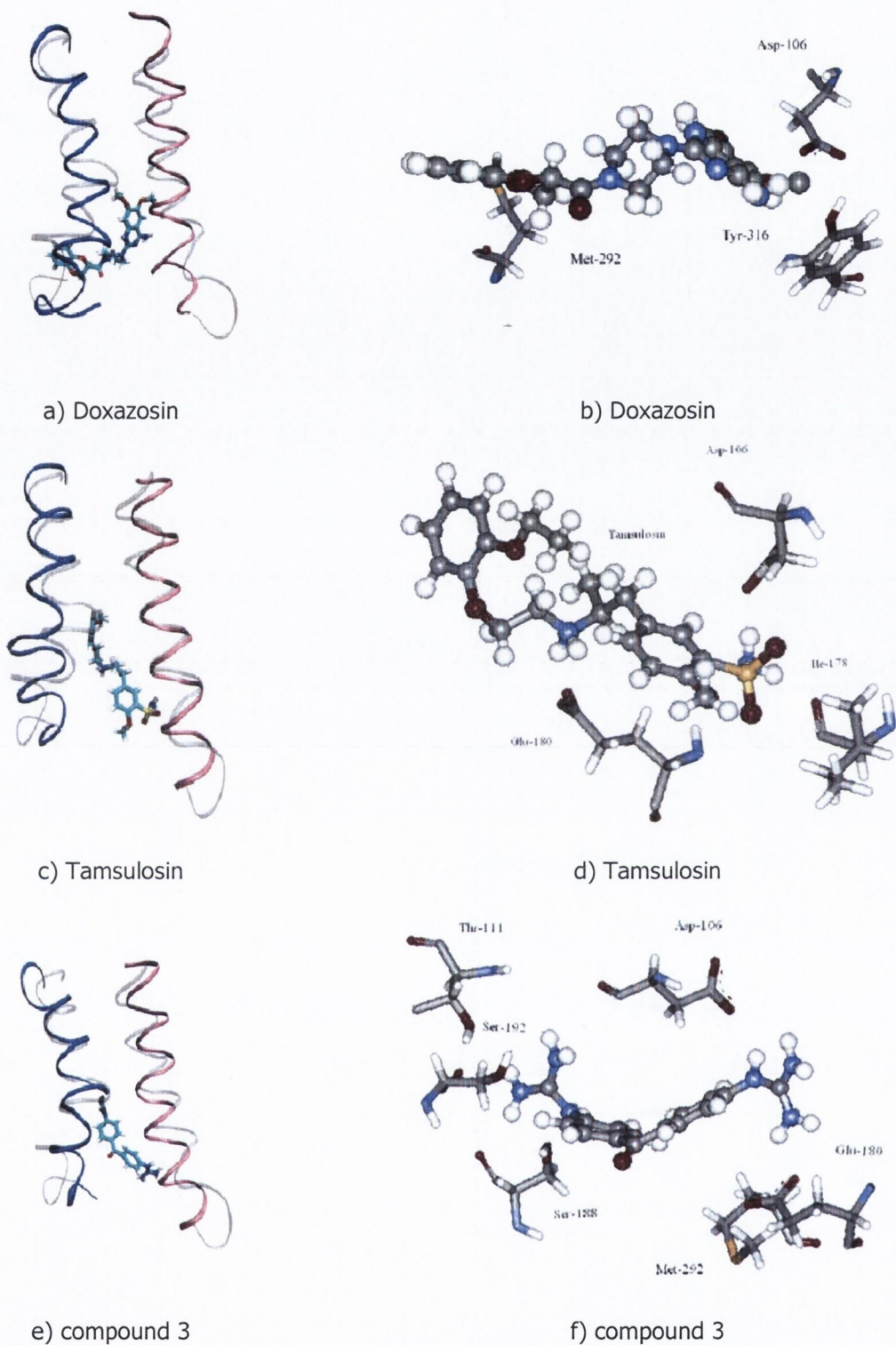
f) compound 3

Figure. 1.31: Docking of doxazosin (a+b) , tamsulosin (c+d) and compound 3 (e+f) in the α_{1A}-AR.[67]
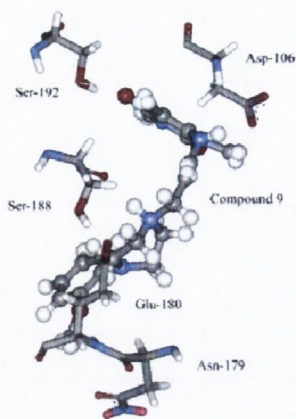
a) Compound 4, upper
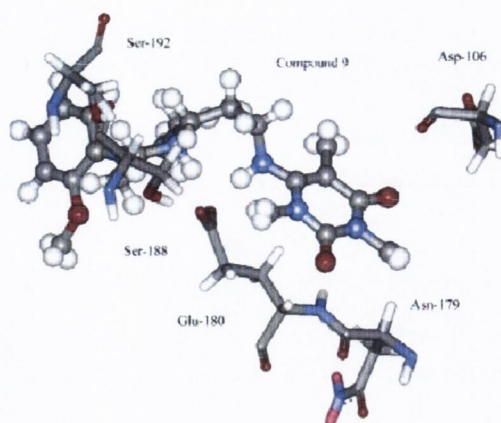
b) Compound 4, middle
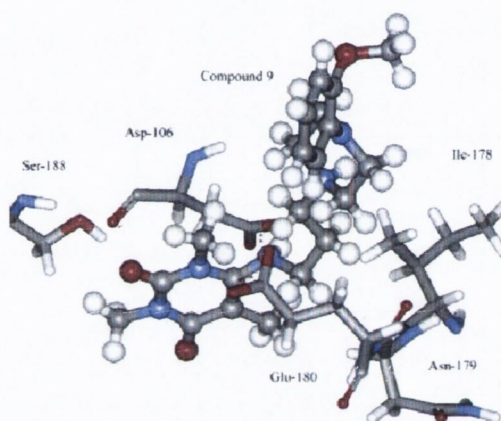
c) Compound 4, lower
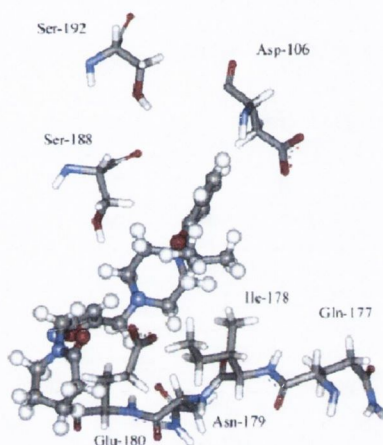
d) Compound 5, upper
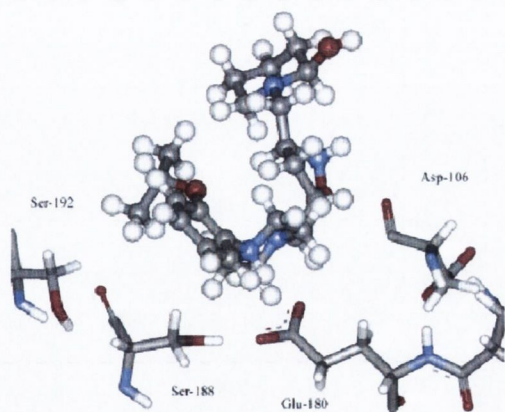
e) Compound 5, middle
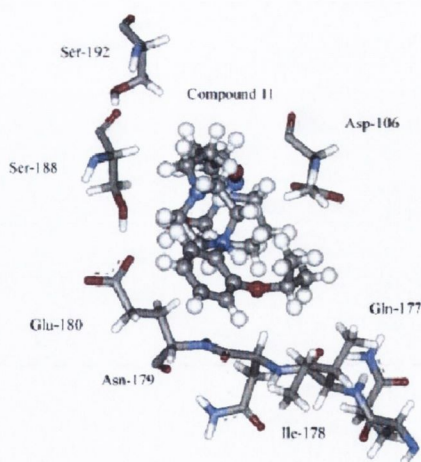
f) Compound 5, lower
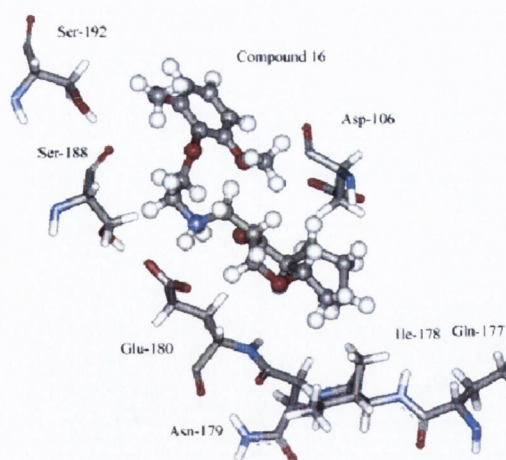
g) Compound 6, upper
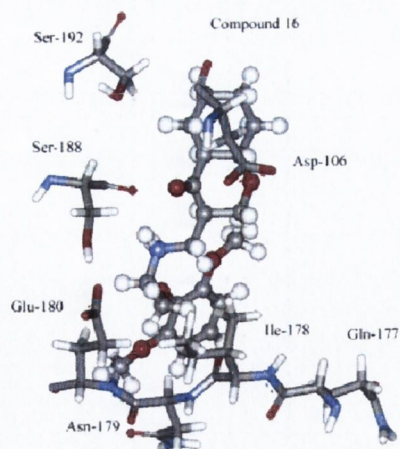
h) Compound 6, middle
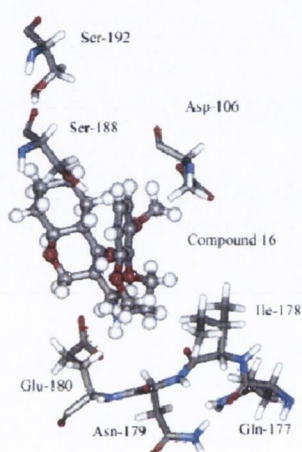
i) Compound 6, lower

Figure 1.32: Docking of compound 4 (a,b,c), compound 5 (d,e,f) and compound 6 (g,h,i) in the α_{1A}-AR in different conformations based on doxazosin (upper conformation) tamsulosin (middle conformation) and compound 3 (lower conformation).[67]

Several models that describe interaction with the α_{1B}-AR have been described in this chapter. As shown previously by our group, it is not clear which conformation is adapted for each compound, since three different binding modes were found for each compound. A different conformation also leads to different interactions with the receptor. It has become clear that residues such as Asp[106], and Glu[180], Ser[188], Ser[192] are commonly involved in binding.

1.4.3.2. Alpha 1B models

The interaction between an antagonist and the α_{1B}-AR model has been studied by Leonardi et al[63], modelling the placement of BMY 7378 in the α_{1B}-AR. They found that there is a tendency of the two phenylpiperazines to cluster in the outer part of the receptor instead of approaching Asp[148]. This is similar in the α_{1A}-AR. Despite the presence of Gly[196] (α_{1B}-AR) instead of a glutamine (α_{1A}-AR) on this position that would favour the

ligand entrance into the binding site, negatively charged amino acids in ECL-II and ECL-III, i.e. Asp^{192}, Glu^{194} and Asp^{327} form an attractive surface for the positively charged nitrogen to interact with an agonist or an antagonist. However, because only one ligand is used, it is impossible to derive $α_{1B}$-AR specific interactions compared to the $α_{1A}$-AR and $α_{1D}$-AR.

1.4.3.3. Alpha 1D models

The interaction of ligands with the $α_{1D}$-AR model was studied by Bautista et al[68] who developed a two state-model consisting of an inactive and an active model. This was achieved by developing a homology model of the $α_{1D}$-AR based on rhodopsin and the docking of agonist adrenaline and antagonist BMY7378. They used the neutral and protonated form of the ligand to deterimine the importance of protonation in ligand binding. On these four complexes an MD simulation of 1 ns followed giving rise to an active state (adrenaline-bound) and an inactive state (BMY7378 bound). These two states were used to redock adrenaline and BMY7378 giving rise to ligand-active receptor and ligand-inactive receptor complexes.

They found that in the case of the amine being protonated there is an increase in hydrogen bonding and aromatic interactions. Protonated adrenaline forms a H-bond with Asp^{176} and has aromatic residues Trp^{172}, Trp^{235}, Phe^{388} and Phe^{388} within 3 Å. Protonated BMY7378 forms H-bonds with Trp^{172} and Lys^{236} and has aromatic residues Trp^{172}, $Trp254$, Phe^{364}, Phe^{384} and Phe^{388} within 3 Å. The interaction of adrenaline with $α_{1D}$-AR is shown in Fig. 1.33. The interaction of BMY7378 with the $α_{1D}$-AR is shown in figure 1.34. It can be seen that only when adrenaline is in its protonated form docked in the active receptor, does it interact with an aspartic acid. For noradrenaline there is no interaction shown with an aspartic acid, which questions if the correct orientation has been obtained. There are ten different residues which play a role in binding in multiple models. These residues are Trp^{172}, Trp^{235}, Tyr^{254}, Lys^{256}, Trp^{361}, Phe^{364}, Phe^{365}, Phe^{384}, Phe^{388}, and Tyr^{392} making these residues most likely to be involved in binding. It can be noted that most of these residues are either Trp, Tyr or Phe, suggesting that interaction with these types of mostly aromatic residues is a necessity for binding. However other residue types should still play an important role in ligand selectivity as different binding profiles are observed in the different models. One conclusion that can be drawn from their work is that protonation of the ligand plays an important role in binding to the $α_{1D}$-AR.
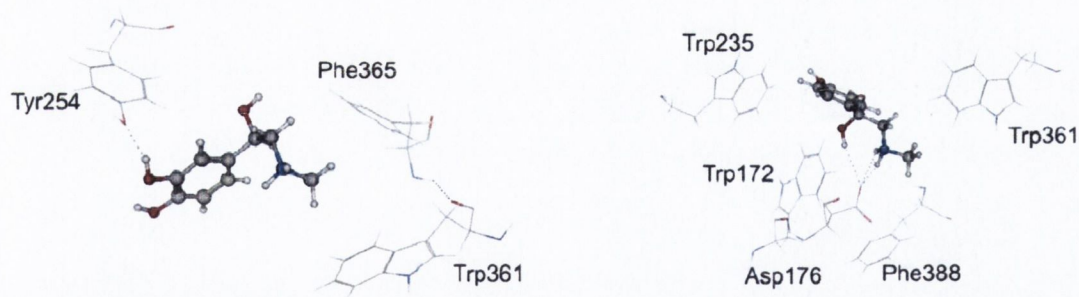
Figure 1.33: Docking of adrenalin into the α_{1D}-AR. The conformations are given for the docked complex of the protonated ligand in either the inactive (left) or the active (right) α_{1D}-AR[68].
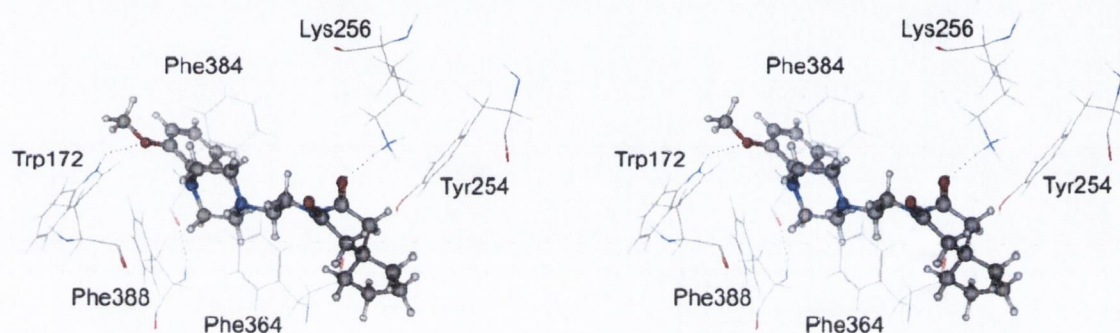
b) Inactive receptor, protonated ligand   d) Active receptor, protonated ligand.

Figure 1.34: Docking of BMY 7378 into the α_{1D}-AR. The conformations are given for the docked complex of the neutral and protonated ligand in either the inactive or the active α_{1D}-AR[68].

Leonardi et al[71] studied the positioning of BMY 7378 in the different α_1-ARs. Their work suggested that the phenypiperazine moiety docks into a site formed by the TM-III, TM-IV, TM-V, TM-VI and ECL-II, but the spiro-cyclic ring of the ligands dock into a site formed by TM-I, TM-II, TM-III and TM-VII. Furthermore, they found that the binding site of the imide moiety does not allow for the simultaneous involvement of the two carbonyl oxygen atoms of BMY7378 in H-bonding interactions. Their docking also suggested that the second and third extracellular loops may act as selectivity filters for the substituted phenylpiperazines. In particular, the positions in ECL-II-III adjacent to the disulfide-bridged cysteine are suggested to contribute to selective binding of the phenylpiperazine derivates, in particular, those that hold the 2,5-dichloro substitution on the phenyl ring. For the α_{1D}-AR it appears that the abilities of the ligands to penetrate into the receptor binding sites, i.e., to approach Asp[176] were better compared with the α_{1A}-AR and α_{1B}-AR. For the two different functional groups, the lowest interaction energy was obtained by ligands involved in charge-assisted H-bonds with Asp[176]. The different behaviour of the α_{1D}-AR compared to the other two α_1-AR subtypes may be due at least in part, to the

synergistic effects of a Phe^{245} and Gly^{247} of the ECL-II-III. The lack of a negatively charged amino acid at Phe^{245} and a bulky side chain at Gly^{247} favours the entrance of the ligand BMY-7378 into the receptor structure.

1.4.3.4. Subtype comparison.

It is clear that a number of residues, such as Asp^{106} in α_{1A}-AR, which are involved in binding of ligands are aligned in all three subtype of the α_1-AR. Most of these have been identified via mutation studies (chapter 1.2.5.). These residues can be used for identification of the binding pocket, but can not explain the binding of each ligand to a subtype. There are a number of residues which are involved in binding of a limited number of ligands, which are not necessarily structurally similar. Therefore, based on the studies that have been described in literature, the phenomenon of subtype specific binding has not be explained and more research is needed.

1.5. Thesis objectives

Chapter two gives an overview of the methods we have used to investigate the binding of antagonists to the α_{1A}-AR, α_{1B}-AR and α_{1D}-AR. As the introduction shows, there are many methods that can be used to investigate the binding of ligands following either a ligand-based or structure-based drug design. We have selected a structure-based drug design (SBDD) approach because it provides more information about how ligands interact with the protein, which can be usefull when designing the next generation of subtype selective α_1-AR inhibitors. Therefore, a selection of SBDD methods has been made including homology modelling, ligand optimization, docking, scoring and molecular dynamics.

Chapter three gives an overview of the homology models that have been generated and describes these different models. In this introduction, information about the binding pocket is described via mutation studies which can be used in the homology modelling process. Furthermore, in this introduction several models are described which have made use of homology models of α_1-ARs.

Chapter four describes the ligand optimization. It is clear from the pharmacophore studies that a positive ionisable group is favourable for ligands to bind to the adrenoceptor and therefore a proton affinity study has been performed.

In chapter five, the molecular dynamics simulations are described. Due to time constraints full conversion to an active or inactive state can not be obtained. But an optimization of the binding of the ligand to the receptor should occur. The binding of several ligands to the the different α_1-AR subtypes is described in our introduction.

<u>Chapter II.</u>

Methodology: An overview of protein modelling, ligand optimization, virtual screening and molecular dynamics simulation.

Computers in the future may weigh no more than 1.5 tons.

Popular Mechanics *(1949)*

2.1. Protein modelling

Several methods are available for modelling proteins such as topology prediction for identifying the residues included in the transmembrane helices, homology modelling for designing a 3D model of a protein and protein analysis for determining the quality of the generated homology models. Molecular dynamics can be used to examine the dynamic motion while docking can be used for positioning a ligand in the binding pocket of a receptor allowing further study of ligand-protein interactions once a protein structure has been obtained.

2.1.1. Topology prediction

Topology prediction is a method by which the transmembrane helices and the loops connecting the helices can be located within the amino-acid sequence. This can be useful for assisting in the alignment of the sequences of a known crystal structure with a protein of an unknown structure because it gives a strong indication of where helices and loops are located and therefore which parts should be aligned.

Every topology prediction method such as Dense Alignment Surface (DAS)[69], Transmembrane prediction with a Hidden Markov Model (TMHMM)[70] and Transmembrane prediction (TMpred)[71] uses several measurements such as hydrophobicity and anisotropic temperature. The transmembrane helices contain a higher concentration of hydrophobic amino acids.

The anisotropic temperature gives a measure of the mobility/uncertainty of a given atomic position. Usually, this value is higher for the loops due to their conformational flexibility. The reliability of the prediction is limited and therefore, these methodologies provide only an indication of where the different helices and loops are located.

2.1.2. Homology modelling

Homology modelling can be used when there is no known three dimensional (3D) structure of the protein under study. The principle of homology modelling[72] is to build a 3D structure based on the crystal structure of a protein family member. The assumption in this process is that all proteins from a certain family share a number of strongly conserved residues and therefore the same basic structure.

The process of homology modelling can be divided in several steps:

1. Select a reference protein with a known 3D structure, belonging to the same family as the target protein.

2. Align the sequences of the target protein with the reference protein.

3. Generate several models for the target protein, the homology model, using the structure of the reference protein as a scaffold or template.

4. Assess the quality of the homology models for selection of one model.

5. Optimise the extra- and intra-cellular loops.

6. Assess the quality of the homology model for selection of one model.

A good source of reference protein structures for the first step is the Research Collaboratory for Structural Bioinformatics (RCSB) protein data bank (www.rcsb.org) which holds most protein crystal structures. The selection of a reference protein should be a protein which is a family member, ideally with a good resolution where the full structure is resolved. The known structure can be obtained via X-ray, NMR or computer based models. The higher the similarity to the reference structure, the more likely structural features are shared amongst the two structures.

The second step is obtaining the residues which possess a similar position in the amino acid sequence of the target protein and that of the reference 3D structure. A similar position in the amino acid sequence means a similar position in both reference and target protein. When residues are similar throughout a family of proteins, they are said to be conserved and are expected to play an important role in the function of the protein. There are on-line tools such as ClustalW[73] and T-coffee[74] that assist in the identification of conserved residues and using these, it is possible to generate an initial alignment. These tools use statistics for comparing the two sequences providing a good starting alignment, but do not necessarily result in an optimal alignment. For this reason, a manual inspection of the alignment should follow to verify a correct alignment or changes in the alignment should be made accordingly.

Another source of information is mutation data. Mutation data are obtained by experimentally changing an amino acid and observing any change in the function of the protein. If such a change causes the loss of a function the assumption is made that this amino acid plays an important role in fulfilling this function. It is assumed that residues which play an important role and are conserved amongst a family of proteins, play an important role in each family member and, therefore, they must be aligned correctly.

In the third step homology models are generated for the target protein based on the reference structure. Using similarities between the amino acid sequence of the target protein and the reference protein the modelling program derives 3D features from the

reference crystal protein structure which are then implemented in the newly created homology models. The reliability of making a good homology model increases when the sequence of the reference and the target protein share more conserved and aligned residues.

In the fourth and sixth steps the analysis of the homology model is performed. Analysis should be performed after the generation of the initial model and after each loop optimisation for the selection of a model for further optimisation. This is performed using a number of different protein analysis programs to obtain a range of validations (section 2.1.3.) and derive a consensus from these findings.

In the fifth step, the loops of the protein are optimized. The similarity in the alignment between the crystal structure of the protein and the target protein is usually lower in the loop-regions. This difference can be due to the type of residues in the amino acid sequences and/or the length of the loop. For short loops this is not a major issue because they have limited conformational flexibility. For longer loops that have no similarity with the target protein and can consist of more than 30 amino acids, the prediction of a realistic conformation can be difficult. The modelling protocol generates several different loop possibilities and protein analysis tools are used to select the best one.

We have used MODELLER[75] to generate our homology models. This program determines features from the reference structure based on the similarities in the alignment between the sequences of the template protein and the target protein. The features consist of spatial restraints derived from the reference structure including Cα-Cα distances, main-chain N-O distances, main-chain and side-chain dihedral angles. The dicision on which on which spatial restraints to use from the reference structure is based on statistical analysis of a database with high quality 3D structures. These restraints are expressed as probability density functions (pdf) and are combined to give a molecular pdf.

A model is produced which satisfies the constraints using a conjugate gradient technique to implement the molecular pdf in the model. This initial model is further optimized using a molecular dynamics / simulated annealing technique resulting in a large number of models.

A separate loop optimization can be performed by restraining the end positions of the loop and optimising the loops by using a molecular dynamics / simulated annealing procedure. The process is initiated by making a few random initial loops. The initial loop conformation is randomized by ~5 Å in each direction. The loops are first optimized considering only the loop atoms, and then the interactions of the loop and the rest of the

system are included. This interaction is based on atomistic distance dependent statistical potentials for non-bonded interactions.[7]

2.1.3. Protein analysis

The quality of the developed models must be examined to determine which will be used for further studies with, e.g. the Ramachandran plot[76], residue environment analysis[77] and non-bonded interaction analysis.[78] These methods can be used individually for scoring the homology models, but since they examine different aspects, none of the programs can give a definitive answer. However, when using all three, a consensus for selection of the best model can be achieved.

2.1.4. Geometrical validation

A geometrical analysis of a protein model can be performed by analysing bond lengths and bond angles. This includes main-chain bond lengths and angles, dihedral angles, chirality, non-bonded interactions, main-chain hydrogen-bond energies and disulfide bond separation length. An example of this is the program Procheck.[76] Procheck has acceptable ranges for features derived from high quality crystal structures (resolution of 2.00 Å or better, and an R factor of 20% or better). The output gives an overview of amino acids that do not meet the criteria and therefore are likely to be poorly modelled.

Particular importance is placed on the relation of the phi/psi dihedral angles in the form of the Ramachandran plot as illustrated in fig. 2.1. The angle denoted phi describes rotation about the C(alpha)-N bond of the amino acid, and the angle psi denotes rotation about the bond linking the

Fig 2.1: Example of a Ramachandran plot for the estrogen receptor (3ERT).

alpha carbon and the carbonyl carbon. The colours in the plot show the different classifications in which a residue can be placed. The darkest regions are the most favourable regions and the lightest region is the disallowed region. The phi-psi torsion angle is measured for each amino acid and is placed on this plot and based on it

placement classified as favourable (red), additional allowed (dark yellow), generously allowed (light yellow) or disallowed (white).

2.1.5. Environmental validation

Environmental validation determines the accuracy of a 3D protein model by comparing the residues position to its surroundings, computed from its atomic coordinates. These findings are compared with known protein structures and a score is given. The program Verify3D[77] makes use of this type of validation.

The criteria used in enviromental evalution are derived from high quality 3D models in which every residue is characterized by its environment and represented by a row of 20 numbers which make up its profile. These numbers are the preferences (called 3D-1D scores) of each of the 20 different possible amino acids which are known for this environment. The environment of the residues is defined by the following parameters; the area of the residue that it is not solvent accessible, the fraction of side-chain area that is covered by polar atoms (O and N) and the local secondary structure. The secondary structure is the local segment such as α-helices of proteins in 3D, but does not take into account the postion of atoms that make up that structure.

As the environment for the residues is statistically determined, a comparison between the measured environments and an optimal environment derived from the sequence can be determined. The sum over all residue positions, of the 3D-1D scores for the amino acid sequence of the protein is the 3D profile score. As the 3D-1D score for each residue is a statistical number, the compatibility of the amino acids with their sequence can be assessed by plotting their 3D-1D score against their sequence number and the average 3D-1D score can be derived. An advantage of using 3D profiles for testing models is that the profiles have not themselves been used in the determination of the structure.

2.1.6. Non-bonded atomic interaction validation

Atomic interaction validation scores of proteins can be based on the number of non-bonded atomic interactions. Interactions between different types of atoms are not randomly distributed in a region and, therefore, a quality assessment can be based on this distribution. An example of this method is the program Errat[78] which produces a confidence value as to which residue can be regarded as incorrectly positioned.

Within the non-bonded atomic interaction validation three atoms types are monitored: carbon (C), nitrogen (N) and oxygen (O) and hence there are six different combinations

of interactions (CC, CN, CO, NN, NO and OO). The number of interactions of each combination is obtained in a nine-residue sliding window along the backbone of the protein, where four residues are included on each side of the centre in the analysis. Reference values have been obtained from known crystal structures. For example, the ERRAT methodology has been validated on 96 crystal structures of proteins which meet the following criteria: a resolution of 2.5 Å or better, an R-factor of less than 25%, a monomeric or homo-oligomeric structure, the exclusion of prosthetic groups and good geometry defined by ω, the dihedral angle made by the peptide bond less than ±15 from ideal. The number of nonbonded interactions is determined with restrictions being applied to those interactions: the distance between the two atoms in space is less than a preset limit, typically 3.5 Å and atoms within the same residue or those that are covalently bonded to each other are not considered. For each protein, the fractions of the number of interactions are calculated (Eq. 2.1). Example f(CC) presents the fraction of all interactions that are of type CC.

$$f(CC) = \frac{n_{CC}}{n_{CC} + n_{CN} + n_{CO} + n_{NN} + n_{NO} + n_{OO}} \qquad (2.1)$$

Where f is the fraction of atomic interactions and n is the number of interactions for each CC, CN, CO, NN, NO or OO. One limitation of this method is that it can not detect severe errors such as misalignments of amino acid with respect to the true position in the fold.[78] The approach is sensitive to the method used for atomic refinement such as molecular dynamics and refined structures show improved scoring. This could result in an inability to identify incorrect model structures that have been extensively refined without experimental constraints.

2.1.7. Optimization of hydrogen bond network

The position of the hydrogens can not be obtained in crystal models as they do not appear in X-ray electron density plots. The positions of the hydrogens on the homology model can be optimized by modelling the different positions and selecting one based on the total hydrogen bond energy.[79] This procedure also predicts the different ionization states of His, Asp and Glu residues in proteins. In addition, the sidechains of His, Gln and Asn are allowed to change by 180 degrees to compensate for crystallographic misassignments and allowing the correct positioning of the hydrogens which are bonded to these residues. A force field is used to find the optimal hydrogen bonding pattern and a definition of the degrees of freedom and the constraints. The force field that is used is

designed for hydrogens only and is an approximation to the effects that can influence the preferred geometry of hydrogen bonds such as electrostatic intersections and steric effects.

The residues His, Gln and Asn are difficult to model on X-ray electron density plots because it can be difficult to differentiate between oxygens and nitrogens if the resolution is not good, which can result in the wrong positioning of the side-chains. Therefore, these residues should be allowed to flip the side-chains for two conformations to be examined. For this, an energy penalty of 1.2 kcal mol^{-1} is subtracted from the interaction energy to increase the certainty of modelling the correct conformation.

The N and O atoms in His, Glu and Asp can be either donor or acceptor, depending on the presence of a bound hydrogen atom. If a hydrogen atom is present on one of these atoms, the carrying atom is only regarded as a donor, if it is absent, the atom is treated as an acceptor. An energy penalty of 4.5 kcal mol^{-1} is subtracted from the interaction energy if one hydrogen atom is present. In this way, only evidence that suggests a strong hydrogen bond is used in the hydrogen optimisation process and limits the possibility of modelling a hydrogen bond when it should not be there.

For many of the hydrogen bond donors (backbone N, Arg side chain etc.) accurate coordinates of the hydrogen atom can be calculated by using simple geometric construction: hydrogen atoms are placed at ideal triangular or tetrahedral positions. Other hydrogen atoms (alcoholic OH, lysine NH_3, N-terminal NH_3) are free to move on a circle, using a fixed bond angle within the functional group. Hydrogen atoms connected to water molecules make a H-O-H angel of 110°, but the orientation of the molecule is free. In all cases hydrogen atoms are placed at a distance of 1.0 Å from the donor atom. An implementation of this method is performed by the program WHAT IF.[80]

2.2. Ligand optimization techniques

Ligands have to be subjected to conformational search to find the minimum energy conformation and atomic charges can be derived using a variety of methods, a selection of which is described below.

2.2.1. Conformational search

A conformational search can be performed by either a systematic or a random approach. With the systematic approach, each dihedral angle or bond angle is rotated in each step until all different angles have been rotated 360° and all conformations are sampled. Each conformation can be followed by a minimization using conjugate gradient to obtain the closest minima. The disadvantage of this method is that it is very time consuming and scales as n^n. An alternative is the random approach where a random change is made (using Monte Carlo algorithms) to the current conformation and hence the system can move from one region of the potential energy surface to another in a single step. This newly obtained conformation is then optimised using the conjugate gradient method. If the optimised conformation has not been previously found, it is stored. If the conformation has been found, it will be discharged. This procedure is continued until a set number of iterations has been performed.

2.2.2. Ab initio methods, Hartree-Fock

The term *ab initio* is used to describe quantum mechanical techniques[81-83] (QM), which are formulated without experimental data. QM methods attempt to solve the Schrodinger equation (Eq. 2.2) and can be used to produce a detailed electron distribution of a molecule. To solve the Schrödinger equation the wavefunction (ψ) is needed which, when operated upon by the Hamiltonian (H), returns the energy of the solution multiplied by the particle energy (E).

$$H\psi = E\psi \tag{2.2}$$

According to the Born Oppenheimer approximation the motion of electrons can be decoupled from the motion of nuclei based on the assumption that electrons move so fast that they see nuclear centres of mass as fixed. This assumption reduces the problem to solving the electronic Schrödinger equation for a set of nuclear geometries (Eq. 2.3).

$$H_e\psi_e = E_e\psi_e \tag{2.3}$$

For an atom that contains a single electron, the Hamiltonian can be written as the kinetic energy of the electrons and potential energy of the nuclei and electron interaction (Eq. 2.4).

$$H_e = -\frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{4\pi\varepsilon_0 r}$$

(2.4)

The kinetic energy depends upon the mass m, Plancks constant $\hbar$ and the second derivative, the Laplacian ∇^2, of the wavefunction. The potential energy depends upon the distance r, between the electron and the nucleus as given by the Coulomb equation, where Z is the nuclear charge. The Schrodinger equation can be solved exactly for a one-electron system. However, when there is more than one electron, there needs to be an electron-electron repulsion term which precludes the exact solution for the Schrödinger equation. Therefore, more approximations need to be introduced.

Within the framework of the approximations used in the Hartree-Fock (HF) approach[83-85], a molecular wavefunction is calculated and various molecular properties are determined. The electronic wavefunction (ψ) is described by the motion of the electrons as the product of the probability distribution of each electron (Eq. 2.5). The exact wavefunction can not be determined, and the energy resulting from an approximation to the true wavefunction will always be greater than the true energy. A better wavefunction has a lower energy and, therefore, the approximation with the lowest energy is choosen as the electronic wavefunction.

$$\psi(x_1....x_N) = \chi_i(x_1)\chi_j(x_2)\chi_k(x_3)......\chi_n(x_N)$$

(2.5)

The Pauli Exclusion Principle states that the wavefunction must be antisymmetric with respect to the electron interchange and the Hartree wavefunction has to satisfy this statement. A Hartree product can be made antisymmetric by adding and subtracting all possible permutations. The resulting HF-wavefunction is the determinant of the system of N electrons and N spin orbitals, termed a Slater determinant (Eq. 2.6) which gives the antisymmetric wavefunction for N indistinguishable particles.

$$\Psi = \frac{1}{\sqrt{N}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_N) & \chi_1(x_1) & \cdots & \chi_N(x_N) \end{vmatrix} \tag{2.6}$$

Each electron is represented as a response to the average potential of the rest of the electrons in the energy expression. This energy expression can be written in a compact form that recognises three types of interaction that produce the total electronic energy of the system. The Fock operator, f_i, (Eq. 2.7), is an one-electron Hamiltonian, comprising of a core Hamiltonian term $H^{core}(1)$, a Coulomb operator $J_j(1)$ and an exchange operator $K_j(1)$ (Eq 2.8):

$$f_i \chi_i(1) = \sum_{j=1}^{N} \varepsilon_{ij} \chi_j(1) \tag{2.7}$$

$$\left[H^{core}(1) + \sum \left\{ J_j(1) - K_j(1) \right\} \right] \chi_i(1) = \sum_{j=1}^{N} \varepsilon_{ij} \chi_j(1) \tag{2.8}$$

The core Hamiltonian (Eq. 2.9) is the average kinetic energy and the potential energy for the electrostatic attraction between the nuclei and the electron. The Coulomb operater (Eq. 2.10) is the potential energy for the electrostatic repulsion between two charge distributions. The exchange operater (Eq. 2.11) is based on the requirement that Ψ should be anti-symmetric with respect to the permutation of the coordinates of any two electrons and is the additional term acquired by using the Slater determinant.

$$H^{core}(1) = -\frac{1}{2} \nabla_1^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{1A}} \tag{2.9}$$

$$J_j(1) = \int dr_2 \chi_j(1) \frac{1}{r_{12}} \chi_j(1) \tag{2.10}$$

$$K_j(1) = \int dr_2 \chi_j(1) \frac{1}{r_{12}} \chi_i(2) \tag{2.11}$$

A systematic way of varying the MO can be introduced and is achieved by representing the MOs as a linear combination of basis functions. The basis functions are similar to atomic orbitals and the method is referred to as a linear combination of atomic orbitals (LCAO). Therefore each molecular orbital can be written as a summation, where ψ_i is a molecular orbital, $c_{\mu i}$ is a coefficient and φ_μ is the basis functions which corresponds to the atomic orbital (Eq. 2.12).

$$\psi_i = \sum_{\mu=1}^{K} c_{\mu i} \phi_\mu \qquad (2.12)$$

The HF equations form a set of pseudo-eigenvalue equations as the Fock operator depends on all the occupied MOs. The derivation of the HF equations for a closed shell system is known as the Roothan-Hall equations (Eq. 2.13).

$$FC = SCE \qquad (2.13)$$

Where F is the matrix representation of the Fock operator, C is the matrix of coefficients $c_{\mu i}$, S is the overlap matrix between two atomic orbitals and E is a matrix of energy eigenvalues. An iterative procedure is used and the SCF procedure is implemented. First, a set of trial solutions for the HF eigenvalue equations are obtained. The HF equations are solved, giving a second set of solutions, which are used in the next iteration. This method gradually refines the structures that correspond to lower total energies until the point is reached at which the results for all the electrons are unchanged and the structure is self-consistent.

The HF energy is not as low as the true energy of the system because the Fock operator treats each electron as though it is moving in a time averaged potential field due to other electrons. However, as the motion of electrons is correlated, they tend to avoid each other more than the HF theory suggests and hence HF overestimates electron-electron repulsion.

2.2.3. Density Functional Theory

The basis for density functional theory[83,86,87] (DFT) is that the ground state electronic energy can be determined from the electron density $\rho(r)$, where the energy E, is a unique function of $\rho(r)$ (Eq. 2.14)

$$E[\rho(r)] = \int V_{ext}(r)\rho(r)dr + F[\rho(r)] \qquad (2.14)$$

The term $V_{ext}(r)$, is based on the interaction of the electrons with an external potential, which is due to the Coulomb interaction with the nuclei. The term $F[\rho(r)]$, is the sum of the kinetic energy of the electrons and the contribution from the electron-electron

interactions. Kohn and Sham, introduced the approximation of $F[\rho(r)]$ as the sum of the kinetic energy $E_{KE}[\rho(r)]$, the Hartree electrostatic energy $E_H[\rho(r)]$ and the contributions from exchange and correlation $E_{xc}[\rho(r)]$ (Eq. 2.15).

$$F[\rho(r)] = E_{KE}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)]$$
(2.15)

$E_{KE}[\rho(r)]$ is defined as the kinetic energy of a system of non-interacting electrons (N) with the same density $\rho(r)$ as the real system (Eq. 2.16). The Hartree electrostatic energy, $E_H[\rho(r)]$, is derived from the interaction between two charge densities, $\rho(r_1)$ or $\rho(r_2)$, which is summed over all possible pairwise interactions (Eq. 2.17).

$$E_{KE}[\rho(r)] = \sum_{i=1}^{N} \int \psi_i(r)\left(-\frac{\nabla^2}{2}\right)\psi_i(r)dr$$
(2.16)

$$E_H[\rho(r)] = \frac{1}{2}\iint \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|}dr_1 dr_2$$
(2.17)

Combining these two terms and adding the electron-nuclear interaction provides the full expression for the energy of a N-electron system within the Kohn-Sham scheme (Eq. 2.18). This equation acts to define the exchange-correlation energy functional $E_{XC}[\rho(r)]$, which contains contributions due to exchange and correlation and a contribution due to the difference between the true kinetic energy of the system and $E_{KE}[\rho(r)]$.

$$E[\rho(r)] = \sum_{i=1}^{N} \int \psi_i(r)\left(-\frac{\nabla^2}{2}\right)\psi_i(r)dr + \frac{1}{2}\sum_{i \neq j}^{N} \iint \frac{\rho(r_1)^2 \rho(r_2)^2}{|r_1 - r_2|}dr_1 dr_2$$
$$+ E_{XC}[\rho(r)] - \sum_{A=1}^{M} \int \frac{Z_A}{|r - R_A|}\rho(r)dr$$
(2.18)

The exchange-correlation functional[24,27,28] $E_{XC}[\rho(r)]$ is key to the DFT approach and this contribution was initially obtained through the local density approximation (LDA). The LDA is based on the uniform gas model in which the electron density is constant throughout all space. Unfortunately, the LDA approximation is known to be inadequate in some cases and, therefore, extensions have been developed.

Besides the density, the gradient of the density should also be included in the functional expressions. If the function depends on the gradients then the functional is referred to as being 'gradient-corrected'. These gradient corrections are typically divided into separate

exchange and correlation terms *i.e.* $E_X[\rho(r)]$ the exchange functional and $E_C[\rho(r)]$ the correlation functional.[88]

The gradient correction to the exchange functional proposed by Becke is popular (B88),[i] while the correction functional of Lee, Yang and Parr (LYP)[ii] is widely used.

One potentially attractive option is to add a correlation energy derived from DFT (e.g. the LDA approximation) to the HF energy. The most common form has been launched under the acronym "B3LYP" and is an empirical mixture of exact exchange, Becke's gradient correction for exchange B88, the Lee-Yang-Parr (LYP) correlation functional (with the gradient term) and the standard local correlation functional due to Vosko, Wilk and Nusair (VWN) (Eq. 2.19).

$$E_{XC} = AE_X^{HF} + (1-A)E_X^{LDA} + B\Delta E_X^{Becke88} + CE_C^{LYP} + (a-C)E_C^{VWN} \qquad (2.19)$$

The constants *A*, *B* and *C* are determined by fitting, and have values of *A*=0.20, *B*=0.72 and *C*=0.81.

2.2.4. Basis sets

An approximation needed in all quantum mechanical methods is the choice of a basis set[83,89,90] which is required to create the wavefunction. Two types of basis functions are normally used for molecular calculations, the Slater type orbital (STO) and the Gaussian type orbital (GTO). Slater type orbitals are physically more realistic, but some of the integrals can not be solved analytically. Therefore Slater orbitals are often replaced by functions based on Gaussians. A Gaussian expansion contains two parameters: the coefficient and the exponent. When in molecular orbital calculations both of these parameters are allowed to vary they are uncontracted or primitive Gaussian type orbitals (PGTO). The Gaussian function has the form $exp(-\alpha r^2)$ where α is the radial extent or spread of the Gaussian. QM calculations use basis functions comprising integral powers of *x*, *y* and *z* multiplied by the exponent (Eq. 2.20).

$$x^a y^b z^c \exp(-\alpha r^2) \qquad (2.20)$$

A minimal basis set is a representation that contains just the number of functions that are required to accommodate all the filled valence orbitals in each atom. A basis set which

doubles the number of functions in the minimal basis set is described as a double zeta (DZ) basis. The doubling of the core orbitals, unlike the valence orbitals, does not has a significant effect on the chemical properties and these vary only slightly from one molecule to the next. Therefore a less computational alternative can be used. A variation of the DZ type basis only doubles the number of valence orbitals, producing a split valence (SV) basis.

In the 6-31G basis set, the core orbitals are a contraction of six PGTOs. The inner part of the valence orbitals is a contraction of three PGTOs and the outer part of the valence is represented by one PGTO. The 6-31G has the same number of contracted basis functions as the 3-21G set but because there are more PGTOs used, the representation of each function is better.

As atoms are brought closer together, a polarization effect occurs, resulting in distortion in their electron density. This can be simulated by adding in basis functions of higher angular momentum than the valence orbitals. For example, the spherical 1s orbital on hydrogen can be distorted by mixing in an orbital with p symmetry. Similarly, the p orbitals can be polarised if an orbital of d symmetry is mixed in. When a polarisation function is introduced to basis sets it is usually denoted using an asterisk '*'. Thus 6-31G*, refers to a 6-31G basis set with polarisation functions on heavy atoms. The addition of another asterisk as in 6-31G** indicates the additional use of polarisation functions on hydrogen and helium. A deficiency of the basis set arises as Gaussian basis functions are rather low far from the nuclei. To compensate, highly diffuse functions can be added to the basis set, indicated by (+). A single + indicates an additional single set of diffuse s- and p- type Gaussian functions, while ++ indicates that the diffuse functions are included for hydrogen as well as for heavy atoms.

2.2.5. Semi-empirical techniques

Further approximations to the HF model can be made leading to semi-empirical methods[83,91,92]. For these calculations only the valence electrons are included while the core electrons are subsumed into the nuclear core. Furthermore, the computation of a large number of the integrals in the HF-SCF calculation is simplified by either neglecting them or replacing them with parameterised equations that can be calculated more easily. The central assumption of semi-empirical methods is the zero differential overlap approximation (ZDOA), in which the overlap between pairs of different orbitals is set to zero for all volume elements. If the ZDOA is applied to all orbital pairs then the Roothan-

Hall equations (Eq. 35) for a closed shell molecule are considerably simplified. However, the major contributors to bond formation are the electron-core interactions between pairs of orbitals and the nuclear cores. Therefore, these interactions are not subjected to the ZDOA approach. The core integrals are parameterised to experimental results, compensating for the ZDOA to some extent.

The semi-empirical methods depend on the exact number of integrals that are neglected and how the parameterisation is performed. The Austin Model 1 (AM1) and Parametric Method Number 3 (PM3) methods have been developed to improve earlier semi-empirical methods, which had a tendency to overestimate repulsions between atoms separated by distances approximately equal to the sum of their van der Waals (vdw) radii.

The AM1 model is a semi-empirical method based on the Modified Neglect of diatomic Differential Overlap method (MNDO). The approach for MNDO is to reproduce experimental data. This was achieved by modifying the parameterization based on experimental data instead of using calculations. If there is a differential overlap between atomic orbitals on different atoms, the term is excluded. All of the parameters are monatomic. The AM1 method has a small modification to the MNDO method in which Gaussian functions are added to modify the core-core interactions to try to compensate for the excessive repulsion between atoms at close distance. This increases the number of parameters per atom from 7 in MNDO to between 13 and 16 in AM1 depending on the number of Gaussians used. This made the parameterization process much more difficult although a significant improvement over the standard MNDO was achieved.

2.2.6. Polarizable continuum model

Quantum chemistry calculations can be performed in gas phase but also in solvent. Instead of using actual solvent molecules the effects of solvation can be mimicked with the use of a solvation model such as the polarizable continuum model[93] (PCM). The cavity surface is divided into a large number of small surface elements, and there is a point charge associated with each surface element. These point charges represent the polarisation of the solvent. In the PCM model, a cavity surface is determined from the van der Waals radii of the atoms. The fraction of each atom's van der Waals sphere which contributes to the cavity is then divided into a number of small surface elements of calculatable surface area. A point charge for each surface element is then calculated from the electric field gradient based on the solute alone (Eq. 2.21).

$$q_i = -\left[\frac{\varepsilon - 1}{4\pi\varepsilon}\right] E_i \Delta S \qquad (2.21)$$

Where q_i is the point charge, ε is the dielectric constant of the solvent medium, E_i is the electric field gradient and ΔS is the area of the surface element. These point charges can then be used in the calculation of the electrostatics of the ligand atoms.

2.2.7. Proton affinities

Depending on the pH of the environment, ligands with many N atoms can be protonated at the different nitrogens.[94] It is, however, likely that at physiological pH only one of these nitrogens is actually protonated. The most likely one can be determined by the difference in enthalpy between the neutral and protonated stated. Computationally, this can be determined by calculating the Gibbs free energy of the neutral state and each of the protonation states where one of the nitrogens is protonated. The unprotonated free energy can then be subtracted from the free energy of the protonated state giving the proton affinity of that state. This can be performed for each protonation state and the state that gives the largest differences between the Gibbs free energies (largest PA) is the protonation site that is most likely to occur.

2.2.8. Calculation of the pK_a

The pK_a is the acid dissociation constant and represents the balance between the ligand in neutral and protonated state. The pK_a has an influence on solubility, membrane permeation and binding to proteins. The pK_a can be determined by calculating the gibbs free energies of the ligand in gas and solvent phase in neutral and protonated form.[95-99] The following thermodynamic cycle (Fig. 2.2) is used to obtain the aqueous solvation free energy (ΔG_{aq}^0) which is needed to calculate the pK_a.

$$\Delta G^0_{gas}(A)$$

$$AH^+ \text{ (gas)} \longrightarrow A \text{ (gas)} + H^+$$

$$\Delta G_S^0(AH^+) \qquad \Delta G_S^0(A) \quad \Delta G_S^0(H^+)$$

$$AH^+ \text{ (aq)} \longrightarrow A \text{ (aq)} + H^+ \text{ (aq)}$$

$$\Delta G^0_{aq}(A)$$
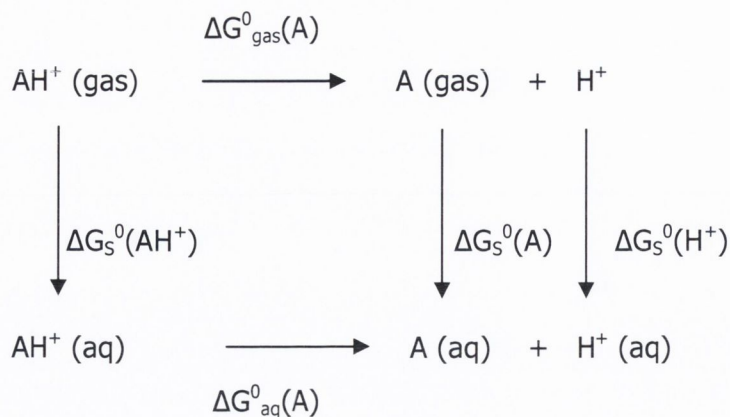
Figure 2.2: Thermodynamic cycle for protonation of a ligand.

From this thermodynamic cycle, equation 2.22 can be derived to obtain ΔG^0_{aq}.

$$\Delta G^0_{aq}(A) = \Delta G^0_S(A) + \Delta G^0_S(H^+) + \Delta G^0_{gas}(A) - \Delta G^0_S(AH^+) \tag{2.22}$$

$\Delta G^0_S(A)$ and $\Delta G^0_S(AH^+)$ can be obtained from our calculations by subtracting the gas-phase free energy from the PCM-phase free energy. The $\Delta G^0_S(H^+)$ is a known experimental value and can be taken from the literature. It was originally defined as 263.8 kcal mol^{-1}[100], but later corrected to 265.9 kcal mol^{-1}. [101] $\Delta G^0_{gas}(A)$ can be obtained using equation 2.23.

$$\Delta G^0_{gas} = G^0_{gas}(AH^+) - G^0_{gas}(A) - G^0_{gas}(H) \tag{2.23}$$

Where $G^0_{gas}(AH^+)$ is the free energy from the protonated state and $G^0_{gas}(H^+)$ is the free energy from the neutral state. $G^0_{gas}(H^+)$ is the free energy from the positive charged hydrogen and can be taken using the Sackur-Tetrode equation[101] resulting in a value of -6.28 kcal mol^{-1}.

After ΔG^0_{aq} has been obtained, the pK_a can be derived from this cycle using equation 2.24.

$$pK_a = \frac{\Delta G^0_{aq}(A)}{RT(\ln 10)} \tag{2.24}$$

Where, R is the universal gas constant and T is the temperature. The value of the universal gas constant is 0.0019872159 kcal K^{-1} mol^{-1} and the temperature is 298 Kelvin at which the free energies are determined.

A PCM model can be used to mimic the solvent effect of water, but a more accurate interaction can be obtained when explicit water molecules are introduced. This can be performed by adding a limited amount of explicit water molecules in addition to the PCM solvent model. When we introduce water molecules into the scheme the following scheme arises (Fig. 2.3).

$$\Delta G^0_{gas}(A{\cdot}H_2O)$$

$$AH^+{\cdot}H_2O \text{ (gas)} \longrightarrow A{\cdot}H_2O \text{ (gas)} \quad + \quad H^+ \text{ (gas)}$$

$$\Delta G_s^0(AH{\cdot}H_2O) \qquad \Delta G_s^0(A{\cdot}H_2O) \qquad \Delta G_s^0(H^+)$$

$$AH^+{\cdot}H_2O \text{ (aq)} \longrightarrow A{\cdot}H_2O \text{ (aq)} \quad + \quad H^+ \text{ (aq)}$$

$$\Delta G^0_{aqua}(A{\cdot}H_2O)$$
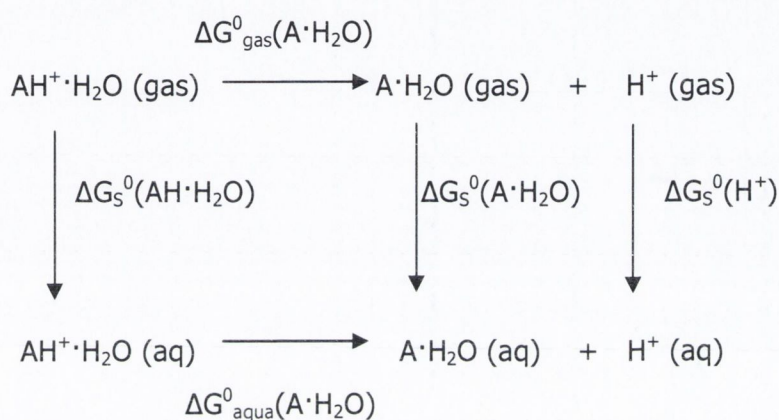
Figure 2.3: Thermodynamic cycle for protonation including one explicit water molecule.

The equations that are used to derive the pKa are similar to the previous scheme (Fig. 2.2).

2.3. Virtual Screening

Virtual screening[103-105] is the combination of docking and scoring. In the docking process the ligand is placed in the binding pocket of the receptor. This ligand-receptor complex can then be scored to determine the interaction between the ligand and the receptor. Virtual screening is commonly used to select active compounds from a database. The resulting ligand-protein complex also allows the determination of the interaction between the ligand and the receptor and provides a visual representation.

2.3.1. Grid based docking

Grid-based docking[106] consists of a grid procedure where a database of grid interaction points is created and a docking procedure where different ligand protein conformations are analysed.

In the grid procedure (Fig. 2.4), a protein or a part of the protein which is of interest, is embedded in a three-dimensional grid and a probe atom placed at each grid point. Grid point spacing can vary from 0.2 Å to 1.0 Å. although the default is 0.375 Å (a quarter of the length of a carbon-carbon single bond). The energy of interaction of this single atom with the protein is assigned to the grid point.

Figure 2.4: Representation of the autogrid method[105]

An affinity grid is calculated for each type of atom in the ligand, typically carbon, oxygen, nitrogen and hydrogen, as well as a grid of electrostatic potential. The energetics of a particular ligand configuration is then obtained by interpolation of affinity values of the eight grid points surrounding each of the atoms of the ligand. The electrostatic interaction is evaluated similarly, by interpolating the values of the electrostatic potential and multiplying by the charge on the atom of the ligand.

With the protein static throughout the simulation, the ligand is moved randomly in the space around the protein. At each step in the simulation, a small random displacement is applied to each of the degrees of freedom of the ligand: translation of its center of gravity, orientation and rotation around each of its dihedral angles. This displacement results in a new configuration, whose energy is evaluated using the grid interpolation procedure. This new energy is compared to the energy of the preceding step. If the new energy is lower, the new configuration is immediately accepted. If the new energy is higher, then the configuration is accepted or rejected based upon the probability expression (Eq. 2.25) which itself is dependent on a defined temperature.

$$P(\Delta E) = e^{(\frac{\Delta E}{k_B T})} \tag{2.25}$$

Where P is the probability of acceptance, ΔE is the difference in energy from the previous step, k_B is the Boltzmann constant and T is the temperature.

At high temperatures, almost all steps are accepted. At lower temperatures, fewer high energy structures are accepted. After a specified number of acceptances or rejections, the next cycle begins with a lower temperature. One software package that makes use of grid-based docking is Autodock.[37]
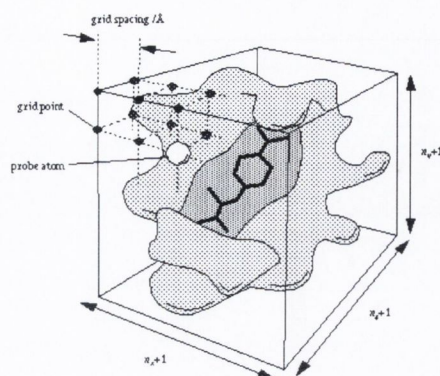
2.3.2. Sphere-based docking

The first step of sphere based docking[107] processes is to identify the potential binding site. Within this binding site, points are identified where ligand atoms could be located when it binds to the protein, by using spheres. To identify possible atom positions a set of overlapping spheres is generated which fill the potential binding site. The sphere centers attempt to capture the shape characteristics of the active site but with using a minimum number of points and without having bias of previously known ligand binding modes. To position the ligand in the active site, the sphere centers are paired with the ligand atoms and several sets of atom-sphere pairs are generated. To limit the number of possible sites of atom-sphere pairs a longest distance heuristic approach is employed. This means that the (long) inter-sphere distances are roughly equal to the corresponding (long) ligand inter-atomic distances. A set of atom-sphere pairs is used to calculate an orientation of the ligand within the binding site. To orientate the entire ligand within the active site, both a translation vector and a rotation matrix are calculated which minimizes the rmsd of the (transformed) ligand atoms and matches the sphere centers of the sphere-atom set.

For the evaluation of the orientation of the ligand in the binding site a shape scoring function and a function approximation of the ligand-protein binding energy is used. All evaluations are performed using a scoring grid to minimize the overall computational time. This is performed by placing the protein in a grid and placing a probe at each grid point. A different probe is used for each atom type in the ligand and the interaction of the probe with the protein can be determined at each grid point. The interaction of one ligand atom with the protein can be determined via interpolation of the surrounding probe atoms. Interpolation of all ligand atoms can be used to score the full ligand-protein interaction.

The shape scoring function is an empirical function resembling the van der Waals attractive energy. The shape score is determined by the position of each ligand atom on the shape scoring grid. The ligand-protein binding energy is taken to be approximately the sum of the van der Waals attractive/dispersive and Coulombic electrostatic energies. To generate the energy score, the ligand atom terms are combined with the receptor terms from the nearest grid point, or combined with receptor terms from a virtual grid

point with interpolated receptor values. The score is the sum of over all ligand atoms for these combined terms.

For the energy score, the orientation of the ligand may be varied slightly to minimize the energy score. After the initial orientation and scoring of the ligand, a grid-based rigid body simplex minimization is used to locate the nearest local energy minimum. The sphere centers themselves are simply approximations to possible atom locations. The orientation generated by the sphere-atom pairing, although reasonable, may not be minimal in energy. An example of sphere-based docking is the docking program Dock.[107]

2.3.3. Scoring ligand-protein interaction

Once ligand-protein complexes have been obtained, these can be assessed to evaluate the binding between the ligand and the receptor and for this purpose scoring functions have been designed. With a good docking algorithm and a good scoring method it is possible to dock sets of ligands into the receptor and evaluate them and therefore make a selection of which ligand would be a 'good' drug or a 'poor' drug.

Several fast and simplified methods have been designed for 'easy' scoring. A drawback of these scoring functions, which are usually implemented in docking software, is that they make various assumptions and simplifications in the evaluation of modelled complexes and do not fully account for a number of physical phenomena that determine molecular recognition, e.g. entropic effects.[108] Therefore the idea has arisen to combine different scoring functions, something that has become known as consensus scoring.[108]

The different scoring methods can be divided into three classes: forcefield based, knowledge based and empirical based.[105]

- Forcefield based: apply classical molecular mechanics energy functions.
- Empirical based: estimate the binding free energy by the summing interaction terms derived from weighted structural parameters.
- Knowledge-based: represent the binding affinity as a sum of protein-ligand atom pair interactions.

2.3.4. Force-field based scoring

Most forcefield scoring functions consider a single protein conformation, resulting in a fixed protein energy and, therefore, make it possible to omit the calculation of internal protein energy, which simplifies the scoring of a ligand-receptor complex. Molecular mechanics force-fields usually quantify the sum of two energies. These two are the receptor-ligand interaction energy and internal ligand energy (such as steric strain induced by binding). Force-field scoring functions are based on a variety of force field parameter sets, but different scoring functions use different parameter sets. For example, G-Score[110] is based on the Tripos force field but Autodock is based on the AMBER force field, however, the resulting parameters are usually similar.

Interactions between ligand and receptor are most often described by van der Waals and electrostatic energy terms. The van de Waals energy term is given by a Lennard Jones potential function. The parameters of the Lennard Jones potential vary depending on the desired 'hardness' of the potential. Higher terms, such as a 12-6 Lennard Jones potential of D-Score,[111] result in increasingly repulsive potentials and will be less forgiving for close contacts between receptor and ligand atoms. If lower terms are used, such as the 8-4 potential of G-score, the potential becomes softer. Electrostatic terms are accounted for by a Coulombic formulation with a distance-dependent dielectric function that lessens the contribution from charge-charge interactions. The functional form of the internal ligand energy is typically very similar to the protein-ligand interactions energy, and also includes van der Waals contributions and/or electrostatic terms.

Some force-field scoring functions have some major limitations. This is because they were originally formulated to model enthalpic gas-phase contributions to structure and energetics, and do not include any solvation and entropic terms. Forcefields like AMBER, CHARMM and OPLS were developed for biomolecular simulations and include the polarization effects of the solvent environment. However, force-field in scoring functions do not tend to describe this effect as the role of solvent in the binding pocket is largely unknown. Force-field based scoring is further complicated by the fact that it generally requires the introduction of cut-off distances for the treatment of non-bonded interactions.

2.3.5. Empirical scoring functions

These scoring functions are designed to reproduce experimental data, such as binding energies, and/or conformations, as a sum of several parameterized functions. The design of empirical scoring functions is based on the idea that binding energies can be approximated by a sum of individual terms which are not correlated to each other. The coefficients of the various terms are obtained from regression analysis using experimentally determined binding energies and, potentially, X-ray structural information. The functional forms are often simpler than force-field scoring functions, although many of the individual contributing terms have counterparts in the force-field molecular mechanics terms. The appeal of empirical functions is that their terms are often simple to evaluate, but they are based on approximations similar to force-field functions. A disadvantage of these methods is their dependence on the molecular data sets used to perform regression analysis and fitting. This often yields different weighting factors for the various terms. As a consequence, terms from differently fitted scoring functions cannot be combined easily into a new scoring function.

In empirical scoring functions, terms accounting for non-bonded interactions can be implemented in rather different ways. For example, in the early LUDI[112] formulation, the hydrogen-bonding term is separated into neutral hydrogen bonds and ionic hydrogen bonds, whereas ChemScore[113] does not differentiate between different types of hydrogen bonds. Furthermore, the LUDI function calculates hydrophobic contributions on the basis of a representation of molecular surface area, whereas ChemScore evaluates contacts between hydrophobic atom pairs. F-Score adds an additional term to account for aromatic interactions.

Empirical scoring functions can include non-enthalpic contributions such as the so-called *roter* term, which approximates entropy penalties on binding from a weighted sum of the number of rotatable bonds in ligands. ChemScore implements ligand rotational entropy in a more complicated form that describes the molecular environment surrounding each rotatable bond. More complex functions begin to address solvation and desolvation effects. However, terms currently used to approximate entropy or desolvation energy provide only incomplete descriptions of these effects on protein-ligand binding. Other scoring functions that are empirically based are X-Score[111] and F-Score[114]

2.3.6. Knowledge-based scoring functions

Knowledge-based scoring functions are designed to reproduce experimental structures rather than binding energies. These experimental structures are usally ligand-bound crystal structures. In knowledge-based functions, protein-ligand complexes are modelled by using relatively simple atomic interaction-pair potentials. A number of atom-type interactions are defined depending on their molecular environment. As is in common with empirical methods, knowledge-based scoring functions attempt to implicitly capture binding effects that are difficult to model explicitly. Popular implementations of such functions include the potential of mean force (e.g. in PMF[115] and DrugScore[116]), which also includes solvent-accessibility corrections to pair-wise potentials.

A major advantage of many knowledge-based scoring functions is their computational simplicity. This allows fast and efficient screening of a single complex and, thus, is suitable for screening a large compound database. A disadvantage of this method is that their derivation is essentially based on information that is implicitly encoded in limited sets of protein-ligand complex structures. Therefore, these methods are not always representative of the studied ligand-receptor complex.

2.3.7. Consensus scoring

Each of the scoring functions usually works well for a limited set of protein-ligand complexes. However, by combining several scoring functions into a consensus score[117] the accuracy and reliability of scoring ligand-protein complexes can be improved. The different approaches to consensus scoring are:

Voting (intersection). Pass-fail criteria are established for each method whereby a cut-off determines a pass or a fail. The final score is based on how many passes a molecule has. These pass-fail criteria are often arbitrary. Good overall enrichment can be achieved but often at the cost of low recovery rates.

Coarse quantiles (positions in ranges) voting. Each scoring function casts a vote if the score falls in the top quantile of the range of values obtained for that scoring function across the dataset of interest. The consensus score is the total number of votes received. When using a single criterion, there is a large number of tied scores, which can be overcome by also referencing another criterion.

Rank voting. Each method has a predefined number of votes for activity and the top ranking compounds using each scoring function are assigned those votes. Performance depends on how many votes each method receives.

Simple sum ranks (rank-sum, also related to average rank, rank-by-rank). Entries are ranked using each scoring function and the ranks are added up (in case of average ranks, this sum is divided by the number of properties).

Deprecated sum-ranks. Entries are ranked based on each scoring function the worst rank for each entry is dropped and the sum of the ranks is calculated from the remaining ranks.

Worst-best rank. Compounds are ranked based on each scoring function and each entry is assigned the second worst rank from these. In case of a higher number of scoring functions, the third or fourth worst rank might be used.

Weighted sum ranks. Ranks from each method are weighted, with the weight reflecting the importance of the given method. The sum of the weighted ranks is then used as the final score.

Regression schemes (multi-linear regression, non-linear regression). The total score is expressed as the linear (or nonlinear) combination of the individual scores. The coefficients are fitted, so that the performance (enrichment, binding energy, etc.) of the equation is optimal.

Unfortunately there is no consensus-scoring function which is regarded as best performing. Therefore, the best method can only be found by experiment and should be validated using a validation set that contains experimental data.

2.3.8. Ligand-protein interaction plot

The hydrogen bonds and hydrophobic interactions between the ligand and the protein can be easily determined using distance measurements[118] Between the hydrogen atom (H), the donor atom (D) and the hydrogen acceptor (A). Furthermore, an angle can be measured between hydrogen and acceptor by including the atom attached to the acceptor (AA). The criteria that can be used for a hydrogen bond between the ligand and the protein are an H-A distance of < 2.7 Å, the D-A distance is < 3.3 Å, the D-H-A angle is > 90° and the H-A-AA angel is > 90°, where the AA atom is the one attached to the acceptor, usually preceding it along the amino acids chain. The non-bonded hydrophobic contacts between are determined just for carbon atoms that are less than 3.9 Å apart.

This method has been implemented in the program LIGPLOT[118] and produces a 2D-plot between the ligand and the protein. This program works by reading the 3D coordinates and identifies the hydrogen bonds and the hydrophobic interactions. Then the rotatable

bonds are identified, the ring groups are flattened, the structure is unrolled, cleaned up and a 2D plot is created.

2.4. Protein simulation

A basic protein simulation procedure consists of several steps including energy minimisation, equilibration and dynamics. The minimisation is applied to obtain a stable point with a minimum of potential energy. This structure needs to be brought up to the temperature of interest and, hence, an equilibration period of dynamics is performed using molecular dynamics[118] where a restraint is placed on the protein backbone and gradually released in several steps to allow the protein to come to equilibration in its environment. The last step is to perform a molecular dynamics run to observe the time-dependent properties of an atomistic model which can be the movement of a protein over time.

2.4.1. Forcefields

Once a structure has been determined, the energetics of conformations, molecular binding and the dynamics of the system can be examined. To do this, a force field[121] which represents the energetics of interactions of the atoms within our system is needed. The energy for a given nuclear configuration can be calculated by writing the total energy as a function of nuclear coordinates. In general the energy is the sum of bond, valence angle, torsional and non-bonded interactions terms (Eq. 2.28).

$$E_{tot} = E_{bond} + E_{angle} + E_{torsion} + E_{v/dWaals+elec} \tag{2.28}$$

These terms are characterised by several analytical functions, examples of which are below (Eq. 2.29).

$$E_{tot} = \sum_{bonds} K_R (R_{ij} - R_o)^2 + \sum_{angles} K_\theta (\Theta_{ij} - \Theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\Phi - y)] +$$
$$\sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{4\pi\varepsilon R_{ij}} \right) \tag{2.29}$$

Where K are the harmonic constants, R_{ij} is the bond length, Θ_{ij} is the bond angle, R_0 and Θ_o are reference values, V_n is the torsional barrier to rotation, Φ is the torsion angle, y is the phase factor, A and B are atom constants, q is the atomic charge and ε_0 is the environment dielectric constant.

The bonded terms represent the energy penalty for distorting away from an equilibrium distance using a harmonic function. The angle term similarly uses a harmonic term to give rise to a penalty for deviation from the equilibrium angle. The torsional potential gives rise to the energy as a function of the torsional angle and is hence cyclic. The periodicity of this function is given by n, which can be adjusted to give minima every 180 (n=2), 120 (n=3) etc. Torsional angles are particularly important in proteins where this is the primary flexibility for conformational change.

The non-bonded interactions are characterized by the electrostatic and the van der Waals interactions and calculation is generally time consuming. The van der Waals term consists of a short range repulsive and a longer range attractive component and are described using the 6-12 Lennard-Jones form. The interaction between point charges is given by the Coulomb potential. In order to treat the long range nature of the electrostatic interactions within a periodic system, the Ewald sum method is used (section 2.4.4.4.).

Several force fields have been derived for DNA and proteins, e.g. Assisted Model Building with Energy Refinement (AMBER)[122], Chemistry at HARvard using Molecular Mechanics (CHARMM)[123] or the Groningen machine for Chemical Simulations (GROMACS)[124]. GROMACS uses a united atom force field where hydrogen atoms are treated implicitly, but AMBER and CHARMM allow an all-atom force field where hydrogen atoms are treated explicitly.

2.4.2. Energy minimisation

The lowest energy positions of the atoms corresponds to a stable state of the system and, therefore, provides an ideal starting point for molecular dynamics. With structural energy minimization the geometry is adjusted until a stationary point on the potential surface is found. The minimisation technique algorithms[120] differ from each other in the way in which they use the gradient of the energy as well as in their robustness and search efficiency.

The steepest descent method uses the first derivative to determine the search direction towards the minimum. It is combined with a line search which optimises the potential energy along the search direction. This continues until a minimum is found (gradient is zero) (Eq. 2.26).

$$x_{k+1} = x_k + \gamma_k \, S_k \qquad \text{where } S_k = -g_k / |g_k| \qquad (2.26)$$

Where x_k are the coordinates, γ_k is the distance along the search direction determined from line optimization, S_k is the gradient unit vector and g_k is the gradient at point k. The drawback of this method is that a line optimisation can re-optimise the same directions as a previous line optimisation and hence 'spoil' previous steps. Because of this, the rate of convergence is also slow near the minimum. With the conjugate gradient method (Eq. 2.27) each line search is not along the current gradient but along the line constructed so that it is conjugated to the previous search direction.

$$ V_k = -g_k + \gamma_k V_{k-1} \qquad \text{where } \gamma_k = \frac{g_k \cdot g_k}{g_{k-1} \cdot g_{k-1}} \qquad (2.27) $$

Where V_k is the direction, g_k is the gradient at point k, γ_k is the scalar constant, V_{k-1} is the previous direction and g_{k-1} is the previous gradient at point k. The gradients at each point are orthogonal but the directions are conjugate and hence subsequent line optimisations do not spoil previous line optimisations.

2.4.3. Molecular dynamics simulation

To solve the time dependent motion, Newton's second law of Motion for updating the momentum (velocities) requires the forces on the used atoms (Eq. 2.30).

$$ F_i = m_i a_i \qquad (2.30) $$

Where F is the force, m is the mass and a is the acceleration of atom i. The primary requirement of any MD[119] program is thus to calculate the forces on the atoms. The force on the atoms can be calculated by the derivative of the potential energy with respect to the atom positions which is equated directly to the acceleration. The potential energy for each atom is calculated using the force field.

The equations are integrated using a Verlet Leapfrog algorithm using ½ time steps. The positions at time t are used to calculate the accelerations. Analysis of the trajectory can be performed using root-mean-square deviation (RMSD) which is an average distance measurement of a selection of residues/atoms between different frames. The difference in distance between the starting frame and the frames in the trajectory can be plotted against the time step of each frame and a movement over time plot is created. Other

features that can be measured along the MD trajectory are potential, kinetic and total energy.

2.4.3.1 Verlet Leap frog algorithm

The position, velocity and acceleration needs to be updated after each step. Several methods are available of which the Verlet leap frog algorithm is commonly used.[125]In the Verlet leap frog algorithm (Fig. 2.5) the position of all atoms is performed at half time steps with respect to the velocity and is numerically equivalent to the original Verlet scheme. The positions at time t are used to calculate the accelerations at t. The velocity is updated from $t-\frac{1}{2}\delta t$ to $t+\frac{1}{2}\delta t$ using the accelerations at time t. This will effectively be the average accelerations over the velocities time step. The velocities can then be used to update the positions in a similar way for t to $t+\delta t$.
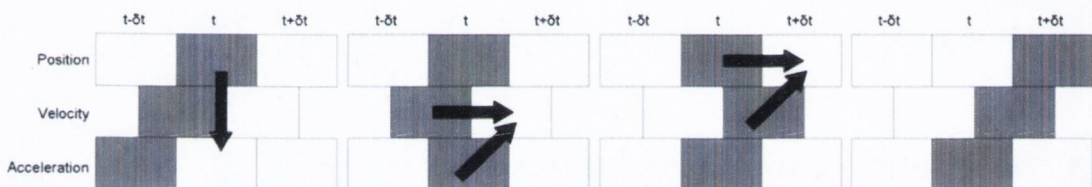
Figure 2.5: Overview of the Verlet leap frog algorithm.

2.4.3.2. Periodic boundaries and minimum image convention

With periodic boundaries conditions[126] (PBC) the simulation cell is surrounded by images of itself. When an atom leaves on one side of the box, it enters the simulation box on the other side. Because there are no walls at the boundary and there are no surface atoms, the size of the simulation box should be large enough to prevent the protein within the box having an interaction with itself via the PBC. To perform the short range interaction cutoff a method called the minimum image convention[127] is used which ensures that each atom does not interact with the same atome twice. An atom cannot interact with an atom that is more than half of the lattice vector away and that is the maximum range for the cutoff.

2.4.4.3. Ewald summation

In most MD simulations and energy minimisations, the long-range interactions (Coulomb interactions) are the most time consuming. Ewald summation[128] was introduced as a technique to sum the long range interactions between particles and all of their images in an infinite array of periodic cells. Long range interactions are evaluated as sums that converge extremely slow. The principle behind the Ewald sum is to convert the summation of the potential energy into two series, each of which converges much more rapidly and a constant term (Eq. 2.31). This is done by considering each charge to be surrounded by a neutralizing charge distribution of equal magnitude but of opposite sign. A Gaussian charge distribution is commonly used.

$$U_{Ewald} = U^r + U^m + U^0 \tag{2.31}$$

The sum over point charges is now converted to a sum of the interactions between the charges plus the neutralizing distributions. This part is the real space sum U^r (Eq. 2.32)

$$U^r = \frac{1}{2} \sum_{i,j}^{N'} \sum_n q_i q_j \frac{erfc(\alpha r_{ij,n})}{r_{ij,n}} \tag{2.32}$$

Where n is an integer, q is the charge, α is the width of the Gaussian distribution, erfc is the complementary error function and r_{ij} is the distance. A second charge distribution is added to the system which exactly counteracts the first neutralizing distribution. This summation is performed in the reciprocal space and is termed U^m (Eq. 2.33).

$$U^m = \frac{1}{2\pi V} \sum_{i,j}^{N'} q_i q_j \sum_{m \neq 0} \frac{\exp(-(\pi m / \alpha)^2 + 2\pi i m \cdot (r_i - r_j))}{m^2} \tag{2.33}$$

The self-term U^0 is a correction term that cancels out the interaction of each of the introduced artificial Gaussian counter-charges with itself (Eq. 2.34).

$$U^0 = \frac{-\alpha}{\sqrt{\pi}} \sum_{i=1}^N q_i^2 \tag{2.34}$$

The particle Mesh Ewald Summation[129] (PMES) method divides the potential energy into Ewald's standard direct and reciprocal sums and used the conventional Gaussian charge distributions. The direct sum is evaluated explicitly using cutoffs while the reciprocal sum is approximated using Fast Fourier Transform (FFT) with convolutions on a grid where charges are interpolated in the grid points. PME does not interpolated but rather

evaluated the forces by analytically differentiating the energies and therefore reducing the amount of memory that is needed.

Chapter III

Homology Modelling: The Use of an Intermediate Amino Acid Sequence in Homology Modelling of the α_1-Adrenoceptors

There are many methods for predicting the future. For example,
you can read horoscopes, tea leaves, tarot cards, or crystal balls.
Collectively, these methods are known as 'nutty methods'.
Or you can put well-researched facts into sophisticated computer models,
more commonly referred to as 'a complete waste of time'.
Scott Adams (1957-) The Dilbert Future

3.1 Introduction

In this chapter the design of homology models for the α_{1A}-, α_{1B}- and α_{1D}-ARs will be described and discussed. Homology models need to be produced because there are no crystal structures available for these receptors. Until recently, the only known crystal structure of a GPCR was that of bovine rhodopsin. Recently, the structure of the β_2-AR has been resolved[51] which has a higher sequence identity to that of the α_1-ARs, making this a better template structure for the generation of homology models. However, at the time of our research this structure was not available and the best option was to use the crystal structure of bovine rhodopsin. The generated homology model can then be used as a starting point for the study of ligand-protein interactions.

One essential feature of homology modelling is the assumption that proteins with similar amino acid sequences have similar structural features. This allows models to be based on the structure of related proteins and these structural features will then be translated into the new model. The issue that always arises is how similar structures need to be in order to generate high-quality models. There are several residues known to be conserved amongst all GPCRs. It is commonly accepted that conserved residues are structurally similar and share similar functions. We have taken some of these features and analysed them in a MD simulation. Two of these features are an ionic lock and a salt bridge. As they play a role in stabilisation and activation of the receptor, they should give some indication about the stability of the protein. However, there are few reference values available to these structural features of GPCRs which can confirm the correctness of our homology models. Still, measuring these features does allow the comparison between the different homology models and the known crystal structures and therefore can be used as a quality assessment of the generated adrenoceptors.

Several different homology models of the three α_1-adrenoceptors have been generated by Pedretti *et al*[62], Leonardi *et al*[63], Evers *et al*[64], Kinsella *et al*[65-67], and Bautista *et al*[68] These are described in section 1.5 of the introduction. However, all of these have taken the standard approach for the generation of their homology models. Some of these studies have performed MD simulations on these models but always within a limited time-scale of less than 2 ns.

One of the new aspects to our research is that we have taken different approaches to the design of homology models. We have taken different methods for generating homology

models and compared them to each other. One approach is by using an intermediate amino-acid sequence and building an intermediate model using this sequence. The other approach is to change the alignment and allow other residues within the sequence of the adrenoceptor and rhodopsin to be aligned. Finally an approach can be taken by combining the two methods. To evaluate the structures several protein analysis tools have been used and also MD simulations were performed to evaluate the structural behaviour of the adrenoceptors over time. These simulations were performed in a solvent mimic as modelling the phospholipid bilayer is difficult to do correctly and requires considerably more time.

3.2. Computational approach

The modelling approach that was taken is shown in Fig 3.1. An alignment was made between the amino acid sequence of rhodopsin and the α_1-AR subtypes. This was initially made with ClustalW[130] and subjected to manual inspection to confirm that the residues that are conserved among most of the GPCRs[131] had been aligned. Using this alignment, Modeller[75] was used to generate 500 different models. From the selected model the two loops that were longer than 15 amino acids and did not have a disulphide bond (ICL-III-IV and ICLV-VI) were optimized using the loop optimization function in Modeller. There were 100 models

| Align sequences |
| Generate models |
| Select model |
| Loop optmization |
| Select model |

Figure. 3.1: Flowchart for the generation of homology models

generated and the best model was selected using protein assessment tools. To validate the homology models after the generation of the complete model and after each loop optimisations, the same protein analysis tools have been used.

Three protein quality assessment programs were used to select the best model. ProCheck[76] was used to perform a stereochemical check. Using a Ramachandran plot every amino acid is classified as having a favoured, additionally allowed, generously allowed or disallowed conformation. Verify3D[77] was used to assess the environment of the residues and gives a score of the percentage of amino acids with a favourable surroundings such hydrophobic or hydrophilic environment and solvent accessibility. Errat[78] was used to determine the ratio of non-bonded interactions between carbon, nitrogen and oxygen of each residue in the model and a score is given for the percentage of amino acids with a favourable ratio of non-bonded interactions. For each model these

assessments were performed and a consensus of which model has the highest quality has been derived. Each model was assessed after the initial generation of the model and after the loop optimisation. After each step in the homology modelling process, the best model was selected and taken into the next step.

Two different alignment approaches (Fig. 3.2) and two methods for producing a homology model have been investigated. First was the optimal alignment of the amino acid sequence between the adrenoceptor and rhodopsin. Using this alignment a homology model was generated (**direct**). Secondly, an intermediate approach was taken by not aligning these two sequences directly, but by creating an intermediate step using the sequence of the adrenoceptor and mutating several amino acids to that of the corresponding amino acid in rhodopsin (**mutation**). This gives the intermediate sequence a higher similarity to rhodopsin and could function as a better template for the adrenoceptor. The selection of amino acids to be mutated was made such that sequence similarity was maintained along the transmembrane residues. Two criteria were used for the selection of residues to be mutated: firstly, no stretch of more than 4 different amino acids lacking similarity in the alignment and secondly, the mutation results in an even spread of identical residues along the whole aligned sequence.

Instead of using the initial alignment (**no-gap**), another approach was taken by introducing gaps in the amino acids sequences (Fig. 3.2) of the adrenoceptor and rhodopsin to obtain an alignment which had a higher amino acid similarity (**gap**). There were three transmembrane helices (TM-II, TM-IV and TM-VII) that would allow for the introduction of gaps, creating a residue alignment with a higher similarity. Gaps were not introduced in the loops due to the difference in length which automatically results in a limited quality of alignment for these structures. This change of alignment results in a different relationship between the amino acid sequence of the adrenoceptor and the x-ray structure of rhodopsin. A final approach was taken by combining the method where an intermediate sequence is introduced and the method where gaps are introduced in the residue chain.

TM-I

rhodopsin	LAAYMFLLIMLGFPINFLTLYVT
α1A	LGVILGGLILFGVLGNILVILSV
α1B	VGLVLGAFILFAIVGNILVILSV
α1D	LGVILGGLYMFGVLGNILVILSV
α1A Mut	LGVILGGLILFGVLGNILVILSV
α1B Mut	LGLVLGAFILFGIVGNILVILSV
α1D Mut	VGVFLAAYMLMGFAGNLLVILSV

TM-II

rhodopsin	ILLNLAVADLFMVFGGFTTTLYTS
α1A	YIVNLAVADLLLTSTVLPFSAIFE
α1B	FIVNLAMADLLLSFTVLPFSAALE
α1D	FIVNLAVADLLLSATVLPFSATME
α1A Mut	YIVNLAVADLLLTFGVLPFSLYFE
α1B Mut	FIVNLAMADLLLSFTVLTTSAYLE
α1D Mut	FIVNLAVADLLLSFGGLPFSLYME

TM-III

rhodopsin	LEGFFATLGGEIALWSLVVLAI
α1A	IWAAVDVLCCTASIMGLCIISI
α1B	IWAAVDVLCCTASILSLCAISI
α1D	VWAAVDVLCCTASILSLCTISV
α1A Mut	IWAFVDVLCCTIAIMGLCIISI
α1B Mut	IWAAFAVLGCTASILSLCAISI
α1D Mut	LWAAVDVLCCTASILSLCVLSV

TM-IV

rhodopsin	AIMGVAFTWVMALACAAPPLVGWSR
α1A	GLMALLCVWALSLVISIGPLFGWRQ
α1B	AILALLSVWVLSTVISIGPLLGWKE
α1D	AAAILALLWVVALVVSVGPLLGWKE
α1A Mut	GLMALACVWALSLVIAIGPLFGWRQ
α1B Mut	AILGVLSVWVLSLVIAIGPLLGWKE
α1D Mut	AAAILALLWVVALAVSVGPLLGWKE

TM-V

rhodopsin	IYMFVVHFIIPLIVIFFCYGQL
α1A	LFSALGSFYLPLAIILVMYCRV
α1B	LFSSLGSFYIPLAVILVMYCRV
α1D	VFSSVCSFYLPMAVIVVMYCRV
α1A Mut	LFSAVVSFYLPLAIILVMYCRV
α1B Mut	LFMSLVHFYIPLAVILVMYCRV
α1D Mut	IFSSVVHFYLPMAVIVVMYCRV

TM-VI

rhodopsin	MVIIMVIAFLICWLPYAGVAFYIF
α1A	TLGIVVGCFVLCWLPFFLVMPIGS
α1B	TLGIVVGMFILCWLPFFIALPLGS
α1D	TLAIVVGVFVLCWFPFFFVLPLGS
α1A Mut	MLGIVVGCFLLCWLPFFLVMPIGS
α1B Mut	MLGIVVGMFILCWLPFFIALPLFS
α1D Mut	MLAIVVGVFVLCWFPFFFVLPLGS

TM-VII

rhodopsin	MTIPAFFAKTSAVYNPVIYIMMN
α1A	FKIVFWLGYLNSCINPIIYPCSS
α1B	FKVVFWLGYFNSCLNPIIYPCSS
α1D	FKVIFWLGYFNSCVNPLIYPCSS
α1A Mut	FKIVFWFFYLNSCINPIIYPCMS
α1B Mut	FKVVPALGaFNSCLNPIIYPCMS
α1D Mut	FKVIFWFFYFNSCVNPLIYIMSS

TM-II (gap)

rhodopsin	-ILLNLAVADLFMVFGGFTTTLYTS
α1A	Y-IVNLAVADLLLTSTVLPFSAIFE
α1B	F-IVNLAMADLLLSFTVLPFSAALE
α1D	F-IVNLAVADLLLSATVLPFSATME
α1A Mut	Y-IVNLAVADLLLTFGVLPFSLYFE
α1B Mut	F-IVNLAMADLLLSFTVLTTSAYLE
α1D Mut	F-IVNLAVADLLLSFGGLPFSLYME

TM-VI (gap)

rhodopsin	MVIIMVIAFLICWLPYAGVAFYIFT
α1A	TLGIVVGCFVLCWLPFFLVM-PIGS
α1B	TLGIVVGMFILCWLPFFIAL-PLGS
α1D	TLAIVVGVFVLCWFPFFFVL-PLGS
α1A Mut	MLGIVVGCFLLCWLPFFLVM-PIGS
α1B Mut	MLGIVVGMFILCWLPFFIAL-PLFS
α1D Mut	MLAIVVGVFVLCWFPFFFVL-PLGS

TM-VII (gap)

rhodopsin	FMTIPAFFAKTSAVYNPVIYIMMN
α1A	FKIVFWLGYLNSC-INPIIYPCSS
α1B	FKVVFWLGYFNSC-LNPIIYPCSS
α1D	FKVIFWLGYFNSC-VNPLIYPCSS
α1A Mut	FKIVFWFFYLNSC-INPIIYPCMS
α1B Mut	FKVVPALGaFNSC-LNPIIYPCMS
α1D Mut	FKVIFWFFYFNSC-VNPLIYIMSS

Figure 3.2: Alignment of bovine rhodopsin with the different α₁-adrenoceptor subtypes (α₁A, α₁B and α₁D) for the direct, mutation, no-gap and gap approach.

For each of the α₁-AR subtypes (α₁A, α₁B and α₁D) four models were generated; **direct-no-gap, direct-gap, mutation-no-gap and mutation-gap**. For each of these models a molecular dynamics simulation was performed with a production run of 4 ns to establish

the structural stability of the protein conformation over time. To prepare our homology models, the program WHAT IF[79] was used to add hydrogens. There is literature reporting the protonation of the Asp in the strongly conserved E/DRY motif as a requirement for activation.[11,15,17,20] We do not expect our model to convert to an active conformation during our MD simulation, but when in a further stage ligand-bound MD simulations are being performed they may be induced to a ligand-activated conformation. Therefore, to observe the stability of the receptor with a possibility of obtaining an active conformation, this amino acid was protonated (α_{1A} Asp^{123}, α_{1B} Asp^{142}, α_{1D} Asp^{193}). The disulphide bond between TM-III and the extracellular loop-IV-V was also defined explicitly in our MD simulations. To simulate the phospholipid bilayer in which the adrenoceptor is normally positioned, a solvent bilayer mimic[132] was used (Fig. 3.3). This mimic represents the hydrophobicity of the phospholipids and still allows the use of water as a solvent for the intra- and extra-cellular compartments. The homology models were positioned in this multilayer box of solvent with the transmembrane helices covered in the chloroform layer representing the phospholipid bilayer, and the intra- and extra-cellular loops in the water layer that representing the extra- and intra-cellular space. The water (TIP3P) and chloroform molecules are part of the AMBER package. The thickness of the chloroform layer was based on the transmembrane helices which should be completely immersed in the chloroform layer. An example of the positioning of α_{1D}-AR in the $H_2O/CHCl_3/H_2O$ bilayer mimic is shown in Fig. 3.3.
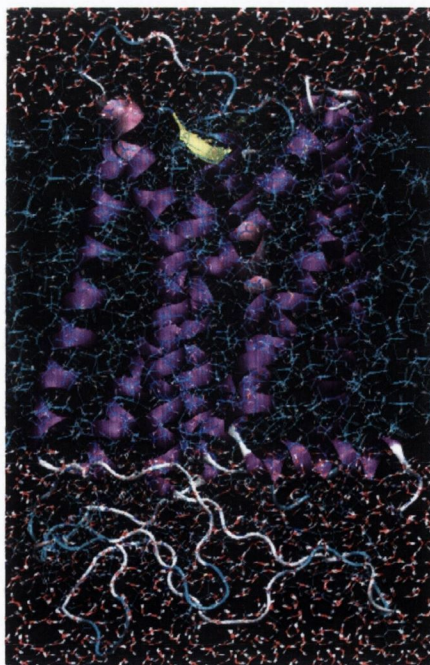
Figure 3.3: Positioning of the α_{1D}-AR in the solvent model consisting of chloroform (blue) and water (red)

A distance of 15 Å between the protein and the edge of the simulation box was used to prevent interactions between receptor periodic images during the MD simulation. Furthermore, the whole system was charge neutralised by adding Cl- atoms using Amber. The MD simulations were performed using the Particle Mesh Ewald Molecular Dynamics (PMEMD) module of AMBER 9.0[122,133] using periodic boundary conditions and PME for evaluation of electrostatics. The ff99sb forcefield that was chosen is a modification of the original ff99 forcefield which prevents overstabilisation of the helices.[134] The structure was minimized using steepest descent for 10.000 steps, followed by a minimisation using conjugate gradient with a maximum of 50.000 steps. The system was heated over 20 ps to a temperature of 300 K using constant volume. The system was then equilibrated at constant pressure using 5 steps of 10 ps each with a restraint of 20, 15, 10, 5 and 0 kcal mol^{-1} on the protein. If a stable energy level has been reached which shows only fluctuation around a particular energy level and not a change to a different energy level, the movement of the protein over time was simulated in a production run of 4 ns that has been performed at constant pressure (isotropic position scaling) and a temperature of 300K (Langlevian dynamics) for each of the models. Snapshots of the adrenoceptors were obtained at 2 and 4 ns and an energy minimisation was performed on these structures.

To analyse our results, the quality of the homology models was scored using ProCheck, Verify3D and Errat as protein assessment tools. This scoring was performed on the initial model generated in the homology modelling process and on the conformations that were obtained after 2 ns and 4 ns of MD simulation. From the production phase of the MD simulation the RMSD over time from helix one to eight (Table 3.1) was extracted to examine the stability of the adrenoceptor models.

Table 3.1: Overview of the amino acids for transmembrane helix one to seven and helix eight that were selected for the RMSD analysis of each of the subtypes of the α_1-AR.

	α_{1A}-AR	α_{1B}-AR	α_{1D}-AR
TM-I	27 - 51	46 – 69	97 – 121
TM-II	64 – 88	80 – 107	134 – 158
TM-III	97 – 127	116 – 146	167 – 197
TM-IV	141 – 163	160 - 181	211 - 232
TM-V	185 – 208	202 – 227	253 – 278
TM-VI	274 – 296	291 - 317	338 - 360
TM-VII	317 – 334	332 – 359	380 – 398
TM-VIII	338 - 346	354 - 363	402 - 410

Analysis of both an ionic lock[17] and salt bridge[135] was performed to determine the stability of the protein and to observe any changes in conformation. One of the key events in the activation of family A GPCRs involves protonation of the aspartic acid in the highly conserved E/DRY motif. The hypothesis behind this activation is that there is an ionic interaction for the α_{1A}-AR between the Asp^{123} and Glu^{267} preventing the protein from moving into an active conformation by constraining the Arg^{124} into a lock. Protonation of the Asp^{123} in TM-III breaks this ionic interaction and allows the Arg^{124} to shift out of the polar pocket leading to cytoplasmic exposure of buried sequences in the second and third intracellular loops allowing a conformational change to an active state.[15] The distances between the three residues should indicate if a stable conformation has been reached or indicate if a change to a conformation that represents a more active conformation takes place.

A salt bridge is described between an Asp in helix three and a Lys in helix seven[135]. During agonist-dependent activation of the α_1-AR, this salt bridge is disrupted to allow the protein to move into the active conformation. Measuring the distance between the sidechains of the Asp^{106} and Lys^{309}, for α_{1A}-AR should give an indication of the behaviour of this salt bridge. Because our MD simulations are performed without a ligand docked in the receptor, conversion to a different state is not likely and therefore limited change of the distance within the saltbridge should be observed.

Reference distances within the ionic lock can be obtained from the crystal structures of rhodopsin (pdb:1u19) and the β_2-AR (pdb:2rh1)[51] (Table 3.2) by measuring the distances between the sidechains of those residues. Rhodopsin has a Glutamic acid (E) positioned in the E/DRY motif while β_2-AR has an Aspartic acid (D) positioned in this motif. To avoid confusion of which oxygen of the carboxylate-group of the amino-acid interacts, the distances between the center carbon atoms of the carboyxlate-groups are measured.

Table 3.2: Distances within the ionic lock for rhodopsin (left) and the β_2-AR (right)

Rhodopsin		β_2-AR	
Glu^{134}-Arg^{135}	4.70 Å	Asp^{130}-Arg^{131}	4.84 Å
Glu^{134}-Glu^{247}	7.20 Å	Asp^{130}-Glu^{255}	14.80 Å
Arg^{135}-Glu^{247}	4.29 Å	Asp^{131}-Glu^{255}	13.02 Å

For the saltbridge there is no Aspartic acid in TM-III in rhodopsin and there is also no Lysine positioned at the top of TM-VII. This makes the presence of this saltbridge in rhodopsin questionable. For the β_2-AR however, these two residues Lys and Asp are placed in the same position as our homology models and therefore a reference distance of 14.83 Å between the sidechains of those two residues can be obtained. Similar to the ionic lock, distances between the center carbon atoms of the carboyxlate-groups are measured.

3.3. Results

The four different homology models generated for each of the three different α_1-AR subtypes (α_{1A}-, α_{1B}- and α_{1D}-AR) will be discussed below.

3.3.1. The α_{1A}-AR homology model

Following the Procheck stereochemical analysis (Table 3.3) little difference emerges between each of the models. The initial mutation-gap model has no residues classified as disallowed, but five residues classified as generously allowed. When summing up both, it has the highest number of residues classified in these categories. The other three models provide similar scores with a variation of only one residue classified more or less in the categories disallowed or generously allowed. Of those three models, the mutation-no-gap model would be the most favourable. When observing the models over time, the direct-no-gap model provides the best result. Both of the direct models showed an improvement of the Procheck score after 4 ns of MD simulation compared to the initial homology model. This is in contrast to both mutation models where a worsening in score is obtained at 2 or 4 ns in comparison with the starting structure.

Verify3D shows a larger range of scores (Table 3.3) for each model with a difference of ~20% between the models. The best results for the initial homology models are obtained for the mutation-no-gap and the direct-gap model. The MD simulation appears to result in a decrease of the Verify3D score over time. The decrease in score differs for each model. After 4 ns the mutation-no-gap and mutation-gap models obtain the most favourable scores.

Errat (Table 3.3) provides similar scores for all models with little variation at the start of the MD simulation. The mutation-no-gap model has the best score of all models. During the MD simulation there is a strong increase in the score by Errat with different results

obtained at 2 and 4 ns, the mutation-gap model obtaining the most favourable score. However, the large increase in score for favourable non-bonded interactions over a short period of time suggests that small changes in structure can have a large influence on the Errat score and therefore a selection based only on this score is not recommended.

Table 3.3: Analysis of the α_{1A}-AR homology models performed with ProCheck, Verify3D and Errat on the initial homology models on models extracted from the MD simulation after 2 and 4 ns.

	Direct, no gap			Mutation, no gap		
	Initial	2 ns	4 ns	Initial	2 ns	4 ns
ProCheck, favoured:	254	241	246	254	240	246
ProCheck, additional allowed:	24	39	32	24	38	30
ProCheck, generously allowed:	1	1	2	2	2	4
ProCheck, disallowed:	2	0	1	1	1	1
Verify3D (%):	41.61	44.10	31.37	52.48	48.76	41.93
Errat (%):	67.00	85.39	87.84	70.10	85.05	83.46

	Direct, gap			Mutation, gap		
	Initial	2 ns	4 ns	Initial	2 ns	4 ns
ProCheck, favoured:	256	239	231	259	242	243
ProCheck, additional allowed:	21	36	47	17	32	33
ProCheck, generously allowed:	2	2	1	5	3	2
ProCheck, disallowed:	2	4	2	0	4	3
Verify 3D:	53.42	42.86	33.23	48.76	47.52	40.06
Errat:	65.25	74.54	82.35	64.88	88.89	93.71

The RMSD analysis (Fig. 3.4) shows stable helices with some fluctuation for all of the models except the mutation-no-gap model. This model shows a drift of the helices, particularly after 2 ns. None of the helices that show movement stabilize after a set amount of MD simulation, but keep moving during the whole simulation. This indicates that the mutation-no-gap model, in comparison to the other models, does not reach a stable protein structure and therefore this homology model should not be selected. This is in agreement with ProCheck which also shows a decrease in quality of the model after 2 ns.

a) direct, no-gap

b) mutation, no-gap

c) direct, gap

d) mutation,gap
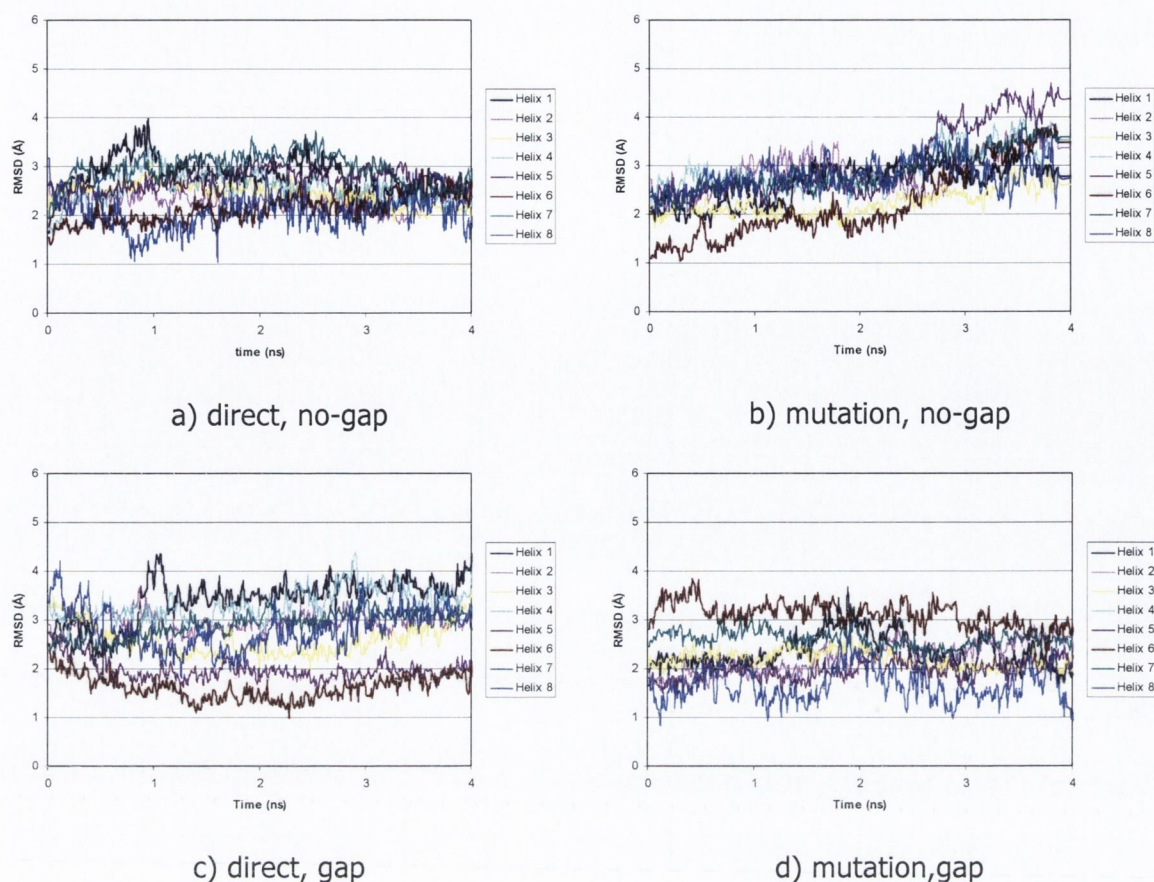
Figure 3.4: RMSD analysis of the MD simulation of the α_{1A}-AR homology models. The RMSD for helix 1-8 is shown as a function of the production run of 4 ns. Structures were initially fit to the starting structure.

Analysis of the MD simulations can be performed by observing the distances within the ionic lock for the residues Asp[123], Arg[124] and Glu[267] over time (Fig. 3.5). The distances between Asp[123]-Glu[267] and Arg[124]-Glu[267] are highly correlated and both show a similar pattern of movement over time. As a change in this ionic lock is needed for changing conformation, these data suggest that a conformational change does not occur during the MD simulation. The distance between Asp[123]-Arg[124] shows no changes. However, the difference in distance of Asp[123]-Glu[267] and Arg[124]-Glu[267] shows a different pattern for the no-gap models (dir-no-gap and mut-no-gap) than for the gap (dir-gap and mut-gap) models. The final distance between Asp[123]-Glu[267] for the gap models is similar to that of the β_2-AR, but for the no-gap models similar to that of rhodopsin. This indicates that the gap-models are more similar to the β_2-AR but it should be noted that the β_2-AR is ligand bound with the partial inverse agonist carazolol and our homology models are not ligand complexed.
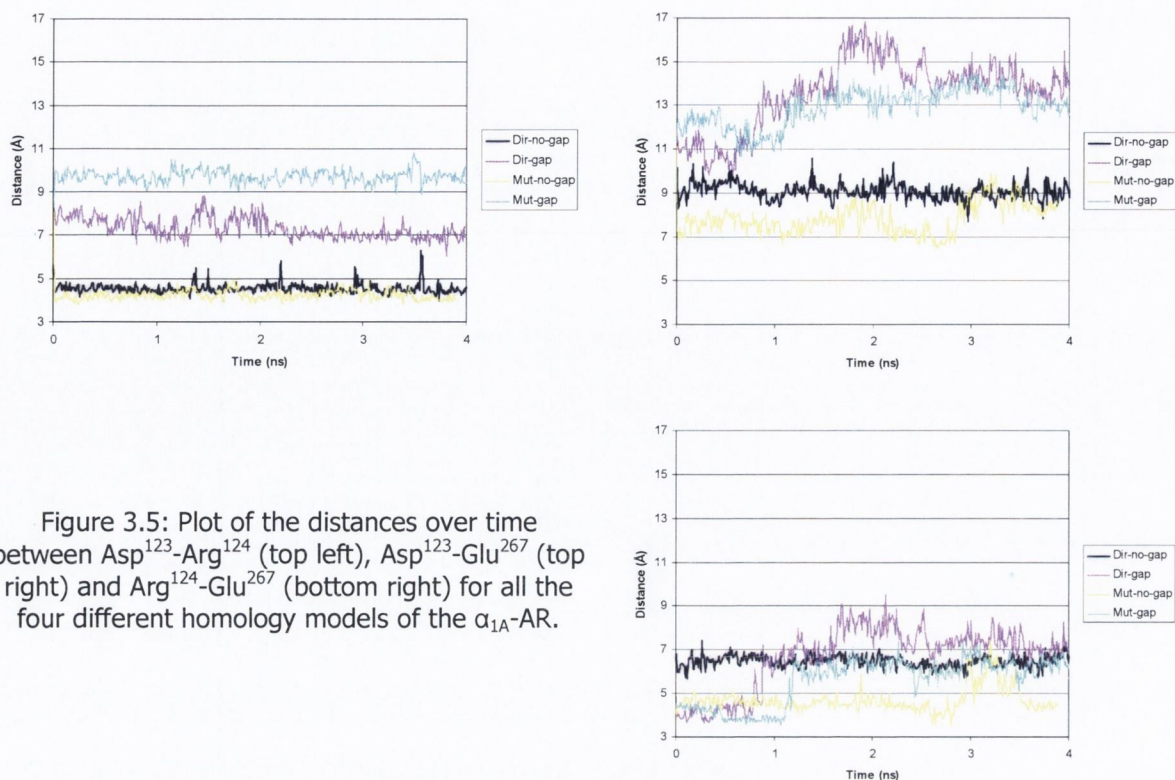
Figure 3.5: Plot of the distances over time between Asp^{123}-Arg^{124} (top left), Asp^{123}-Glu^{267} (top right) and Arg^{124}-Glu^{267} (bottom right) for all the four different homology models of the α_{1A}-AR.

Another feature that can be compared is a saltbridge between residue Asp^{106} and Lys^{309} (Fig. 3.6). This saltbridge contributes to the stability of the protein. The dir-no-gap model shows the most consistent distance between the two residues over time. The other three models (dir-gap, mut-no-gap and mut-gap) show changes in the distance of the salt bridge over time with the mut-gap model showing an increase of 4 Å in distance between 2 and 4 ns. All the models seem to produce a final saltbridge distance of less than 14.83 which is observed for the β_2-AR. However, all models show an interaction close to this distance during the MD simulation, concluding that all models are within this reference distance.
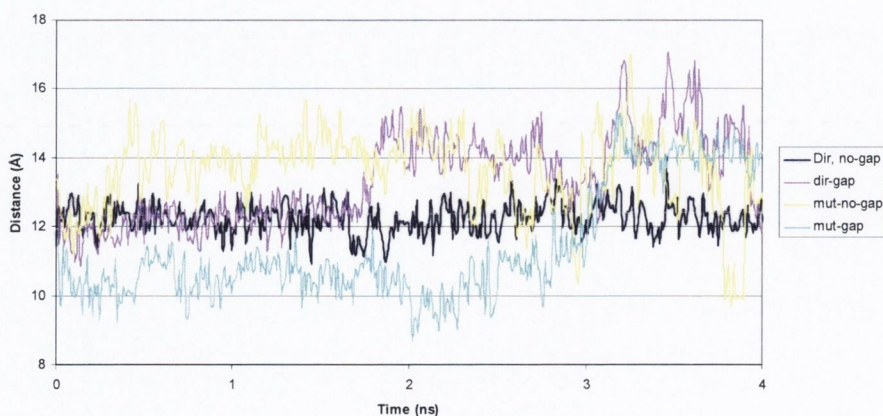
Figure 3.6: Plot of the salt bridge over time between Asp^{106} and Lys^{309}

Taking all results together it can be summarized that the different scoring tools Procheck, Verify3D and Errat, do not provide a similar result for each model and therefore for the selection of the most favourable model more data must be used. Analyses of the helix movement during the MD simulation provides little clarity either, as with the exception of the dir-gap model, three models show a stable structure over time.

More interesting results can be obtained when analyzing structural features such as the ionic lock and the salt bridge that are known to exist in the adrenoceptor. It is clear that a change in the alignment has an influence on these structural features of the homology model. The salt bridge shows the most stable behaviour in the direct-no-gap model making this the best choice of homology model. Given that there is no ligand bound to the receptor, either a very stable distance over time or an initial change in distance resulting in a stable distance afterwards can be expected.

3.3.2. The α_{1B}-AR homology model

Procheck analysis (Table 3.4) shows small differences between each of the models. At the start of the MD simulation the mutation-gap is the only model with no residues classified as disallowed. But taking both disallowed and generously allowed residues into account this would not be the best model as the combined number is higher than either the mutation-no-gap or the direct-gap model. During the MD simulation the stereochemistry worsens for the direct-no-gap and mutation-gap model, but improves for the mutation-no-gap model. The mutation no-gap model is the only model that after 4 ns showed no residues classified as either disallowed or generously allowed and also has the lowest number of residues assigned as additional allowed.

Verify3D (Table 3.4) shows that the initial homology models of both direct and mutation alignment models obtain a similar score, but during the MD simulation an increase of this score is observed for the direct models, while a decrease of this score is observed for the mutation models. This could suggest a better optimization of the positioning of the side-chains for the direct models during the MD simulation. The best score for the initial homology model is obtained by the mutation-no-gap model and after 4 ns of MD simulation with the direct-gap model.

Errat provides similar scores (Table 3.4) for all models with little variation for the initial homology model. The direct models (gap and no-gap) obtain slightly better scores than the mutation models with the direct-gap model being the most favourable according to

Errat. During the MD simulation there is a strong increase in the Errat score and after 4 ns of MD simulation, the direct-gap model obtains the highest score indicating a higher percentage of favourable non-bonded interactions.

Table 3.4: Model analysis of α_{1B}-AR. Model analysis performed with ProCheck, Verify3D and Errat on the initial homology models and after a MD simulation of 2 and 4 ns.

	Direct, no gap			Mutation, no gap		
	Initial	2 ns	4 ns	Initial	2 ns	4 ns
ProCheck, favoured:	252	246	242	261	255	257
ProCheck, additional allowed:	34	38	43	27	31	33
ProCheck, generously allowed:	2	2	2	1	4	0
ProCheck, disallowed:	2	4	3	1	0	0
Verify3D (%):	42.46	35.69	49.54	50.15	34.46	31.69
Errat (%):	68.35	90.00	90.32	62.94	89.40	89.69

	Direct, gap			Mutation, gap		
	Initial	2 ns	4 ns	Initial	2 ns	4 ns
ProCheck, favoured:	261	244	254	262	236	235
ProCheck, additional allowed:	27	44	34	24	51	52
ProCheck, generously allowed:	1	1	1	4	1	1
ProCheck, disallowed:	1	1	1	0	2	2
Verify3D (%):	37.85	49.85	51.08	37.23	29.23	21.54
Errat (%):	68.99	96.48	94.643	63.26	89.15	93.47

All of the four homology models show a different pattern of behaviour when analysing the RMSD of the helices (Fig. 3.7). The direct-no-gap model shows a stable structure with little helical change except helix four. This helix undergoes a conformational change at 1.5 ns but stabilizes afterwards for the rest of the MD simulation. This suggests that a stable conformation of the homology model has been obtained after 2 ns. The mutation-no-gap model exhibits a stable conformation of the protein with none of the helices making large movements after the start of the MD simulation. The direct-gap model shows a stable structure with little change except a movement of helix 1. This helix makes a conformational change at the start of the production run but stabilizes after 1 ns for the remainder of the MD simulation. This suggests that a stable conformation of the homology model has been obtained after 1 ns. The mutation-gap model shows a stable conformation with no large movement of any of the helices over time.

a) direct, no-gap

b) mutation, no-gap

c) direct, gap

d) mutation, gap
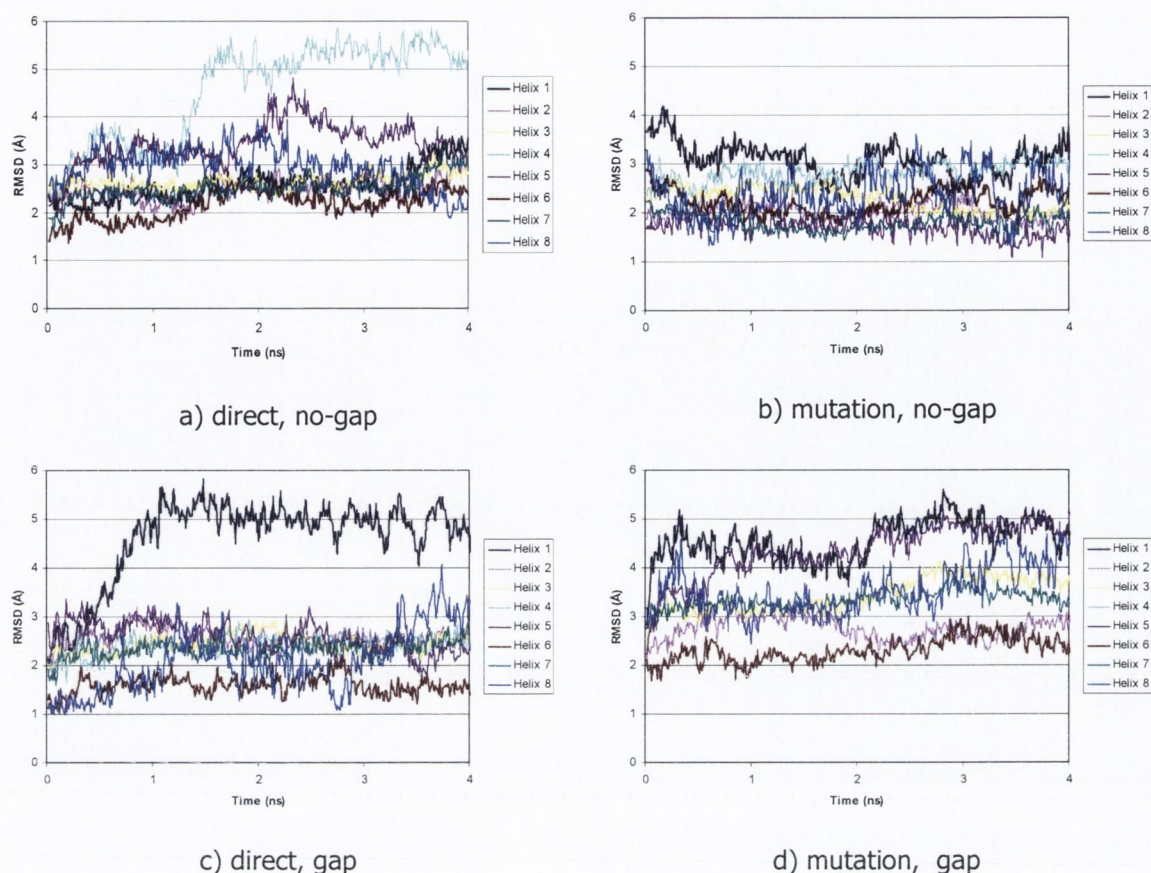
Figure 3.7: RMSD analysis of MD simulation of the α_{1B}-AR homology models. The RMSD for helix 1-8 is shown as a function of the production run of 4 ns. Structures were initially fit to the starting structure.

The ionic lock (Fig. 3.8) is stable for the direct-no-gap and the direct-gap model, with only small changes in distance observed. The mutation-no-gap and mutation-gap model show larger changes. The distance for Asp^{142}-Arg^{143} converges to ~9 Å for all models except the direct-gap model. The distance for Asp^{142}-Glu^{269} converges to ~12 Å and the distance between Arg^{143}-Glu^{269} to ~5 Å for all of the models, except the mutation-gap model. The comparison of these distances with the crystal structure of rhodopsin and the β_2-AR shows for the distance between Asp^{142}-Glu^{269}, a similarity to rhodopsin (7.2 Å) for the mutation-gap model and the other three models, a Asp^{142}-Glu^{269} distance similar to that of the β_2-AR (14.8 Å). However, for the distance between Arg^{143}-Glu^{269} the models dir-no-gap and dir-gap are similar to rhodopsin, but no models relate to the distance observed for the β_2-AR. The combination of these two findings suggests that none of our homology models correspond to both the distances observed in the crystal structures.
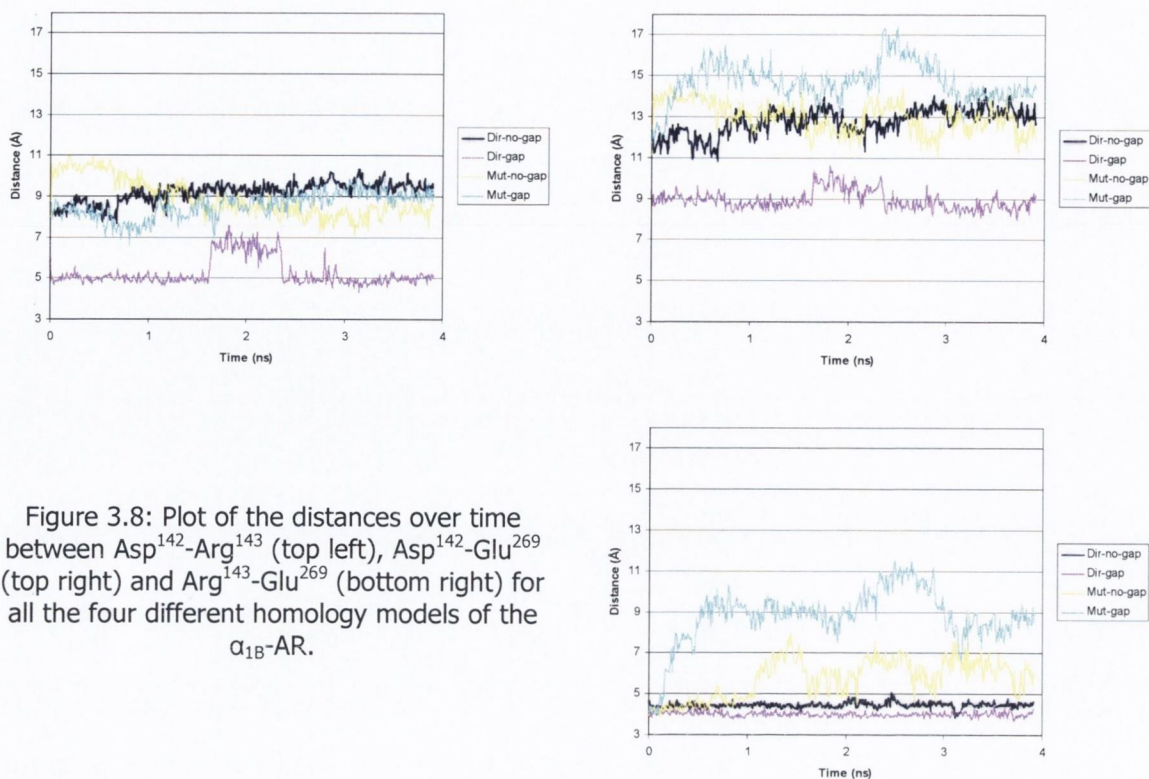
Figure 3.8: Plot of the distances over time between Asp^{142}-Arg^{143} (top left), Asp^{142}-Glu^{269} (top right) and Arg^{143}-Glu^{269} (bottom right) for all the four different homology models of the α_{1B}-AR.

The analysis of the salt bridge between residue Asp^{125} and Lys^{331} (Fig. 3.9) shows that all of the models except the mutation-no-gap model converge towards a larger distance of the saltbridge during the MD simulations. Large fluctuations of this distance are observed during the production run. Both of the mutation models show some movement during the MD simulation, but overall show little movement. The models dir-no-gap and dir-gap would come closest to the distance of 14.83 observed in the β_2-AR.
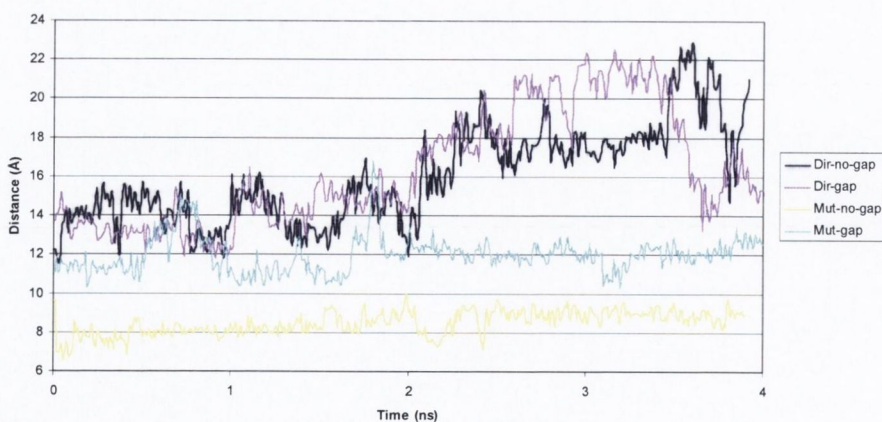
Figure 3.9: Salt bridge between residue Asp^{125} and Lys^{331} for the different α_{1B}-AR homology models.

Taking all results together we can summarize the following about the different models generated for the α_{1B}-AR: analysis with Procheck, Verify3D and Errat shows little differences in scores assigned to each model. The different scoring tools do not provide a similar result for each model and therefore for the selection of the most favourable model more data should be used. Analyses of helical movement during the MD simulation provides little clarity as well, as all of the runs (except model dir-gap) show a stable structure during the simulation with both of the direct models having one helix moving significantly. This helix movement however results into a stable conformation.

Other results can be obtained when analyzing structural features such as the ionic lock and the salt bridge within the homology models. A convergence to a similar distance within the ionic lock and salt bridge is observed for most models. Some models do not show the ability to change the conformations of these features. But in contrast, the direct-no-gap model shows this convergence for both the ionic lock and the saltbridge suggesting that this model has good behaviour during the MD simulation. Therefore the direct-no-gap model was selected.

3.3.3. The α_{1D}-AR homology model

Analysis with Procheck at the start of the MD simulation (Table 3.5) shows that all the models have none or only one residue classified as disallowed. Both the gap models (direct and mutation) have a higher number of residues classified as generously allowed, compared to the no-gap models. During the MD simulation, the score for all models seems to worsen with an increase of the number of residues classified as disallowed rising to two for each model. The mutation models show a stronger increase in residues classified as generously allowed, making the direct models more favourable.

Verify3D (Table 3.5) shows a larger difference between the scores assigned to each model but no model obtains the best score consistently over each of the three time steps. The initial homology model with the best result is the direct-gap model. During the MD simulation the scores decrease for all models except the direct-no-gap model which obtains the most favourable score after 4 ns of MD simulation. The direct-gap model obtains the worst score after 4 ns of MD simulation. This decrease in score indicates that the positioning of the sidechains is not optimal for the direct-gap model.

Errat (Table 3.5) provides similar scores for all models with little variation at the start of the MD simulation and exhibiting the same trend during the MD simulation. The mutation-

no-gap model obtains the best score for the initial homology model, but the direct-no-gap model obtains the best score after 4 ns of MD simulation. During the MD simulation a significant increase in the Errat score is observed for all of the models as stabilising interactions allowed.

Table 3.5: Model analysis of the α_{1D}-AR. Model analysis performed with ProCheck, Verify3D and Errat on the initial homology models and after a MD simulation of 2 and 4 ns.

	Direct, no gap			Mutation, no gap		
	Initial	2 ns	4 ns	Initial	2 ns	4 ns
ProCheck, favoured:	258	249	246	274	255	248
ProCheck, additional allowed:	30	38	42	15	33	37
ProCheck, generously allowed:	1	1	0	1	0	3
ProCheck, disallowed:	1	2	2	0	2	2
Verify3D (%):	48.17	51.22	50.3	52.13	42.68	43.29
Errat (%):	55.17	91.97	90.81	61.69	81.76	84.14

	Direct, gap			Mutation, gap		
	Initial	2 ns	4 ns	Initial	2 ns	4 ns
ProCheck, favoured:	254	244	248	260	253	249
ProCheck, additional allowed:	30	42	39	26	32	35
ProCheck, generously allowed:	6	2	1	4	3	4
ProCheck, disallowed:	0	2	2	0	2	2
Verify3D (%):	58.23	45.12	29.57	57.32	41.46	46.95
Errat (%):	55.03	86.20	89.29	57.05	86.23	85.28

All of the four homology models exhibit a stable helical RMSD pattern (Fig. 3.10). The direct-no-gap model shows a stable structure with little helical change except for helix eight. This helix makes a conformational change from the start of the production run up to 2 ns but stabilizes afterwards for the rest of the MD simulation. Helix eight is not a transmembrane helix and this movement is likely to be due to an adjustment in the bilayer mimic, suggesting an overall stable conformation of the homology model. The mutation-no-gap model shows a stable conformation with none of the helices making large movements after the start of the MD simulation. The direct-gap model shows a stable structure suggesting that a stable conformation of the homology model has been obtained after 2 ns. The mutation-gap model shows a very stable conformation with no large movement of any of the helices over time. The RMSD analysis does not show any significant differences in RMSD behaviour between the direct and mutation models or

between the no-gap and gap models. It does show that all four models obtain a stable structure during the MD simulations.

a) direct, no-gap

b) mutation, no-gap

c) direct, gap

d) mutation, gap
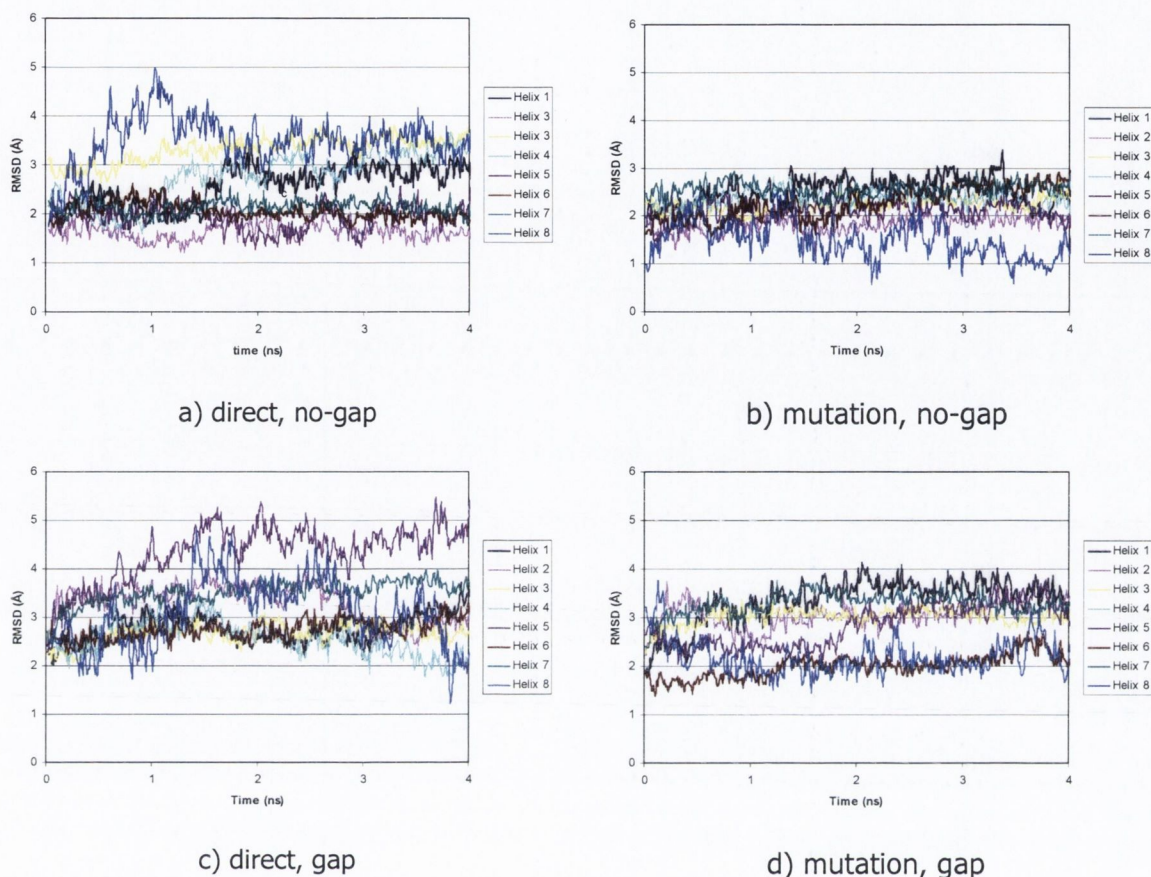
Figure 3.10: RMSD analysis of MD simulation of the α_{1D}-AR homology models. The RMSD for helix 1-8 is shown as a function of the production run of 4 ns. Structures were initially fit to the starting structure.

Distances within the ionic lock between Asp[193]-Arg[194]-Glu[343] are different for each model with the direct-no-gap model having the largest distances for Asp[193]-Arg[194] ($\sim$8 Å), Asp[193]-Glu[343] ($\sim$14 Å) and Arg[194]-Glu[343] ($\sim$7 Å) compared to the other three homology models which show shorter distances, Asp[193]-Arg[194] ($\sim$5-7 Å), Asp[193]-Glu[343] ($\sim$9 Å) and Arg[194]-Glu[343] ($\sim$4 Å). For each model, convergenc to similar distances is observed over time (Fig. 3.11). During the MD simulation, the distance between Asp[193]-Arg[194] converges to $\sim$9-12 Å, the distance between Asp[193]-Glu[343] to $\sim$11 Å and the distance between Arg[194]-Glu[343] to approx. 4 Å. All the models therefore converge to a similar conformation of the ionic lock and hence, this criteria can not be used for selection of the model. Compared to the crystal structure of rhodopsin all of the distances observed for Asp[193]-Glu[343] are longer in the homology models and but are similar for the distance between Arg[194]-Glu[343]. Compared to the crystal structure of β_2-AR, all of our homology models show shorter Asp[193]-Glu[343] and Arg[194]-Glu[343] distances.
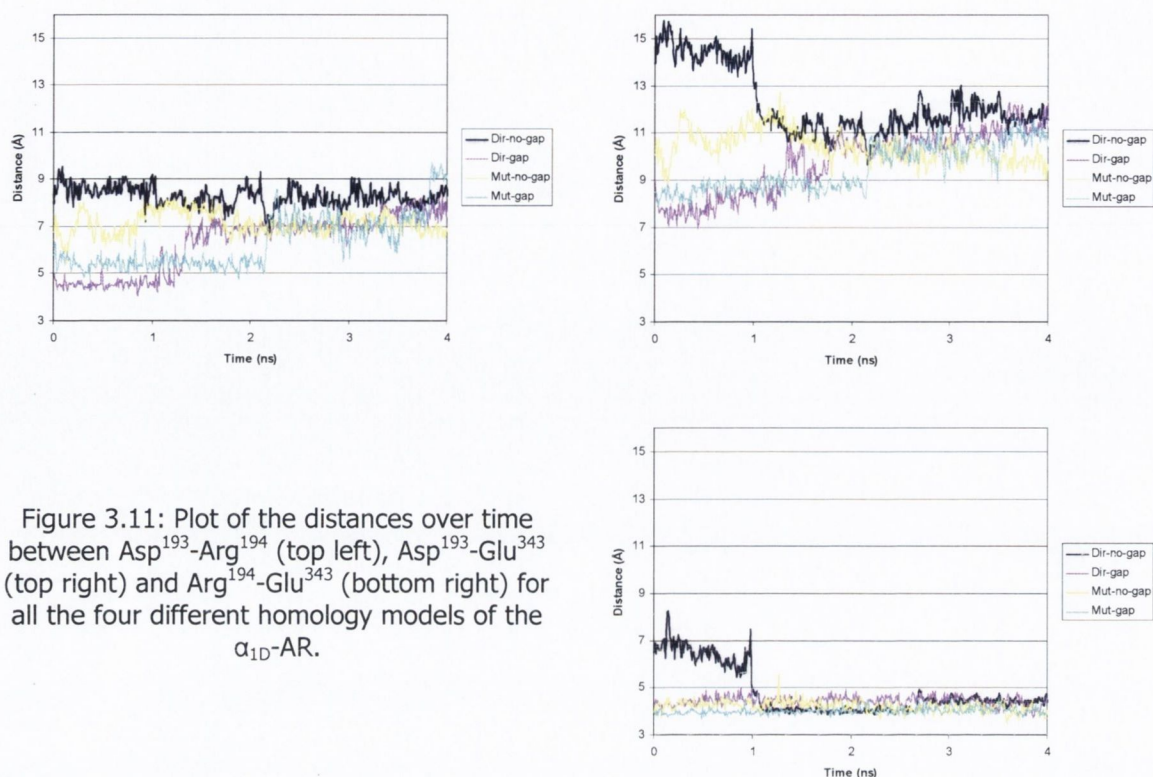
Figure 3.11: Plot of the distances over time between Asp^{193}-Arg^{194} (top left), Asp^{193}-Glu^{343} (top right) and Arg^{194}-Glu^{343} (bottom right) for all the four different homology models of the α$_{1D}$-AR.

The salt bridge between Asp^{176} and Lys^{385} during the MD simulation (Fig. 3.12) shows different distances at the start of the MD simulation for all models with the direct models having a significant shorter distance than the mutation models between those two residues. During the MD simulation a convergence to a distance of ~18 Å is seen for all models except the direct-gap model which shows an unchanged distance of ~4 Å throughout the MD simulation. This indicates that the direct-gap model is in either a stable state or is not able to move to a different state.
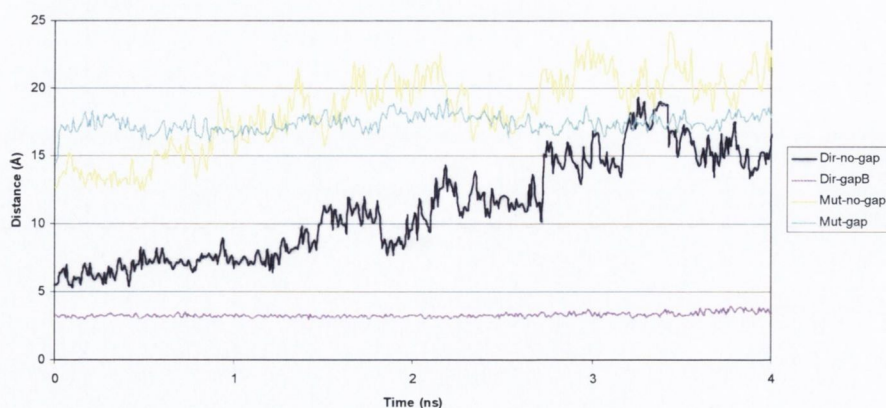
Figure 3.12: Plot of the distance of the salt bridge between Asp^{176} and Lys^{385} for the different α$_{1D}$-AR homology models during 4 ns of MD simulation.

Taking all results together we can summarize that analysis with Procheck, Verify3D and Errat show little differences in scores assigned to each model. Analyses of the helix

movement during the MD simulation provides little clarity as well, as most of the runs (except model dir-gap) show stable transmembrane helices over time. When analyzing structural features such as the ionic lock and the salt bridge a convergence of the distances within these structural features occurs for most models. Within the ionic lock, all models converge to the same distance, but for the salt bridge all models except the direct-gap model converges to a similar distance. The combination of limited structural movement and the convergence to similar distances in the ionic lock and saltbridge leads us to believe that the three models direct-no-gap, mutation-no-gap and mutation-gap, despite having different starting structures evolve into a similar structure when a MD simulation is performed. Therefore the most straightforward method which generates the direct-no-gap models should be selected.

3.3.4. Comparison of the homology models, the crystal structure of rhodopsin and the β-adrenergic receptor.

The initial homology models of the three α_1-AR subtypes are very similar (Fig. 3.13), with all of the helices showing similar positions. This can be expected as they all use rhodopsin as a template and the residues that are conserved amongst the adrenoceptor and rhodopsin are mainly found in the transmembrane helices. The largest difference between the homology models is observed in the extracellular loop between TM-IV and TM-V (ECL-IV-V). This is due to the limited similarity with rhodopsin but also due to the difference in length of the loop between the three subtypes. A manual alignment of the 3d structures was performed because the difference in loops makes the application of automated alignment more difficult.
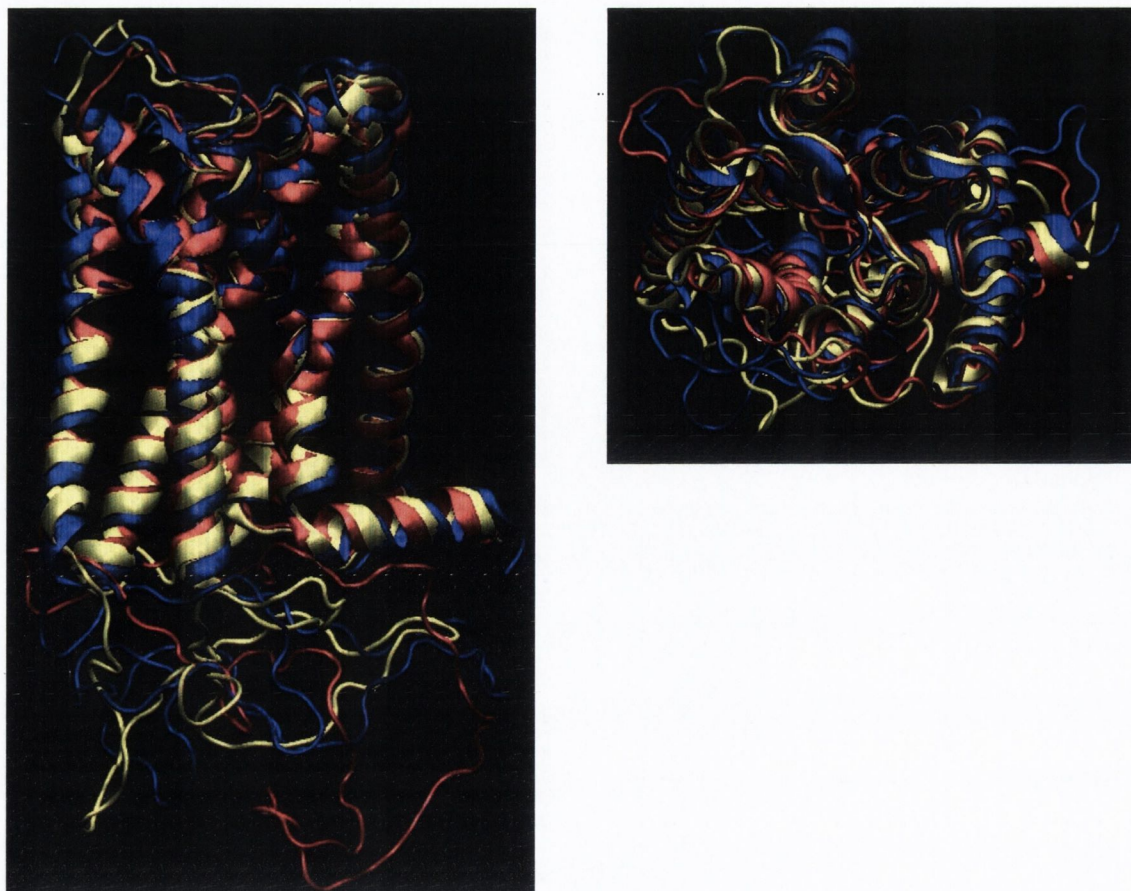
Figure 3.13: Manual alignment of the three direct-no-gap homology models showing α_{1A}-AR (red) α_{1B}-AR (yellow) and α_{1D}-AR (blue) seen from the side (left) and the top (right)

The initial homology model and the model obtained after 4 ns of the α_{1A}-AR (Fig. 3.14), α_{1B}-AR (Fig 3.15) and α_{1D}-AR (Fig 3.16) were aligned to observe the structural differences that have accumulated during the MD simulation. For all the three adrenoceptors, a movement in TM-I is observed in which the extra-cellular part of this helix moves outwards. These movements suggest a widening of the binding pocket. For the α_{1B}-AR (Fig 3.15) a movement of TM-III is observed in contrast to the other two AR-models. Furthermore, the extra- and intra-cellular loops of all models move during the MD simulation. Due to the poor similarity with the template model rhodopsin and the in general being more flexible loops, this was expected.
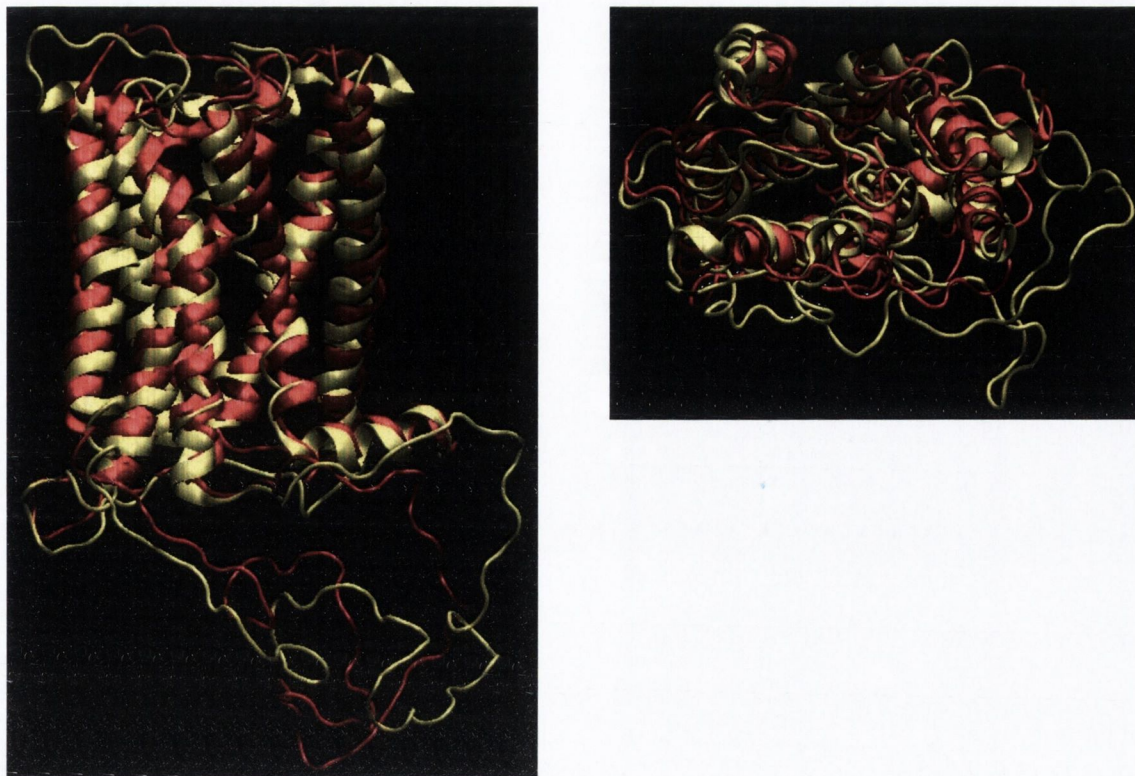
Figure 3.14: Manual alignment of the α_{1A}-AR for the homology model (red) and after being subjected to 4 ns of MD simulation (yellow) seen from the side (left) and from the top (right).
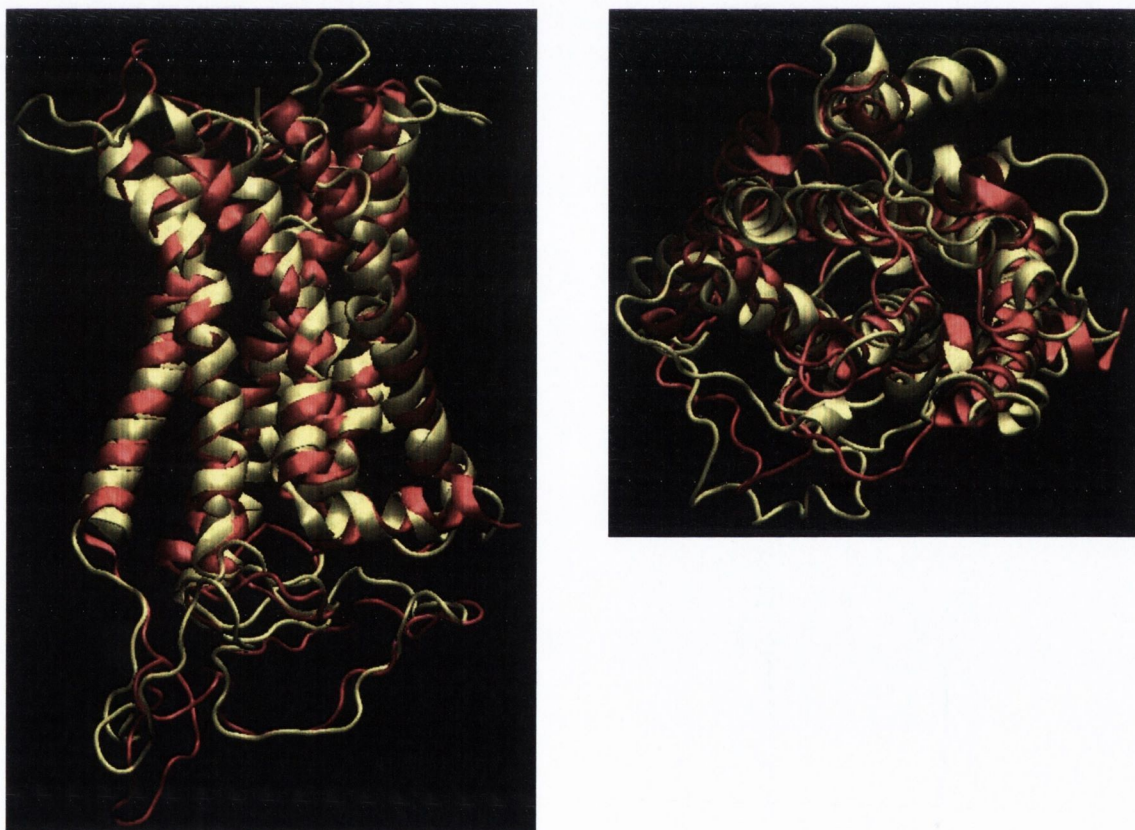
Figure 3.15: Manual alignment of the α_{1B}-AR for the homology model (red) and after being subjected to 4 ns of MD simulation (yellow) seen from the side (left) and from the top (right).
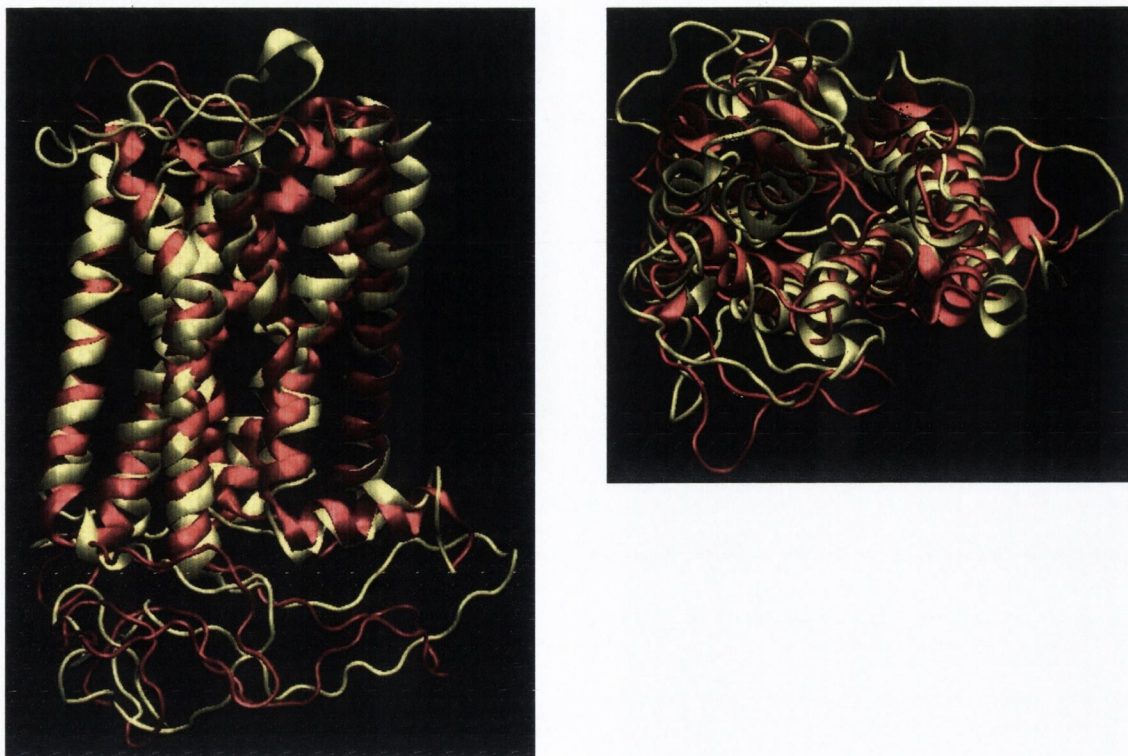
Figure 3.16: Manual alignment of the α_{1D}-AR for the homology model (red) and after being subjected to 4 ns of MD simulation (yellow) seen from the side (left) and from the top (right).

The structure of the three α_1-AR subtypes are similar to each other, but also to bovine rhodopsin (Fig. 3.17) as the overlay of the α_{1B}-AR and rhodopsin shows. For clarity the α_{1A}-AR and α_{1D}-AR were not included in this picture, but figure 3.13 shows that the three AR models are similar. This similarity between the α_1-ARs and rhodopsin is expected because rhodopsin was used as the template in their construction. The largest difference between the homology models and the crystal structure of rhodopsin is observed in ECL-IV-V. This is due to the limited similarity with rhodopsin but also because a separate loop optimization has been performed which is not based on a template structure.
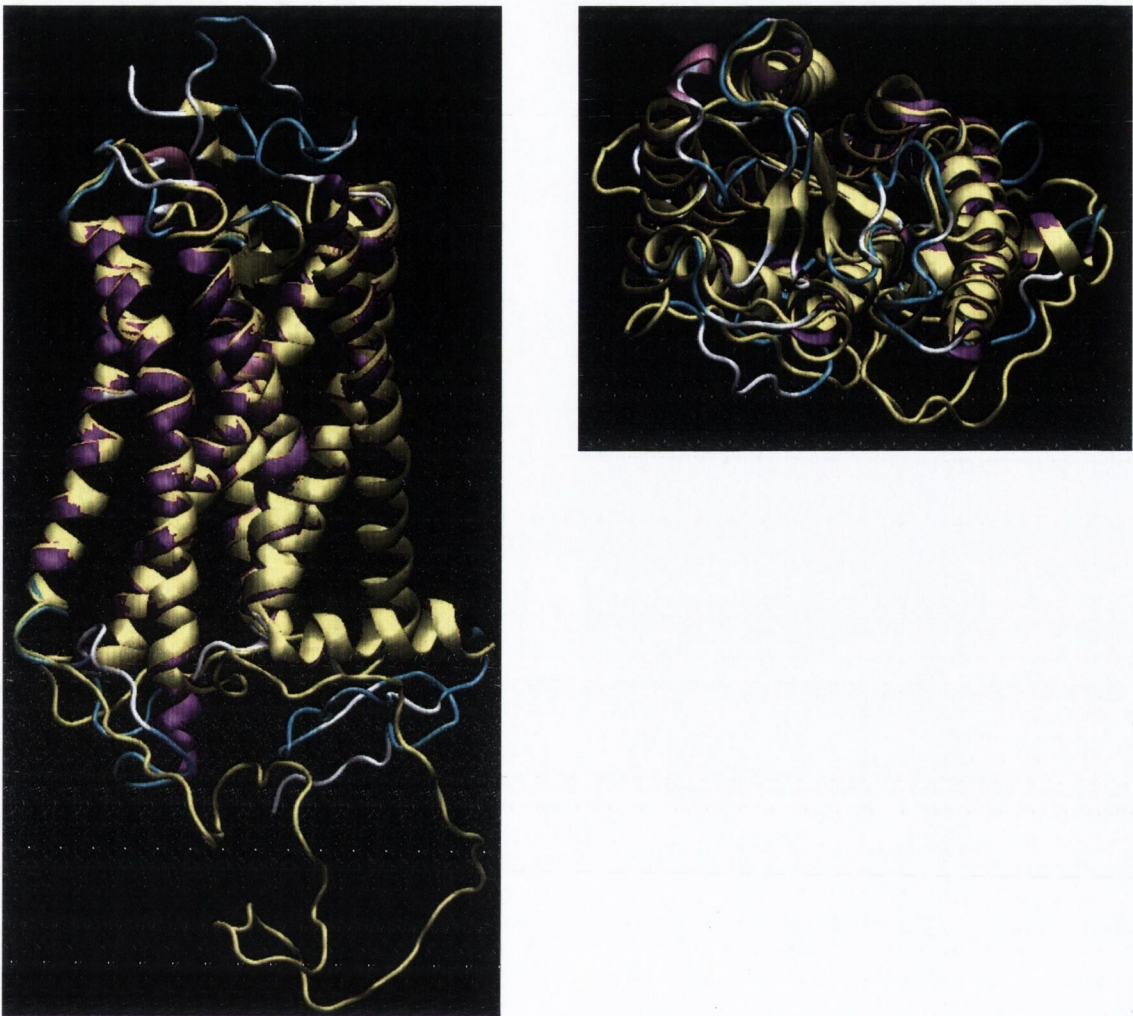
Figure 3.17: Manual alignment of rhodopsin (purple) with the direct-no-gap homology model of α_{1B}-AR (yellow). Because the structures of the three α_1-AR are similar, for visualisation purposes only the α_{1B}-AR is shown.

Several crystal structures of the β_2-adrenoceptor have been recently published.[51,136] As crystal 'pdb:2RH1' has the full structure of the β_2-AR, a structure alignment with the direct-no-gap models, was made by overlaying the crystal structure and the homology model manually (Fig. 3.18). The positioning of most helices within both structures is similar, with the exception of TM-I and TM-III. In the crystal structure of β_2-AR, these two helices are pointing outwards from the centre of the protein when compared to the homology models of the α_1-ARs. The inwards positioning of TM-I and TM-III decreases the volume of the binding pocket and, therefore, make successful docking more difficult. The loops are considerably different, especially the longer loops of ECL-IV-V and the ICL-V-VI. The ECL-IV-V shows the formation of a helix in the crystal structure of the β_2-AR, whereas our homology models do not. This is because this helix is not observed in rhodopsin on which our homology models are based. The intracellular loop between TM-V and TM-VI of the β_2-AR and our homology model show no conformational similarity. This

is because there is a great variety of the length of this loop and of residues within this loop amongst the different adrenoceptor subtypes.
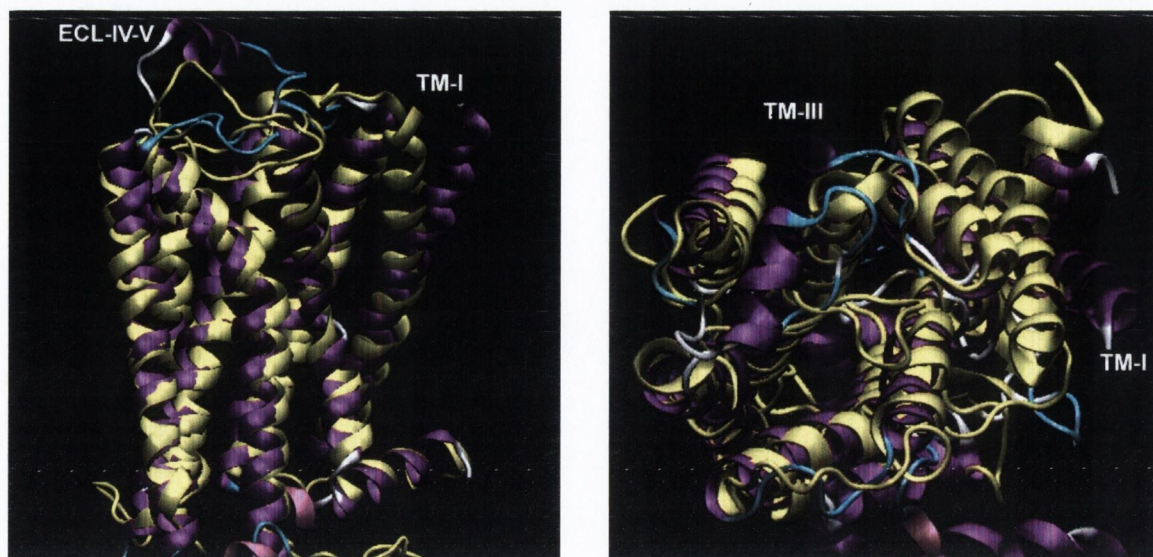
Figure 3.18: Alignment of the homology model of the α_{1B}-AR (yellow) and the crystal structure of the β_2-AR (purple) seen from the side (left) and from the top (right).

The alignment of the initial homology model and the model obtained after 4 ns of MD simulations has shown a movement of TM-I. A similar change was observed when comparing the crystal structure of the β_2-AR with that of the three initial homology models. The alignment of the α_{1A}-AR obtained after 4 ns of MD simulation and the β_2-AR (Fig. 3.19) shows that the adrenoceptor model has moved to a structure more similar to that of β_2-AR. The positions of TM-I in both structures do not align as in the β_2-AR it is still pointed more outwards that the α_{1A}-AR. The position of TM-III has converged to as similar position as in the β_2-AR. The space within the binding pocket formed within TM-I, TM-II, TM-III, TM-VI and TM-VII is still larger in the β_2-AR than in the α_{1A}-AR. This could be because there is a ligand bound in the β_2-AR, but not in our adrenoceptors. However, most helices are positioned in a similar position, suggesting that a more correct α_1-AR conformation has been obtained. The extracellular loop ECL-IV-V appears to adopt a helix conformation in the β_2-AR, something which is not observed in the α_{1A}-AR. The intracelllar loop between TM-V and TM-VII is shown to be different in the α_{1A}-AR and β_2-AR. Furthermore, this loop is different within each adrenoceptor and, even with a known crystal structure of the β_2-AR, would be hard to predict.
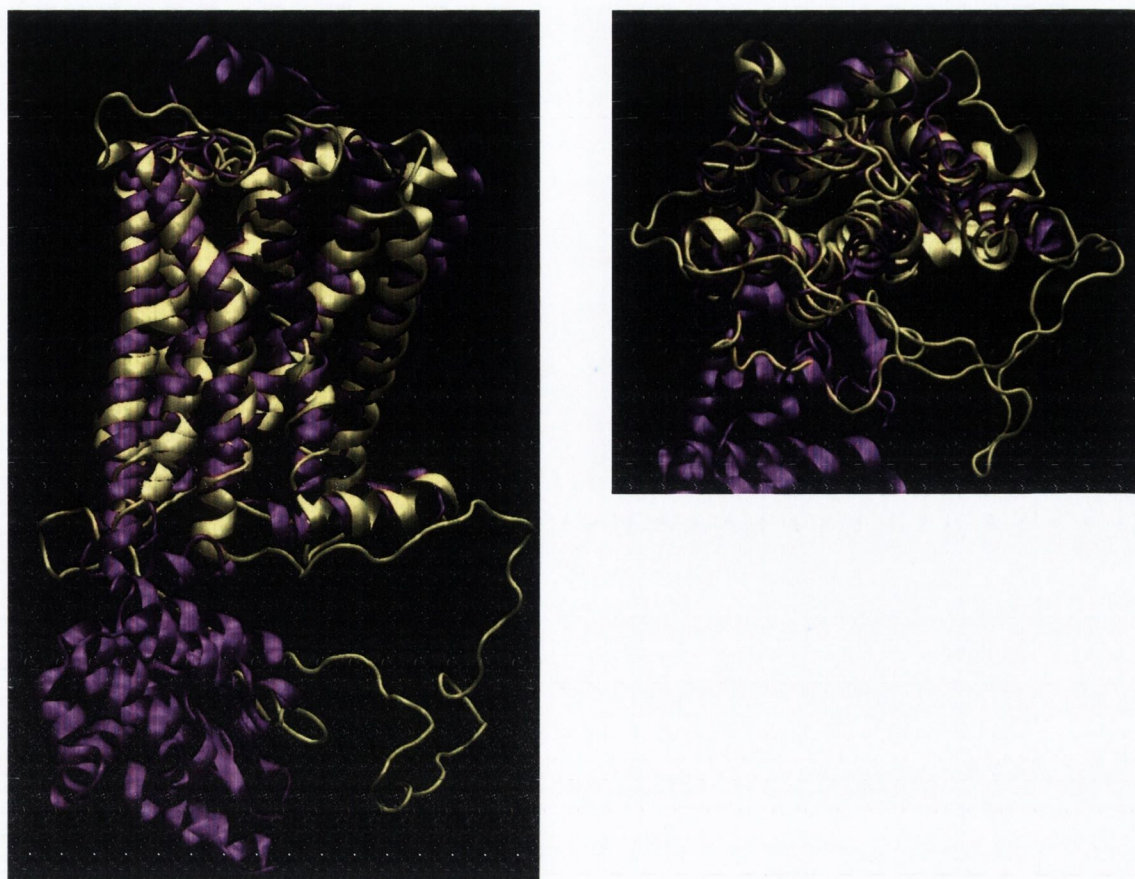
Figure 3.19: Alignment of the model of α_{1A}-AR obtained after 4ns of MD simulation (yellow) and the crystal structure of the β_2-AR (purple) seen from the side (left) and from the top (right).

3.4. Conclusions

The analysis of the generated homology models and the models obtained after 2 and 4 ns of MD simulation indicates that according to the protein assessment tools, the quality of the homology model does not always improve during the MD simulation. After 4 ns of MD simulation ProCheck classifies more amino acids as disallowed or generously allowed. One of the reasons for this could be that the ProCheck method is based on crystal structures which are likely to be in a more static form than models extracted from a MD simulation. Verify3D does not differentiate between the initial homology models and the models extracted from the MD simulation. However, it does not provide a similar score for all the models after performing a MD simulation. For different models, changes in the scores are variable. We are not able to determine any cause for this change in score. When scoring the initial model and the model extracted from the MD simulation with Errat, the latter obtains a significantly higher score. An explanation for this increase in score is that the Errat method is based on non-bonded interactions. The forcefield used during the MD simulation describes the non-bonded interactions and uses them to predict the movement

of the atoms, thereby, optimizing the non-bonded interactions in the model. This leads to a state that is favoured by Errat.

The RMSD analysis of the helices does not reveal much in terms of quality of the homology models, although serious drifts of helix conformations are easily identified. Hence, this method can be used to rule out structures lacking conformational stability. However, the RMSD analysis is not able to determine the quality of models that exhibit generally structural behaviour over time.

We found that known structural features such as an ionic lock or a salt bridge are better ways to assess protein stability. It is observed in several of the MD simulations that distances between certain residues optimize to a set distance. This could be used as a tool for assessing the homology model. However, it is hard to obtain the correct reference value for this measurement as no crystal structures of the modelled α_1-ARs are available. Some reference values can be derived from the crystal structures of bovine rhodopsin and the β_2-AR. However, these differ considerably between the two structures and there is no clear similarity of either one with the results obtained from the MD simulation of our homology models.

The comparison of the different models showed that a good result was always obtained with the direct-no-gap approach which is the most commonly used approach in homology modeling. However, it should be kept in mind that the different methods presented in this chapter can lead to different structures with different properties and alternative models can be generated when the conventional approach produces models with an error.

Our results obtained with ProCheck, Verify3D and Errat show that the mutation model can lead to a model with increased scores. The initial generated homology models and the conformation of these models obtained after the MD simulation show the same trend. The analysis of the trajectories of the MD simulations give very limited results that can be used to assess the quality of the models

The comparison of our homology models to rhodopsin shows very similar positioning of the helices. Some structural differences such as the ionic lock do not always match, but this could be due to a structural difference between rhodopsin and the adrenoceptor. The comparison with the crystal structure of the β_2-AR shows a difference in the positioning of helix TM-I and TM-III which is likely to have on effect on further studies using these

homology models when this model is to be used for docking of antagonists. The binding pocket in our homology models is smaller when comparing them to the binding pocket of the β_2-AR. MD simulations show a widening of this binding pocket, but do not converge fully to the structure of the β_2-AR. This suggests that for generation homology models of the α_1-ARs the β_2-AR would make a better template then rhodopsin, the α_1- and β_2-AR do not necessarily share an exact similar structure.

The quality of homology models of GPCRs is always questionable due to the limited similarity between the GPCR itself and rhodopsin (less than 30 %). However, taking different approaches and optimizing loops separately a reasonable result can be obtained. Further, conformations are obtained by running a MD simulation and observing a stable conformation. While we have shown that there are multiple ways of generating homology models, a straightforward approach appears to be the most successful. The discovery of the crystal structure of the β_2-AR will also contribute significantly to increase the quality of homology models of GPCRs and in particularly adrenoceptors.

The initial homology models that have been generated in this research will be used for docking of antagonists followed by MD simulations of the complexes. This allows the identification of the correct binding of the antagonists to the different adrenoceptor subtypes as the MD simulation should optimize the interaction between ligand and receptor. It also allows the study of conformational changes of the receptors upon binding of an antagonist.

Chapter IV

Ligand Optimization: Proton Affinities And Basicity Of α_1-Adrenoceptor Ligands In Aquous Solutions.

In a few minutes a computer can make a mistake
so great that it would have taken many men, many
months to equal it.

Unknown

4.1 Introduction

In this chapter we will describe and discuss the optimization of the studied ligands. From these optimized ligands, the proton affinities and the acid dissociation constant (pK_a) can be determined. The proton affinities can be used to determine which nitrogen is protonated. The pK_a is a physicochemical property which can influence the binding of the ligand, solubility and membrane permeation. As a ligand to interact with the adrenoceptor it needs to be protonated as it has shown to bind more strongly.[62] Ionized compounds are also more soluble which is beneficial for drugs. However, ionized compounds are less able to penetrate membranes, which make it more difficult to reach the target receptor.

It is important that we obtain minimum energy conformations and the correct protonation state for these ligands to accurately model the interaction with the different α_1-AR subtypes. These optimized conformations can then be used to derive accurate charges. If charges and protonation state are not correctly modelled, this will give rise to incorrect ligand-protein interactions and, therefore, result in less than optimal docking and molecular dynamics studies.

One of the most important steps in the present work was to perform an extensive conformational search followed by a thorough optimization of a large number of possible conformations. This step is time consuming and often simplified by limiting the number of conformations that are optimized. However, the correct minimum energy conformation should allow a better modelling of ligand-protein interactions and, therefore, this is an important step. We have chosen an approach that selects a number of low energy conformations based on structural flexibility and allows further refinement of each conformation. From the collection of optimized conformers the minimum energy conformation can be selected. This minimum energy structure can, then, be used for a proton affinity study where the preference of a nitrogen to be protonated can be calculated. Given that most ligands at physiological pH are protonated, knowledge of this state is crucial as it has an important effect on the interaction with the target protein.

A number of ligands are known to interact with the α_1-adrenoceptors, both agonists and antagonists. The natural agonists for the adrenoceptor are adrenaline and noradrenaline. These are relatively small molecules and, therefore, allow a quick optimization. More complicated structures are observed for the antagonists. And several of them are known to interact with the α_1-ARs, some of them being in clinical use. For four of these

compounds (prazosin, alfuzosin, doxazosin and terazosin) there are experimental pK_a data available[99] and therefore they make a good validation set for our methodology. To complete our dataset, more antagonists were taken from the literature.[53]

Similar work has been carried out previously in our group[99] using a number of ligands, similar to the dataset used in the present work. However, for simulating the solvation effect of water the Onsager model was used, which gave poor results. Instead, in this work, a PCM solvent model is used. The advantage of the Onsager solvent model is that it works more quickly. However, it is also less accurate as shown for some compounds.[99] To maintain a high standard of accuracy in our research, we use a PCM model to mimic the aqueous environment.

The effect of the addition of one explicit water molecule in addition to the PCM model has additionally been studied for our ligand data set. A similar study, where one or two explicit water molecules have been added to an Onsager solvent model has been performed by Campillo *et al.*[137] As B3LYP/6-31G* calculations are time consuming we limit our experiments by including one explicit water molecule in addition to the ligand in both gas phase and PCM.

4.2. Methodology

A selection of antagonists has been taken from the work of Bremner *et al.*[53] The agonists adrenaline and noradrenaline were added to give a total of 28 compounds (Table 4.1, Fig. 4.1) covering a wide range of binding affinities for the different α_1-ARs subtypes. Of these ligands, only six compounds, adrenaline, noradrenaline, prazosin, alfuzosin, doxazosin and terazosin have a known pK_a value.[99] These compounds will serve as a validation set for the determination of computational pK_a values.

Table 4.1: Affinity constants for the different compounds in our data set for each α_1-AR subtype.

	Ligand	Ki (nM) α_{1A}	Ki (nM) α_{1B}	Ki (nM) α_{1D}	Reference
1	Adrenaline	-	-	-	-
2	Noradrenaline	-	-	-	-
3	Prazosin	0.2	0.25	0.32	137
4	Cyclazosin	12	0.13	3.2	138
5	Abanoquil	0.04	0.08	0.04	139
6	REC-15/2615	1.9	0.3	2.6	140
7	Alfuzosin	10	10	3.16	141
8	Doxazosin	3.16	1.0	4.0	141
9	Terazosin	6.3	2.0	2.5	141
10	BMY-7378	250	630	6.3	141
11	SNAP-8719	294	191	1.6	142
12	NAN-190	2.0	15	0.8	143
13	WAY-100635	144	186	63	143
14	RS-100,975	1.0	79	100	141
15	Uropidil	288	1320	1660	143
16	Phentolamine	1.6	7.9	7.9	144
17	SKF-104856	44	63	5.2	143
18	Discretamine	616	360	25	145
19	Corynanthine	142	517	253	146
20	WB-4101	0.16	2.5	0.25	137
21	KMD-3213	0.04	20	2.0	139
22	Benzoxathian	0.2	4.0	0.4	147
23	RS-17053	0.6	16	16	141
24	JHT-601	0.4	1.2	1.2	141
25	Spiperone	7.9	0.5	13	148
26	Indoramin	4.0	40	160	141
27	SNAP-1069	16	200	790	141
28	A-131701	0.22	6.95	0.97	137
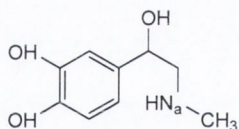
1. Adrenaline
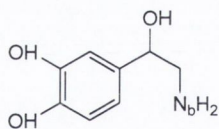
2. Noradrenaline
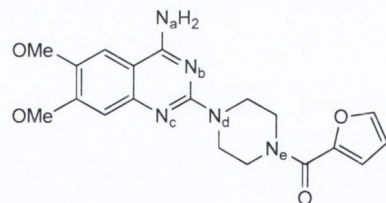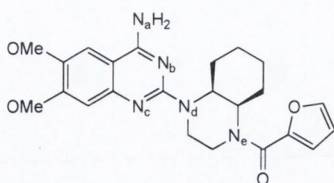
3. Prazosin

4. Cyclazosin
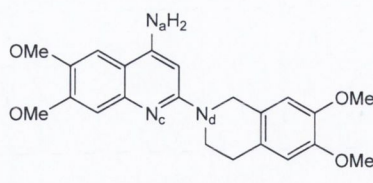
5. Abanoquil
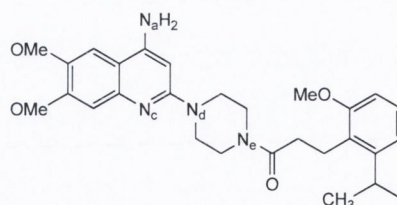
6. REC-15/2615
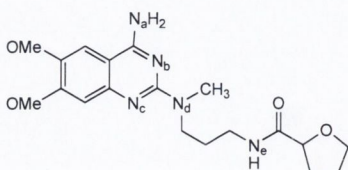
7. Alfuzosin

8.Doxazosin

9. Terazosin

10. BMY-7378

11. SNAP-8719

12. NAN-190

13. WAY-100635

14. RS-100,975

15. Uropidil

16. Phentolamine

17. SKF-104856

18. Discretamine

19. Corynanthine

20. WB-4101

21. KMD-3213

22. Benzoxathian

23. RS-17053

24. JHT-601

25. Spiperone

26. Indoramin

27. SNAP-1069

28. A-131701
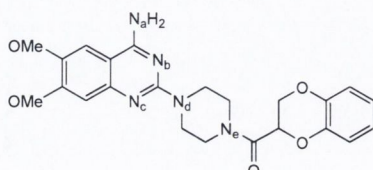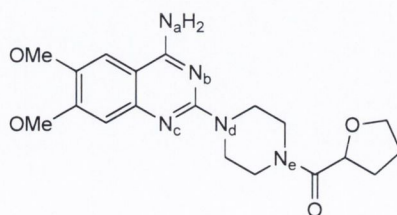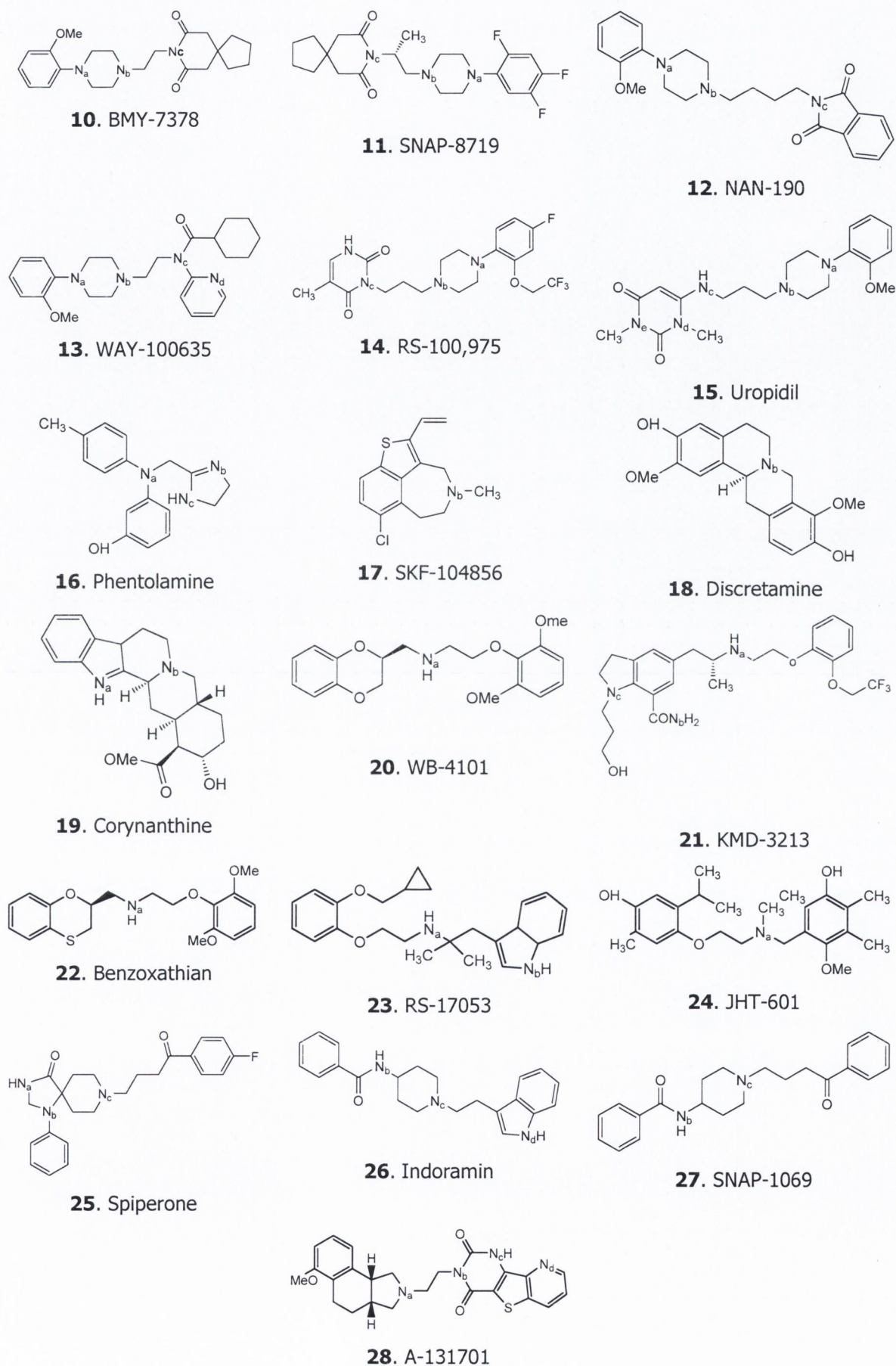
Figure 4.1: Structures of the ligands in our dataset with a letter indication for the different nitrogens.

These ligands were subdivided into six different groups based on their structural similarity. Structural groups used for this classification were the following: catechol; aminopyrimidine; phenylpiperazine; position of N atoms in an aliphatic chain; aminophenyl and piperidine (Fig. 4.2). Hence, a common scaffold was determined and can subsequently be used to derive the site of protonation for similar ligands. The ligand A-131701 (**28**) was not placed in any of the groups due to its structural difference form the other ligands.

Group 1

Group 2

Group 3

Group 4
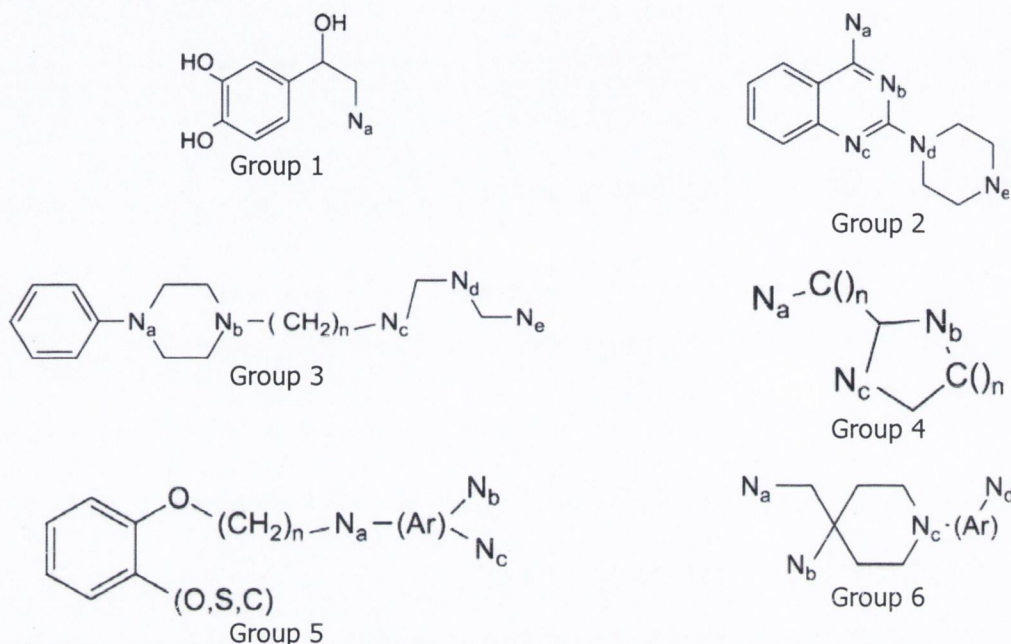
Group 5

Group 6

Figure 4.2: Overview of the different scaffolds of group 1 (**1,2**), group 2 (**3-9**), group 3 (**10-15**), group 4 (**16-19**), group 5 (**20-24**) and group 6 (**25-27**)

To evaluate the different conformations a random conformational search using a Monte Carlo method (Chapter 2.2.1.) was performed with Sybyl7.2[150] applying the minimisation method conjugate gradient, a maximum iterations set of 1000, a gradient of 0.0005, using the Tripos force-field and the Gasteiger-Hückel charges to sample different conformations of the ligands. This gives rise to 200-1500 different conformations for each compound. Each conformation was further optimised using Gaussian03[151] with the AM1 semi-empirical method. From each conformation the energy and the dihedral angles were extracted to determine unique conformations. Two different approaches were used, with the one selected based on the number of dihedral angles that could account for ligand flexibility. When the molecule had five or less rotatable dihedral angles, each pair of dihedrals was plotted against each other, which led to the formation of clusters. Each dihedral angle of the minimized structures produces either two, four, six or eight possible angles. From each cluster the lowest energy conformation was selected.

With more than five dihedral angles accounting for flexibility in the ligand, the energy was plotted against one dihedral angle giving rise to several clusters of which the lowest energy conformation was selected. The selection of lowest energy conformations gave rise to a smaller set of 10-25 conformations for each compound.

Each of the selected AM1 conformations was further optimised with Gaussian03 using the B3LYP hybrid method with the 6-31G* basis set. From the different optimized conformations the electronic energies were extracted and the conformation with the lowest energy was selected. To determine the physiological protonation state, each of the nitrogens of this ligand was individually protonated. Optimization of the neutral and the different protonated conformations was performed using B3LYP/6-31G* level of theory in gas phase and in the PCM solvent model.

Additionally, vibrational frequency calculations were performed at the B3LYP/6-31G* level, to characterise the stationary points as minima. The variation in zero-point energies (ZPE) and thermal corrections from zero degrees to 298K have been considered in the calculations. No scaling factor for the ZPE values has been taken into account.

An alternative strategy was pursued by adding one explicit water molecule[137] (Fig. 4.3) to the ligands interacting through hydrogen bonds. For the unprotonated state the hydrogen of the water molecule was orientated towards the nitrogen of the ligand. As an explicit water molecule has a different effect on the ligand than the PCM model, this should explore if it is possible to find a different minimum energy conformation and, hence, a more accurate pK_a. If multiple nitrogens were present in the ligand, multiple calculations were performed each time, with the water molecule orientated towards a different nitrogen. For the protonated form, the oxygen of the water molecule was orientated towards the hydrogen that was responsible for the protonation of the nitrogen.

Figure 4.3: Positioning of the water molecule that was added for the unprotonated state (left) and the protonated state (right) of the ligand.

4.3. Results

The results are subdivided into several sections describing firstly the conformational search followed by the determination of proton affinities in gaseous phase or using the PCM solvent model. Next is the study of the addition of explicit water molecules and their effect on the energy. The last section shows the use of minimum protonation energies of each ligand to derive pK_a values.

4.3.1. Conformational search

An example of a compound with five or less dihedral angles rotated is prazosin (**3**). This compound has three dihedral angles which are mainly responsible for conformational changes (Fig. 4.4).
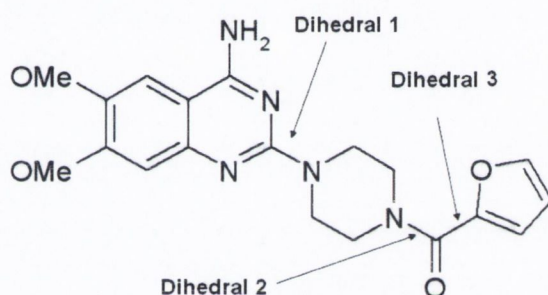
Figure 4.4: Dihedral angles of prazosin that have
the largest influence on different conformations

These three dihedral angles can be plotted against each other giving rise to three different plots (Fig. 4.5). Each plot contains 10-12 clusters representing different conformations. One conformation was chosen from each cluster based on energy. Using this approach, some conformations will be sampled twice in the different plots. For prazosin 20 unique low energy conformations were found which were then used for further optimization using B3LYP/6-31G* level of theory.

a)

b)
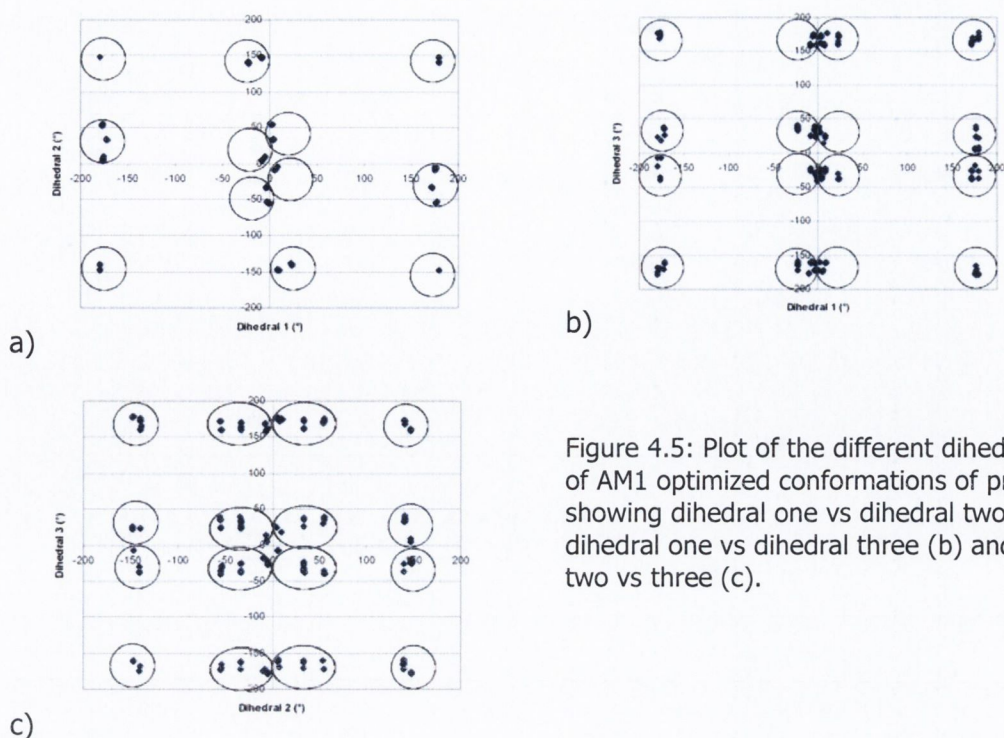
Figure 4.5: Plot of the different dihedral angels of AM1 optimized conformations of prazosin (**3**) showing dihedral one vs dihedral two (a), dihedral one vs dihedral three (b) and dihedral two vs three (c).

c)

A different approach was taken when there were more than five dihedral angles to be rotated in the molecule. If the approach of selecting a conformation from clusters formed by two dihedral angles would have been taken, it would result in a significant increase in the number of low energy conformations. If each of these conformations needs to be optimized using B3LYP/6-31G* either considerable more computer power or time is needed and this is not feasible. Therefore, an alternative approach was taken by plotting each of the flexible dihedral angles against the AM1 energy of the conformation. An example is the study of the conformational space of the compound SNAP-1069 (**27**) which has eight main dihedral angles (Fig. 4.6) accounting for flexibility.
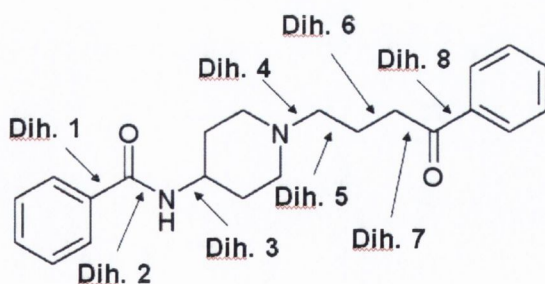
Figure 4.6: Dihedral angles (Dih) of SNAP-1069 that have the largest influence on different conformations

Each dihedral angle is plotted against the energy giving rise to two, four or eight different clusters (Fig. 4.7). The lowest energy conformations can be sampled from each cluster. For SNAP-1069 18 low energy conformations were found which are used for further optimization using B3LYP/6-31G* level of theory.
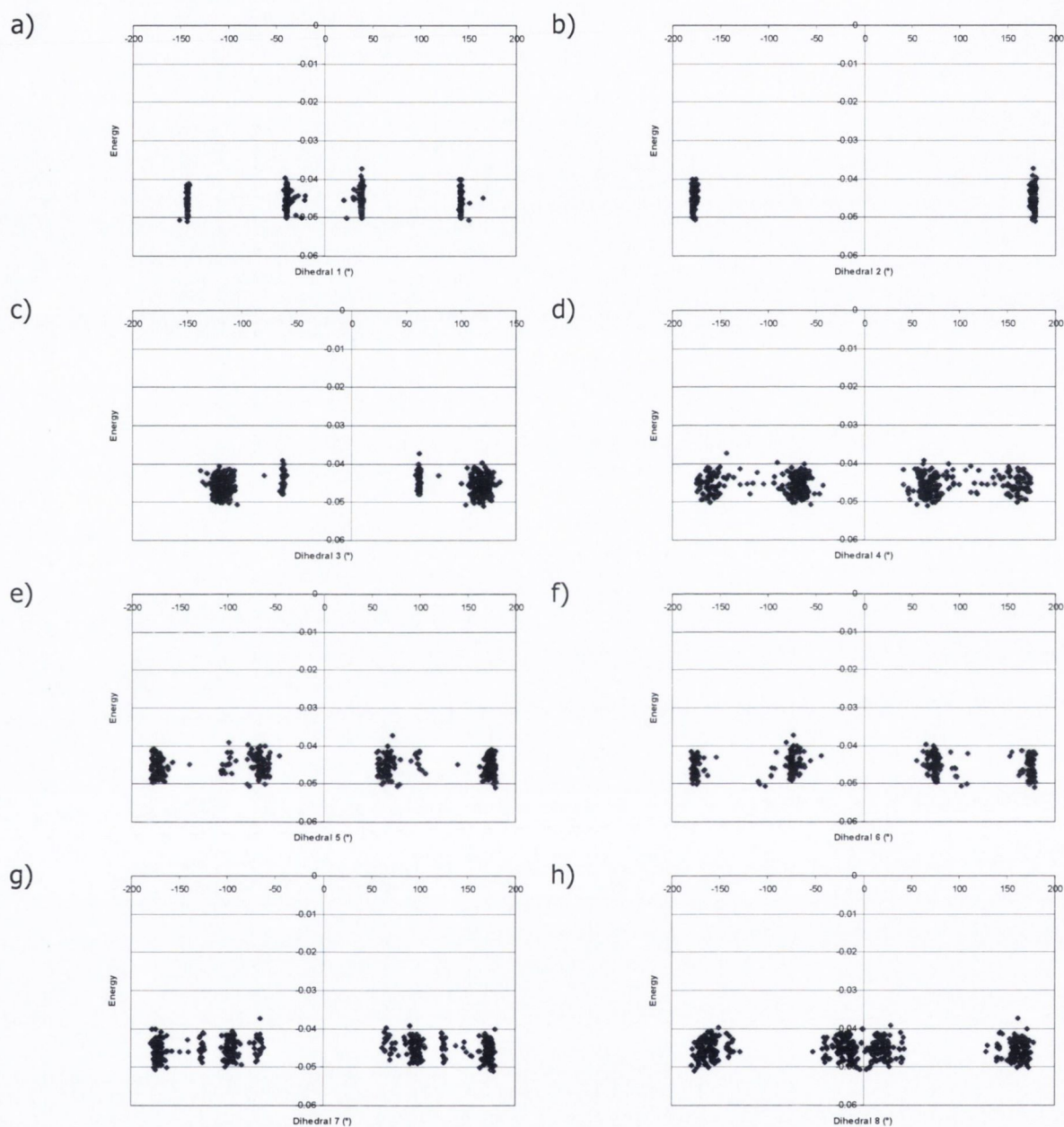
Figure 4.7: Plot of the AM1 energy vs dihedral angles of AM1 optimized conformationhs of (a) SNAP-1069 (**28**) showing dihedral one, (b) dihedral two, (c) dihedral three, (d) dihedral four, (e) dihedral five, (f) dihedral six, (g) dihedral seven and (h) dihedral eight.

4.3.2. Determination of the proton affinities (PAs)

The proton affinity for each compound was calculated as the difference in energy between the neutral and protonated species. The natural agonists adrenaline (**1**) and noradrenaline (**2**) only have one nitrogen in the molecular structure making this the protonation site of the ligand. The PAs (Table 4.2) show a decrease in energy of ~9 kcal mol^{-1} for the 298 K thermally corrected structures compared to the 0 K structure. This observation is made for both the gas phase structures and the PCM solvent optimized structures. The difference between the gas phase optimizations and PCM optimizations at both 0K and 298K is ~53 kcal mol^{-1} for adrenaline and ~57 kcal mol^{-1} for noradrenaline.

Table 4.2: Proton affinities of the ligands in group 1 that share the following structure.
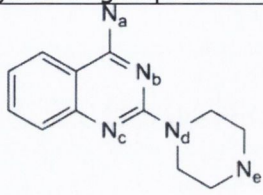
Ligand	Nitrogen	In Vacuo		PCM	
		PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})	PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})
1 Adrenaline	Na	241.460	232.452	294.153	285.410
2 Noradrenaline	Na	236.386	227.517	293.547	284.490
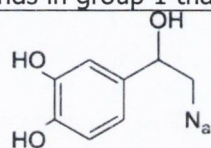
The results in the second group (compounds **3** to **9**) for the gas phase and the (PCM) solvent models at 0 K and 298 K show similar PA results for six out of seven compounds with the nitrogen Nc being favoured for protonation (Table 4.3). The difference with the second nitrogen favoured for protonation ranges from 5 to 15 kcal mol^{-1} with an average of 8 kcal mol^{-1} difference in PA to the second favoured site for protonation. The smallest difference is observed for prazosin with a PA difference of ~5 kcal mol^{-1} for nitrogen Nc over Nb. Alfuzosin (**7**) is the only compound in this group showing different protonation states for the gas phase and the PCM phase. There is a protonation preference for nitrogen Ne in the gas phase with a difference of ~5 kcal mol^{-1} to nitrogen Nc and a preference for nitrogen Nc in the PCM phase with a difference of ~8 kcal mol^{-1}. The compounds in this group have consistent preference for protonation of nitrogen Nc in the PCM solvent phase. Considering that solvation effects are always present, we can assure that the PCM results would be the most accurate, and that we consider nitrogen Nc as the most likely atom to be protonated. The PAs show similar results when comparing the 0 K and 298 K structures with a decrease of ~8 kcal mol^{-1} in energy for the optimized structures in the gaseous phase and a decrease of ~9.5 kcal mol^{-1} for the PCM optimized structures. The difference between the gas phase optimizations and PCM optimizations at 0 K is ~38.5 kcal mol^{-1} and for 298K is ~37 kcal mol^{-1}. It should be noted that the

difference in energy is larger between the gaseous phase and PCM phase optimized structures than between the 0 K and 298 K optimized structures. Given the consistency of the results in this group there is a clear trend that the nitrogen N_c is the selected nitrogen for protonation when the ligand is in water.

Table 4.3: Proton affinities of the ligands in group 2 that share the following scaffold.
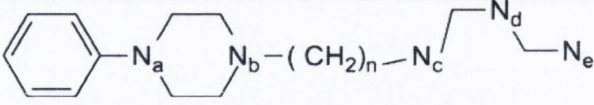
Ligand	Nitrogen	In Vacuo		PCM	
		PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})	PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})
3 Prazosin	Na	222.469	214.060	270.265	262.704
	Nb	243.707	236.572	285.860	279.347
	Nc	255.563	247.284	292.424	284.190
	Nd	242.683	234.559	279.994	270.742
	Ne	230.405	222.357	270.233	262.769
4 Cyclazosin	Na	224.478	216.691	271.350	261.717
	Nb	248.149	240.949	287.140	278.598
	Nc	256.518	249.086	293.372	284.450
	Nd	247.398	239.826	281.055	270.209
	Ne	234.646	227.064	268.566	257.652
5 Abanoquil	Na	226.873	218.266	275.196	265.278
	Nc	261.089	253.601	295.961	286.278
	Nd	249.045	241.329	284.717	274.285
6 REC-12/2615	Na	224.401	216.906	275.125	265.409
	Nc	257.639	250.580	295.805	286.760
	Nd	240.281	232.972	280.879	271.497
	Ne	229.809	221.131	267.851	257.296
7 Alfuzosin	Na	242.348	233.337	272.797	263.841
	Nb	253.210	244.899	285.778	275.863
	Nc	259.609	250.370	292.897	283.574
	Nd	240.401	230.232	276.492	265.827
	Ne	264.739	255.979	270.279	261.386
8 Doxazosin	Na	220.986	212.982	270.136	260.212
	Nb	241.922	235.143	285.573	276.800
	Nc	253.238	245.850	292.106	282.901
	Nd	240.063	232.391	279.256	269.623
	Ne	214.491	207.571	259.571	251.264
9 Terazosin	Na	223.322	215.122	270.409	260.857
	Nb	244.894	237.520	286.000	275.500
	Nc	255.960	247.951	292.509	283.207
	Nd	243.513	234.885	280.273	268.483
	Ne	222.436	214.468	265.257	255.813

There is an overall preference for nitrogen Nc to be protonated for all of the compounds in group 3 (compounds **10** to **15**) when observing the PCM based models (Table 4.4). The compounds NAN-190 (**12**) and RS-100,975 (**14**) show a preference for nitrogen Na in the gas phase state. Compound WAY-100635 (**13**) shows a preference for nitrogen Nc in most cases except for the calculated gas phase at 298 K. For these compounds in the gas phase only a difference of less than 3 kcal mol^{-1} is observed for favouring nitrogen Na over Nb. In the PCM phase the difference for NAN-190 (**12**) is ~2 kcal mol^{-1} for nitrogen Nb over Na making nitrogen Nb the preferred site for protonation. For RS-100,975 (**14**) a difference of 10 kcal mol^{-1} is observed favouring nitrogen Nb over Na. This means that solvent stabilizes the protonation of nitrogen Nb. A different result is obtained for uropidil (**15**) which shows a preference for nitrogen Nb in all calculations except in the PCM phase at 298 K where a preference is shown for nitrogen Na. This difference is only 0.03 kcal mol^{-1} and, therefore, both protonation states are likely to be present.

A lower energy for the 298 K structure is observed in comparison with the 0 K structure with a decrease of ~8.5 kcal mol^{-1} for the gas phase structures and a decrease of ~8 kcal mol^{-1} for the PCM structures. The energy difference between the gas phase and the PCM phase results in an increase of ~41 kcal mol^{-1} for the 0 K and 298 K structures.
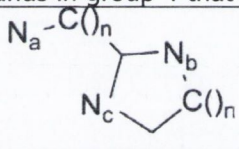
Table 4.4: Proton affinities of the ligands in group 3 that share the following structure.

Ligand	Nitrogen	In Vacuo		PCM	
		PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})	PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})
10 BMY-7378	Na	249.995	240.630	288.687	278.739
	Nb	252.369	244.520	289.460	279.813
	Nc	199.264	192.131	246.072	239.813
11 SNAP-8719	Na	236.618	238.179	280.565	271.229
	Nb	246.721	238.179	285.197	277.923
	Nc	199.694	192.0159	246.502	240.781
12 NAN-190	Na	251.097	242.523	288.769	281.538
	Nb	248.902	239.408	293.257	283.774
	Nd	205.436	198.216	249.873	243.177
13 WAY-100635	Na	252.037	243.570	288.515	277.488
	Nb	252.914	242.145	291.298	280.208
	Nc	221.316	213.84	263.116	253.703
	Nd	236.563	228.221	280.802	271.215
14 RS-100,975	Na	247.873	237.938	283.089	274.435
	Nb	246.789	236.465	293.074	283.924
	Nc	193.337	189.165	247.358	241.374
	Nd	192.761	185.989	246.329	238.110
15 Uropidil	Na	244.154	232.956	288.182	277.802
	Nb	245.650	236.086	288.621	277.775
	Nc	245.647	236.068	268.193	258.735
	Nd	213.702	207.286	251.832	246.325
	Ne	215.971	209.393	255.647	247.701

The next set of compounds is presented in group 4 and they show a consistent preference of nitrogen Nb to be protonated in both the gaseous phase and in the PCM solvent model. SKF-104856 (**17**) and discretamine (**18**) have only one nitrogen which can be protonated and phentolamine (**14**) and corynanthine (**19**) have multiple nitrogens but a difference of 19-20 kcal mol^{-1} is oberserved for nitrogen Nb over Na making nitrogen Nb most stable for protonation. The PAs show a decrease of ~9 kcal mol^{-1} in energy for the 298 K systems with for the PCM compared to the 0 K optimized systems. The difference is observed for gas phase and PCM phase optimized structures. The difference between the gas phase optimizations and PCM optimizations is ~46 kcal mol^{-1} for both 0 K and 298 K optimizations. Despite the limited number of compounds in this group and only two of them having multiple nitrogens, the large energy difference clearly indicates that nitrogen Nb is the one to be protonated.

Table 4.5: Proton affinities of the ligands in group 4 that share the following structure.

Ligand	Nitrogen	In Vacuo		PCM	
		PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})	PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})
16	Na	232.998	224.508	277.289	267.282
Phentolamine	Nb	251.400	243.819	297.845	289.646
	Nc	226.855	218.379	275.349	266.484
17 SKF-104856	Nb	241.628	231.907	290.342	280.446
18 Discretamine	Nb	247.875	238.342	292.057	281.995
19 Corynanthine	Na	220.995	212.018	265.580	256.866
	Nb	249.115	239.970	294.331	285.864

The ligands in group 5 (compounds **20-24**) show consistent results at 0 K or 298 K either in gas phase or in the PCM solvent model with all compounds showing a preference for nitrogen Na as the choice for protonation (Table 4.6). WB-4101 (**20**), Benzoxathian (**22**) and JHT-601 (**24**) only have one nitrogen which can be protonated. KMD-3213 (**21**) shows a difference of ~ 7 kcal mol-1 favouring nitrogen Na over Nc RS-17053 (**23**) shows a difference of ~28 kcal mol^{-1} favouring nitrogen Na over Nb, making this site the favourable for protonation. The PAs exhibit similar results when comparing the 0 K and 298 K results with a decrease of ~10 kcal mol^{-1} in energy for the gas phase models and a decrease of ~11.5 kcal mol^{-1} for the PCM optimized structures. The difference between the gas phase optimizations and PCM optimizations is ~42 kcal mol^{-1} for the 0 K optimizations and ~40 kcal mol^{-1} for the 298K optimizations. Given the difference of energy with other protonated nitrogens, there is a clear indication that the nitrogen Nc is preferred for protonation when the ligand in water.

Table 4.6: Proton affinities of the ligands in group 5 that share the following scaffold.
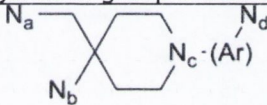
Ligand	Nitrogen	In Vacuo		PCM	
		PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})	PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})
20 WB-4101	Na	250.263	240.258	290.796	278.861
21 KMD-3213	Na	252.085	241.291	291.709	281.651
	Nb	226.458	217.626	268.800	260.857
	Nc	234.746	223.785	283.150	274.548
22 Benzoxathian	Na	247.342	237.515	290.045	278.808
23 RS-17053	Na	258.415	246.965	296.308	283.921
	Nb	220.614	209.703	265.945	255.771
24 JHT-601	Na	248.951	239.922	287.727	275.692
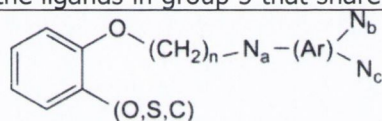
The ligands in the group 6 (compounds **25**-**27**) show consistent results at 0 K or 298 K in both the gas phase and PCM solvent model (Table 4.7). The difference in PA between nitrogen Nc over other nitrogens ranges from 13 to 24 kcal mol^{-1} resulting in nitrogen Nc being selected for protonation. Despite there only being a few ligands in this group this is a consistent result.

When comparing the 0 K and 298 K structures with a decrease of ~8.5 kcal mol^{-1} in energy for the gas phase models and a decrease of ~10.5 kcal mol^{-1} for the PCM optimized structures. The difference between the gas phase optimizations and PCM optimizations is ~41 kcal mol^{-1} for both the 0 K optimizations and ~39 kcal mol^{-1} for the 298 K optimizations. Given the difference in PA, nitrogen Nc over other nitrogens and the consistency of the results in this group there is a clear trend that nitrogen Nc is the selected for protonation when the ligand is in a hydrophilic solvent environment.
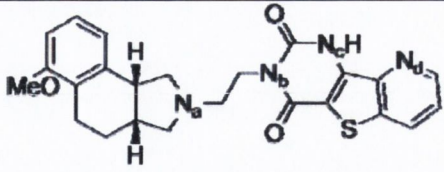
Table 4.7: Proton affinities of the ligands in group 6 that share the following structure.

Ligand	Nitrogen	In Vacuo		PCM	
		PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})	PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})
25 Spiperone	Na	218.103	209.634	264.1829	254.838
	Nb	237.021	226.811	282.262	271.806
	Nc	256.006	245.402	295.970	284.107
26 Indoramin	Nb	226.712	221.899	269.901	258.482
	Nc	246.319	238.07	293.399	282.711
	Nd	233.925	225.921	267.821	256.155
27 SNAP-1069	Nb	224.348	214.604	267.241	257.232
	Nc	259.298	250.037	297.398	287.106

The structure of A-131701 (**28**) shares little similarity with any other of the compounds in our set of ligands. When looking at the results of of A-131701 (Table 4.8), a preference for protonation is shown in nitrogen Na with a difference of ~8 kcal mol-1 over Nd and more than 34 kcal mol^{-1} over Nb and Nc in the PCM phase. Similar results are obtained for all calculations. A decrease of ~9 kcal mol^{-1} from 0 K to 298 K and a decrease of ~42 kcal mol^{-1} from gas phase to the PCM models is in agreement with results obtained for other ligands in this dataset.

Table 4.8: Proton affinities of compound 28.

Ligand	Nitrogen	In Vacuo		PCM	
		PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})	PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})
28 A-131701	Na	255.920	246.610	292.967	278.554
	Nb	211.770	203.449	253.975	244.491
	Nc	190.463	183.243	237.840	229.634
	Nd	227.025	219.492	280.990	270.205

The ligands that have been used in our dataset have different structures. When observing the PAs in each group of ligands, we were able, to identify which nitrogen is most likely to be protonated, with the exemption of uropidil (**15**) which shows a small difference in PA when comparing two protonation states and therefore both protonation states could occur. The structural difference of the compounds does not allow the comparison of protonation states of all the ligands and, therefore, no general rule can be formulated for determining the favoured protonation site in all ligands. However, our results show that there is clear indication that in structurally similar ligands, the same nitrogen is preferred for protonation. Therefore, the assumption can be made that ligands which are grouped according to their similarity in structure, as described in this chapter, have the same preferred nitrogen for protonation. This can be used when modelling a serie of similar compounds, as only for a limited number the protonation state needs to be determined, which, when similar results are obtained, can be applied to all similar compounds.

4.3.3. Addition of water

A different approach to the calculation of PA has been taken by introducing one explicit molecule of water properly orientated, interacting with a selection of molecules from our set, and performing calculations at B3LYP/6-31G* level in gas phase and with the PCM solvation model. The addition of this water molecule was performed to investigate if explicit water molecules can influence the PA values. Water molecules can form H-bonds with the nitrogen or hydrogen attached to the nitrogen when protonated and, therefore, could change the PA of the ligand. The results obtained for PAs are shown in table 5.9. In general, very good agreement is found between the results with an explicit water molecule and without it, and in all cases (except for compound **13**), the same nitrogen is preferred to be protonated. The results obtained with the PCM solvation model and the

explicit water molecule show, as expected, larger PAs than in the gas phase but, in general, very similar to the PAs obtained without the water of molecule and the PCM solvation model.

The decrease in energy when correcting from 0 K to 298 K is ~9 kcal mol^{-1} in the gas phase and the PCM models. The difference between the gas phase and the PCM model is an average of ~40 kcal mol^{-1} for the 0 K models and ~31 kcal mol^{-1} for the 298 K models. This is in agreement with the optimizations without the use of an explicit water molecule.

Table 4.9: Proton affinities derived from the B3LYP/6-31G* calculations using the inclusion of an explicit water molecule

	Ligand	Nitrogen	In Vacuo + explicit water molecule		PCM + explicit water molecule	
			PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})	PA (0 K) (kcal mol^{-1})	PA (298 K) (kcal mol^{-1})
1	Adrenaline	Na	248.079	241.907	294.823	289.007
2	Noradrenaline	Na	244.395	237.786	294.409	288.230
9	Terazosin	Na	234.640	226.439	279.736	273.159
		Nb	252.827	245.029	290.216	280.686
		Nc	260.036	252.864	294.969	288.170
		Nd	253.356	245.755	285.604	275.113
		Ne	236.443	228.122	273.344	267.680
13	WAY-100635	Na	268.789	255.463	291.164	282.102
		Nb	259.571	251.228	293.452	284.103
		Nc	233.658	225.224	271.087	263.172
		Nd	245.630	238.044	286.387	277.143
16	Phentolamine	Na	248.626	236.940	285.084	274.589
		Nb	259.545	253.383	299.673	294.325
		Nc	227.768	222.055	279.197	271.229
17	SKF-104856	Nb	251.315	242.978	293.426	285.922
18	Discretamine	Nb	256.196	247.716	294.925	286.954
23	RS-17053	Na	262.237	253.710	297.570	285.655
		Nb	231.710	220.573	275.425	265.584
24	JHT-601	Na	256.197	247.710	293.488	285.360
25	Spiperone	Na	229.988	221.595	269.097	261.147
		Nb	247.202	239.017	286.626	278.063
		Nc	261.242	250.664	296.462	285.547

As mentioned, the addition of a water molecule does not change the preferred nitrogen for protonation and this comparison has been summarized in table 4.10. The proton affinity values itself appear to be higher for the ligands with the explicit water molecule. However, there is not a fixed amount of energy that can be added to the ligand-water complex suggesting that the water molecule does have an effect on the conformation of the ligand. Our work shows that the influence of a water molecule does not yield a

different protonation state of the ligand and, therefore, the PCM is as reliable for determining which nitrogen is most likely to be protonated.

Table 4.10: Comparison between the PCM model and the water + PCM model (combo) (298 K)

Ligand no.	Name	Protonated nitrogen (PCM)	PA PCM (kcal mol^{-1})	Protonated nitrogen (combo)	PA combo (kcal mol^{-1})
1	Adrenaline	Na	285.410	Na	289.007
2	Noradrenaline	Na	284.490	Na	288.230
9	Terazosin	Nc	283.207	Nc	288.170
13	WAY-100635	Nb	280.208	Nc	284.103
16	Phentolamine	Nb	289.646	Nc	294.325
17	SKF-104856	Na	280.446	Na	285.922
18	Discretamine	Na	281.995	Na	286.954
23	RS-17053	Na	283.921	Nc	285.655
24	JHT-601	Na	275.692	Nc	285.360
25	Spiperone	Nc	284.107	Nc	285.547

4.3.4. Determination of the pK_a of the ligands studied

The acidity of the ligands, and therefore their pK_a, is critical in the ligand-receptor interaction. However, the experimental determination of the pK_a is not always possible and there are computational alternatives to determine this parameter. To validate our prediction of the pK_a we have selected several compounds for which an experimental pK_a had been previously determined (Table 4.11). These compounds include adrenaline and noradrenaline and four compounds which are currently commercially available as antagonists for the α_1-adrenoceptor. Applying the thermodynamic scheme in chapter 2.2.8. and using the energies obtained at 298 K for the different states (neutral, protonated, gas, PCM) the pK_a has been calculated. The results in table 4.11 and figure 4.5 show both the computed and experimental pK_a values and the correlation between them. In general, the computed values are slightly higher than the experimental ones. Prazosin (**3**) shows the largest difference between experimental pK_a and the computed pK_a. When excluding this compound from our validation set, a noticeable increase in the correlation coefficient (R^2) was be observed (Fig. 4.5).

One limitation of this validation set is that the range of pK_a values covered by the six compounds is limited. This is because these compounds are commercially in use for treatment of BPH and, therefore, have pK_a values within a favourable range to provide physicochemical properties that allow the ligand to be distributed and still allow the interaction with the α_{1A}-ARs in the prostate. Therefore, care should be taken with compounds that have pK_a values outside the range that is covered in our validation set.

Our results clearly indicate that acceptable values of pK_a are obtained using this theoretical method.

Table 4.11: Overview of the experimental[15] and computational pK_a values.

	Experimental pK_a	Computational pK_a
1 Adrenaline	8.59	9.70
2 Noradrenaline	8.58	9.03
3 Prazosin	7.0	8.81
7 Alfuzosin	8.1	8.36
8 Doxazosin	6.9	7.86
9 Terazosin	7.1	8.09

$y = 0.62\ pK_a\,[\exp] + 3.82$
$R^2 = 0.55$
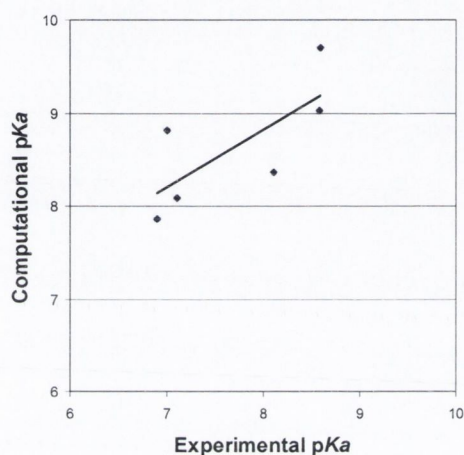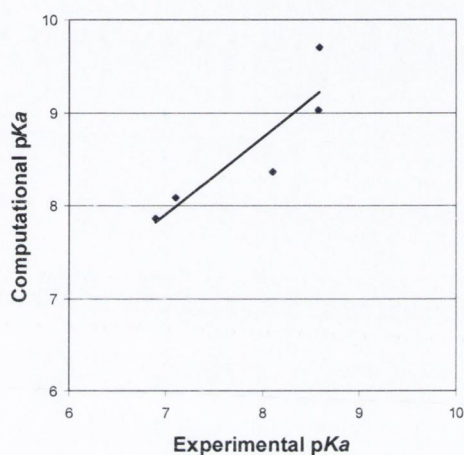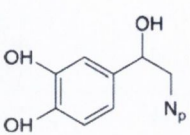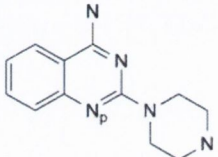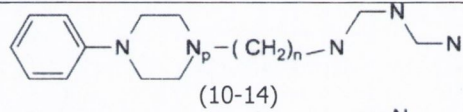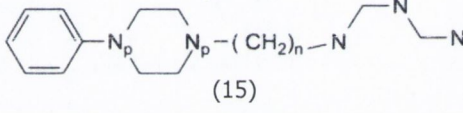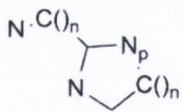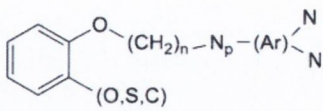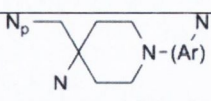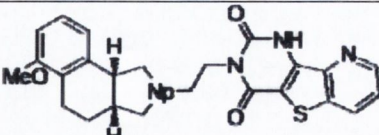
$y = 0.83\ pK_a\,[\exp] + 2.12$
$R^2 = 0.79$

Figure 4.5: The correlation plot for all six compounds shown in the left. The correlation plot for all compounds except Prazosin (**3**) is show in the right. For each plot, a trendline is added with the correlation coefficient (R^2).

For all of the compounds in our dataset we have determined the pK_a value using the free energies of the neutral and protonated conformations. These results are presented in table 4.12 which show a considerable range of pK_a values. This range exceeds that of the pK_a values used in our validation set, but these values provide a good indication of the acid dissociation constant. The higher the pKa, the more acid and they will be easily protonated. Most compounds show a pKa ranging from 7 to 9 and, therefore, these compounds or the ones with higher pKa can be treated as protonated at physiological pH. Five compounds BMY-7378 (**10**), uropidil (**15**), WB-4101 (**20**), benzoxathian (**22**), JHT-601 (**24**) and A-131701 (**28**) have a lower pK_a. However, these compounds are still potent against the α_1-ARs meaning that their low pKa does not limit their ability to interact with their target. Considering that the natural agonist and four ligands which are in clinical use have a pK_a ranging from 7 to 8.6, it is likely that other properties, such as

solubility and membrane penetration, require optimization of pK_a to be effective in clinical use.

Table 4.12: Predicted pK_a values for the compounds in our dataset using computational methods.

Compound no.	Preferred nitrogen for protonation (N_p)	Name	Predicted pK_a	Proton affinity (kcal mol^{-1})
1		Adrenaline	9.70	285.410
2		Noradraline	9.03	284.490
3		Prazosin	8.81	284.190
4		Cyclazosin	9.00	284.450
5		Abanoquil	10.34	286.278
6		REC-15/2615	10.69	286.760
7		Alfuzosin	8.36	283.574
8		Doxazosin	7.86	282.901
9		Terazosin	8.09	283.207
10		BMY-7378	5.60	279.813
11		SNAP-8719	4.21	277.923
12		NAN-190	8.50	283.774
13		WAY-100635	5.89	280.208
14		RS-100,975	8.61	283.924
15		Uropidil	4.11	277.802
16		Phentolamine	12.81	289.646
17		SKF-104856	6.06	280.446
18		Discretamine	7.20	281.995
19		Corynanthine	10.04	285.864
20		WB-4101	4.90	278.861
21		KMD-3213	6.95	281.651
22		Benzoxathian	4.86	278.808
23		RS-17053	8.61	283.921
24		JHT-601	2.58	275.692
25		Spiperone	8.75	284.107
26		Indoramin	7.72	282.711
27		SNAP-1069	10.95	287.106
28		A-131701	4.67	278.554

The results in table 4.13 show that for the three compounds which have the pK_a determined with all three methods, there is little correlation with the pK_a obtained with explicit water molecules and the experimental or the PCM pK_a values. When comparing the explicit water pK_a values with the experimental ones there is a large difference between the two sets of values. Comparing the explicit water pK_a values with the PCM

values there are three trends that are visible. The first observation is that the pK_a calculated with an explicit water molecule shows pK_a values that are not realistic (negative value) for terazosin (**9**) and phentolamine (**18**). The second observation is that the pK_a values obtained with calculations including the explicit water molecule are larger than those obtained using the PCM model. The last observation is that the range of pK_a values calculated with the combined PCM-explicit water molecule is limited. This is clearly illustrated by JHT-601 (**24**) which produces a value of 2.58 for the PCM model and 9.67 for the combined PCM-water molecule approach.

One explanation for this difference between the PCM models and the models which include an explicit water molecule could be that a different conformation of the ligand is obtained due to the inclusion of that explicit water molecule. This could lead to a larger difference in the solvent-exposed surface of the protonated chemical group and, therefore, give a change in the electrostatic contribution to the hydration free energy. It is difficult to exactly show these contributions, but if one water molecule can change the conformation of the ligand, it is likely that multiple water molecules[147] need to be added to obtain the correct conformation of the ligand.

Table 4.13: pK_a values of the calculations with an explicit water molecule as compared to experimental values and the computational values obtained with an implicit solvent model PCM

	Ligand	pK_a (experimental)	pK_a (PCM)	pK_a (explicit water)
1	Adrenaline	8.59	9.70	12.34
2	Noradrenaline	8.58	9.03	11.77
9	Terazosin	7.1	8.09	-15.37
13	WAY-100635	-	5.89	8.74
16	Phentolamine	-	12.81	-22.77
17	SKF-104856	-	6.06	10.07
18	Discretamine	-	7.20	10.83
23	RS-17053	-	8.61	9.88
24	JHT-601	-	2.58	9.67
25	Spiperone	-	8.75	9.80

A note can be made on the calculation of the pK_a. In this chapter a complex model has been used which could be simplified. When correlating the PA and pK_a values, a linear relation can be observed (Fig. 4.6) and the pK_a can be determined using Eq. 4.1:

PA=1.363 pK_a +272.18
$(r^2=1)$
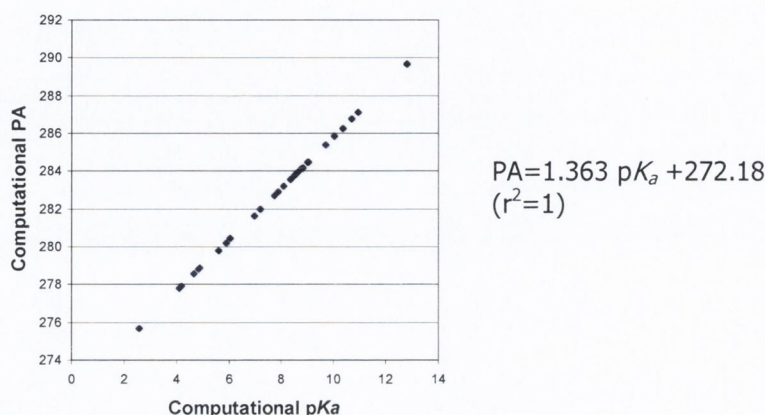
Fig 4.6: Correlation between computed PA and pK_a.

$$pKa = \frac{PA_{PCM} - 272.18}{1.363}$$

(Eq. 4.1)

The PA$_{PCM}$ is the largest proton affinity obtained from the calculations in the PCM solvent model and means that gas phase calculations are not neccesary to obtain the pK_a. For the determination of PAs, solvent based calculations are preferred as solvent can influence the preferred protonation site, as shown in our calculations. The number -272.18 and 1.363 are not randomly choosen, but are actually used in our calculations. The value -272.18 represents the free energy assiociated with H^+_{aq}. This value can be determined substracting the H^+_{aq} which is calculated to be -6.28 kcal mol^{-1} from ΔG^0_S (H^+) which is experimentally determined to be 265.98 kcal mol^{-1}. The value 1.363 is the result of RT ln(10) where R is the universal gas constant and the T = 298 K.

4.4. Conclusions

Methods for determining the proton affinity and pK_a have been around for some time and our results confirm the applicability of these methods. Similar work has been carried out previously by our group[94,99] and now we find improved results for the proton affinities[94]

but different results for the pK_a values.[99] We have made an extensive conformational search and optimization to find the minimum energy conformation to be more stable (lower energy) minimum.

The determination of the proton affinity shows in several cases different results when optimizing in gas phase or using the PCM solvent model. Due to the fact that ligands, when interacting with a receptor are normally not in the gas phase but in an aquous environment, the determination of the proton affinity is to be considered more reliable when using the PCM solvent model. The proton affinities obtained at 0 K and 298 K gives different values, but show similar favoured protonation states.

Lower protonation energies were obtained for the models at 298 K compared to these at 0 K, with a decrease of ~8-10 kcal mol^{-1} for the gas phase calculations and a decrease of ~8-12 kcal mol^{-1} for the PCM ones. In addition, an increase in energy of the PCM models was observed compared to the gas phase models with the increase being ~41-46 kcal mol^{-1} for both the 0 K and 298 K structures.

The preferred nitrogen for protonation in each ligand according to the proton affinities differs in each group of compounds and therefore is is difficult to derive a general rule to determine which nitrogen in all ligands will be protonated. Within one group, consistent results for a preferred nitrogen for protonation were observed. This suggests that when a larger group of structurally similar ligands needs to be studied, the site for protonation of these molecules can be determined by examining only a small number of these, and then select this nitrogen for protonation for all ligands within the group.

The protonation of ligands that target the α_1-ARs is commonly accepted as is shown by different pharmacophore models (Chapter 1.4.1.)[52-55] and an MD study.[62] To what extent these ligands protonate can be determined using the pK_a which we have evaluiated.

The addition of an explicit water molecule to our calculations did not improve the accuracy of either PA or pK_a. The same nitrogen for protonation was preferred with or without the explicit water molecule, although different proton affinity values were obtained. However, when the the the pK_a was determined, different values were obtained and the addition of one explicit water molecule gave results that were less well correlated to

experimental pK_a. Because PA and pK_a are correlated, this should not be observed and this shows that the PCM model provides more reliable results.

The addition of the water molecule does not provide a consistent change in energy on the ligand, because the proton affinities were not consistently higher or lower. This leads to the conclusion that the addition of one single explicit water molecule does not improve the optimization of the conformation and the energies associated with this conformation. Further studies to improve this situation could be performed by adding multiple water molecules.

The determination of pK_a values can be performed using the scheme presented in the methodology chapter of this thesis. The results that were obtained showed good correlation with experimental values despite having reference compounds which were limited in number and in pK_a range. The accurate prediction of the pK_a confirms that we have found good minima for the ligands. Using these minima it is easy to derive accurate charges for each compound and hence, use these ligands in docking and molecular dynamics simulation studies when complexed with the adrenoceptor models.

Chapter V

Molecular Dynamics: Simulation Of Antagonist Bound α_1-Adrenoceptor Antagonist Bound Complexes In A Phospholipid Membrane Mimic.

What happens if a big asteroid hits Earth? Judging from
realistic simulations involving a sledge hammer
and a common laboratory frog, we can assume
it will be pretty bad.
Dave Barry (1947 -)

5.1. Introduction

The interaction between a number of ligands and the three adrenoceptor subtypes α_{1A}-AR, α_{1B}-AR and α_{1D}-AR will be described and discussed in this chapter. Understanding of the interaction between the antagonist and the receptor is key to the development of new drugs in a structure based drug design approach.

The homology models as developed in chapter 3 provide a starting point for docking studies for the ligands which have been optimized in chapter 4. The homology models that were generated using the direct method and a set of 12 antagonists based on their binding affinity were selected. The docking of the ligands into each of the adrenoceptor subtypes is not a straightforward process as different ligand-protein complexes can be obtained. Therefore, a careful selection of complexes needs to be made.

The docking of multiple conformations in the homology models allows the generation of different ligand-protein complexes. Analysis based on the Autodock energy score, a visual inspection of the complex and a plot which shows the interaction between the antagonist and the adrenoceptor subtype allows the selection of a complex. This complex can be further optimized using a MD simulation. If analysis of the trajectory highlights unwanted effects such as the destabilisation of the protein in the form of helices drifting away far from its starting structure or the ligand being forced out of the protein, this shows that the initial complex is not correct and needs optimization. A different starting antagonist-AR complex was then selected and optimized in an MD simulation.

Several research projects have been undertaken before by our group (Kinsella et al)[65-67] and other groups such as Pedretti *et al*[62], Leonardi *et al*[63], Evers *et al*[64] and Bautista *et al*[68] to simulate the behaviour of a ligand-AR complex. However, most of these studies either use a limited time-scale of one or two ns for the MD simulation or model only the interaction of a ligand with one subtype or use very few ligands, which limits their ability to derive common features of interaction.

In this research, longer MD simulations of 5 ns have been introduced for the three different α_1-AR subtypes and twelve antagonists which are structurally different and cover a range of binding affinities for each subtype, have been used.

5.2. Methodology

As several homology models have been generated, a selection from these models was made. Two approaches for selecting the model can be taken; selection of the best model for each of the adrenoceptor subtype or selections of a similar model for each subtype of the adrenoceptor. The second approach was taken as the direct-no-gap model showed good results when analyses were performed and never showed a significant error when compared to some of the other models.

We have a dataset containing 28 optimized ligands. Due to limitations of time and computer power a selection form this dataset had to be made. The criteria that were used for the selection of the ligands were, firstly, the ligands should cover a range of binding affinities for each of the adrenoceptor (Table 5.1) and secondly, there should be structural diversity among the selected ligands (Fig. 5.1).

The range of binding affinities should allow the study of the effect of weak and strong binders in terms of interaction with the receptor. The structural diversity of the ligands should allow the study of how similar or different antagonists interact with the adrenoceptor. Also this could produce different conformations of the adrenoceptor as it is not uncommon for a receptor to adopt a different binding pocket conformation when a different ligand is used.[67]

Table 5.1: Overview of the ligands in the dataset with activity data for the different adrenoceptor subtypes.

	Ligand	Ki (nM) α_{1A}	Ki (nM) α_{1B}	Ki (nM) α_{1D}	Reference
4	Cyclazosin	12	0.13	3.2	137
7	Alfuzosin	10	10	3.16	141
10	BMY-7378	250	630	6.3	141
11	SNAP-8719	294	191	1.6	142
12	NAN-190	2.0	15	0.8	143
13	WAY-100635	144	186	63	143
14	RS-100,975	1.0	79	100	141
15	Uropidil	288	1320	1660	143
18	Discretamine	616	360	25	145
19	Corynanthine	142	517	253	146
21	KMD-3213	0.04	20	2.0	139
27	SNAP-1069	16	200	790	141

4. Cyclazosin

7. Alfuzosin

10. BMY-7378

11. SNAP-8719

12. NAN-190

13. WAY-100635

14. RS-100,975

15. Uropidil

18. Discretamine

19. Corynanthine

21. KMD-3213

27. SNAP-1069
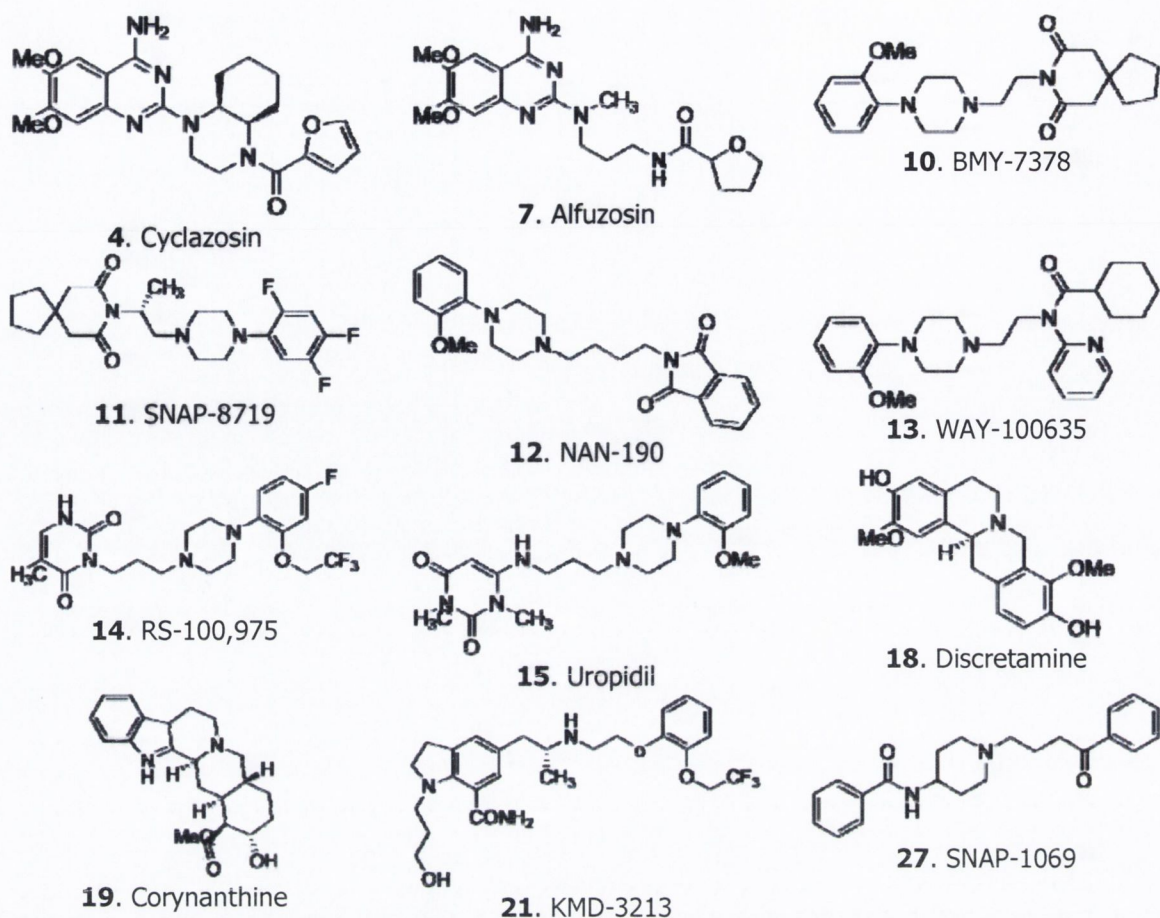
Figure 5.1: Antagonists that were docked in the three α_1-ARs and subjected to MD simulation.

Docking was performed using Autodock.[32] A centre for the grid procedure was choosen by selecting the middle of several residues (Table 5.2) that are known to interact with adrenoceptor antagonists.[25-29] Despite Autodock making use of ligand flexibility during the docking procedure there is still an interest in docking different conformations into the adrenoceptor which generate different ligand-protein complexes. Different conformations are taken from our ligand optimisation scheme. These ligands were subsequently similarly protonated and charges were calculated and added. ESP-charges were taken from the B3LYP/6-31G* optimized ligands and subsequently copied to each of the different conformations.

Table 5.2: Residues used as center for docking

α_{1A}-AR	α_{1B}-AR	α_{1D}-AR
Asp^{106}	Asp^{125}	Asp^{176}
Ile^{178}	Val^{197}	Ile^{248}
Ser^{188}	Ser^{208}	Ser^{258}
Ser^{192}	Ser^{212}	Ser^{262}
Trp^{285}	Trp^{307}	Trp^{361}
Phe^{288}	Phe^{310}	Phe^{365}

Each conformation was docked and ten different possible ligand-receptor complexes were generated. For each initial conformation, the lowest energy conformer was selected leading to the same number of ligand-protein complexes as we had conformations. Analysis was further performed using manual inspection and by dermining the hydrogen bonds and hydrophobic interactions using the program LigPLot.[118] Manual inspection was used to observe if the ligand was positioned in the binding pocket. LigPlot was used to provide a schematic overview of the non-bonded interactions and hydrogen bonds between the ligand and the receptor. Based on those three binding parameters: binding energy, manual inspection and interaction with amino acids, a plausible ligand-protein complex was selected. If no plausible complex could be achieved, the ligand would be re-docked into the adrenoceptor but with the centre for autogrid randomly positioned in a different place in the binding pocket.

The ligand-protein complex was subjected to an MD simulation. To simulate the phospholipid bilayer in which the adrenoceptor is normally located a solvent mimic was used. The homology models were positioned in a multilayer of solvent consisting of water, chloroform and water. In this mimic the chloroform represents the lipophilic area of the phospholipid bilayer and the water represents the extra- and intra-cellular space. The thickness of the chloroform layer was based on the transmembrane helices which should be completely in the chloroform layer with the intra- and extracellular loops surrounded by water (Fig. 5.2). A distance of 15 Å was kept around the protein to be occupied with solvent to prevent interaction of the receptor with its periodic images during the MD simulation. The molecular dynamics simulations were performed using the Particle Mesh Ewald Molecular Dynamics (PMEMD) module of AMBER 9.0[122,133] MD package using periodic boundary conditions. The forcefield that was chosen was the ff99SB which is a modification of the original ff99 forcefield with adjustment to prevent the overstabilisation of helices.[134] The structure was minimized using steepest descent for 10000 steps, followed by a minimisation using conjugate gradient with a maximum of 50000 steps. The system was heated over 20 ps to a temperature of 300 K at constant volume. The system was then equilibrated at constant pressure using 5 steps of 10 ps each with a restraint of 20, 15, 10, 5 and 0 kcal mol^{-1} on the protein. If a stable energy has been reached, showing only fluctuation around a particular energy, the movement of the protein over time was simulated in a production run of 4 ns at constant pressure and a temperature of 300K (Langlevian dynamics) for each of the models.
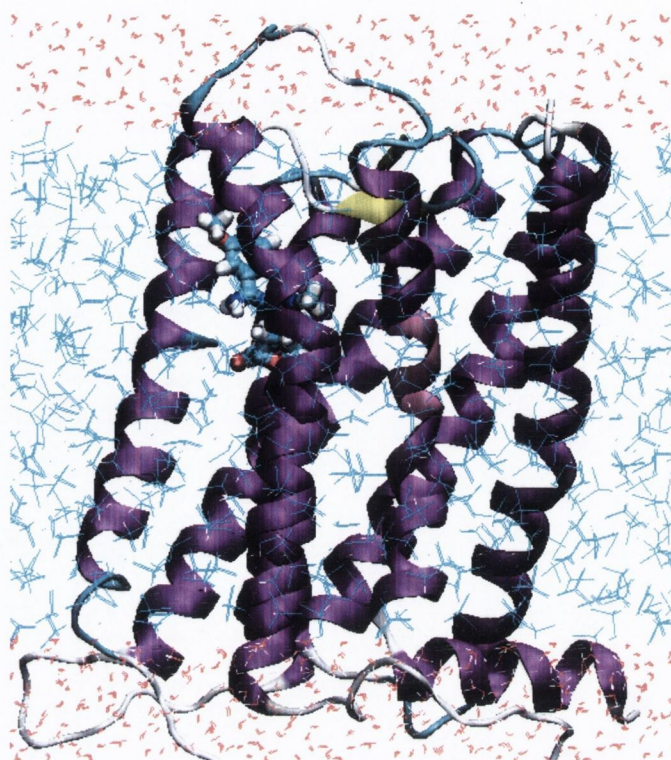
Figure 5.2: The system for the MD simulation showing the α_{1A}-AR complexed
With alfuzosin surrounded by water (red) and chloroform (blue).

Analysis of the thermal stability of the complexes was performed by extracting the kinetic, potential and total energy from the MD simulation. Succesfull equilibration was established when these energies were stable over time. Analysis on the trajectory after equilibration was performed to observe the influence of the ligand on the structure of the homology model. This was achieved by measuring the helixes movement (RMSD) and observing the distances within two features of the adrenoceptors, the ionic lock and a salt bridge (Table 5.3). To avoid confusion of which oxygen of the carboxylate-group of the amino-acid interacts, the distances between the center carbon atoms of the carboyxlate-groups are measured.

Table 5.3: Residues involved in the (left) ionic lock and (right) salt bridge for the three different α_1-ARs.

Ionic lock				Salt bridge		
α_{1A}-AR	Asp^{123}	Arg^{124}	Glu^{267}	α_{1A}-AR	Asp^{106}	Lys^{309}
α_{1B}-AR	Asp^{142}	Arg^{143}	Glu^{269}	α_{1B}-AR	Asp^{125}	Lys^{331}
α_{1D}-AR	Asp^{193}	Arg^{194}	Glu^{343}	α_{1D}-AR	Asp^{176}	Lys^{385}

To determine the interaction of the ligand with the protein, a snapshot was taken after 5 ns of MD simulation followed by an energy minimization and analysis using the program LigPlot to determine hydrogen bonds or hydrophobic interactions.

The results of our docking/MD studies have been subdivided into four sections. First, the docking of the antagonists in each of the α_{1A}-ARs is discussed. Second the MD simulation, third, the obtained structure after MD simulation and fourth the discussion of the ligands in each of the three binding pockets.

5.3. Docked structures

To analyze the docking of the ligand in the homology models, scoring tests have been performed on each complex using three functions: London dG, Affinity dG and Alpha HB. London dG estimates the free energy of binding, Affinity dG the enthalpic contribution to the free energy and Alpha HB is the contribution of both. A low energy score given by each scoring function represents a strong binding interaction. In addition, a ranking of the compounds according to their scores can be made. These scores and ranks were compared to the experimental binding affinities (K_i) to verify if the initial models are suitable for docking studies. Furthermore, as hydrogen bonds are commonly formed in protein-ligand interactions, they were determined to establish if all ligands form strong interactions with specific residues in the receptor.

5.3.1. α_{1A}-Adrenoceptor

The scoring of the docked ligands in the α_{1A}-AR homology models (Table 5.4) shows no correlation with their binding affinity (Table 5.5) when either the scores are compared to the binding affinities or when a comparison between the ranking of the binding affinities and ranking of the scores is made. Uropidil (**15**) shows scores which are different to the range covered by the other compounds and therefore should be regarded as wrong. These results are expected as scoring functions are useful for differentiating between active and non-active compounds, but are known to lack accuracy when ranking a set of active compounds, as is the case in our dataset. The misscoring of uropidil (**15**) is likely due to an unfavourable interaction between the ligand and the protein.

Table 5.4: Scoring and ranking using London dG, Affinity dG and Alpha HB for each compound docked in the homology model of α_{1A}-AR.

| Ligand | Score | | | | Ranking | | | |
	K_i (nM)	London dG (kcal mol^{-1})	Affinity dG (kcal mol^{-1})	Alpha HB (kcal mol^{-1})	K_i rank	London dG	Affinity dG	Alpha HB
4	12	-14.45	-3.53	-103.80	5	2	1	3
7	10	-20.36	-2.34	-108.08	4	1	6	2
10	250	-10.98	-0.89	-90.37	9	9	7	7
11	294	-10.21	-0.67	-89.84	11	11	8	8
12	2	-11.73	-3.12	-90.50	3	7	4	6
13	144	-12.64	-3.28	-112.59	8	3	3	1
14	1.0	-11.46	0.19	-100.95	2	8	9	5
15	288	0.28	66.19	58.59	10	12	12	12
18	616	-11.85	-3.41	-82.29	12	6	2	10
19	142	-11.88	1.11	-75.45	7	5	10	11
21	0.04	-11.91	1.40	-103.70	1	4	11	4
27	16	-10.81	-2.90	-88.27	6	10	5	9

Table 5.5: Correlation (R^2) for each scoring function and ranking with the experimental K_i values for the docking in the homology model of α_{1A}-AR.

Scoring method	Correlation with score/rank	Correlation (R^2)
London dG	score	0.125
	rank	0.182
Affinity dG	score	0.048
	rank	0.016
Alpha HB	score	0.116
	rank	0.298

When analyzing the formed H-bonds (Table 5.6), it can be seen that only Trp^{262} seems to be common to most of the interactions established with the different ligands, but only in four out of twelve complexes. In case of uropidil (**15**) and SNAP-1069 (**27**) two H-bonds are formed. Although uropidil has two H-bonds formed with the α_{1A}-AR, it is a weak binder and therefore other interactions (e.g. hydrophobic) can be important as well for good binding to the α_{1A}-AR.

Table 5.6: Residues involved in the formation of h-bonds with α_{1A}-AR antagonists.

Residue	Ligand
Thr^{88}	SNAP-1069 (**27**)
Arg^{143}	NAN-190 (**12**), RS-100,975 (**14**)
Ser^{165}	KMD-3213 (**21**)
Ser^{169}	Alfuzosin (**7**), Uropidil (**15**)
Trp^{262}	Uropidil (**15**), Discretamine (**18**), KMD-3213 (**21**), SNAP-1069 (**27**)

Furthermore, H-bonding with Asp[106] is not observed, but is known to play an important rol in activation. As shown in the homology modelling chapter, the crystal structure of the β_2-AR indicates that the size of the binding pocket in our homology models is likely to be smaller than the size of the actual α_{1A}-AR. Therefore. a tighter fit of the ligand in the binding site is expected. This shows that the binding between the antagonists and the α_{1A}-AR is not optimal, but does provide a starting point for MD studies which leads to further optimization of the interactions in the ligand-protein complex.

5.3.2. α_{1B}-Adrenoceptor

The scoring of the docked ligand in the α_{1B}-AR homology models (Table 5.7) shows no correlation with their binding affinity (Table 5.8) when either the scores itself are compared or a ranking of the scores is compared. Examples of inaccurate scoring are cyclazosin (**4**) and BMY-7378 (**10**). Cyclazosin (**4**) is a strong binder, corretly predicted according to London dG and Alpha HB, but affinity dG predicts it as a weak binder. BMY-7378 (**10**) is a weak binder, correctly predicted by London dG and Alpha HB, but incorrectly according to Affinity dG.

Table 5.7: Scoring and ranking using London dG, Affinity dG and Alpha HB for each compound docked in the homology model of α_{1B}-AR.

	Score				Ranking			
Ligand	K_i (nM)	London dG (kcal mol^{-1})	Affinity dG (kcal mol^{-1})	Alpha HB (kcal mol^{-1})	K_i rank	London dG	Affinity dG	Alpha HB
4	0.13	-13.88	-4.81	-117.14	1	1	9	3
7	10	-12.10	-3.71	-104.59	2	9	11	6
10	630	-11.68	-6.34	-90.90	11	10	5	10
11	191	-13.35	-7.29	-127.90	7	3	3	1
12	15	-12.67	-8.80	-117.83	3	7	1	2
13	186	-12.27	-5.72	-86.57	6	8	8	12
14	79	-13.20	-5.76	-113.82	5	4	7	4
15	1320	-11.59	-6.11	-93.35	12	11	6	8
18	360	-12.73	-4.66	-90.52	9	5	10	11
19	517	-12.72	-7.87	-104.52	10	6	2	7
21	20	-11.54	-1.86	-92.87	4	12	12	9
27	200	-13.44	-7.25	-109.66	8	2	4	5

Table 5.8: Correlation (R^2) for each scoring function and ranking with the experimental K_i values for the docking in the homology model of α_{1B}-AR.

Scoring method	Correlation with score/rank	Correlation (R^2)
London dG	score	0.216
	rank	0.053
Affinity dG	score	0.047
	rank	0.132
Alpha HB	score	0.184
	rank	0.200

Eight residues of the α_{1B}-AR are shown to form H-bonds with the antagonists (Table 5.9). Most of these show only the formation of a hydrogen bond with one or two antagonists, suggesting that different poses have been obtained. The residue Ser[165] is commonly involved in binding as it forms H-bonds with 5 antagonists, which suggests that this residue is important in binding of ligands to the α_{1B}-AR.

Table 5.9: Residues involved in the formation of h-bonds with α_{1B}-AR antagonists.

Residue	Ligand
Val[84]	RS-100,975 (**14**)
Lys[143]	Discretamine (**18**)
Tyr[161]	RS-100,975 (**14**)
Ser[165]	BMY-7378 (**10**), SNAP-8719 (**11**), NAN-190 (**12**), RS-100,975 (**14**), SNAP-1069 (**27**)
Ser[169]	RS-100,975 (**14**), Discretamine (**18**)
Trp[265]	Uropidil (**15**)
Phe[269]	Cyclazosin (**4**), Afluzosin (**7**)
Lys[289]	WAY-100635 (**13**), SNAP-1069 (**27**)

Similar to our docking in the α_{1A}-AR, binding to Cys[165] is observed, meaning that a similar positioning of the ligands is obtained in both models.

5.3.3. α_{1D}-Adrenoceptor

The scoring of the docked ligand in the α_{1D}-AR homology models (Table 5.10) shows no correlation with their binding affinity (Table 5.11). The obtained scores also are inconsistent with each other. For example, discretamine (**18**) is a medium binder, but is ranked strongest according to London dG, medium by Affinity dG and as a weak binder by Alpha HB. This limits the use of scoring functions for predicting binding affinity by using the homology model directly for docking studies.

Table 5.10: Scoring and ranking using London dG, Affinity dG and Alpha HB for each compound docked in the homology model of α_{1D}-AR.

	Score				Ranking			
Ligand	K_i (nM)	London dG (kcal mol^{-1})	Affinity dG (kcal mol^{-1})	Alpha HB (kcal mol^{-1})	K_i rank	London dG	Affinity dG	Alpha HB
4	3.2	-12.43	-7.14	-144.22	5	6	4	1
7	3.16	-12.10	-3.97	-134.24	4	10	12	2
10	6.3	-12.18	-5.84	-111.47	6	8	8	8
11	1.6	-13.87	-9.34	-122.78	2	2	1	4
12	0.8	-12.10	-5.67	-113.24	1	11	9	6
13	63	-12.68	-9.29	-132.50	8	5	2	3
14	100	-12.79	-5.02	-84.51	9	4	10	12
15	1660	-12.14	-6.06	-116.50	12	9	7	5
18	25	-14.08	-6.10	-96.48	7	1	6	11
19	253	-11.85	-4.49	-108.21	10	12	11	10
21	2.0	-13.11	-6.83	-112.41	3	3	5	7
27	790	-12.32	-8.08	-108.43	11	7	3	9

Table 5.11: Correlation (R^2) for each scoring function and ranking with the experimental K_i values for the docking in the homology model of α_{1D}-AR.

Scoring method	Correlation with score/rank	Correlation (R^2)
London dG	score	0.095
	rank	0.074
Affinity dG	score	0.000
	rank	0.016
Alpha HB	score	0.008
	rank	0.010

The α_{1D}-AR forms H-bonds via five residues (Table 5.12). Compared to the α_{1A}-AR and α_{1B}-AR there is much more consistency in the number of antagonists that form H-bonds with a particular residue, suggesting that the ligands bind similarly in the α_{1D}-AR. Lys^{143} forms H-bonds with three antagonists, Ser^{169} with three and Trp^{268} with seven. The results for docking in the α_{1D}-AR are much more consistent and possibly could lead to better or similar optimization during the MD simulation.

Table 5.12: Residues involved in the formation of h-bonds with α_{1D}-AR antagonists.

Residue	Ligand
Cys^{87}	Discretamine (**18**)
Lys^{143}	BMY-7378 (**10**), SNAP-8719 (**11**), NAN-190 (**12**)
Ser^{165}	Alfuzosin (**7**)
Ser^{169}	WAY-100635 (**13**), RS-100,975 (**14**), RS-100,975 (**15**)
Trp^{268}	Cyclazosin (**4**), SNAP-8719 (**11**), NAN-190 (**12**), WAY-100635 (**13**), Discretamine (**18**), KMD-3213 (**21**), SNAP-1069 (**27**)

In this model, similar to the α_{1A}-AR and α_{1B}-AR, H-bonding with Asp[176] in the α_{1D}-AR is not observed. Furthermore Ser[165] shows only interactions with one antagonsist. The residue Trp[268] interacts with seven antagonists which is more than for the α_{1A}-AR (Trp[262], 4 ligands) and the α_{1B}-AR (Trp[265], 1 ligand). This could even suggest that the binding pockets of the α_{1A}-AR and α_{1D}-AR are more similar compared to the α_{1B}-AR.

5.3.4. Binding modes in the α_1-Adrenoceptor

Scoring the initial complexes can not be used for predicting binding affinity, but some inidication of how ligands bind can still be derived. Several residues in all three α_1-ARs are shown to be involved in binding which are Lys[A-166/B-186/D-236], Ser[A-188/B-208/D-258], Ser[A-192/B-212/D-262] and Trp[A-285/B-307/D-361]. These residues, except Lys[A-166/B-186/D-236] have been mentioned in the literature and can be linked back to our theoretical binding pocket derived for agonists, but not for antagonists (chapter 1.2.5.). This suggests that most residues that are involved in binding in agonists are also involved in binding of antagonists. Residue Asp[A-106/B-125/D-176] is known to be common in binding of ligands and the formation of a H-bond is likely to be a key interaction of activation/inhibition of the α_1-ARs. This is not observed in our docked structures. Still, several known interactions have been observed and further optimization of the binding between antagonist and α_1-AR is likely to occur during MD simulations.

5.4. Molecular dynamics simulations

To monitor the stability of the protein-ligand complex during the MD simulation, the RMSD of the helices as a function of time was monitored. This gives an indication of the stability of the protein and can show conformational changes associated with the dynamical optimization of the binding pocket of ligand-protein interactions. After 5 ns of MD simulation, a snapshot of the ligand-protein complex was taken and overlayed with the crystal structure of β_2-AR, the crystal structure of which was determined during this work, to determine similarity. This analysis was performed because the β_2-AR is more closely related to the α_1-ARs than is rhodopsin, and the final MD models would be expected to show a greater structural similarity to that of the β_2-AR than rhodopsin. Furthermore, distances within the ionic lock and salt bridge can be measured and plotted against time to monitor changes in conformation.

5.4.1. α₁ₐ-Adrenoceptor

RMSD analysis of the helices (example in Fig. 5.3) shows small changes (<3 Å) for the helices meaning that the complexed α_{1A}-ARs have a stable conformation. Similar observations are made in all ligand-α_{1A}-AR complexes. There is a change for TM-I where the upper part of the helix moves outwards of the centre of the protein. This is observed in most of the cases. In addition, the top part of TM-IV and TM-V moves outward, while the bottom part stays fixed in its position. Helix TM-VIII also moves which is likely to be an adjustment to the solvents as it is bordered between water and chloroform. The movement of TM-IV and TM-V can be explained as a widening of the binding pocket which provides the ligand with a larger binding pocket. A similar observation can be made in the uncomplexed α_{1A}-AR homology model. The movement of TM-I, TM-IV and TM-V can be explained when it is aligned to the crystal structure of the β_2-AR (Fig. 5.4), since it occupies a similar position. However, there is no perfect fit and each ligand bound complex obtains a slightly different conformation after 5 ns of MD simulation which is known as a ligand-induced conformation. This suggests that homology models based on rhodopsin do converge to a conformation similar to the crystal of the β_2-AR, but small differences occur due to the influence of the ligand.
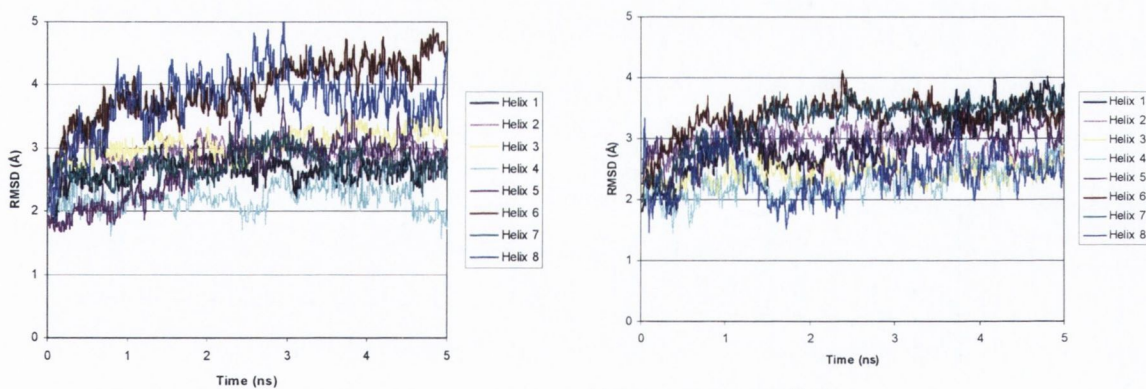
a) Alfuzosin (**7**)

b) Discretamine (**18**)

Figure 5.3: RMSD helix movement of the α_{1A}-AR complexed with (a) alfuzosin and (b) discretamine during a 5 ns MD simulation.
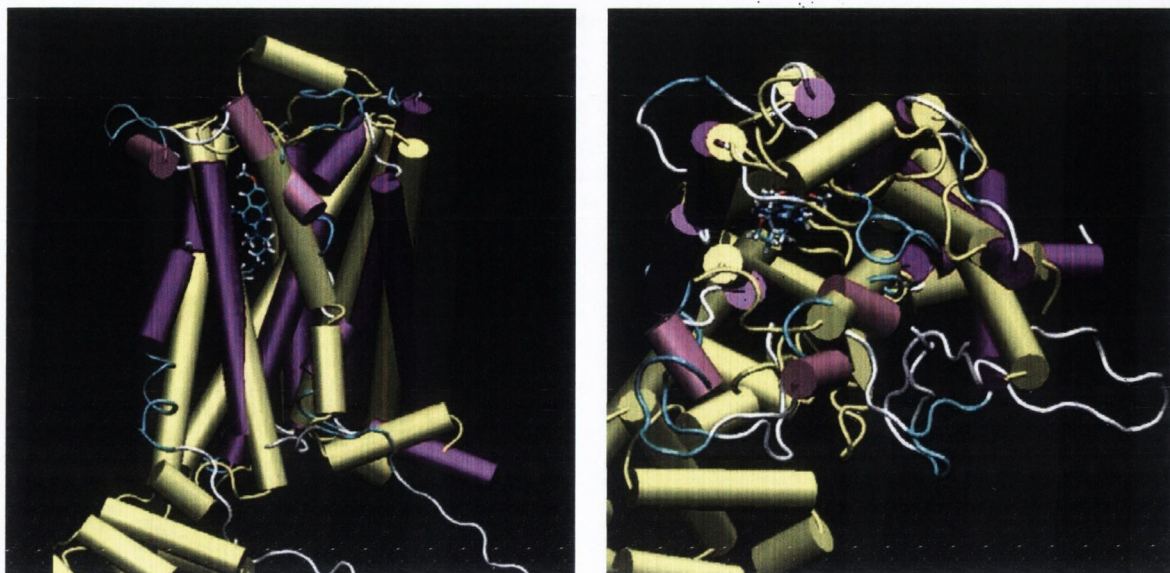
Figure 5.4: Alignment of alfuzosin-α_{1A}-AR complex (purple) after 5 ns of MD simulation with the crystal structure of β_2-AR (yellow).

Looking at the ionic lock, the distance between Asp[123] and Arg[124] appears to either stay stable or increase during the MD simulation (example in Fig. 5.5). The starting distance is 4 Å, increases to 7, 8 or 9 Å and then it then usually stabilizes, compared to 4.84 Å in the β_2-AR and ~4.5 Å in the α_{1A}-AR homology model after MD. The distance between Asp[123] and Glu[267] increases over time for most of the MD simulations. It can be observed that the distance between Asp[123] and Arg[124] changes to either 8 Å to 10 Å or more than 14 Å. A change of more than 14 Å suggests a change to a different conformation of the α_{1A}-AR, compared to 14.8 Å in the β_2-AR and ~9 Å in the α_{1A}-AR homology model after MD. The distance between Arg[124] and Glu[267] either stays stable, increases or fluctates. When there is an increase in the distance, this change is from 6 Å to either 8 Å or more than 9 Å, compared to 13.02 Å in the β_2-AR and ~7 Å in the α_{1A}-AR homology model after MD.

Changes in the ionic lock occur in most of the simulations, but can not be linked to the strength of antagonists binding affinities. It can be observed that an increase in any of the distances within the ionic lock occurs in most cases. This means that a different conformation is obtained.
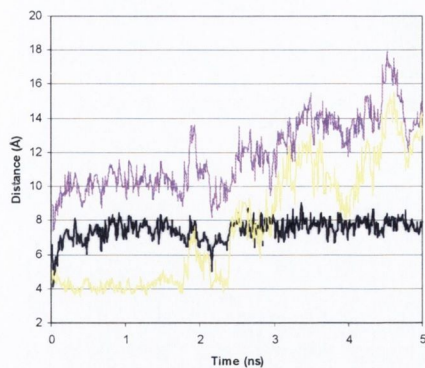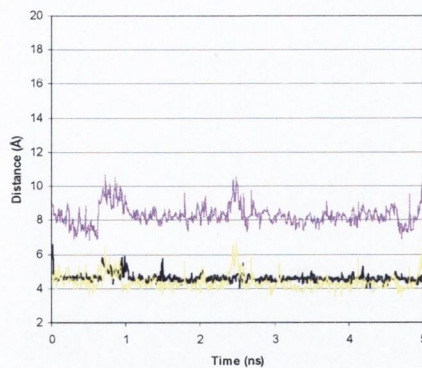
a) Cyclazosin (**4**)

b) Discretamine (**18**)
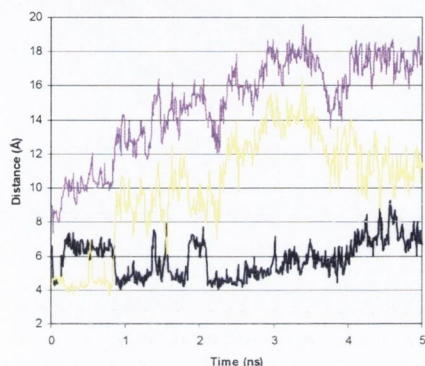
c) WAY-100,635 (**13**)
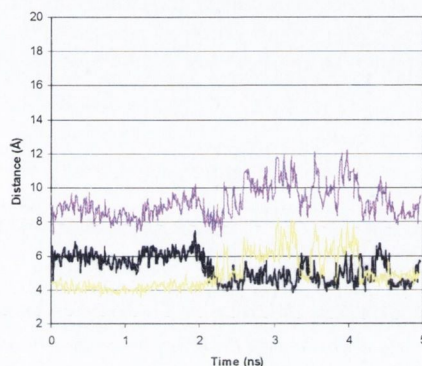
d) Uropidil (**15**)

Figure 5.5: Examples of behaviour of the ionic lock between Asp^{123}, Arg^{124} and Glu^{267} in the α_{1A}-AR for (a) cyclazosin, (b) discretamine, (c) WAY-100,635 and (d) uropidol, showing the different patterns of distances between the three residues that have been observed in the MD simulations.

The salt bridge between Asp^{106} and Lys^{309} shows four different types of behaviour during the MD simulation (Fig. 5.6). The distance can either stay stable (**4**), increase (**7**, **12**, **13**, **14**, **18**, **19** and **30**), fluctuate (**10** and **21**) or decrease (**11**). There limited correlation between this behaviour and the binding affinity. For example, SNAP-8719 (**11**) can be regarded as a weak binder with a K_i value of 294, and shows a decrease in the distance of the salt bridge.

The distance of the salt bridge after 5 ns appears to vary between 13 Å and 18 Å which except for SNAP-8719 (**11**) is an increase in the distance compared to the start of the MD simulation. Most compounds show an increase in the in the salt bridge distance, only three compounds having a salt bridge of less than 15 Å, three compounds in the range of 15-17 Å and six compounds having a salt bridge longer than 17 Å. This compares to a distance of 14.83 Å in the crystal structure of the β_2-AR and ~12 Å in the uncomplexed homology model after MD simulation. These data suggest that a longer distance of the salt bridge could be linked to the antagonist inhibited state of the α_{1A}-AR. However, the different patterns of behaviour of the salt-bridge between Asp^{106} and Lys^{309} can not be

linked to any weak or strong binding antagonists. Furthermore, it could even be assumed that after binding of the ligand, the role of the salt bridge in stabilizing the α_{1A}-AR plays a less significant role than in the unbound state as a ligand-induced conformation is obtained.

a) Cyclazosin (**4**)

b) Alfuzosin (**7**)

c) KMD-3213 (**21**)

d) SNAP-8719 (**11**)
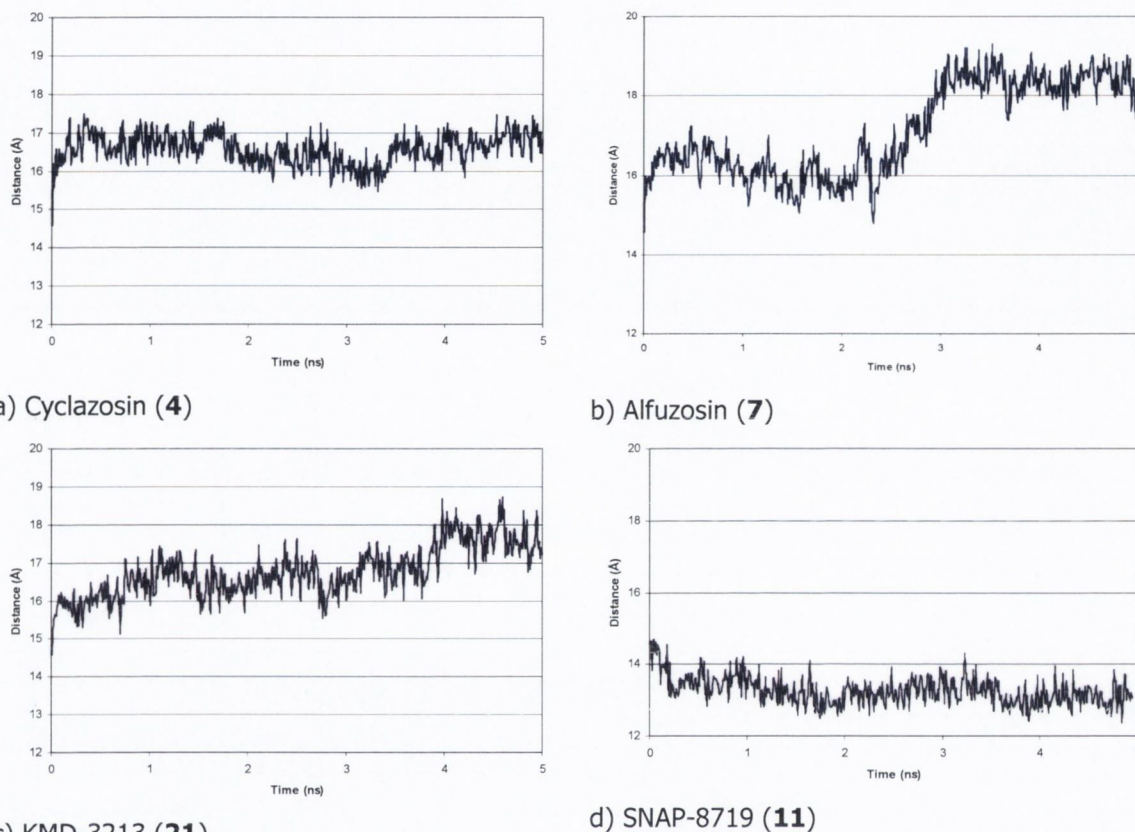
Figure 5.6: Examples of behaviour of the salt bridge between Asp^{106} and Lys^{309} in the α_{1A}-AR for a selection of compounds showing (a) stable, (b) increase, (c) fluctuation or (d) decrease in the distance between the residues Asp^{106} and Lys^{309}.

5.4.2. α_{1B}-Adrenoceptor

The RMSD plots of the helices during the MD simulation (examples in Fig. 5.7) show stable proteins, with a change of less than 2 Å observed for most complexes. Similar observations are made in most ligand-α_{1B}-AR complexes However, some helices such as the case in complexes with cyclazosin (**4**), alfuzosin (**7**) and SNAP-8719 (**11**) show a significant movement of one helix. In each case a different helix moves. Visualization shows that this is due to the repositioning of the ligand in the receptor and not a drift of the helix. A similar observation to the α_{1A}-AR is made where an opening of the binding pocket occurs by the outwards shift of TM-IV and TM-V. However, TM-I occupies a similar position at the start and at the end of the MD simulation.
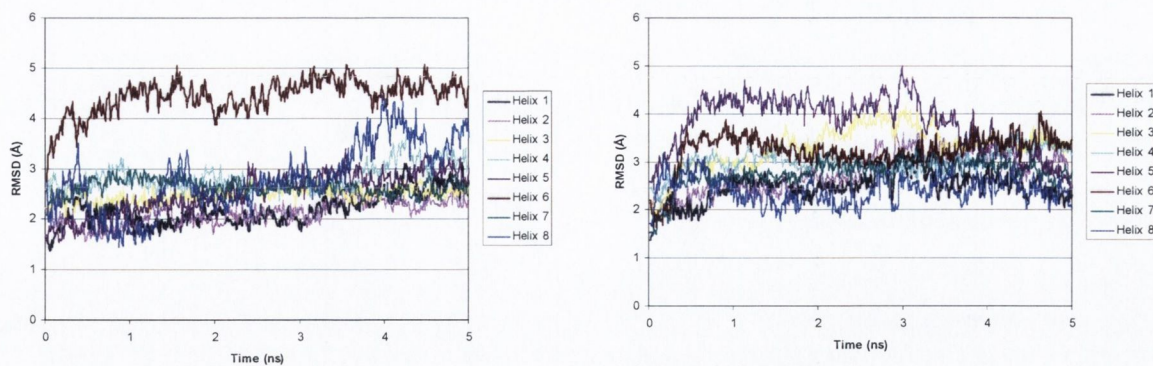
a) Alfuzosin (**7**)

b) WAY-100635 (**13**)

Figure 5.7: Helix RMSD of the α_{1B}-AR interacting with (a) alfuzosin and (b) WAY-100635 during 5 ns of MD simulation.

Alignment of the complexed structures of the α_{1B}-AR after 5 ns of MD simulation to the crystal structure of β_2-AR (Fig. 5.8) shows most of the helices, except TM-I are aligned. This is observed in multiple cases and suggests that the binding pocket is similar to that of the β_2-AR.

Figure 5.8: Alignment of alfuzosin-α_{1B}-AR complex (purple) with the crystal structure of β_2-AR (yellow) seen from the (left) side and (right) top.

The interaction between Asp^{142} and Arg^{143} shows either a stable distance or a decreased one (Fig. 5.9). The range of distances covered after the MD simulation is 6 to 9 Å with the most common distance being ~7 and ~8 Å compared with a distance of 8.7 Å for the initial homology model. This can be compared to 4.84 Å in the β_2-AR and ~9 Å in the α_{1B}-AR homology model after MD.

The interaction between Asp^{142} and Glu^{289} (Fig. 5.9) during the MD simulation shows three different patterns with either a stable distance, a decrease and for one case (**21**) an increase in distance. When there is an increase in the distance between the two residues,

the range of final distances is 10 to 17 Å compared to a starting distance of ~12.3 Å. The phases of minimization, heating and equilibration can reduce this distance to ~10 Å or increase it to ~14 Å. The most common distance after five ns of MD simulation is ~10 or ~11 Å. This is an increase when compared to 4.84 Å in the β_2-AR and ~9 Å in the α_{1B}-AR homology model after MD.

The interaction between Arg[143] and Glu[289] (Fig 5.9) shows two different kinds of behaviour during the MD simulation. All of the models show a decrease of this distance of ~1 Å during minimization, heating and equilibration. The production runs show either a stable or an increase of the Arg[143] and Glu[289] distance over time (Fig. 5.12). The range of distances after 5 ns goes from 4 to 9 Å with a stable distance of 4 Å being the most common. This can be compared to 13.02 Å in the β_2-AR and ~4.5 Å in the α_{1B}-AR homology model after MD.

a) Cyclazosin (**4**)

b) Corynanthine (**19**)

c) SNAP-8719 (**11**)
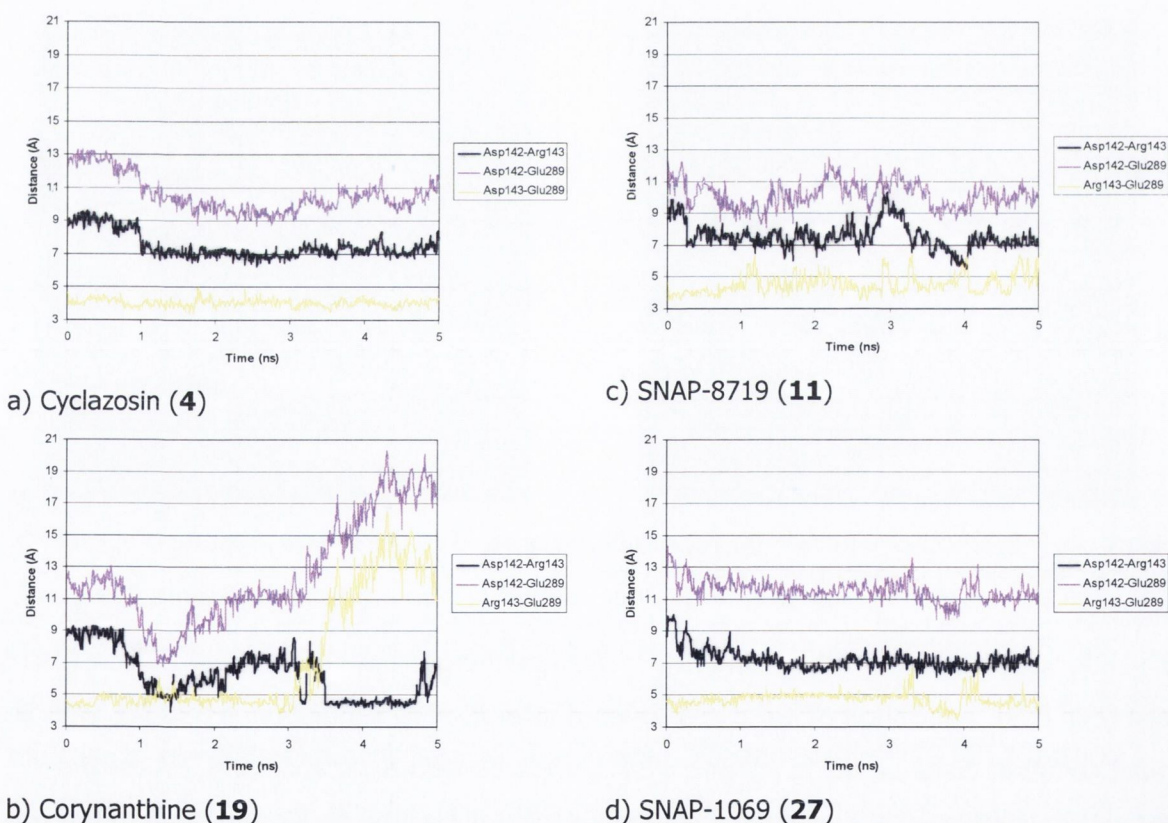
d) SNAP-1069 (**27**)

Figure 5.9: Examples of behaviour of the ionic lock between Asp[142], Arg[143] and Glu[289] in the α_{1B}-AR for (a) cyclazosin, (b) SNAP-8719, (c) Corynanthine and (d) SNAP-1069 showing the different patterns of distances between the three residues that have been observed in the MD simulations.

A change in the ionic lock is associated with conformational changes, but, generally, it can not be correlated to the activity of the antagonists, although different behaviour of the ionic lock is observed for each antagonist.

The distance of the salt bridge between Asp[125] and Lys[331] in the α_{1D}-AR increases for most of the antagonists, except for SNAP-8719 (**11**). Four different patterns of distance change (Fig. 5.10) can be observed: a steadily increase over time, a strong increase at the start of the MD simulation followed by a stabilisation, a fluctuation of the salt bridge and a decrease in distance during the MD simulation. The distance of the salt bridge after 5 ns appears to result in range of 13 Å and 20 Å compared to a starting distance of 14 Å. This compares to a distance of 14.83 Å in the crystal structure of the β_2-AR and ~16 Å in the uncomplexed homology model after MD simulation. These data suggest that an increase in distance, compared to the starting structure, of the salt bridge during a MD simulation of the complexed α_{1B}-AR is expected after the binding of the antagonist as it is observed for eleven out of twelve compounds.
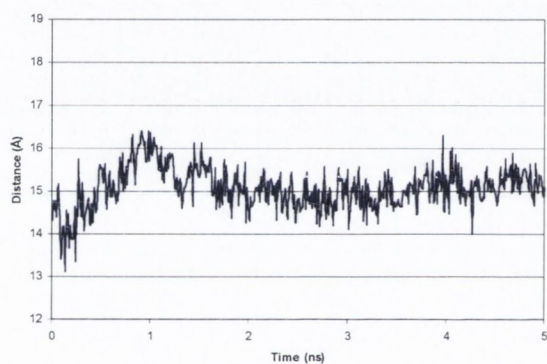
a) Alfuzosin (**7**)
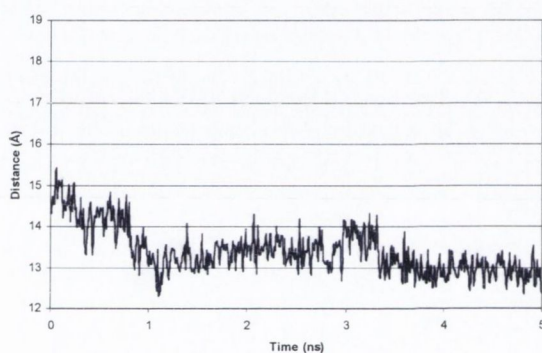
b) SNAP-8719 (**11**)
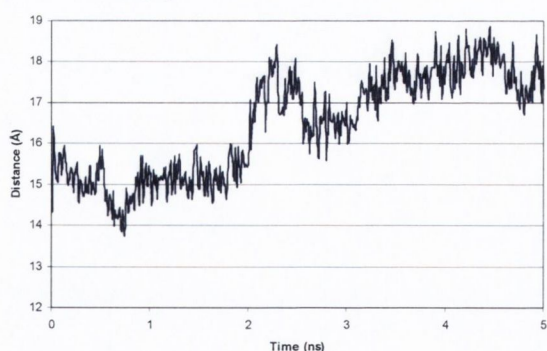
c) RS-100,975 (**14**)
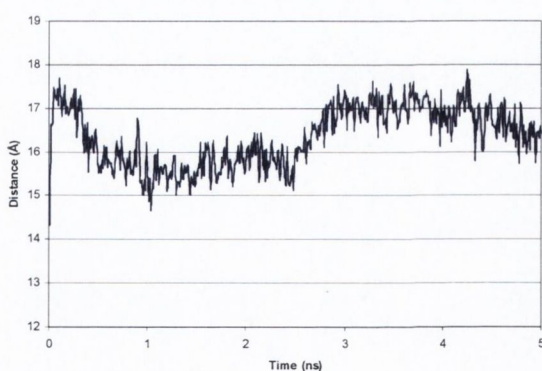
d) Uropidil (**15**)

Figure 5.10: Examples of behaviour of the salt bridge between Asp[125] Lys[331] in the α_{1B}-AR or a selection of compounds showing (a) a strong increase followed by stabilization, (b) a descrease followed by stabilization, (c) an increase over time and (d) a fluctuation in the distance of the ionic lock.

5.4.3. α_{1D}-Adrenoceptor

Helix RMSD analysis of the α_{1D}-AR antagonist bound structures shows stable proteins with limited movement of the helices which can be regarded as optimization (example in Fig. 5.11). Individual helices appear to move specifically in reaction to the binding of a ligand. The overall movements are similar to that of the α_{1A}-AR and α_{1B}-AR with the binding pocket formed by TM-III, TM-IV, TM-V and TM-VI opening up. There is also a movement of TM-I as observed for the α_{1A}-AR. Alignment with the β_2-AR (Fig. 5.12) shows similar structures, although not as similar as observed for the α_{1A}-AR. Despite this, the overlay of α_{1D}-AR with β_2-AR show that α_{1D}-AR moves towards a conformation that is more similar to β_2-AR than it is to rhodopsin.
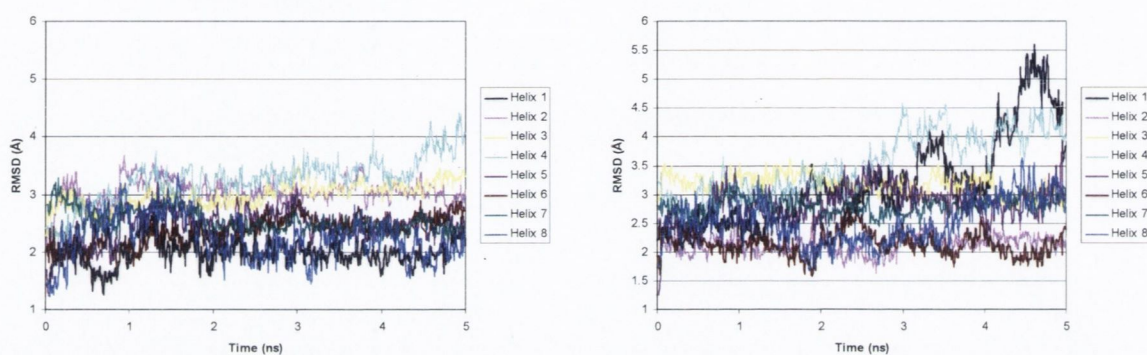
a) BMY-7378 (**10**)   b) WAY-100635 (**13**)

Figure 5.11: Helix RMSD analysis of the α_{1D}-AR complexed with (a) BMY-7378 and (b) WAY-100635 during a 5 ns MD simulation.
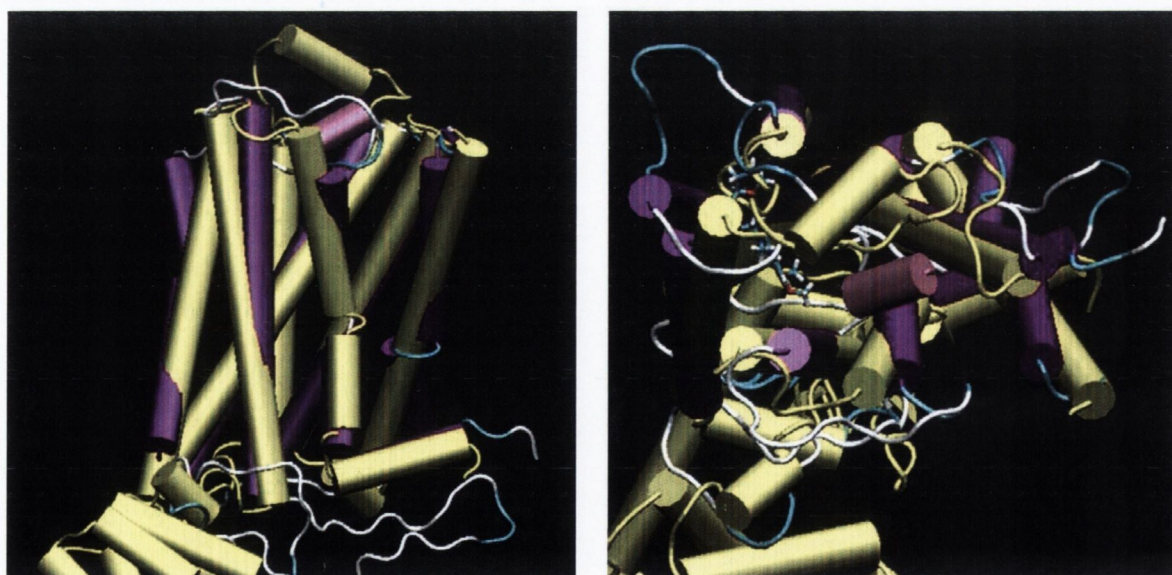
Figure 5.12: Alignment of alfuzosin-α_{1A}-AR complex (purple) with the crystal structure of β_2-AR (yellow) seen from the (left) side and (right) top.

The interaction between Asp^{193} and Arg^{194} shows either a stable or a decrease in distance (Fig. 5.13). The range of distances covered after the MD simulation is 6 to 10 Å with the most common distance being 7 Å. This can be compared to 4.84 Å in the β_2-AR and ~8 Å in the α_{1D}-AR homology model after MD.

The interaction between Asp^{193} and Glu^{343} (example in Fig. 5.13) shows three different patterns with either a stable distance during the trajectory, an increase in distance or a decrease in distance over time. When there is an increase in the distance between the two residues, the range of final distances is 10 to 16 Å compared to a starting distance of 8 Å. The most common distance after five ns of MD simulation is 11 or 12 Å. This can be compared to 14.8 Å in the β_2-AR and ~12 Å in the α_{1B}-AR homology model after MD.

The interaction between Arg^{194} and Glu^{343} (example in Fig. 5.13) shows only two different kinds of behaviour. All of the models show a decrease of distance of ~1 Å during minimization, heating and equilibration followed by a production run that shows either a stable distance or an increase of the distance over time. The range of distances after 5 ns is from 4 to 9 Å with either 4, 7 or 9 Å being the most common distances. There is no convergence in the final distance in the protein-ligand complexes. This can be compared to 13.02 Å in the β_2-AR and ~4.5 Å in the α_{1B}-AR homology model after MD. The different patterns of behaviour in the ionic lock show no correlation to weak, medium or strong binding antagonists.

c) WAY-100635 (**13**)
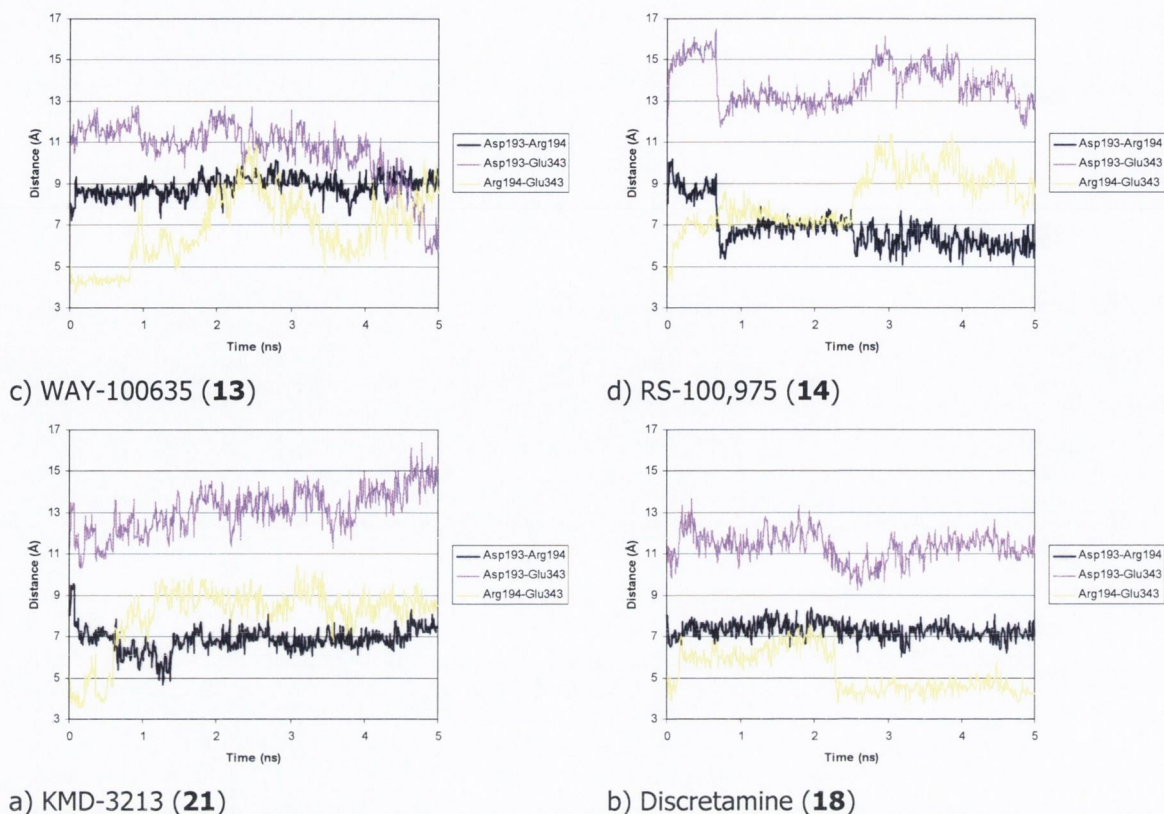
d) RS-100,975 (**14**)

a) KMD-3213 (**21**)

b) Discretamine (**18**)

Fig 5.13: Examples of behaviour of the ionic lock between Asp[193] Arg[194] and Glu[343] in the α_{1D}-AR for (a) Way-100635, (b) RS-100,975, (c) KMD-3213 and (d) Discretamine showing the different patterns of distances between the three residues that have been observed in the MD simulations.

The distance of the salt bridge between Asp[176] and Lys[385] in the α_{1D}-AR increases for most of the antagonists except SNAP-8719 (**11**) and NAN-190 (**29**). Three different patterns of changes in the distance can be observed (example in Fig. 5.14): a steady increase over time, a strong increase at the start of the MD simulation followed by a stabilisation of this distance and a stable distance during the MD simulation. The distance of the salt bridge after 5 ns appears to fall in range of 15 Å and 22 Å compared to a starting distance of 15 Å. The most common distance after the MD simulation was found to be 20 Å. This compares to a distance of 14.83 Å in the crystal structure of the β_2-AR and ~15 Å in the uncomplexed homology model after MD simulation. These data suggest that a longer distance of the salt bridge is correlated to the binding of the antagonist as it is observed for eleven out of twelve compounds.

a) Cyclazosin (**4**)

b) BMY-7378 (**10**)

c) NAN-190 (**12**)

d) WAY-100635 (**13**)
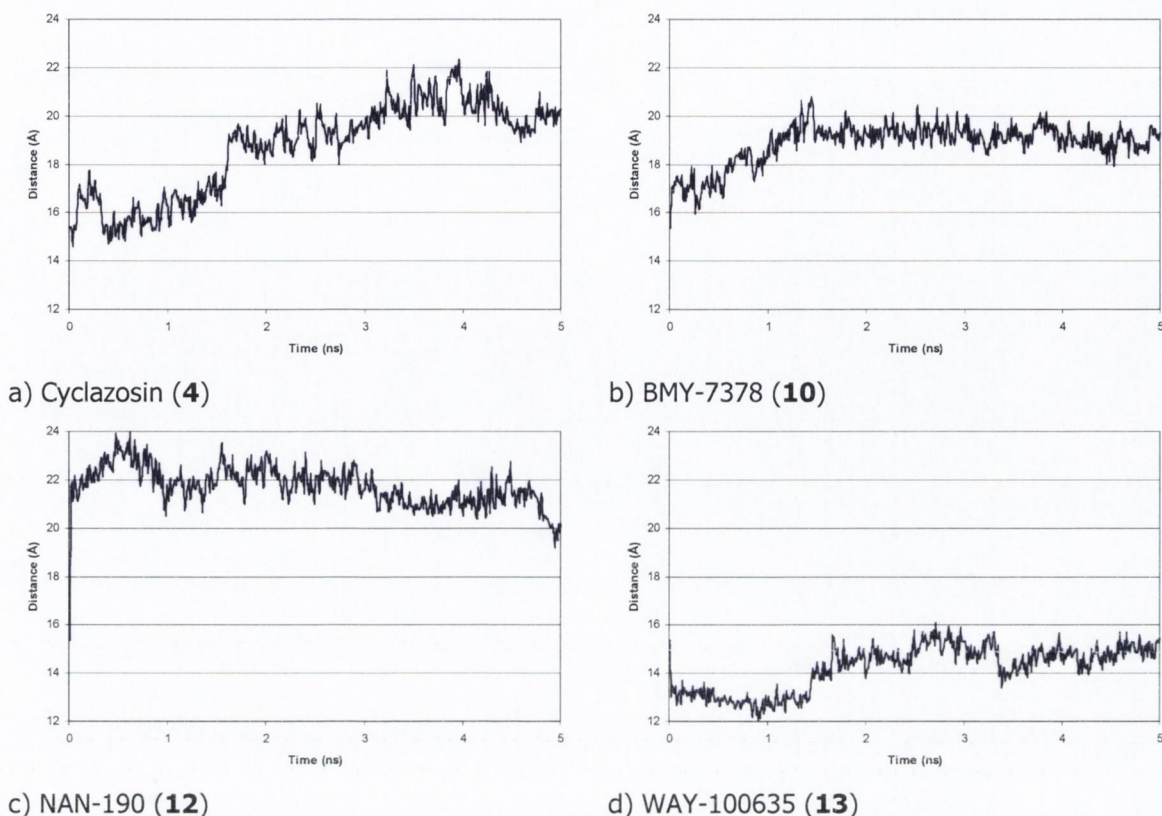
Figure 5.14: Examples of behaviour of the salt bridge between Asp^{176} Lys^{385} in the α_{1D}-AR for a selection of compounds showing (a) increase, (b+d) increase followed by stabilization and (c) decrease of the distances over time.

Two interesting cases are NAN-190 (**12**) and WAY-100635 (**13**). NAN-190 (**12**), which is the most potent α_{1D}-AR antagonist shows a gradual decrease in the salt bridge distance and WAY-100635 (**13**), which is selective for α_{1D}-AR even though it has a medium binding affinity, shows an increase in the salt bridge distance. However the initial homology model has a distance of 15.3 Å between Asp^{176} and Lys^{385}. Examination of the plot shows that the complex with NAN-190 (**12**) an initial increase to 21 Å which occurs during energy minimization, heating and equilibration. The production run shows a decrease of this distance but it is still is larger than in the starting structure. The complex with WAY-100635 (**13**) shows an opposite pattern with a decrease during the phases of minimization, heating and equilibration. The production run shows an increase in distance from its starting point resulting in a distance that is equal to the starting distance between the two residues.

5.4.4. Dynamics of α_1-ARs

During the MD simulation several changes occur in the α_1-ARs. The conformation changes form a 'rhodopsin-like' conformation to a 'β_2-AR-like' conformation. This is expected because the α_1-ARs share a higher sequence similarity to the β_2-AR than is is to rhodopsin. This also shows that MD simulations are suitable for optimizing ligand-receptor complexes.

Changes such as in the ionic lock and the saltbridge can not be linked to common processes such as antagonist-binding but are more likely to be ligand specific. They still can be used to determine if a different conformation has been obtained.

5.5. Final structures

After the 5 ns of MD simulation a snapshot of the antagonist complexed α_1-AR subtypes was taken and subjected to energy minimization. The residues that are involved in binding of the antagonists to the receptor can be determined using LigPlot. A binding pocket for each ligand can be derived and by comparing the binding pockets for the different antagonists, a common binding pocket for each of the α_1-AR subtypes was derived. A score for the interaction between the ligand and receptor can be obtained using London dG, Affinity dG and Alpha HB. In addition, a ranking, based on the scores can be made. These ranks and score were compared to the available experimental K_i values.

5.5.1. α_{1A}-Adrenoceptor

Residues in TM-III, TM-V, TM-VI, TM-VII and ECL-IV-V are found to be involved in the binding of our set of α_1-AR antagonists (Table 5.13). The residues Asp^{106}, Val^{107}, Cys^{110}, Ile^{178}, Ser^{188}, Ser^{192}, Tyr^{285}, Phe^{288} and Tyr^{316} are found to be the main residues involved in binding of most of the ligands and can be regarded as the binding pocket of the α_{1A}-AR.

Hydrogen bonding between the receptor and the antagonists occurs with the residues Asp^{106}, Cys^{110}, Ile^{178}, Ser^{192}, and Tyr^{285}. Of these, Asp^{106} is most commonly involved in hydrogen bonding with 8 out of 12 antagonists showing this interaction. Out of the other four, three show a hydrophobic interaction. Therefore, this Asp^{106} can be regarded as essential in interation with antagonists. In comparison with the initial complexes, before

MD simulation, this interaction is not observed and suggests that MD simulation optimizes the interaction between ligand and receptor.

As hydrogen bonding via the protonated nitrogen is expected, the interaction with a specific residue can be determined. In most cases this interaction is with Asp[106]. Discretamine (**18**) and corynanthine (**19**), which are structurally related, form hydrogen bonds with Cys[110]. These two antagonists have a significant different structure than the other ligands as they have no long aliphatic chains, but consist of connected rings, which make them conformationally restricted and therefore more rigid. The compound Cyclazosin (**4**) interacts with Ser[188] and WAY-100635 (**13**) with Ser[192]. The only compound that does not form a hydrogen bond via its protonated nitrogen is NAN-190 (**12**), although a non-bonded interaction with Asp[106] is observed. Perhaps a longer MD simulation could show the formation of a hydrogen bond.

Compared to the literature (chapter 1.2.5) we find similarities in the antagonist binding pocket with the interacting residues: Asp[106], Cys[110], Ile[178], Phe[193], Phe[288], Phe[312] and Tyr[316]. Interaction of the residues Gln[177], Asn[179] and Phe[308] are not observed in our ligand-protein complexes. The interaction with the binding pocket described in literature, despite the lack of antagonists interaction with some residues, suggests that the ligands are well positioned in the binding pocket of our MD optimized complexes. This is an improvement compared to our initially generated complexes obtained before MD simulation. A number of those residues are shown, in the literature, to interact with agonists, but according to our studies, they also take part in interactions with antagonists.

Table 5.13: Overview of the residues of the α_{1A}-AR which are involved in binding of the selection of compounds via (1) hydrogen bonds or via (2) hydrophobic interactions with the residue. Ligands are coloured by strong binders (green), medium binders (orange) and weak binders (blue).

	TM-III							ECL-IV-V		TM-V					
	Ala 103	Asp 106	Val 107	Cys 110	Thr 111	Ile 114	Met 115	Arg 166	Ile 178	Tyr 184	Val 185	Phe 187	Ser 188	Ala 189	Gly 191
4		1	2	2	1								1		
7	2	1	2		1					2	2				
10		1		2										2	
11		1		1									1		
12		2							1				2	2	
13		2	2	2					2		2				
14		1	2	2					2		2		2	1	2
15		1	2	2				1			2		1		
18		2	2	1			2								
19			2	1		2							2		
21		1		2					2		2				
27		1	2	2				2	2		2		2	2	

	TM-V		TM-VI								TM-VII			
	Ser 192	Phe 193	Tyr 194	Pro 196	Phe 281	Val 282	Tyr 285	Phe 288	Phe 289	Met 292	Lys 309	Phe 312	Trp 313	Tyr 316
4	1						2	2		2				
7					2	2	2							
10		2					2	2	2				2	2
11	1	2					2	2	2	2	2		2	
12			2				2	2	2				2	2
13	1	2			2		2	2	2	2				2
14	2						2							2
15					2		2							
18	2	2					2							2
19	2				2	2	2							
21	2	2					1			2	2	2		
27						1	2							2

When comparing the scores and ranks for each compound docked in the α_{1A}-AR to the experimental binding affinites (Table 5.14) it can be observed that there is no correlation between the two (Table 5.15). The three scoring functions do not provide a consistent result which indicates a weak or strong binder. Some compounds are identified as active, such as SNAP-1069 (**21**) with a K_i of 0.04 nM being ranked as a strong binder by all three scoring functions.

Table 5.14: Score and ranking for each compound in the α_{1A}-AR after 5 ns of MD simulation.

| Ligand | Score | | | | Ranking | | | |
	K_i (nM)	London dG (kcal mol^{-1})	Affinity dG (kcal mol^{-1})	Alpha HB (kcal mol^{-1})	K_i Rank	London dG	Affinity dG	Alpha HB
4	12	-13.64	-8.88	-142.02	5	6	4	5
7	10	-13.42	-6.63	-141.73	4	8	11	6
10	250	-13.43	-8.28	-139.27	9	7	8	7
11	294	-15.49	-11.63	-119.84	11	2	1	10
12	2	-11.44	-8.32	-109.01	3	11	7	11
13	144	-12.51	-10.93	-147.30	8	10	2	2
14	1.0	-15.13	-6.80	-144.58	2	4	10	3
15	288	-16.56	-8.25	-143.31	10	1	9	4
18	616	-12.67	-8.78	-128.86	12	9	5	8
19	142	-9.86	-6.24	-95.80	7	12	12	12
21	0.04	-15.47	-10.24	-160.89	1	3	3	1
27	16	-14.35	-8.71	-120.79	6	5	6	9

Table 5.15: Correlation (R^2) for each of the scoring function and ranking with the experimental Ki values.

Scoring method	Correlation with score/rank	Correlation (R^2)
London dG	score	0.000
	rank	0.001
Affinity dG	score	0.045
	rank	0.044
Alpha HB	score	0.016
	rank	0.099

Based on the similarity between the binding described in literature and the identified interaction, we can conclude that MD simulation has generated more realistic conformations of the protein-ligand interactions, resulting in the identification of interactions between ligand and protein. Using our findings a common binding pocket for the α_{1A}-AR can be derived consisting of Asp^{106}, Cys^{110}, Ile^{178}, Ser^{192}, Tyr^{285}, Asp^{106}, Val^{107}, Cys^{110}, Ile^{178}, Ser^{188}, Ser^{192}, Tyr^{285}, Phe^{288} and Tyr^{316}. Scoring these complexes show no correlation with experimental data.

5.5.2. α_{1B}-Adrenoceptor

Residues in all of the helices except TM-I, TM-II and ECL-VI-V are found to be involved in the binding of α_{1B}-AR antagonists (Table 5.16). However, there is only one residue in TM-IV which binds to one antagonist, suggesting that this residue does not play a major role in the binding pocket and, therefore, the α_{1B}-AR binding pocket consists of residues from TM-III, TM-IV, TM-V, TM-VI and ECL-IV-V.

The residues Asp^{125}, Thr^{130} and Ser^{207} are commonly found to interact via hydrogen bond with the antagonists. Besides this, residues Cys^{129}, Lys^{185}, Glu^{200} Tyr^{203}, Ala^{204} and Tyr^{338} are also capable of forming a hydrogen bond with one of the antagonist. These eight residues form the common binding pocket of the α_{1B}-AR. WAY-100635 (**13**) and SNAP-1069 (**27**) show H-bonding interaction with Asp^{125}. KMD-3213 (**21**) shows two possible interactions with Cys^{129} and Thr^{130}. Alfuzosin (**7**) and BMY-7378 (**10**) show a hydrogen bond with Ser^{207}. SNAP-8719 (**11**) forms a hydrogen bond with Glu^{200}. Six ligands (**4, 12, 14, 15, 18** and **19**) show no formation of a hydrogen bond via the protonated nitrogen. Cyclazosin (**4**) is the best ligand and maybe this indicates other binding modes.

Residues that play an important role in binding via non-bonded interactions are Asp^{125}, Val^{126}, Cys^{129}, Tyr^{203}, Ser^{207}, Phe^{310}, Phe^{311}, Trp^{335} and Tyr^{338}. The residue Phe^{310} is involved in binding to all ligands, suggesting that this residue is essential in the interactions with antagonists. Several residues show interaction with only one or two antagonists which questions if these residues are involved in the binding to a class of compounds or just to a particular antagonist.

There does not appear to be a residue which is specifically involved in the binding of ligands with a low binding affinity or a high binding affinity. This suggest that the binding pocket of the α_{1B}-AR consists of several residues commonly involved in binding but very few that are responsible for selective binding properties of antagonists.

Compared to the literature (chapter 1.2.5) we find similarities in the antagonist binding pocket with the interacting residues: Asp^{125}, Cys^{129}, Gly^{196}, Val^{197}, Phe^{310}, Phe^{334} and Tyr^{338}. Interaction of the residues Thr^{198}, Phe^{213} and Phe^{330} are not observed in our ligand-protein complexes. According to the data about the binding pocket found in the literature, it seems that the ligands are well positioned. Because we have optimized our complexes with MD simulations a good ligand-protein complex is obtained. Similar to the α_{1A}-AR, several residues that are supposed to be agonist specific, according to our studies, take part in interaction with antagonists.

Table 5.16: Overview of the residues of the α_{1B}-AR which are involved in binding of the selection of compounds via (1) hydrogen bonds or via (2) hydrophobic interactions with the residue. Ligands are coloured by strong binders (green), medium binders (orange) and weak binders (blue).

	TM-III					TM-IV		ECL-IV-V				TM-V			
	Asp 125	Val 126	Cys 129	Thr 130	Ile 133	Ile 176	Trp 184	Lys 185	Asn 190	Gly 196	Val 197	Glu 199	Glu 200	Tyr 203	Ala 204
4	2	2	2										2	2	1
7		2	2	1											2
10			2												
11						2							1	2	
12			2	1	2	2								2	2
13	1		2						2				2	2	2
14	1	2						2						2	
15	2	2	2							2	2		2	2	
18	2	2	2					2				2		2	
19						2					2		1		
21	1	2	1	1	2			1		2	1			1	
27	1														

	TM-V				TM-VI							TM-VII			
	Phe 206	Ser 207	Ser 208	Phe 212	Phe 303	Trp 307	Phe 310	Phe 311	Ala 313	Leu 314	Gly 317	Lys 331	Phe 334	Trp 335	Tyr 338
4		1		2			2	2	2					2	1
7	2	1					2	2	2	2			2		
10		1		2		2	2	2							2
11							2	2							
12	2	2		2		2	2							2	
13		2	2				2					2	2	2	2
14							2						2	2	2
15		2					2				2	2	2		2
18							2	2							2
19			2			2	2	2		2	2				
21		2			2	2	2	2						2	2
27		2					2	2		2				2	2

When comparing the scores for each compound docked in the α_{1B}-AR to the experimental binding affinites (Table 5.17) it can be observed that there is no correlation between both (Table 5.18). When ranking the experimental binding affinities and the scores (Table 5.17), no correlation is obtained either (Table 5.18). The three scoring functions do not provide a consistent indication which antagonist is a weak or strong binder, although cyclazosin (4) is scored as a strong binder, suggesting that strong binders can be identified. KMD-3213 (21) shows much better scores than any of the other ligands. This is likely due to an overestimation of a binding interaction.

Table 5.17: Score and ranking for each compound in the α_{1B}-AR after 5 ns of MD simulation.

Ligand	Score				Ranking			
	K_i (nM)	London dG (kcal mol^{-1})	Affinity dG (kcal mol^{-1})	Alpha HB (kcal mol^{-1})	K_i Rank	London dG	Affinity dG	Alpha HB
4	0.13	-14.43	-9.42	-187.73	1	3	2	2
7	10	-14.46	-6.13	-132.56	2	2	10	5
10	630	-10.77	-8.03	-119.23	11	11	5	10
11	191	-14.09	-7.98	-148.83	7	4	6	4
12	15	-13.12	-8.46	-125.36	3	7	3	7
13	186	-13.85	-8.31	-156.81	6	5	4	3
14	79	-12.09	-5.80	-124.70	5	8	11	8
15	1320	-11.89	-5.49	-123.70	12	9	12	9
18	360	-11.67	-7.95	-112.08	9	10	7	11
19	517	-13.38	-7.00	-140.96	10	6	8	6
21	20	-20.41	-12.70	-188.79	4	1	1	1
27	200	-10.00	-6.54	-111.61	8	12	9	12

Table 5.18: Correlation (R^2) for each of the scoring function and ranking with the experimental Ki values.

Scoring method	Correlation with score/rank	Correlation (R^2)
London dG	score	0.403
	rank	0.685
Affinity dG	score	0.428
	rank	0.441
Alpha HB	score	0.385
	rank	0.622

Based on our data, a common binding pocket can be derived consisting of Asp^{125}, Thr^{130}, Ser^{207}, Asp^{125}, Val^{126}, Cys^{129}, Tyr^{203}, Ser^{207}, Phe^{310}, Phe^{311}, Trp^{335} and Tyr^{338}. As a whole, no correlation is found between the scoring and experimental data, but some strong binders are well described by these scorings, which is in contrast to the α_{1A}-AR, where this this identification of strong binders is not obtained.

5.5.3. α_{1D}-Adrenoceptor

Residues in all of the helices except TM-I and ECL-VI-V are found to be involved in the binding of α_{1D}-AR antagonists (Table 5.19). However, those residues in TM-II and TM-IV only bind to one antagonist each. This would suggest that they do not play a major role in defining the binding pocket and therefore the binding pocket is formed by residues in TM-III, TM-IV, TM-V, TM-VI and ECL-IV-V.

The residue Asp^{176} is commonly found to interact with seven out of twelve antagonists to interact via the formation of a hydrogen bond to the α_{1D}-AR. This makes this residue

important in binding of antagonists, which is similar to the α_{1A}-AR. In addition, residues Thr^{181}, Ile^{248}, Glu^{251}, Ser^{258}, Ser^{262}, Phe^{357}, Trp^{361}, Leu^{368}, Gly^{391} and Tyr^{392} show the formation of a hydrogen bond with one or two antagonists.

Regarding the interaction with the antagonist protonated nitrogen, the most common hydrogen bond that is formed, with residue Asp^{176}, is observed in six cases. Three other cases show hydrophic intraction with this residue and could potentially lead to the formation of a hydrogen bond if longer MD simulations are performed. This suggests that the most likely interaction for strong binders is to occur via the formation of a hydrogen bond between the protonated nitrogen and Asp^{176}. One different hydrogen bond is formed with residue Ser^{258} for discretamine (**18**).

Residues that play an important role in binding via non-bonded interactions are Asp^{176}, Cys^{180}, Ala^{255}, Ser^{258}, Trp^{361}, Leu^{368} and Tyr^{392}. Especially Cys^{180} and Trp^{361} show interaction with a large number of antagonists, making these residues important in binding. These seven residues can be regarded as the common antagonist binding pocket for the α_{1D}-AR. There are other residues that show interaction with only one or two antagonists meaning that they are not essential for antagonist activity.

Considering the six best compounds in terms of affinity (**4, 7, 10, 11, 12** and **21**) interactions with Cys^{180}, Ala^{255} and Trp^{361} seem to be significant for the antagonist binding.

Compared to the data in the literature (Chapter 1.2.5) similarities are found in the antagonist binding pocket with the interacting residues: Asp^{176}, Cys^{180}, Gly^{247}, Ile^{248}, Thr249, Phe^{263}, Phe^{365} and Tyr^{392}. Interaction of the residues Thr^{198}, Phe^{213} and Phe^{330} are not observed in our ligand-protein complexes. Similar to α_{1A}-AR and α_{1B}-AR agonist-specific residues also interact with antagonists.

Table 5.19: Overview of the residues of the α_{1D}-AR which are involved in binding of the selection of compounds via (1) hydrogen bonds or via (2) hydrophobic interactions with the residue. Ligands are coloured by (green) strong binders, (orange) medium binders and (blue) weak binders.

	TM-III									ECL-III-IV		TM-IV	
	Val 149	Trp 172	Asp 176	Val 177	Cys 179	Cys 180	Thr 181	Ser 183	Ile 184	Val 227	Pro 231	Lys 236	Pro 241
4			1			2							2
7				2		2	1				2		
10			1			2	2	2					
11			1										2
12			2	2		2				2			
13	2	2	1		2	2						2	
14													
15			1	2		2	1						
18			2			2							
19			1			2							
21			1			2							
27								2					

	TM-IV						TM-V						
	Phe 245	Gly 247	Ile 248	Thr 249	Glu 250	Glu 251	Tyr 254	Ala 255	Ser 258	Ser 259	Ser 262	Phe 263	Phe 357
4	2	2	2			1		2					
7							2	2	2		1	2	2
10				2		2		2					
11	2					2		2	2	2			
12	2	2	1								2	2	
13													
14							2	2			2	2	1
15			2			2		2					
18			2							1	2	2	
19	2		2				2						
21			2						2	2		2	
27				2	2	2	2						

		ECL-V-VI					TM-VI				
	Val 358	Cys 360	Trp 361	Phe 364	Phe 365	Leu 368	Gly 371	Trp 389	Gly 391	Tyr 392	Phe 393
4				2		2	2				
7			2								
10			2	2	2	2					
11			2	2		2	2	2		2	
12			2							2	
13		2		2		2		2	1	1	2
14			2								
15	2		1			2					
18			2					2		2	
19			2	2				2		2	
21			2			2				1	
27			2								

When comparing the scores and ranks for each compound docked in the α_{1D}-AR to the experimental binding affinites (Table 5.20) it can be observed that there is no correlation between the two (Table 5.21). The three scoring functions do not provide a consistent result that indicates which antagonist is a weak or strong binder. However, two weak binders RS-100,975 (**14**) and SNAP-1069 (**27**) are ranked low. Unfortunately the dataset is too small and there are too many inconsistencies to validate this model as being able to filter out weak binding ligands.

Table 5.20: Score and ranking for each compound in the α_{1D}-AR after 5 ns of MD simulation.

	Score				Ranking			
Ligand	K_i (nM)	London dG (kcal mol^{-1})	Affinity dG (kcal mol^{-1})	Alpha HB (kcal mol^{-1})	K_i Rank	London dG	Affinity dG	Alpha HB
4	3.2	-16.79	-8.41	-154.82	5	1	5	2
7	3.16	-11.65	-5.87	-145.23	4	9	10	3
10	6.3	-15.29	-9.44	-131.56	6	3	3	7
11	1.6	-15.15	-9.90	-140.20	2	4	1	5
12	0.8	-11.05	-8.44	-128.80	1	10	4	8
13	63	-15.06	-9.57	-169.50	8	6	2	1
14	100	-10.76	-5.08	-113.12	9	11	11	10
15	1660	-15.08	-7.70	-138.43	12	5	8	6
18	25	-12.69	-5.89	-106.85	7	8	9	11
19	253	-13.25	-8.38	-144.20	10	7	6	4
21	2.0	-16.27	-8.37	-127.76	3	2	7	9
27	790	-8.91	-4.18	-93.98	11	12	12	12

Table 5.21: Correlation (R^2) for each of the scoring function and ranking with the experimental Ki values.

Scoring method	Correlation with score/rank	Correlation (R^2)
London dG	score	0.008
	rank	0.082
Affinity dG	score	0.058
	rank	0.233
Alpha HB	score	0.027
	rank	0.022

Based on the final complexes obtained from the MD simulations, a common binding pocket can be derived and consists of Asp^{176}, Cys^{180}, Ile^{248}, Asp^{176}, Cys^{180}, Ala^{255}, Ser^{258}, Trp^{361}, Leu^{368} and Tyr^{392}. The scoring of our complexes as a whole, shows no correlation with experimental data.

5.5.4. Binding of antagonstist in optimized ligand-α_1-AR complexes

For the α_{1A}-AR, α_{1B}-AR and α_{1D}-AR different binding pockets have been derived and were aligned to each other (table 5.22). These binding pockets are similar, but there is a number of residues that show subtype specific interaction. This suggests that the mode of binding of each ligand to the different α_1-AR subtypes is similar. Therefore, functional groups attached to a 'core' part of the ligand can explain ligand-subtype specificity.

Table 5.22: Residues that form the general binding pockets of the three α_1-AR subtypes. The shaded residues shows the residues which are aligned in each of the three α_1-AR subtypes.

	α_{1A}-AR	α_{1B}-AR	α_{1D}-AR
Hydrogen bond	Asp^{106}	Asp^{125}	Asp^{176}
	Cys^{110}	Thr^{130}	Cys^{180}
	Ile^{178}		Ile^{248}
	Ser^{192}	Ser^{207}	
	Tyr^{285}		
Hydrophobic interaction	Asp^{106}	Asp^{125}	Asp^{176}
	Val^{107}	Val^{126}	
	Cys^{110}	Cys^{129}	Cys^{180}
	Ile^{178}	Tyr^{203}	Ala^{255}
	Ser^{188}	Ser^{207}	Ser^{258}
	Ser^{192}		
	Tyr^{285}		
	Phe^{288}	Phe^{310}	
		Phe^{311}	
		Trp^{335}	Trp^{361}
			Leu^{368}
	Tyr^{316}	Tyr^{338}	Tyr^{392}

Scoring of the different complexes has been performed but shows no correlation with experimental data. Scoring functions are known to be inaccurate and, therefore, using a small dataset, a correlation between the experimental binding affinity and the computational scores should not be expected.

5.6. Comparison of binding of ligands in α_{1A}-AR, α_{1B}-AR and α_{1D}-AR.

A comparison of how each of the twelve ligands bind to each of the α_1-AR subtypes has been made based on the formation of H-bonds and hydrophobic interactions using a 2D plot. If possible, functional groups of the ligand and residues of the α_1-AR subtype which can account for selectivity of the ligand for each α_1-AR subtype were identified.

5.6.1. Binding of Cyclazosin

Cyclazosin (**4**) is a compound with a medium binding affinity to the α_{1A}-AR (K_i =12 nM), a strong binding affinity for the α_{1B}-AR (K_i=0.13 nM) and a medium-strong binding affinity for the α_{1D}-AR (K_i =3.2 nM). The correlation of the scores with the experimental data is limited as it is only correctly predicted by Alpha HB (Table 5.23).

Table 5.23: Ranks (experimental and computated) and scores (computed) for the binding affinity of Cyclazosin (**4**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	3	-13.64	-8.88	-142.02	2	2	3
α_{1B}-AR	1	-13.43	-9.42	-187.73	3	1	1
α_{1D}-AR	2	-16.79	-8.41	-154.82	1	3	2

In each receptor (Fig. 5.15), three H-bonds are formed with this ligand. Only Ser[188] can be aligned in the α_{1B}-AR (Ser[207]) and α_{1D}-AR (Ser[258]). The other two resdidues involved in H-bonding are subtype specific. The protonated nitrogen interacts with Ser[188] in the α_{1A}-AR and Asp[176] in the α_{1D}-AR. For the α_{1B}-AR, no interaction with the protonated nitrogen is observed and suggests that cyclazosin (**4**) binds different in α_{1B}-AR. The α_{1B}-AR does interact through Ser[207] with the other nitrogen in the piperazine group which is not observed for the α_{1A}-AR and α_{1D}-AR. Furthermore, the nitrogen of the NH$_2$ group of the pyrimidine ring, shows interactions with each of the three α_1-ARs, although with non-aligned residues.The methoxy oxygen atoms do not form H-bonds. The oxygen positioned between the pyrazine and furan group shows interaction with all three receptors, although with three different residues.

The α_{1A}-AR, compared to the other two subtypes shows a limited number of non-bonded interactions. The residue Met[292] (α_{1A}-AR) interacts with the ligand, but no Met residues of the other α_1-AR subtypes show interaction. The residue Trp[285] in the α_{1A}-AR, and Trp[389] in

the α_{1D}-AR interact with the ligand, but the α_{1B}-AR shows no similar interaction is shown in the α_{1B}-AR. The cyclohexane ring attached to the piperazine show most hydrophobic interactions in the α_{1D}-AR.

This compound is potent for each of the three α_1-AR subtype and in each modelj which can be linked to the large number of observed interactions, suggesting that this model is correct. Our models show a different binding of the cyclazosin (4) for the α_{1B}-AR compared to the α_{1A}-AR and α_{1D}-AR.

Figure 5.15: Analysis of binding of cyclazosin (4) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

5.6.2. Binding of Alfuzosin

Alfuzosin (**7**) is a compound with a strong binding affinity to the α_{1A}-AR (K_i =10 nM) and the α_{1B}-AR (K_i =10 nM) and a strong binding affinity for the α_{1D}-AR (K_i =3.16 nM). This compound is not selective for any of the adrenoceptors as the difference in binding affinities is limited and there is strong interaction with all three α_1-AR subtypes. The correlation between the computed scores and the experimental binding affinity is not predicted correctly by any of the three scoring functions (Table 5.24), although Alpha HB predicts the strongest affinity correct. Because α_{1A}-AR and α_{1B}-AR have similar binding affinities, a ranking is likely not to perform well.

Table 5.24: Ranks (experimental and computated) and scores (computed) for the binding affinity of Alfuzosin (**7**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	2	-13.42	-6.63	-142.02	2	1	2
α_{1B}-AR	2	-14.46	-6.13	-132.56	1	2	3
α_{1D}-AR	1	-11.65	-5.87	-145.23	3	3	1

Two residues in each AR subtype are involved in binding which are aligned, being for α_{1A}-AR Thr[111] and Ser[192], for α_{1B}-AR, Thr[130] and Ser[207], for α_{1D}-AR Thr[181] and Ser[262] (Fig. 5.16). This Ser residue is aligned, for the α_{1A}-AR and α_{1B}-AR, but they show different interactions, where in the α_{1A}-AR, a H-bond is formed with the NH_2 group attached to the quinazoline group whereas, in the α_{1B}-AR, the aligned serine residue forms a H-bond with the nitrogen on the quinazoline group (the protonated nitrogen). Furthermore, a residue in the α_{1A}-AR (Thr) and α_{1B}-AR (Ser) shows interaction with the nitrogen in the chain that link the rings in the molecule. The α_{1D}-AR only forms hydrogen bonds with the nitro group of the quinazoline ring and not with the protonated nitrogen.

There is a number of other residues showing hydrophobic interactions with alfuzosin (**7**) which differ for each subtype. For example, the residue Cys[110] does not interact with the ligand in the α_{1A}-AR but similar residues in the α_{1B}-AR, Cys[129], and in the α_{1D}-AR, Cys[180], do show interaction with alfuzosin (**7**).

Summarizing, in the case of the alfuzosin (**7**) complexes, for the α_{1A}-AR and α_{1B}-AR a similar binding mode is observed which is different to α_{1D}-AR. Given the H-bonds formed in each adrenoceptor and the number of hydrophic interaction the binding of alfuzosin (**7**)

can be explained. Given the potency of alfuzosin (**7**) this model should be used in further studies.

a)

b)

c)
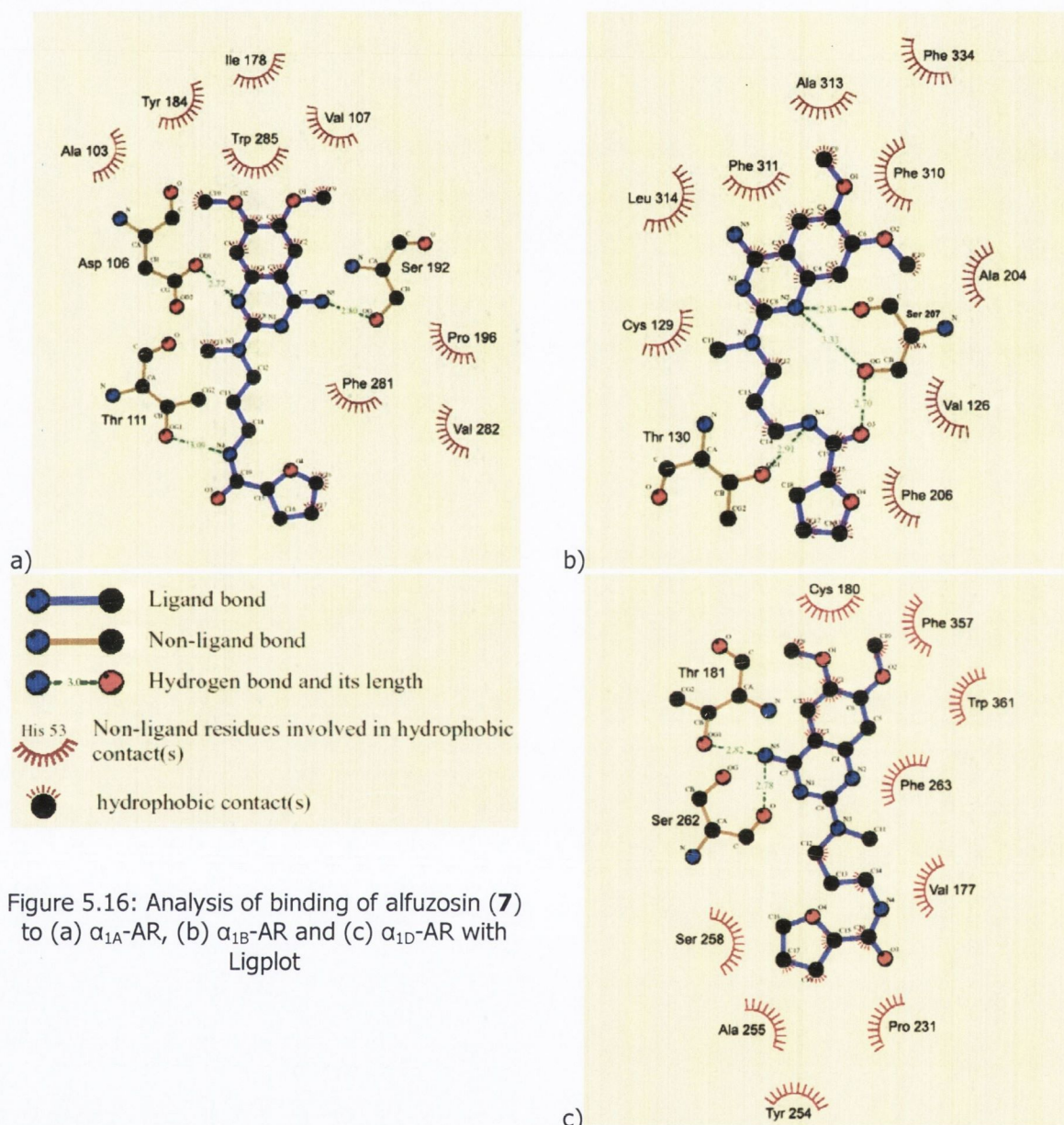
Figure 5.16: Analysis of binding of alfuzosin (**7**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

5.6.3. Binding of BMY-7378

BMY-7378 (**10**) is a compound with a weak binding affinity to the α_{1A}-AR (K_i =250 nM) and α_{1B}-AR (K_i =630 nM) and a strong binding affinity for the α_{1D}-AR (K_i =6.3 nM). This compound is selective for the α_{1D}-AR with a 250-fold stronger affinity over the α_{1A}-AR. Ranking for experimental binding affinity is not correctly predicted by any of the three scoring functions (Table 5.25).

Table 5.25: Ranks (experimental and computated) and scores (computed) for the binding affinity of BMY-7378 (**10**) to the three α_1-AR subtypes.
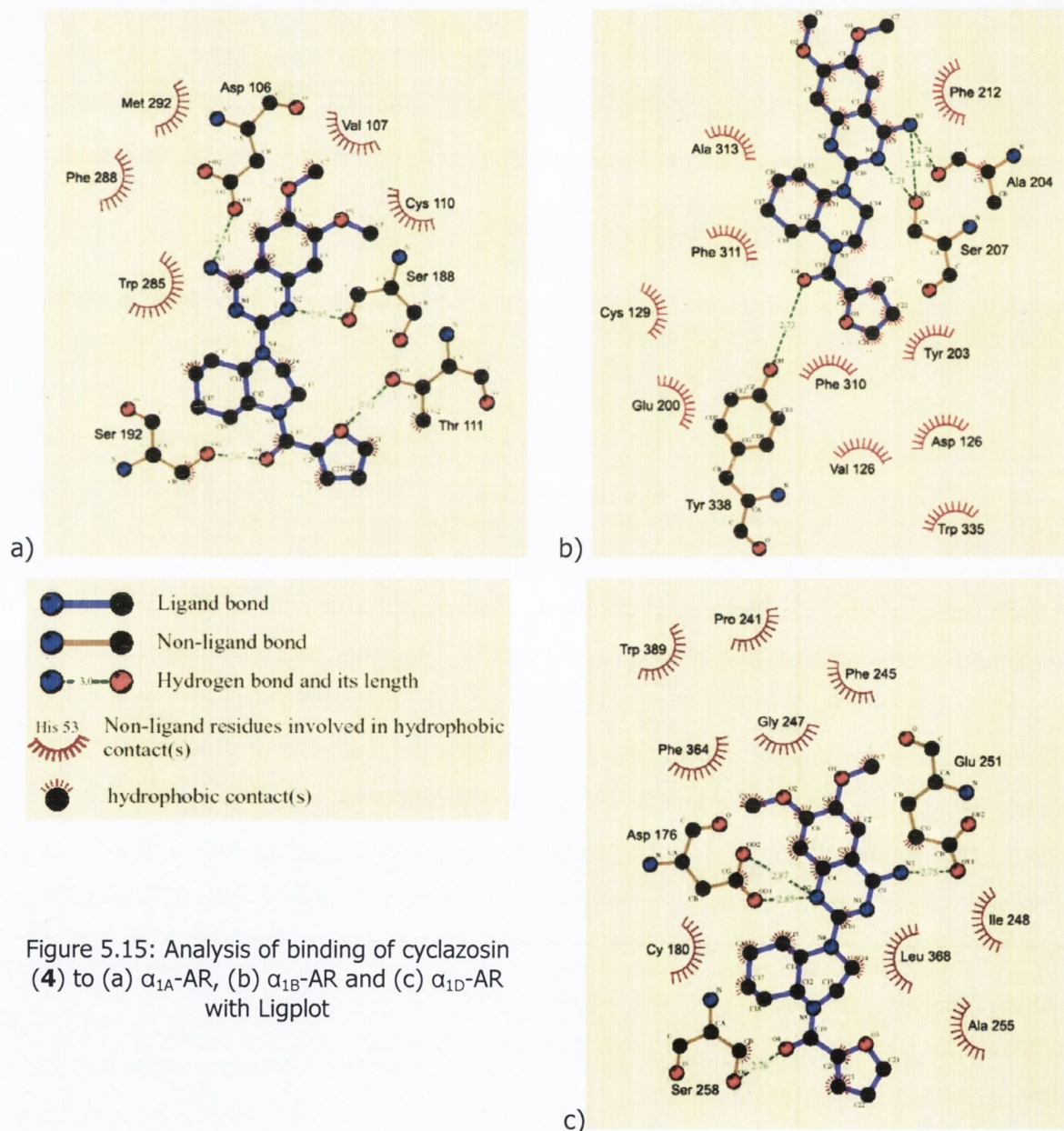
	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	2	-13.43	-8.28	-139.27	2	2	1
α_{1B}-AR	3	-10.77	-8.03	-119.23	3	3	3
α_{1D}-AR	1	-15.29	-9.44	-131.56	1	1	2

Formation of a hydrogen bond between the ligand with the three adrenoceptor subtypes (Fig. 5.17) is observed with the protonated nitrogen in the pyrazine group. This nitrogen interacts with residue Asp[106] for the α_{1A}-AR, Ser[207] for the α_{1B}-AR and Asp[176] for the α_{1D}-AR. Both aspartic acid residues in the α_{1A}-AR and α_{1D}-AR are aligned.

Furthermore, a H-bond is formed with Ser[262] in the α_{1D}-AR. Thus, based on H-bonding, the stronger affinity to the α_{1D}-AR can be explained. It can be observed that the rings formed by carbon 19-20-21-22 and 14-15-16-17-18 show a hydrophobic interaction with the α_{1A}-AR and α_{1D}-AR but not with the α_{1B}-AR. Furthermore there are more residues showing hydrophobic interactions with the phenyl rings in the α_{1B}-AR complex and very little interaction with the rest of the ligand.

The binding of BMY-7378 (**10**) to the α_{1B}-AR can be explained by the small number of interactions established. The selectivity for the α_{1D}-AR subtype could be explained not only because the H-bond formed with Asp[176], but also because the H-bond with Ser[212]. Moverover, more non-bonded interactions are observed in the α_{1D}-AR complex with BMY-737.

a)

Ligand bond

Non-ligand bond

Hydrogen bond and its length

His 53   Non-ligand residues involved in hydrophobic contact(s)

hydrophobic contact(s)
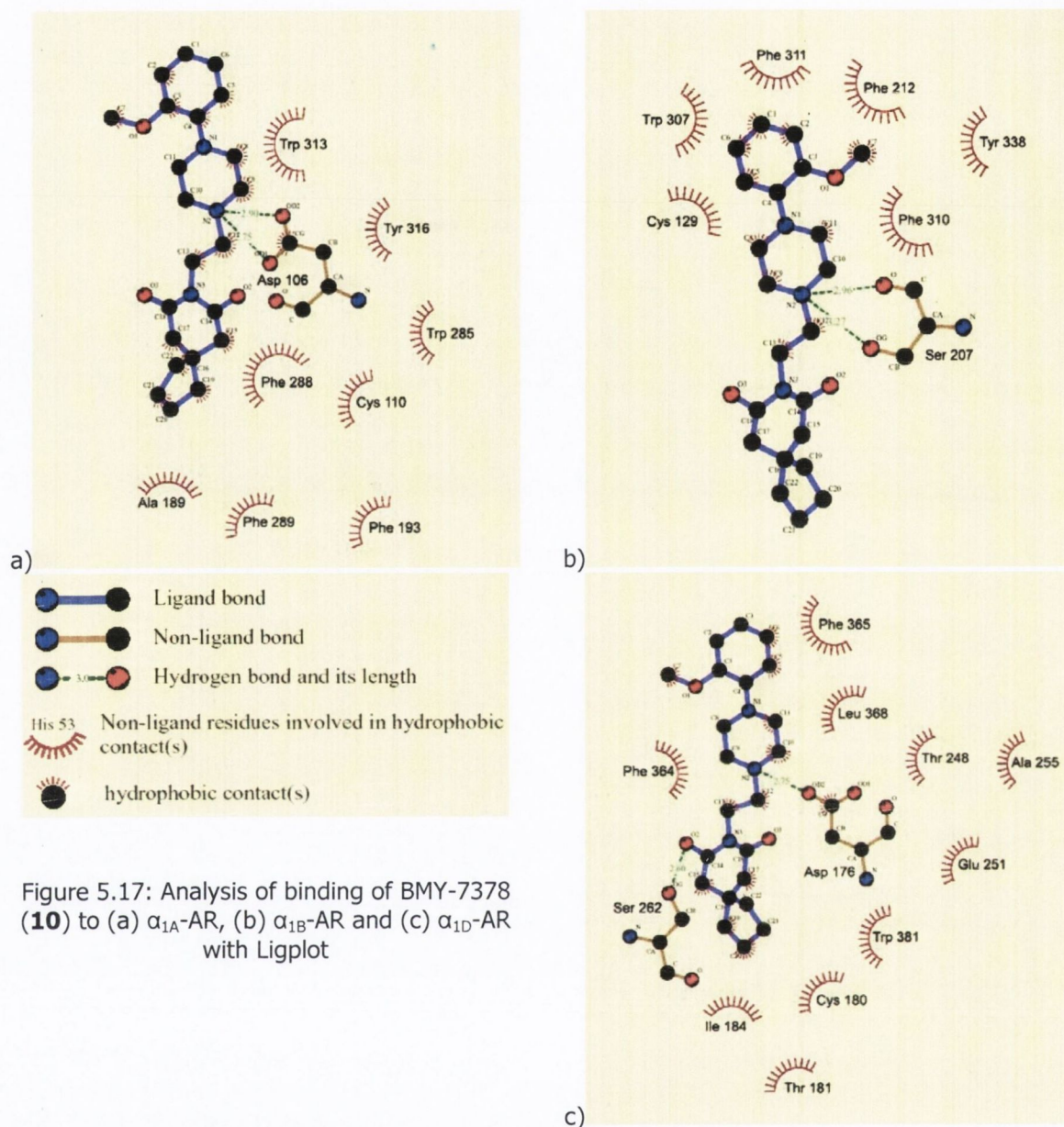
Figure 5.17: Analysis of binding of BMY-7378 (**10**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

b)

c)

5.6.4. Binding of SNAP-8719

SNAP-8719 (**11**) is a compound with a weak binding affinity to the α_{1A}-AR (K_i =294 nM) and α_{1B}-AR (K_i =191 nM) and a strong affinity for the α_{1D}-AR (K_i =1.6 nM). Therefore, this compound is selective for the α_{1D}-AR. The ranking of the experimental binding affinties is correctly predicted by London dG (Table 5.26). However, it is not predicted by the other two scoring functions.

Table 5.26: Ranks (experimental and computated) and scores (computed) for the binding affinity of SNAP-8719 (**11**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	3	-11.44	-8.32	-119.84	3	2	3
α_{1B}-AR	2	-14.09	-7.98	-148.83	2	3	1
α_{1D}-AR	1	-15.15	-9.90	-140.20	1	1	2

H-bonds are formed with the α_{1A}-AR via Asp^{106}, Cys^{110}, Ser^{188} and Ser^{192}. For the α_{1B}-AR via Glu^{200} and the α_{1D}-AR via Asp^{176}. Asp^{106} (α_{1A}-AR) and Asp^{176} (α_{1D}-AR) are aligned in both receptors and also interact with the same nitrogen which is protonated. This protonated nitrogen interacts in the α_{1B}-AR via residue Glu^{200}.

Considering the common H-bonds formed between SNAP-8719 (**11**) and acidic residues such as Asp^{106} (α_{1A}-AR), Glu^{200} (α_{1B}-AR) and Asp^{176} (α_{1D}-AR), the selectivity for the α_{1D}-AR could only be explained by the large number of hydrophobic interactions formed in the bindig pocket of this subtype (Fig 5.18c). It should be noted that the α_{1B}-AR has a limited number residues showing hydrophobic interactions.

a)

b)

Ligand bond

Non-ligand bond

Hydrogen bond and its length

His 53   Non-ligand residues involved in hydrophobic contact(s)

hydrophobic contact(s)
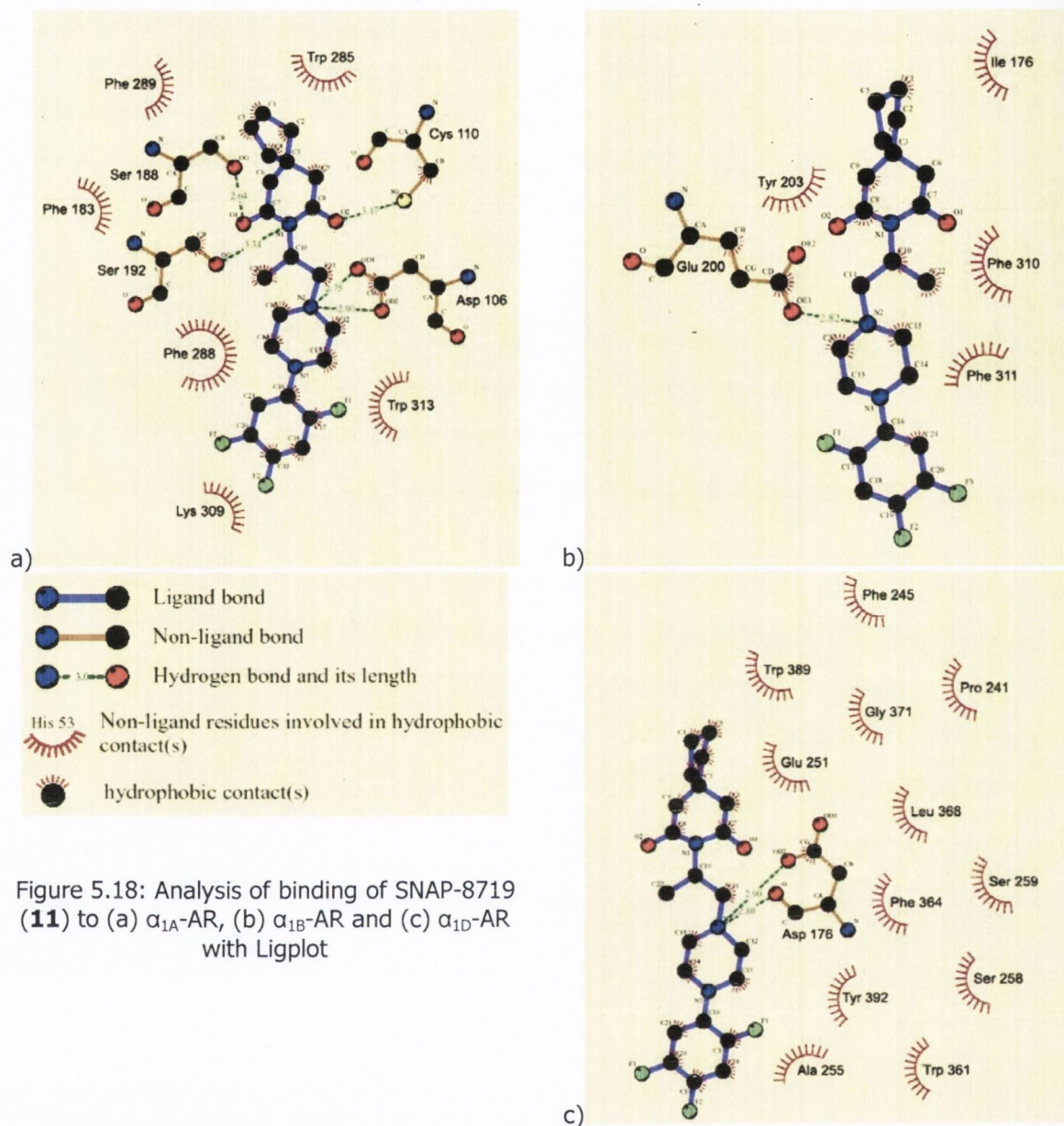
Figure 5.18: Analysis of binding of SNAP-8719 (**11**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

c)

5.6.5. Binding of NAN-190

NAN-190 (**12**) is a compound with a strong binding affinity to the α_{1A}-AR (K_i =2.0 nM), a medium binding affinity for the α_{1B}-AR (K_i =15 nM) and a strong binding affinity for the α_{1D}-AR (K_i =0.8 nM). This compound shows a better activity for α_{1A}-AR (7.5) than the α_{1B}-AR, it can be regarded as non-selective for the α_{1A}-AR because of the potency for the α_{1B}-AR, resulting in little difference in clinical use. The ranks of experimental binding affinities are incorrectly predicted by all three scoring functions (Table 5.27), limiting their use in practice.

Table 5.27: Ranks (experimental and computated) and scores (computed) for the binding affinity of NAN-190 (**12**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	2	-11.44	-8.32	-109.01	2	3	3
α_{1B}-AR	3	-13.12	-8.46	-125.36	1	2	2
α_{1D}-AR	1	-11.05	-8.44	-128.80	3	1	1

For each AR subtype, only one H-bond is observed with NAN-190 (**12**). For the α_{1A}-AR and α_{1D}-AR this is via an oxygen and for the α_{1B}-AR via a nitrogen (Fig. 5.19). None of the α_1-AR subtypes show interaction with the protonated nitrogen.

Several hydrophobic interactions are also involved in the three complexes and very few residues are specific for NAN-190. The phenyl rings on both ends of the molecule show hydrophobic interaction in each of the three α_1-AR subtypes. With a potent compound, more H-bonds, especially with the protonated nitrogen would be expected. Although a number of hydrophobic interactions are observed these structures, further refinement is needed.
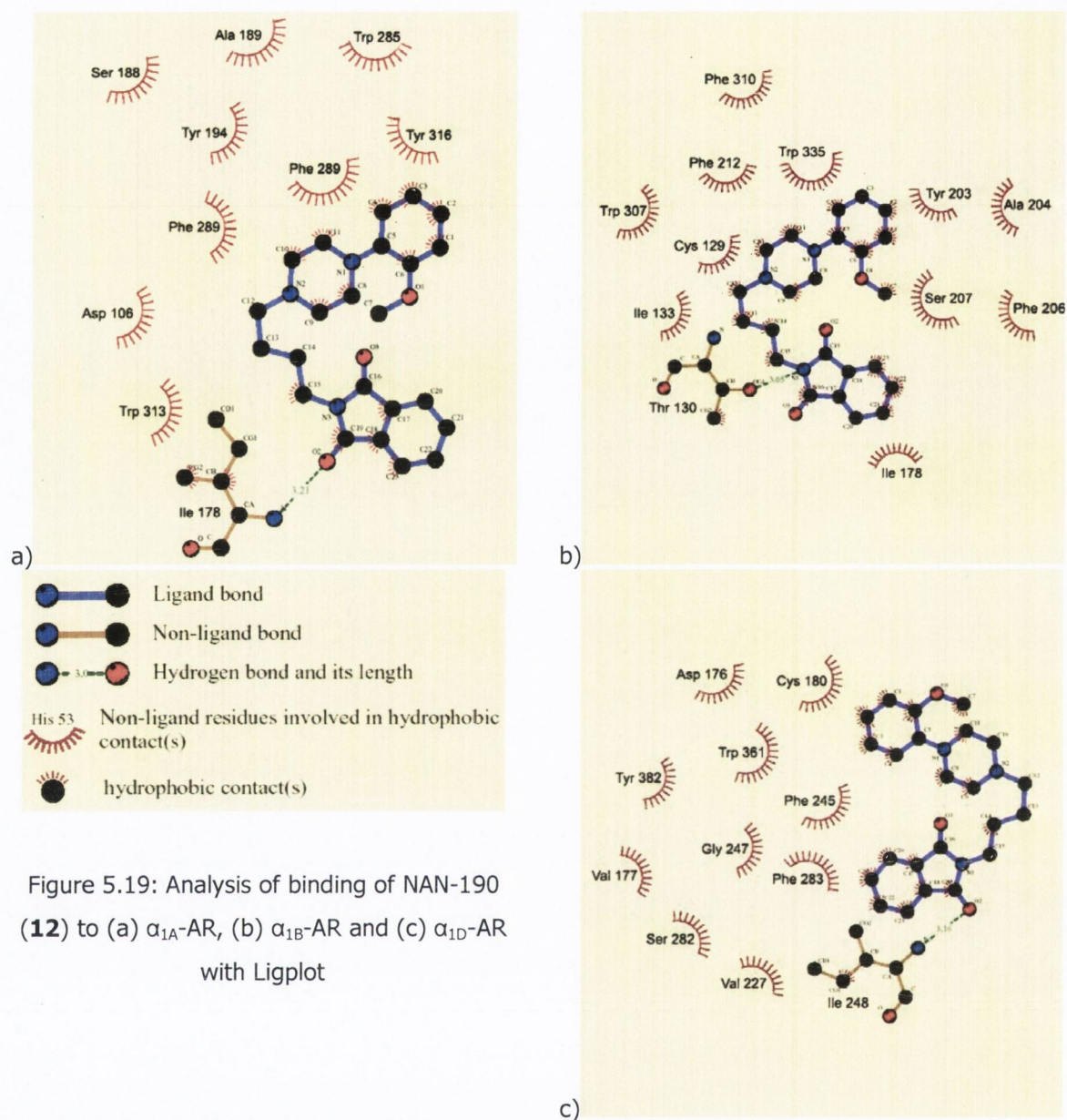
Figure 5.19: Analysis of binding of NAN-190 (**12**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

5.6.6. Binding of WAY-100635

WAY-100635 (**13**) is a compound with a weak binding affinity to the α_{1A}-AR (K_i =144 nM) and α_{1B}-AR (K_i =186 nM) but has a medium binding affinity for the α_{1D}-AR (K_i =63 nM). There is limited correlation between the ranking of the experimental binding affinities and the computed ranks (Table 5.28), although, the best ranked affinity is identified by London dG and Alpha HB.

Table 5.28: Ranks (experimental and computated) and scores (computed) for the binding affinity of WAY-100635 (**13**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	2	-12.51	-10.93	-147.30	3	1	3
α_{1B}-AR	3	-13.85	-8.31	-156.81	2	3	2
α_{1D}-AR	1	-15.06	-9.57	-169.50	1	2	1

The protonated nitrogen of ligand WAY-100635 forms a H-bond with the α_{1A}-AR via residue Ser[192], with the α_{1B}-AR via residue Asp[125] and with the α_{1D}-AR via residue Asp[176] (Fig. 5.20). The two aspartic acid residues occupy the same position in the α_{1B}-AR and α_{1D}-AR. The α_{1D}-AR shows in contrast to the other two the formation of a hydrogen bond with an oxygen via residue Gly[391] and Tyr[392].

A common hydrophobic interaction is observed via Cys[110] in the α_{1A}-AR which corresponds to Cys[129] in the α_{1B}-AR and Cys[180] in the α_{1D}-AR. Furthermore, several hydrophobic interactions are found among the whole ligand with each of the three subtypes suggesting no functional group to be subtype specific.

The binding of WAY-100635 to all three α_1-ARs is explained by the hydrogen bond that is formed with the protonated nitrogen and the large number of hydrophobic interactions with the ligand. The limited potency of the compound is due to the lack of hydrogen bonds. Limited selectivity is because each α_1-AR forms a common H-bond and a number of hydrophobic interactions with the whole ligand and, therefore, no functional groups are specifically involved in binding to a specific α_1-AR.
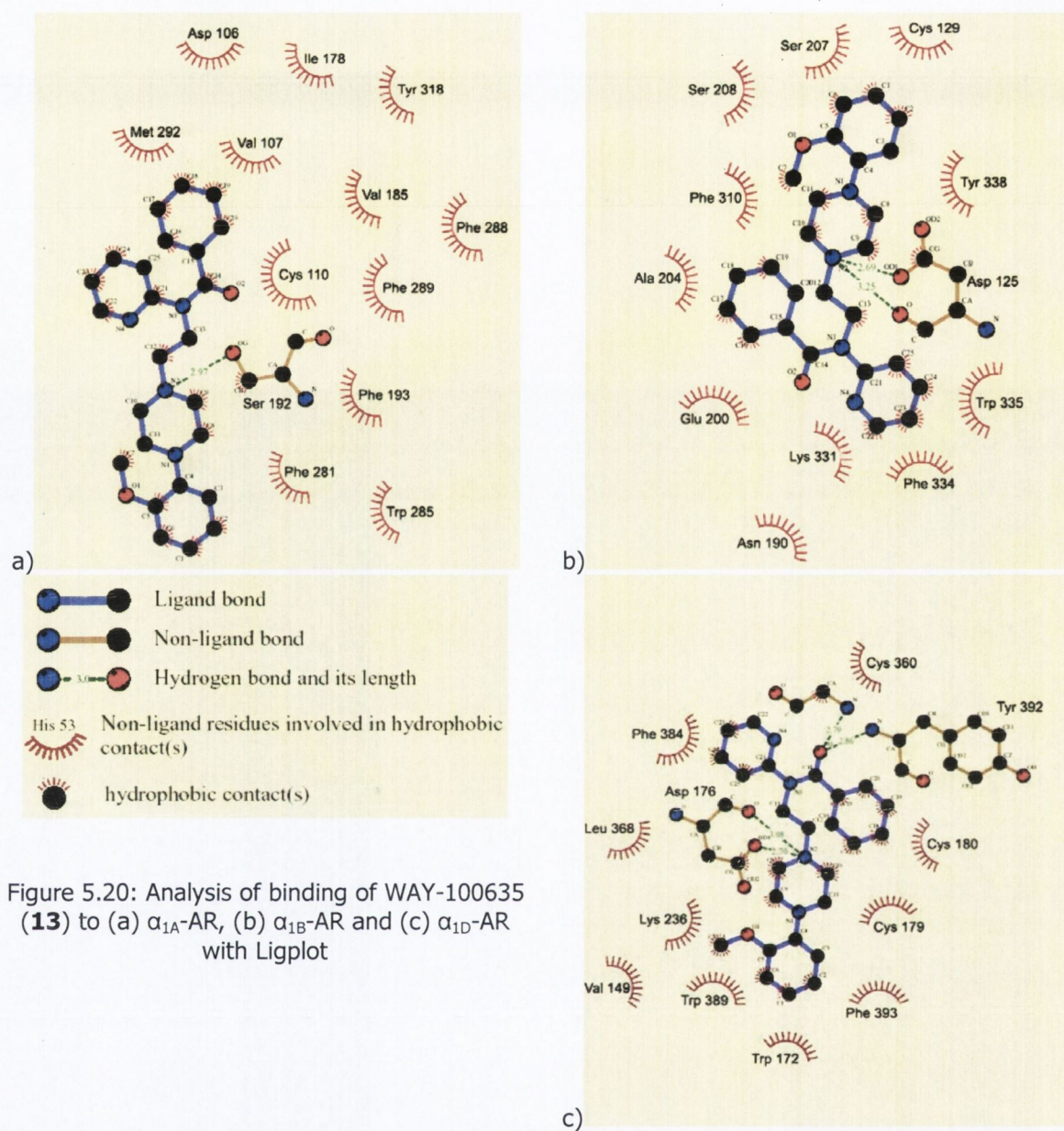
Figure 5.20: Analysis of binding of WAY-100635 (**13**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

5.6.7. Binding of RS-100,975

RS-100,975 (**14**) is a compound with a strong binding affinity to the α_{1A}-AR (K_i =1.0 nM), a medium binding affinity for the α_{1B}-AR (K_i =79 nM) and a weak binding affinity for the α_{1D}-AR (K_i =100 nM). The compound is selective for α_{1A}-AR as it binds 79-100 fold stonger compared to the other two α_1-AR subtypes. The ranking of the experimental binding affinities are correctly predicted by all three scoring functions (Table 5.29).

Table 5.29: Ranks (experimental and computated) and scores (computed) for the binding affinity of RS-100,975 (**14**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	1	-15.13	-6.80	-144.58	1	1	1
α_{1B}-AR	2	-12.09	-5.80	-124.70	2	2	2
α_{1D}-AR	3	-10.76	-5.08	-113.12	3	3	3

Interaction of RS-100,975 (**14**) with the three α_1-AR subtypes (Fig. 5.21) is observed via H-bonding. The protonated nitrogen in the piperazine group interacts only with the α_{1A}-AR via residue Asp[106]. The α_{1B}-AR and α_{1D}-AR show no interaction via this nitrogen but through one of the nitrogens in the pyrimidine ring with the α_{1B}-AR via Asp[125] and α_{1D}-AR via Phe[367].

There is a common hydrophobic interaction via the residue Ser[188] (α_{1A}-AR), Ser[207] (α_{1B}-AR) and Ser[258] (α_{1D}-AR) which are aligned residues. There is no interaction with the trifluoromethyl group. The α_{1B}-AR has a limited number of hydrophobic interactions. However, there is still an interaction with Asp[125] which is also described to interact with other ligands.

The α_{1D}-AR shows a limited number of residues that have hydrophobic interactions with RS-100,975 (**14**), but, especially interaction with Cys[180] is not observed. Also the formation of a H-bond with Asp[176], as shown for most ligands, is not observed. This suggests that the binding of RS-100,975 (**14**) to the α_{1D}-AR needs further optimization.
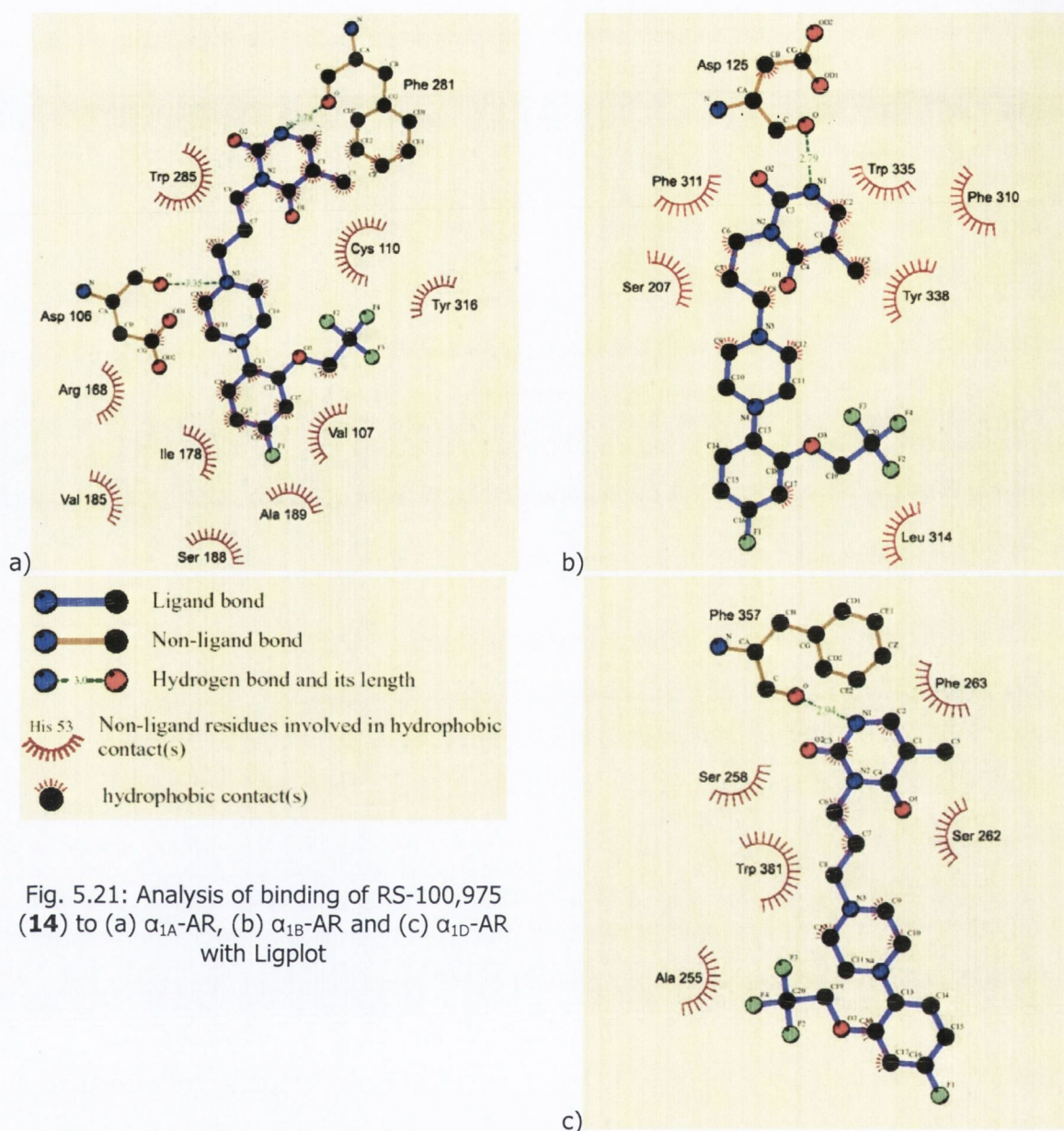
Fig. 5.21: Analysis of binding of RS-100,975 (**14**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

5.6.8. Binding of Uropidil

Uropidil (**15**) is a compound with a weak binding affinity for the α_{1A}-AR (K_i =288 nM), α_{1B}-AR (K_i =1320 nM) and the α_{1D}-AR (K_i =1660 nM). Despite the general poor affinites of this compound, it has been described as the first α_{1A}-AR selective antagonist. The correlations between experimental binding affinity ranks and the ranks by the scoring functions is limited (Table 5.30), although, London dG and Affinity dG predict the best affinity for the α_{1A}-AR correctly.

Table 5.30: Ranks (experimental and computated) and scores (computed) for the binding affinity of uropidil (**15**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	1	-16.56	-8.25	-134.31	1	1	2
α_{1B}-AR	2	-11.89	-5.49	-123.70	3	3	3
α_{1D}-AR	3	-15.08	-7.70	-138.43	2	2	1

H-bonding was observed in the α_{1A}-AR via residues Asp[106], Arg[166] and Ser[188], is not observed in the α_{1B}-AR, but are formed in the α_{1B}-AR via Asp[176], Thr[181] and Trp[361] (Fig 5.22). H-bonding with the protonated nitrogen is only observed in the α_{1A}-AR via Asp[106]. The oxygens attached to the pyrimidine group form H-bonds bonds with the α_{1A}-AR (Arg[166]) and α_{1D}-AR (Trp[361]).

Both the α_{1A}-AR and α_{1D}-AR show interaction with large parts of the ligand which can explain the binding of uropidil (**15**). More H-bonds are observed in the α_{1A}-AR than in the two other α_1-AR subtypes which can explain the selectivity of uropidil (**15**).

Although uropidil (**15**) binds weakly to the α_{1B}-AR a number of interactions would be expected as the compound is not completely inactive. It can be assumed that given the number of residues that show interaction, residues that interact with this ligand have been identified, but binding is not necessarily fully optimized as H-bonding is expected.

a)

b)

Legend:

- Ligand bond
- Non-ligand bond
- 3.0 — Hydrogen bond and its length
- His 53 — Non-ligand residues involved in hydrophobic contact(s)
- hydrophobic contact(s)
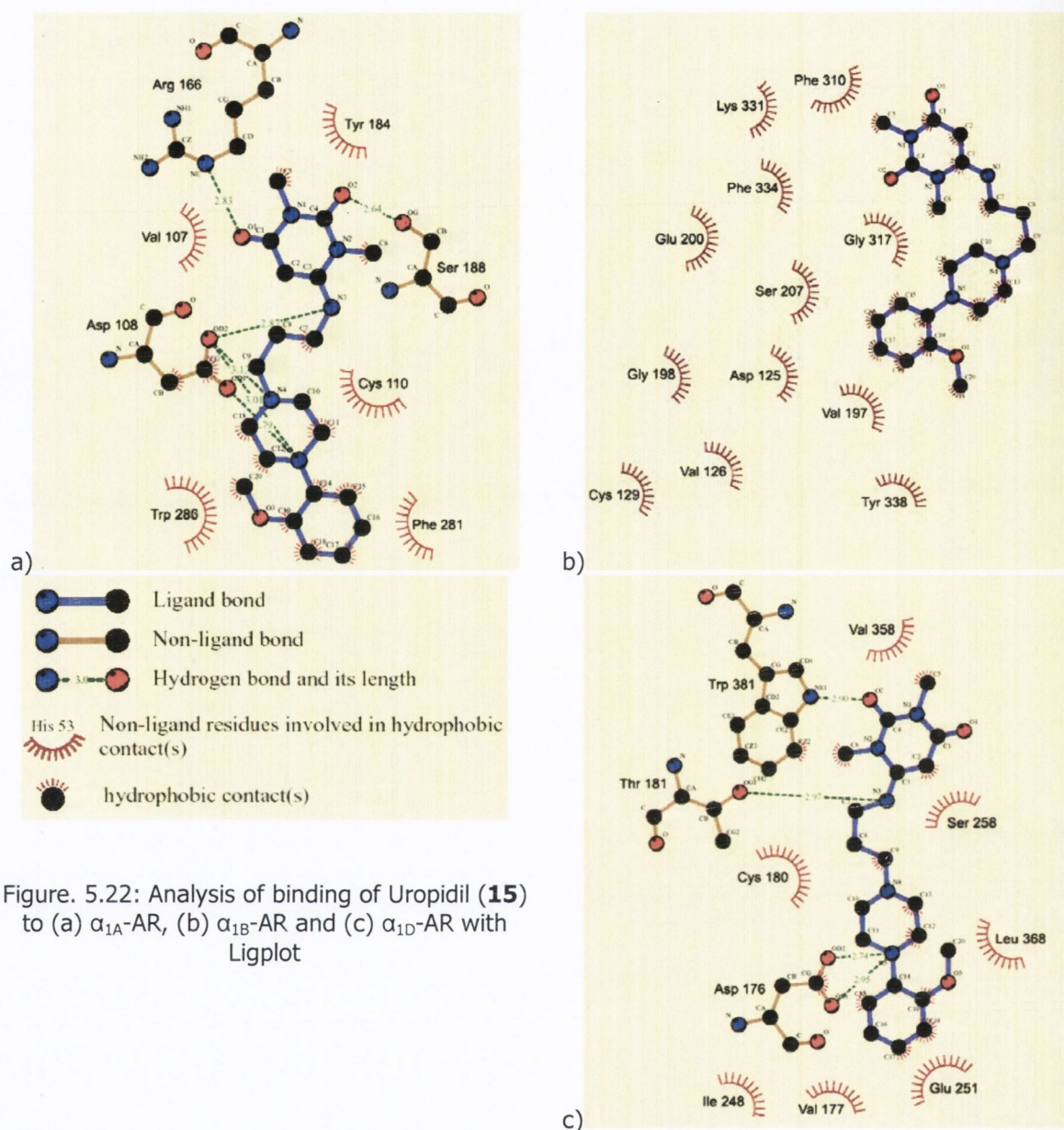
Figure. 5.22: Analysis of binding of Uropidil (**15**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

c)

The ligand-protein conformations obtained after 5ns of MD simulation allow the identification of the ligand-protein interactions. This enables the identification of the interactions that are responsible for a ligand binding to a receptor. In addition, when a limited number of interactions are observed, although the ligand is known to bind strongly, then an incorrect ligand-protein complex can be identified. When trying to predict the binding affinity with scoring functions, in general, poor results are obtained. Because scoring functions are commonly used to assess ligand-protein complexes, it should be suggested that scoring functions are better in the classification of binders and non-binders. Therefore, using these protein ligand interaction plots provides more insight into binding than scores.

5.6.9. Binding of Discretamine

Discretamine (**18**) is a compound with a poor binding affinity to the α_{1A}-AR (K_i =616 nM) and α_{1B}-AR (K_i =360 nM) and a medium binding affinity for the α_{1D}-AR (K_i =25 nM). This compound is selective for the α_{1D}-AR. None of the scoring functions rank the binding affinities correctly (Table 5.31).

Table 5.31: Ranks (experimental and computated) and scores (computed) for the binding affinity of discretamine (**18**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	3	-12.67	-8.78	-128.86	2	1	1
α_{1B}-AR	2	-11.67	-7.95	-112.08	3	2	3
α_{1D}-AR	1	-12.69	-5.08	-113.12	1	3	2

The protonated nitrogen shows the formation of hydrogen bonds (Fig. 5.23) with the α_{1A}-AR via residue Cys^{110} and the α_{1D}-AR via residue Ser^{258}. The formation of a hydrogen bond with α_{1B}-AR is not observed. There are common interactions via three aligned residues in the α_{1A}-AR (Asp^{106}, Cys^{110} and Tyr^{316}) in the α_{1B}-AR (Asp^{125}, Cys^{129} and Tyr^{338}) and in the α_{1D}-AR (Asp^{176}, Cys^{180}, Tyr^{392}).

The limited number of H-bonds and the similarity in the residues that are involved in binding of discretamine (**18**) in each of the three adrenoceptors can not explain the difference in binding affinities. Therefore, an optimal conformation for ligand-protein interaction is likely not obtained, which could limit the use of these structures in future studies.
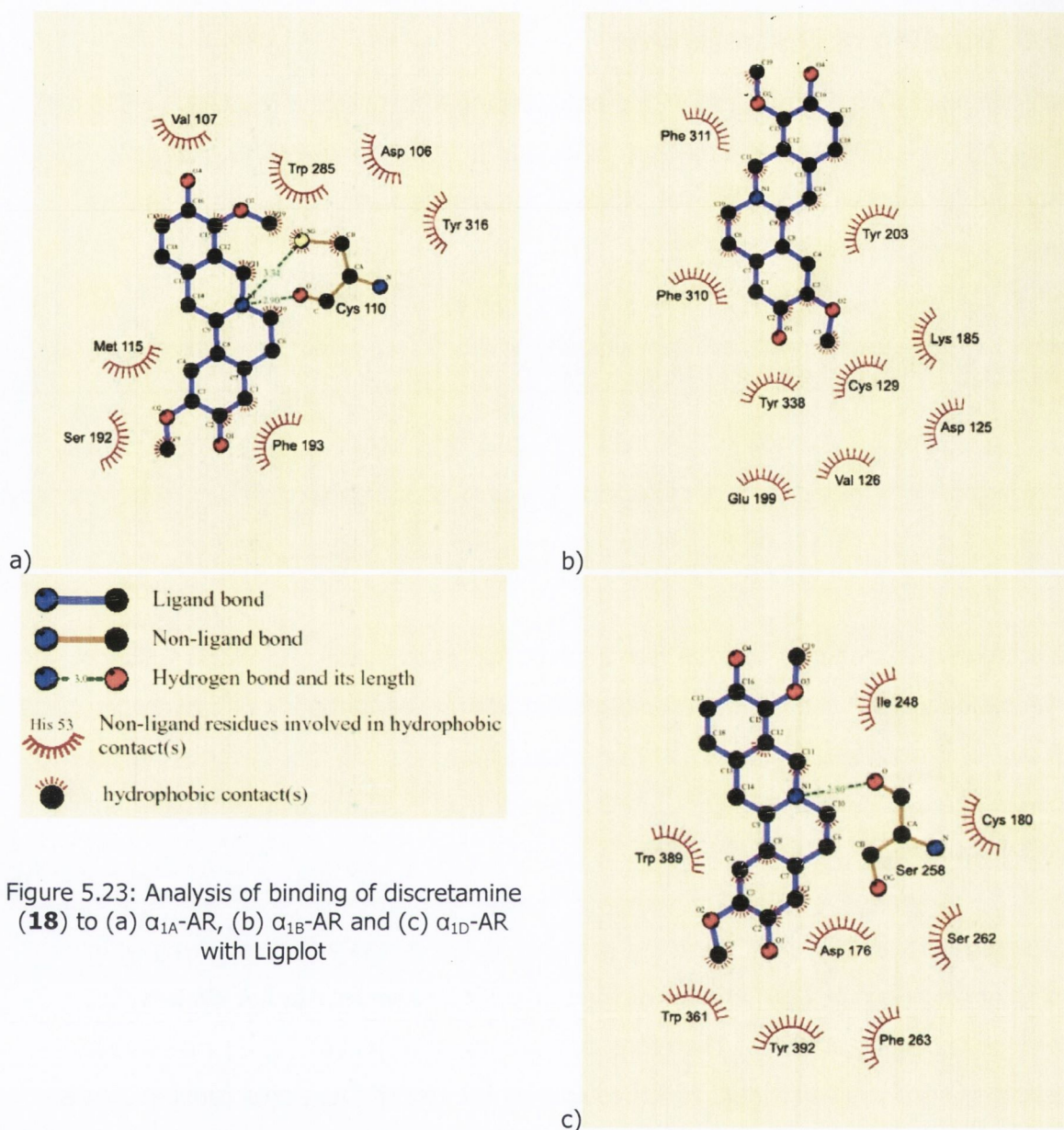
Figure 5.23: Analysis of binding of discretamine (**18**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

5.6.10. Binding of Corynanthine

Corynanthine (**19**) is a compound with a weak binding affinity for the α_{1A}-AR (K_i =142 nM), α_{1B}-AR (K_i =517 nM) and α_{1D}-AR (K_i =253 nM). This compound is not selective to any of the three α_1-ARs. There is no correlation between the ranking of the scoring functions and the ranking of the experimental binding affinities (Table 5.32).

Table 5.32: Ranks (experimental and computated) and scores (computed) for the binding affinity of corynanthine (**19**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	1	-9.86	-6.24	-95.80	3	3	3
α_{1B}-AR	3	-13.38	-7.00	-140.96	1	2	2
α_{1D}-AR	2	-13.25	-8.38	-144.20	2	1	1

Two nitrogens form a hydrogen bond with the adrenoceptors (Fig. 5.24). The protonated nitrogen shows the formation of a H-bond with α_{1A}-AR via the residue Asp^{110} and with the α_{1D}-AR via the residue Asp^{176}. The α_{1B}-AR shows interaction via the other nitrogen in corynantine (**19**) and forms a H-bond with Glu^{200}.

None of the oxygens show interactions with any residues of the adrenoceptors either via a hydrogen bond or via hydrophobic interactions. The benzene ring of the indole system shows hydrophic interaction with all three α_1-AR subtypes.

The binding profiles to the three α_1-AR subtypes suggest that interaction via the protonated nitrogen is important for binding to the α_{1A}-AR and α_{1D}-AR. Furthermore, there is a number of other residues showing hydrophobic interactions with the same functional group of the ligand. The weak binding of corynanthine (**19**) to each α_1-AR can be explained by only one H-bond being formed in each receptor and the limited number of hydrophobic interactions.

a)

b)

c)

Legend:

- ●━━● Ligand bond
- ●━━● Non-ligand bond
- ●╍╍● Hydrogen bond and its length
- His 53 / Non-ligand residues involved in hydrophobic contact(s)
- hydrophobic contact(s)
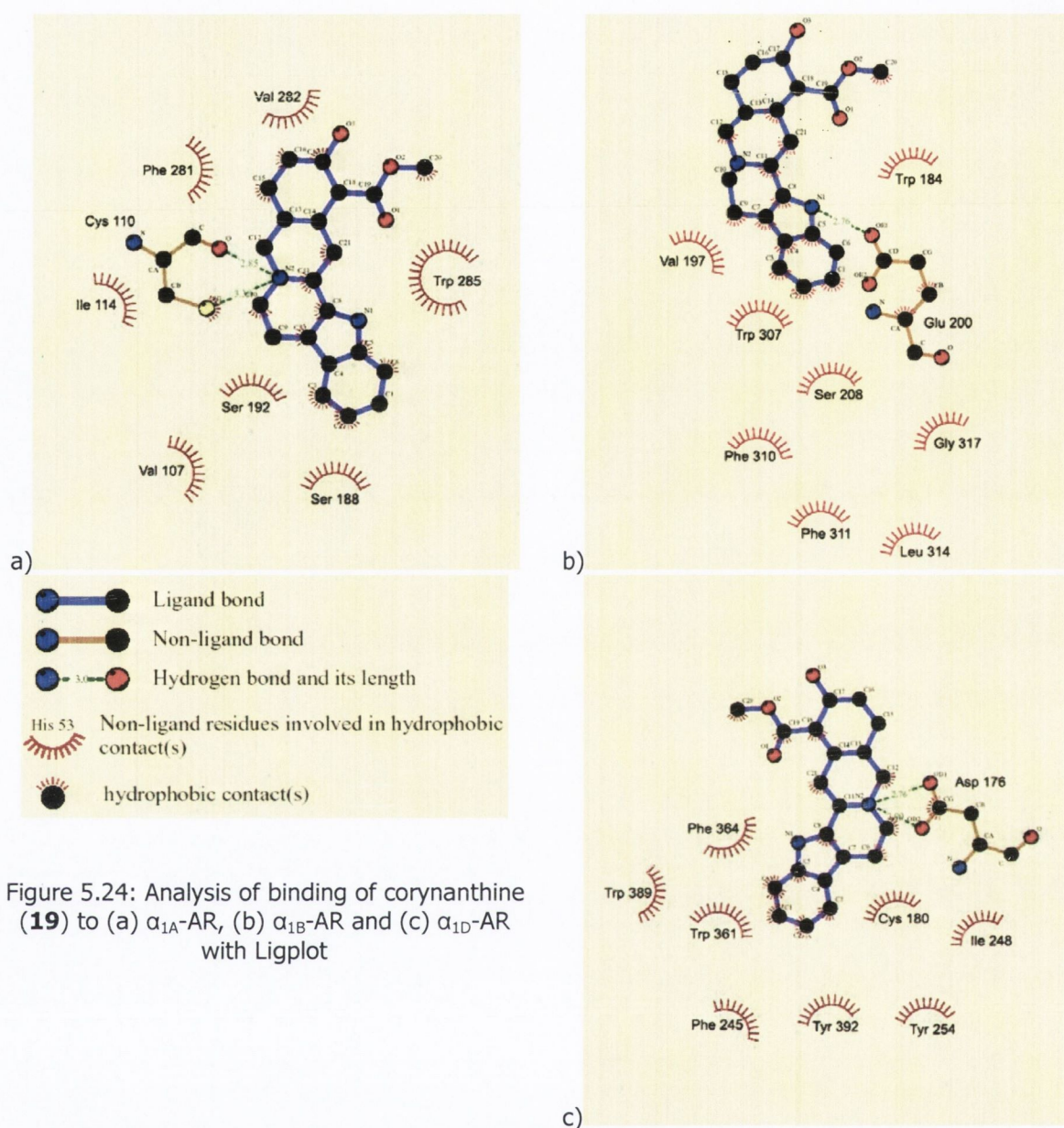
Figure 5.24: Analysis of binding of corynanthine (**19**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

5.6.11. Binding of KMD-3213

KMD-3213 (**21**) is a compound with a very strong binding affinity to the α_{1A}-AR (K_i =0.04 nM), a medium binding affinity for the α_{1B}-AR (K_i =20 nM) and a strong binding affinity for the α_{1D}-AR (K_i =2.0 nM). This compound can be regarded as α_{1A}-AR selective due to its 50-fold stronger binding affinity over the other receptors and a relative weak binding affinity to the α_{1B}-AR and α_{1D}-AR, and in fact, is the antagonist with the best K_i towards any α_1-AR subtype. There is no correlation between the ranking of the scoring functions and the ranking of the experimental binding affinities (Table 5.33).

Table 5.33: Ranks (experimental and computated) and scores (computed) for the binding affinity of Cyclazosin (**4**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	1	-15.47	-10.24	-160.89	3	2	2
α_{1B}-AR	3	-20.41	-12.70	-188.79	1	1	1
α_{1D}-AR	2	-16.27	-8.37	-127.76	2	3	3

All of the three subtypes interact via the protonated nitrogen (Fig. 5.25): α_{1A}-AR via Asp[106], α_{1B}-AR via Thr[130] and Cys[129] and α_{1D}-AR via Asp[176] and Tyr[392]. Futhermore, H-bonding is observed via the $CONH_2$ group for α_{1A}-AR and α_{1B}-AR. In the α_{1A}-AR complex, the amide group only acts as a hydrogen bond donor, but in the α_{1B}-AR complex it acts as a hydrogen bond donor and acceptor to three residues. Furthermore, H-bonding with the alcohol group is observed for the α_{1B}-AR, but not for the other two α_1-AR subtypes. This could inditate why the binding affinity is better for the α_{1A}-AR and α_{1D}-AR

The CF_3 group shows interactions with the α_{1B}-AR, but not with the α_{1A}-AR and α_{1D}-AR which could not explain a higher binding affinity to the α_{1A}-AR and α_{1D}-AR. This model identifies interactions that make the ligand bind to each of the α_1-AR subtypes.
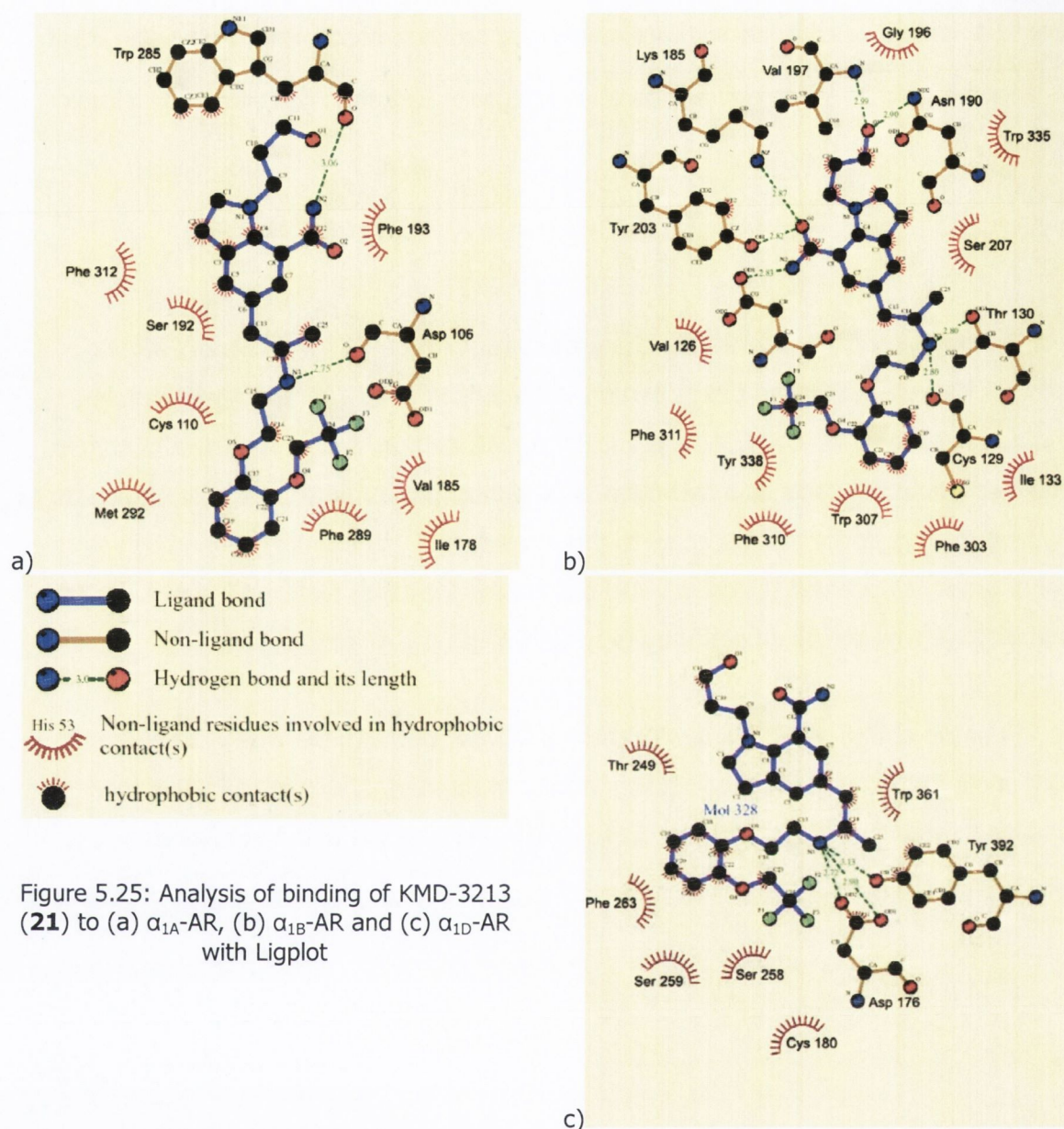
Figure 5.25: Analysis of binding of KMD-3213 (**21**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

5.6.12. Binding of SNAP-1069

SNAP-1069 (**27**) is a compound with a medium binding affinity to the α_{1A}-AR (K_i =16 nM), and a weak binding affinity for the α_{1B}-AR (K_i =200 nM) and α_{1D}-AR (K_i =790 nM). This compound is selective for the α_{1A}-AR. The three scoring functions provide the same ranking, but this ranking does not match the ranking of the experimental binding affinities (Table 5.34).

Table 5.34: Ranks (experimental and computated) and scores (computed) for the binding affinity of SNAP-1069 (**27**) to the three α_1-AR subtypes.

	Rank (exp)	Score London dG	Score Affinity dG	Score Alpha HB	Rank London dG	Rank Affinity dG	Rank Alpha HB
α_{1A}-AR	1	-14.35	-8.71	-120.79	2	2	2
α_{1B}-AR	2	-20.41	-12.70	-188.79	1	1	1
α_{1D}-AR	3	-8.91	-4.18	-93.98	3	3	3

The formation of a H-bond between the ligand and the adrenoceptors (Fig. 5.26) is observed in the complexes with α_{1A}-AR and α_{1B}-AR. In the α_{1A}-AR two H-bonds are observed where the Asp[106] interacts with the protonated nitrogen and Ser[188] interacts with the other nitrogen in SNAP-1069 (**27**). The α_{1B}-AR does not interact with the protonated nitrogen, but Asp[125] interacts with the other nitrogen in SNAP-1069 (**27**). For the α_{1D}-AR no H-bonds are observed.

In the complex with α_{1A}-AR two H-bonds are formed with Ser[188] and Asp[106] and many hydrophobic interactions are established. In the complex with α_{1B}-AR only one H-bond is observed with Asp[125] ond other residues establish hydrophobic interaction with other parts of the molecule. No H-bonds are observed in the complex with α_{1D}-AR. Further optimization of the α_{1D}-AR complex is needed.

The strong binding for α_{1A}-AR can be explained by the number H-bonds and hydrophobic interactions, which, compared to the α_{1B}-AR and α_{1D}-AR are larger in number.

a)

b)

c)
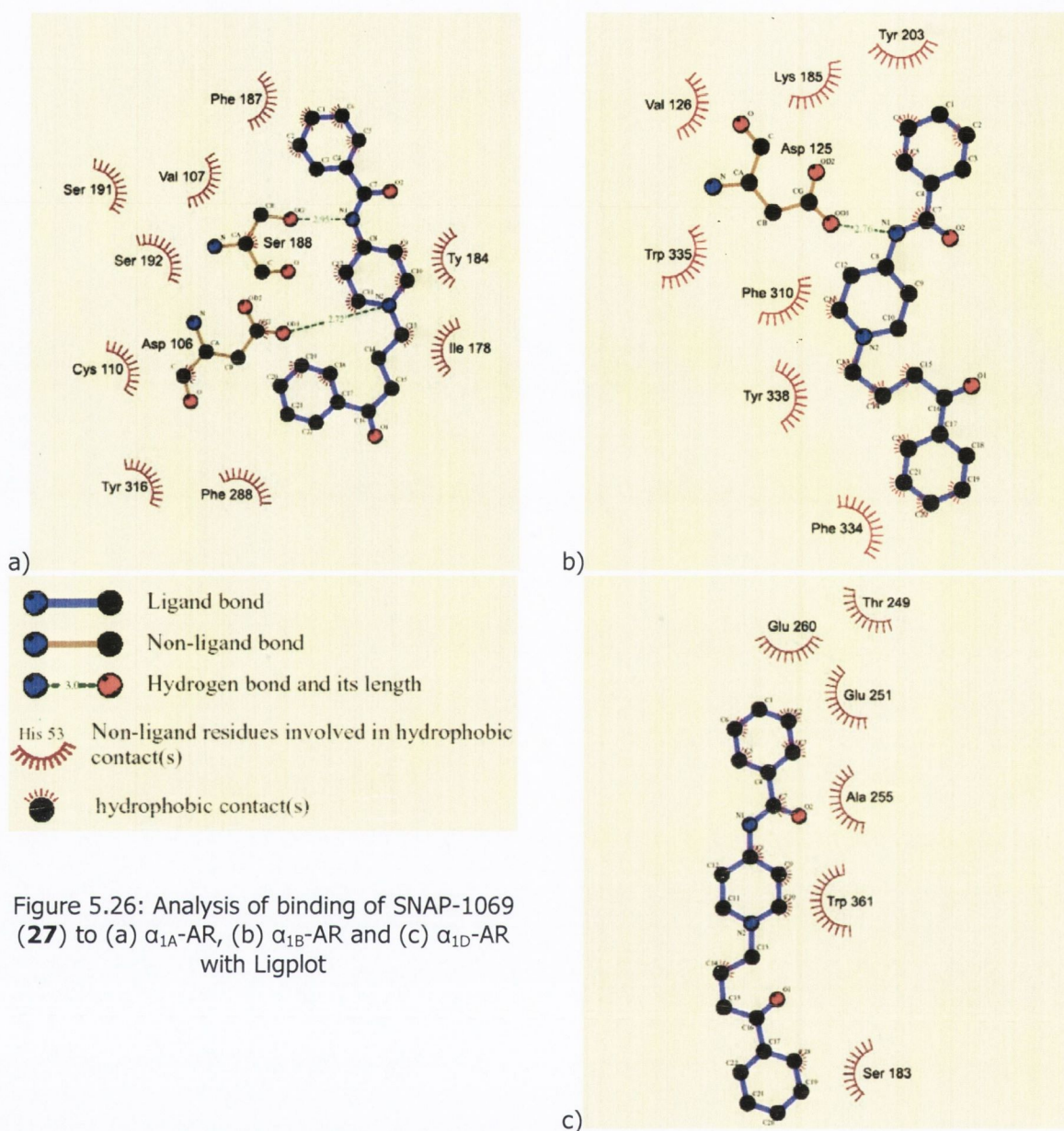
Figure 5.26: Analysis of binding of SNAP-1069 (**27**) to (a) α_{1A}-AR, (b) α_{1B}-AR and (c) α_{1D}-AR with Ligplot

5.7. Selection of models for virtual screening

Based on the binding modes of the ligands as described in the previous section, ligand-protein complexes can be selected when an correlation between the interactions and the binding affinites has been obtained, for use in further studies. As it has become evident, not all models have converged to a state that represents an optimal binding of the antagonist-AR complex. However using the following criteria several models show an interaction which can be described as an optimized antagonist-adrenoceptor complex The criteria are: (1) the presence of hydrogen bonds, especially with the protonated nitrogen, (2) the hydrophobic interaction with a number of residues and (3), the number of observed interactions need to correspond to the binding affinity. Four different structures

of each subtype are proposed which are most likely to be succesfull in further research. The four models that represent optimal binding in all three α_1-AR sybtypes are the receptors complexed with cyclozosin (**4**), alfuzosin (**7**), BMY-7378 (**10**) and WAY-100635 (**13**).

For the α_{1A}-AR a common interaction within all four antagonists (Fig 5.27) is observed with Asp[106]. In three cases this is a hydrogen bond and in the other case (WAY-100635) hydrophobic interaction occurs. Furthermore, strong interactions with the residues Val[107], Cys[110], Tyr[285] and Phe[288] are observed. It can also be noticed that there are a number of residues which interact with only one or two antagonists, determining selectivity. Thus, H-bond interactions with both Asp[106] and Trp[285] together, could determine the α_{1A}-AR selectivity, since, those are present at the same time and only in the α_{1A}-AR complexes.

a)   b)

c)   d)

Figure 5.27: Binding of (a) cyclazosin (**4**) , (b) alfuzosin (**7**), (c) BMY-7378 (**10**) and (d) WAY-100635 (**13**) to the α_{1A}-AR.

For the α_{1B}-AR (Fig. 5.28) common interactions for binding of four antagonists are observed via Cys[129], Ser[207] and Phe[310]. Residues that show interaction with three out of four antagonists are Ala[204], Phe[311] and Tyr[338], suggesting they play an important role in binding. Hydrogen bonds are formed with residue Ser[207] in three cases. In contrast to the α_{1A}-AR where the protonated nitrogen interacts mainly with Asp[106], the α_{1B}-AR forms this interaction mainly with Ser[207]. This suggests a different binding mode in which the interaction with a Ser (α_{1B}-AR), instead of an Asp (α_{1A}-AR and α_{1D}-AR), in which the Ser is not aligned to the Asp residue, allowing the rest of the ligand to interact with different residues, therefore, obtaining receptor selectivity. It can be suggested that targeting Ala[204] and Tyr[338] simultaneously can also introduce α_{1B}-AR selectivity.
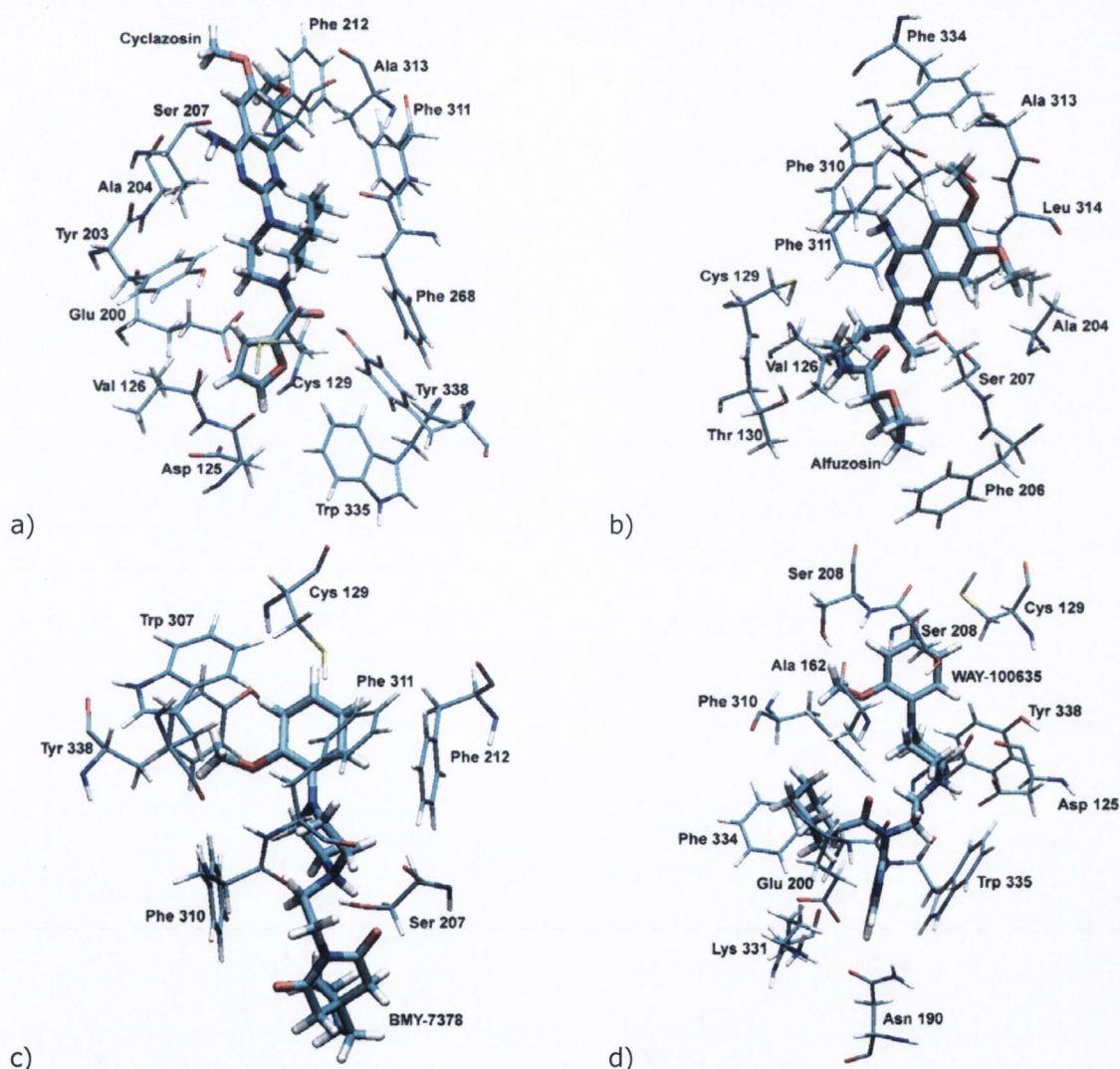
Figure 5.28: Binding of (a) cyclazosin (**4**) , (b) alfuzosin (**7**), (c) BMY-7378 (**10**) and (d) WAY-100635 (**13**) to the α_{1B}-AR.

For the α_{1D}-AR (Fig. 5.29) common interactions that are involved are Asp[176], Ala[255], Phe[364] and Leu[368]. Hydrogen bonds are formed commonly via residue Asp[176] and are also observed via residue Thr[181], Glu[251], Ser[262], Gly[391] and Tyr[392], suggesting that residue Asp[176] is common in forming a hydrogen bond with an antagonist, and several other hydrogen bonds can be formed which, in addition to non-bonded interactions, contribute to the binding affinity. The binding mode for the α_{1D}-AR is similar to the α_{1A}-AR. It can be suggested that targeting Ala[176] and Tyr[392] simultaneously can introduce α_{1D}-AR selectivity.
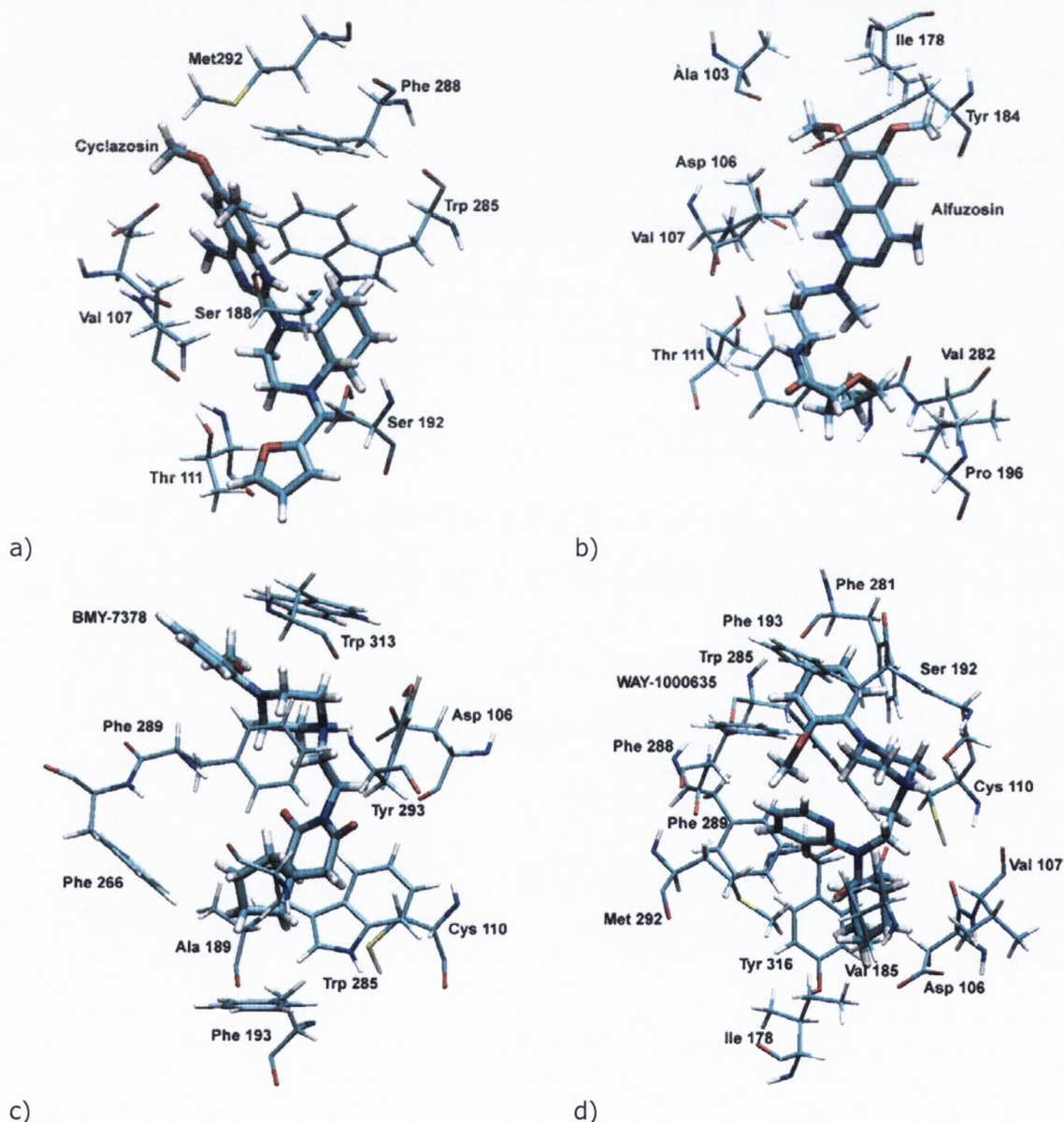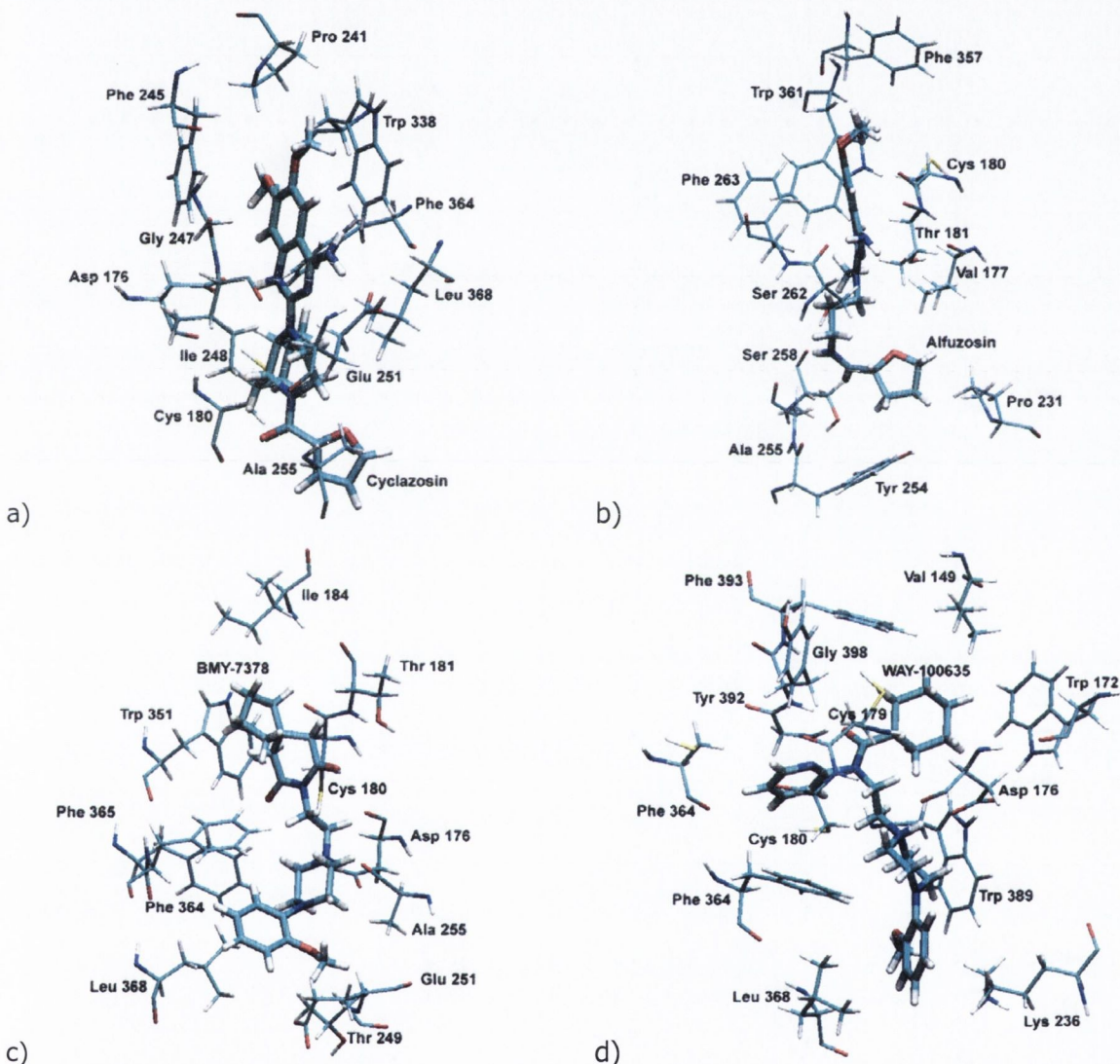

Figure 5.29: Binding of (a) cyclazosin (**4**) , (b) alfuzosin (**7**), (c) BMY-7378 (**10**) and (d) WAY-100635 (**13**) to the α_{1D}-AR.

5.8. Conclusions

In earlier times, the interaction between a ligand and the target receptor has been described as a lock and key mechanism. However, it soon became clear that the ligand-protein interaction is a dynamic process. This is also observed in our different interaction models. There are several residues which appear to be key to interaction with these residues and can almost be described as a necessity for binding of each ligand to the receptor to occur. Besides these, there appear to be many other residues involved in the binding of a ligand which are ligand-specific, but not α_1-AR subtype specific. For each ligand in our dataset a number of residues were identified that can interact with this ligand. Because of a small set of compounds that were selected, this proposition can not be fully established.

During the MD simulation a conformational change of the adrenoceptor occurs which can be attributed to two factors. First, because our model is based on bovine rhodopsin, an optimization will occur to that of the adrenoceptor state. Second, the ligand has an effect on the conformation and each final structure can be considered as an antagonist-induced conformation. During the MD simulation, the binding pocket opens up with the top part of the helices TM-IV and TM-V moving away form the centre. This allows a better fit of the ligand in the binding pocket. Another conformational change can be observed for TM-I, mainly in the α_{1A}-AR and α_{1D}-AR, where this helix moves away from the centre of the protein. A similar conformation is observed in the recently obtained crystal structure of the β_2-AR. Alignment of the ligand-induced conformations of the α_1-AR with that of the crystal structure of the β_2-AR show similarity in the positioning of the helices. This suggests that the β_2-AR would make a better template for the design of homology models, but it also means that homology models based on bovine rhodopsin are capable of converging to the native state of the receptor.

We have identified a binding pocket with the residues which are important for interaction for each of the three α_1-AR subtypes. These consist of a hydrogen bond with one residue and a hydrophobic interaction with several residues. When these three binding pockets are compared (Table 5.35) it can be observed that they are very similar with only a few residues being different in each binding pocket. These general binding pockets can be used to identify where a ligand will bind, and some residues could give some indication of ligand selectivity.

Table 5.35: Residues that form the general binding pockets of the three α_1-AR subtypes. The shaded residues shows the residues which are aligned in each of the three α_1-AR subtypes.

	α_{1A}-AR	α_{1B}-AR	α_{1D}-AR
Hydrogen bond	Asp^{106}	Asp^{125}	Asp^{176}
	Cys^{110}	Thr^{130}	Cys^{180}
	Ile^{178}		Ile^{248}
	Ser^{188}		
		Ala^{204}	
	Ser^{192}	Ser^{207}	
	Tyr^{285}	Tyr^{338}	Tyr^{392}
Hydrophobic interaction	Asp^{106}	Asp^{125}	Asp^{176}
	Val^{107}	Val^{126}	
	Cys^{110}	Cys^{129}	Cys^{180}
	Ile^{178}	Tyr^{203}	Ala^{255}
	Ser^{188}	Ser^{207}	Ser^{258}
	Ser^{192}		
	Tyr^{285}		
	Phe^{288}	Phe^{310}	
		Phe^{311}	
		Trp^{335}	Trp^{361}
			Leu^{368}
	Tyr^{316}	Tyr^{338}	Tyr^{392}

Although, most of the residues that form the common binding pocket are similar, a different binding mode was observed for the α_{1B}-AR compared to the α_{1A}-AR and α_{1D}-AR. The protonoted nitrogen usually interacts with an Asp located in TM-III, in the α_{1A}-AR and α_{1D}-AR, whereas, in the α_{1D}-AR it interacts with a Ser which is located in TM-V.

The MD simulations have been shown to optimize the binding between the antagonist and the receptor to an antagonist induced conformation. However, full optimization is not always obtained when performing a 5 ns simulation. Based on the interaction via hydrogen bonds and the number of residues which show hydrophobic interactions, four AR-complexed antagonists cyclazosin (**4**), alfuzosin (**7**), BMY-7378 (**10**) and WAY-100635 (**13**), have reached a close to optimal binding for each subtype. Based on these binding profiles four different binding pockets are created which can be used in further studies.

Chapter VI

Summary and further studies

I do not fear computers, I fear the lack of them.

Isaac Asimov (1920-1992)

6.1. Summary

A detailed structure based drug design study has been performed on the three α_1-AR subtypes. The aim of this work was to obtain models of the three AR subtypes α_{1A}-AR, α_{1B}-AR and α_{1D}-AR and identify binding interactions that can be used in the discovery of antagonists that are selective for α_{1A}-AR and α_{1D}-AR, with application in the treatment of BPH. A computational protocol was developed which integrated the prediction of the structure of each of the three α_1-AR subtypes, optimization of ligands that are known to bind to these receptors and the performance of MD simulations to optimize the interaction of antagonist- α_1-AR complexes.

Homology modelling was used to develop 3D models of the three subtypes of α_1-AR. These models are based on the alignment of the aminoacid sequences of bovine rhodopsin and those of the three adrenoceptor subtypes. Several different alignment approaches within the homology modelling framework were investigated to explore the possibilities of increasing the quality of the resulting homology models. Initially, models were created using the 'standard' method of aligning the amino acid sequence of the adrenoceptor and that of rhodopsin. The second approach involved an intermediate step implementing a new amino acid sequence. This new sequence is essentially that of the adrenoceptor sequence with several residues mutated to the corresponding ones in the rhodopsin sequence. This sequence is then used to generate an intermediate model by aligning it to the rhodopsin sequence. This intermediate model is subsequently used for alignment with the adrenoceptor sequence to generate the final α_1-adrenoceptor models. Another approach was taken by changing the alignment between rhodopsin and the adrenoceptor sequence. As a result, 'gaps' of non-aligned residues occur in the alignment. This increases the similarity between the sequences of rhodopsin and the adrenoceptors when they are aligned. Following the four approaches, models were generated for all of the α_1-AR subtypes (α_{1A}-AR, α_{1B}-AR and α_{1D}-AR). Each model was subjected to a molecular dynamics (MD) simulations with a production run of 4 ns. Models were extracted at 2 and 4 ns and analyzed using several protein assessment tools. Our research shows that the standard method leads to very acceptable results and there is always a model generated with limited errors. In addition, the standard method always provides stable structures during MD simulations. Models that make use of an intermediate mutated amino acid sequence were shown to improve or decrease, depending on the case, the quality of a homology model and, therefore, the standard approach is the preferred method.

The stability of the homology models also validates the used of the membrane mimic consisting of water/chloroform/water layers to mimic the normal environment of membrane bound receptors. However, it should be noted that we have performed relatively short MD simulations with a production run of 4 ns. Longer MD simulations for other proteins have been published (100 ns) but performing these with the solvent mimic has not been validated. Therefore, if these longer simulations would be used a large number of artefacts could be introduced resulting in an incorrect interpretation of the results.

As part of modelling, an extensive ligand optimization protocol and the study of the interactions between the antagonists and the α_{1A}-, α_{1B}- and α_{1D}-adrenoceptors an extensive ligand optimization were performed. We have further optimized the minimum energy conformation of each ligand, determined their proton affinity and were able to derive accurate charges. As part of this process we determined the acid dissociation constants (pK_a) for each antagonist. Firstly, a conformational search was performed followed by DFT calculations using B3LYP/6-31G* level of theory. For each group of ligands there was a preferred nitrogen to be protonated. Based on this protonation state, we calculated the pK_a for a set of 29 ligands. Of these ligands, six have a known pK_a and comparison between the experimental and computational values showed good correlation. This leads us to believe that our methods are accurate for determining the protonation state and pK_a. In addition, we tried an alternative method using explicit water molecules with the polarizable continuum model (PCM) solvent model. This resulted in different energies compared to the calculations without the explicit water molecules. The obtained pK_a values did not correlate as well with the experimental pK_a values as the PCM calculations alone.

For examining the interaction between antagonists and adrenoceptor subtypes we have selected twelve antagonists which cover a range of binding affinities for each of the three α_1-ARs and are structurally different. Each ligand was docked individually into the three AR-subtypes giving rise to 36 complexes. All of these complexes, embedded in a membrane mimic model consisting of water/chloroform/water were subjected to a MD simulation of 5 ns. This membrane model was shown to provide stable homology models in MD simulations and, therefore, can also be used in ligand-protein complex simulations.

The docking of the ligands into the adrenoceptors is not a straightforward task as different modes of binding can be considered. Therefore, the docking of multiple conformations of each ligand was performed producing a number of ligand-protein complexes. From these ligand-protein complexes one was selected and was subjected to a MD simulation.

The docking was performed using different conformations of the ligand, which were taken from the conformational search in the ligand optimization step. This produced a number of conformations that needed to be assessed for binding. Three parameters were used for this analysis: visual inspection, Autodock score and an interaction analysis using LigPlot. The visual inspection allowed the elimination of complexes that did not bind properly, e.g. a ligand that sticks out of the receptor. This is not necessarily shown in the binding score by Autodock, and interaction with a number of other residues can still be observed. Because we used a static protein in the docking process, a fully optimized binding of the antagonist to the adrenoceptor will not be obtained. Therefore, a large difference in binding probably indicates a different binding mode with a theoretical preference for the lowest energy complex. A small difference in energy indicates a different fit of the ligand in the receptor with no preference for either of these two binding possibilities, and it could converge to the same ligand-receptor complex when refined further with MD simulations. The LigPlot analysis allows the identification of existing interactions between the antagonists and the adrenoceptor subtypes. The starting structure is important as it provides a basis for interaction in the MD simulation. Therefore, docking of each ligand into each α_1-AR subtype needs to be carefully analyzed and a consensus based on the three observations needs to be derived to select a ligand-receptor complex.

When performing the analysis of the MD simulation several structural features were monitored such as the positioning of the helices, the behaviour of an ionic lock and the behaviour of a salt bridge. The ionic lock plays a role in activation/deactivation of the receptor and the salt bridge a role in the stability of the receptor.

RMSD analysis of the helices has shown conformational changes in the binding pocket with the helices TM-IV and TM-V influencing the shape of the binding pocket the most. Another change that was observed was the position of TM-I which move outward of the binding pocket. These conformational changes corresponded to a structure that shows more similarity to the β_2-AR than bovine rhodopsin on which the homology models are built. This conformational change is likely due to two causes. The first is that our

adrenoceptors are based on the crystal structure of bovine rhodopsin which has a lower sequence similarity than the β_2-AR. This means that the structure of the β_2-AR is more similar to that of the α_1-AR and when MD simulations are performed, a structural convergence to the β_2-AR is expected. However, this effect is not fully observed for the uncomplexed receptors, suggesting that the binding of a ligand allows the receptor to evolve to a different conformation. Secondly, different conformational changes and conformational states were observed when structurally different ligands were docked into the same receptor. This suggests that there is a ligand-induced conformation and, therefore, one crystal structure can not take into account the binding of all its possible binders. However, the obtained structure can be used to screen similar ligands. In our study, we have optimized several ligand-protein complexes which allow the screening of a diverse dataset.

Observations for the ionic lock show a change in the distances between the three residues involved (α_{1A}-AR: Asp^{123}-Arg^{123}-Glu^{267}, α_{1B}-AR: Asp^{142}-Arg^{143}-Glu^{289} and α_{1D}-AR: Asp^{193}-Arg^{194}-Glu^{343}). This suggests that a conformational change occurs upon complexation and a movement to a different conformational state is observed. A change in the ionic lock is known to occur in the receptor activation process, but because we are modelling antagonists a conformational change to an active state should not occur, but a ligand-induced conformational change was expected. The same result can be noted for the salt bridge (α_{1A}-AR: Asp^{106}-Lys^{331}, α_{1B}-AR: Asp^{125}-Arg^{331} and α_{1D}-AR: Asp^{176}-Lys^{385}). Because of the limited timescale of the MD simulation (5 ns) a full convergence to a ligand induced conformation is not expected. This conformational change of the protein does not mean that on optimized binding between ligand and receptor has not occurred and, therefore, these models are suitable for screening virtual libraries.

More important is the interaction of the ligand with the receptor which can be studied using the program LigPlot which identifies residues that form hydrogen bonds with the ligand or establish hydrophobic interactions. When comparing the ligand-protein interactions in the initial docked homology models and after 5 ns of MD simulation a different interaction pattern is observed. Especially the interaction between the protonated nitrogen and the aspartic acid becomes evident in the structures obtained after the MD simulation and shows the importance of optimization. Analysis of all 12 complexes for the three α_1-AR-subtypes showed that most of the residues involved in binding are located in TM-III, TM-IV, TM-V, TM-VI and ECL-IV-V. This was observed for each subtype and a general binding pocket for the three subtypes was derived based on

the number of ligands showing interaction with a particular residue. This allowed the definition of a binding pocket for each subtype. These three binding pockets are remarkably similar with only 3 residues of difference between each binding pocket. It was observed that the different α_1-AR-subtypes interact differently with the protonated nitrogens. In the α_{1A}-AR and α_{1B}-AR this interaction is observed with an Asp, whereas, in the α_{1B}-AR this interaction is observed with a Ser. This suggests that a different binding mode exists for the α_{1B}-AR compared to the α_{1A}-AR and α_{1D}-AR which can be further exploited in structure based drug design and indicates that antagonists that target the α_{1A}-AR and α_{1D}-AR selectively is possible.

To select adrenoceptor models which can be used in receptor subtype-specific drug design, models need to have the ability to identify interactions with a range of ligands. Based on the formation of hydrogen bonds and other interactions with different residues and based on the similarity to other bound ligands, a selection of four antagonists was made which were likely to have a close to optimal interaction with each of the three subtypes α_{1A}-AR, α_{1B}-AR and α_{1D}-AR. Due to the structural difference of these four ligands, these models represent different binding modes which is confirmed by the different residues that are involved in each binding pocket. This allows screening of compounds which bind in a different binding mode and can initiate the discovery of a new class of $\alpha_{1A/D}$-AR selective antagonists.

6.2 Further studies

With the publication of the crystal structure of the β_2-AR the molecular modelling of the interactions between GPCRs and their ligands will change dramatically. Until now, most AR models were based on the bovine rhodopsin crystal structure and are, due to the higher amino acid sequence identity, likely to be replaced by models built on the β_2-AR. Although, similarity to the β_2-AR is higher, structural optimization is likely to be needed as related proteins are not copies of each other.

Another way of improving existing models is performing the MD simulations in a phospholipid bilayer which should be more accurate.[152] Several bilayer models are available and, therefore, selection of the optimal will be crucial. Consensus on which bilayer model would perform best has not been achieved and each model will have its advantages and drawbacks. The advantage of solvent-based models, as used in our study, is that the calculations are quick to perform compared to phospholipid bilayers. A comparison of the effect of the phospholipid and solvent models on the adrenoceptor structure could be made.

A feature which is largely unexplored is the formation of dimers. We know this exists for several GPCRs, but there is little understanding of how exactly it is manifested. The knowledge of which receptors dimerize in homo- or hetero-form and which residues are involved in connecting those receptors is unknown. To examine dimers, several configurations could be simulated, but a determination of which configuration is favourable is unlikely due to the lack of reference data. It can also be questioned if this approach would produce better results than the approach taken in this thesis and, therefore, other approaches may be considered first.

In this thesis several models have been presented for ligand-protein interaction. Although every model in this study can be used for further studies, we suggest four antagonists-$\alpha_{1A/B/D}$-AR complexes which show a known h-bond with an aspartate mentioned in literature. These complexes can be further refined using Quantum Mechanics/Molecular Mechanics (QM/MM) to obtain a very accurate ligand-protein complex. This is a combination of methods that treat different parts of the systems at different levels of theory. Quantum mechanics can model chemical reactions and, therefore, the part of interest is subjected to this method. MM methods are much quicker to perform and,

therefore, used on the rest of the structure and provides a biochemical background for a reaction/interaction to occur.

Four binding pockets have been proposed in each of the three α_1-AR subtypes which could be used for virtual screening studies. Firstly, the twelve compounds that were used in our MD simulations should be docked to observe if it is possible to reproduce the binding that is shown for each simulation with one docking run. Furthermore, a small dataset (28 compounds) is available from our ligand optimization which could be used as an initial dataset to validate the use of the complexed-structures in docking studies. We have taken an extended approach in optimization of our ligand dataset which, due to restrictions of time, is not possible on larger datasets. This problem needs to be solved by replacing this with a quick optimization method of the new and larger ligand dataset. A program that is commonly accepted for converting 3D structures is Corina.[153]

Docking studies should be performed using a larger dataset. To expand the dataset, two considerations should be made. There are more known ligands which interact with the different α_1-ARs and can be added to the dataset. Also, inactive compounds should be added which are known as decoys.[154] A database design strategy needs to be employed to ensure that it contains enough active and inactive compounds which cover a significant amount of structural diversity. When docking and retrieving the active ligands is succesful the screening of a large database such as ZINC[155] can be performed. Due to the large size of the ZINC database, it is recommendable to select properties beneficial to the α_{1A}-AR and filter out compounds.

The scoring of ligand-protein complexes would be a large part of this virtual screening. Numerous functions are available, but the reliability of these is limited. Therefore, a number of scoring functions should be used and these should be combined into consensus scoring functions.

Finally, a new approach that can taken is the use of protein-ligand interaction fingerprints.[156,157] These fingerprints do not score the ligand-protein complex according to a binding energy, but use the interaction with selected residues as a determination for differentiation between binders and non-binders. These methods do require a crystal structure as a reference for interacting residues. The obtained complexes from our MD simulations provide these reference ligand-protein interactions.

References

Want to make your computer go really fast?

Throw it out a window.

Anonymous.

References

1   Benign prostatic hyperplasia: a progressive disease of aging men
     M. Embeton, G.L. Andriole, J. de la Rosette, B. Djavan, K. Hoefner, R.V. Navarette,
     J. Nordling, C. Roehrborn, C. Schulman, P. Teillac, A. Tubaro, J.C. Nickel
     Urology, 61 (2003) 267-273

2   Impact of lower urinary tract symptoms on quality of life using functional
     assessment cancer therapy scale
     J.E. Batista-Miranda, B. Molinuevo, Y. Pardo
     Urology 69 (2007) 285-288

3   The origin and development of benign prostatic hyperplasia an age-dependent
     process
     J.E. Oesterling
     J. Androl. 12 (1991) 348-355

4   EAU 2004 Guidelines on assessment, therapy and follow-up of men with lower
     urinary tract symptoms suggestive of benign prostatic obstruction (BPH guidelines)
     S. Madersbacher, G. Alivizatos, J. Nordling, C.R. Sanz, M. Emberton, J.J. de la
     Rosette
     Eur. Urol. 46 (2004) 547-554

5   5-α-Reductase inhibitors prevent the progression of benign prostatic hyperplasia
     C.G. Roehrborn
     Rev. Urol. 5 (2003) 12-21

6   The evolution of alpha-blockers for the treatment of benign prostatic hyperplasia
     H. Lepor
     Rev. Urol. 8 (2006) 3-9

7   Alpha-adrenoceptors subtypes
     B.C. Calzada, A.A. de Artinano
     Pharmacol. Res 44 (2001) 195-208

8    $α_1$ Adrenoceptor subtypes: a synopsis of their pharmacology and molecular biology
     A.A. Hancock
     Drug Develop. Res. 39 (1996) 54-107

9   A molecular mechanism for the cleavage of a disulfide bond as the primary function of agonist binding to G-protein-coupled receptors based on theoretical calculations supported by experiments.

W. Brandt, A. Golbraikh, M. Täger , U. Lendeckel

Eur. J. Biochem. 261 (1999) 89-97

10   Activation of the α_{1B}-adrenergic receptor is initiated by disruption of an interhelical salt bridge constraint

J. Porter, J. Hwa, D. Perez

J. Biol. Chem. 271 (1996) 28318-28323

11   Uncovering molecular mechanisms involved in activation of G protein-coupled receptors

U. Gether

End. Rev. 21 (2000) 90-113

12   The second extracellular loop of the dopamine D_2 receptor lines the binding-site crevice

L. Shi, J. Javitch

Proc. Natl. Acad. Sci. 101 (2004) 440-445

13   The third extracellular loop of the β_2-adrenergic receptor can modulate receptor/G-protein affinity

M. Zhao, R. Gaivin, D. Perez

Mol. Pharm. 53 (1998) 524-529

14   The third extracellular loop of G-protein-coupled receptors: more than just a linker between two important transmembrane helices

Z. Lawson, M. Wheatly

Biochem. Soc Trans. 32 (2004) 1048-1050

15   The activation process of the α_{1B}-adrenenergic receptor: potential role of protonation and hydrophobicity of a highly conserved aspartate

J, Scheer, F. Fanelli, T. costa, P. Benedetti, S. Cotecchia

Proc. Natl. Acad. Sci. 94 (1997) 808-813

16   Mutational analysis of the highly conserved arginine within the Glu/Asp-Arg-Tyr motif of the α_{1B}-adrenergic receptor: effects on receptor isomerization and activation

J. Scheer, T. Costa, F. fanelli, P. Benedetti, S. Kodja, L. Abuin, M. Tosato, S. Cotecchia

Mol. Pharm. 57 (2000) 219-231

17   Activation of the β2-adrenergic receptor involves disruption of an ionic lock
        between the cytoplasmic ends of transmembrane segments 3 and 6
        J.A. Ballesteros, A.D. Jensen, G. Liapakis, S.G.F. Rasmussen, L. Shi, U. Gether, J. A.
        Javitch
        J. Biol. Chem. 276 (2001) 29171-29177

18   Phe[310] in transmembrane VI of the α_{1B}-adrenergic receptor is a key switch residue
        involved in activation and catecholamine ring aromatic bonding
        S. Chen, M. Xu, F. Lin, D. Lee, P. Riek, R. Graham
        J. Biol. Chem. 274 (1999) 16320-16330

19   The activation process of the α1B-adrenergic receptor: potential role of protonation
        and hydrophobicity of a highly conserved aspartate
        J. Scheer, F. Fanelli, T. Costa, P.G. de Benedetti, S. Cotecchia
        Proc. Natl. Acad. Sci, 94 (1997) 808-813

20   Molecular dynamics simulations of bovine rhodopsin: influence of protonation
        states and different membrane-mimicking environments
        B. Schlegel, W. Sipppl, H.D. Holtje
        J. Mol. Model. 12 (2005) 49-64

21   Phe[303] in TMVI of the α_{1B}-adrenergic receptor is a key residue coupling TM helical
        movements to G-protein activation
        S. Chen, F. Lin, M. Xu, R. Graham
        Biochemistry, 41 (2002) 588-596

22   Phe-308 and Phe-312 in transmembrane domain 7 are major sites of α_{1A}-adrenergic
        receptor antagonist affinity
        D. Waugh, R. Gaivin, M. Zuscik, P. Gonzalez-Cabrera, S. Rosss, J. Yun, D. Perez
        J. Biol. Chem. 276 (2001) 25366-25371

23   Phe[310] in transmembrane VI of the α_{1B}-adrenergic receptor is a key switch residue
        involved in activation and catecholamine ring aromatic bonding
        S. Chen, M. Xu, F. Lin, D. Lee, P. Riek, R. Graham
        J. Biol. Chem. 274 (1999) 16320-16330

24   Novel aromatic residues in transmembrane domains IV and V involved in agonist
        binding at α_{1A}-adrenergic receptors
        D. Waugh, M. Zhao, M. Zuscik, D. Perez
        J. Biol. Chem. 275 (1999) 11698-11705

25   α_1-adrenergic receptors: new insights and directions
        M. Piascik, D. Perez
        J. Pharmacol. Exp. Ther. 298 (2001) 403-410

26   Binding site analysis of full-length α_{1A} adrenergic receptor using homology
        modelling and molecular docking
        J. Pedretti, M. Silva, L. Villa, G. Vistoli
        J. Pharmacol. Exp. Ther. 298 (2001) 403-410

27   Novel aromatic residues in transmembrane domains IV and V involved in agonist
        binding at α_{1A}-adrenergic receptors
        D. Waugh, M. Zhao, M. Zuscik, D. Perez
        J. Biol. Chem. 275 (1999) 11698-11705

28   Phe^{303} in TMVI of the α_{1B}-adrenergic receptor is a key residue coupling TM helical
        movements to G-protein activation
        S. Chen, F. Lin, M. Xu, R. Graham
        Biochemistry, 41 (2002) 588-596

29   Phe-308 and Phe-312 in transmembrane domain 7 are major sites of α_{1A}-adrenergic
        receptor antagonist affinity
        D. Waugh, R. Gaivin, M. Zuscik, P. Gonzalez-Cabrera, S. Ross, J. Yun, D. Perez
        J. Biol. Chem. 276 (2001) 25366-25371

30   G Protein-coupled receptors
        U. Gether, B.K. Kobilka
        J. Biol. Chem. 273 (1998) 17979-17982

31   G-protein-coupled receptors at a glance
        W. K. Kroze, D. J. Sheffler, B. L. Roth
        J. Cell. Sci. 116 (2003) 4867-4869

32   The conformation of neurotensin bound to its G protein-coupled receptor
        S. Luca, J.F. White, A. K. shoal, D.V. Filippov, J.H. van Boom, R. Grisshammer, M.
        Baldus
        Proc. Natl. Acad. Sci. 100 (2003) 10706-10711

33   Structure of bovine rhodopsin in a trigonal form
        J. Li, P.C. Edwards, M. Burghammer, C. Villa, G..F Schertler
        J. Mol. Biol. 343 (2004) 1409-1438

34   Interaction of class A G protein-coupled receptors with G proteins
        R. Slusarz, J. Ciarkowski
        Acta Biochim. Pol. 51 (2004) 129-136

35   The ternary complex model
        K.A. Wreggett, A. de Lean
        Mol. Phar. 26 (1983) 214-227

36   Drug efficacy at G protein-coupled receptors

      T. Anakin

      Annu. Rev. Pharacol. Toxicol. 42 (2002) 49-79

37   RAMPs: the past, present and future

      N. Parameswaran, W.S. Spielman

      Trends Biochem Sci. 31 (2006) 631-638

38   The receptor activity modifying protein family of G protein coupled receptor
      accessory proteins

      M. Udawela, D. L. Hay, P.M. Sexton

      Seminars in cell & developmental biology 15 (2004) 299-308

39   Arresting developments in heptahelical receptor signalling and regulation

      S. J. Perry, R. J. Lefkowitz

      Trends Cell Biol. 12 (2002) 130-137

40   Dimerization of α_1-adrenoceptors

      G. Milligan, J. Pediani, M. Fidock, F. Gimenez

      Biochem. Soc Trans. 32 (2004) 847-850

41   Subtype specific dimerization of α_1-adrenocpetors: effects on receptor expression
      and pharmacological properties

      M.A. Uberti, R.A. Hall, K.P. Minneman

      Mol. Pharmacol. 64 (2003) 1379-1390

42   X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in
      lipidic cubic phases.

      E. Pebay-Peyroula, G. Rummel, J.P. Rosenbusch, E.M. Landau

      Science, 277 (1997) 1676-1681

43   First principles predictions of the structure and function of G-protein-coupled
      receptors: validation for bovine rhodopsin

      R. Trabino, S.E. Hall, N. Vaidehi, W.B. Floriano, V.W.T. Kam, W.A. Goddard

      Biophys. J. 86 (2004) 1904-1921

44   Identification of core amino acids stabilizing rhodopsin

      A.J. Rader, G. Anderson, B. Isin, H.G. Khorana, I. Bahar, J. Klein-Seetharaman

      Proc. Natl. Acad. Sci. 101 (2004) 7246-7251

45   Rhodopsin: structural basis of molecular physiology

      S.T. Menon, M. Han, T.P. Sakmar

      Physiol. Rev. 81 (2001) 1659-1688

46   The crystallographic model of rhodopsin and its use in studies of other G- protein-coupled receptors

S. Filipek, D.C. Teller, K. Palczewski,k R. Stenkamp

Annu. Rev. Biophys. Biomol. Struct. 32 (2003) 375-397

47   Crystal structure of rhodopsin: a G protein-coupled receptor

K. Polczewski, T. Kumusaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, M. Miyano

Science 289 (2000) 733-734

48   Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography

T. Okada, Y. Fujiyoshi, M. Silow, J. Navarro, E.M. Landau, Y. Shichida

Proc. Natl. Acad. Sci. 99 (2002) 5982-5987

49   The retinal conformation and its environment in rhodopsin in light of a new 2.2Å crystal structure

T. Okada, M. Sugihare, A.N. Bonder, M. Elstner, P. Entel, V. Buss

J. Mol. Biol. 342 (2004) 571-583

50   Rhodopsin: a structural primer for G-protein coupled receptor

R.E. Stenkamp, D.C. Teller, K. Palczewski

Arch. Pharm. Chem. Life Sci. 338 (2005) 209-216

51   Crystal structure of the human beta2 adrenergic G-protein-coupled receptor

S.G. Rasmussen, H.J. Choi, D.M. Rosenbaum, T.S. Kibilka, F.S.Thian, P.C. Edwards, M. Burghammer, V.R. Ratnala, R. Sanishvili, R.F. Fischetti, G.F. Schertler

Nature 15 (2007) 383-387

52   Synthesis, biological evaluation, and pharmacophore generation of new pyridazinone derivatives with affinity toward a_1- and a_2-adrenoceptors

R. Barbaro, L. Betti, M. Botta, F. Corelli, G. Giannaccini, L. Maccari F. Manetti, G. Strappaghetti, S. Corsano

J. Med. Chem. 44 (2001) 2118-2132

53   Ligand design for a_1-adrenoceptor subtype selective antagonists

J.B. Bremner, B. Coban, R. Griffith, K.M. Groenwoud, B.F. Yates

Bioorg. Med. Chem 8 (2000) 201-214

54   Selective pharmacophore design for a_1-adrenoceptor subtypes

I.J.A. MacDougall, R. Griffith

J. Mol. Graph. 25 (2006) 146-157

55   Pharmacophore identification of α_{1A}-adrenoceptor antagonists

M. Li, K. Tsai, L. Xia

Bioorg. Med. Chem. 15 (2005) 657-664

56   New pyrimido[5,4-b]indoles as ligands for α_1-adrenoceptor subtypes

G. Romeo, L. Materia, F. Manetti, A. Cagnotto, T. Mennini, F. Nicoletti, M. Botta, F. Russo, K.P. Minneman

J. Med. Chem. 46 (2003) 2877-2894

57   Phenylpiperazinylalkylamino substituted pyridazinones as potent α_1 adrnocpetor antagonists

D. Barlocco, G. Cignarella, V.D. Piaz, M.P. Giovannoni, P.G. de Benedetti, F. Fanelli, F. Montesano, E. Poggesi, A. Leonarid

J. Med. Chem. 44 (2001) 2403-2410

58   A genetic-function-approximation-based QSAR model for the affinity of arylpiperazines toward α_1 adrenoceptors

L. Maccari, M. Magnani, G. Strappaghetti, F. Corelli, M. Botta, F. Manetti

J. Chem. Inf. Model. 46 (2006) 1466-1478

59   α_1 adrenoceptor subtype selectivity 3D-QSAR models for a new class of α_1 adrenoceptor antagonists derived from the novel antipsychotic sertindole

T. Balle, K. Andersen, K. K. Soby, T. Liljefors

J. Mol. Graph. 21 (2003) 523-534

60   Self-organizing molecular field analysis on α_{1a}-adrenoceptor dihydrophyridine antagonists

M. Li, L. Du, B. Wu, L. Xia

Bioorg. Med Chem 11 (2003) 3945-3951

61   QSAR study for a novel series of ortho monosubstitued phenoxy analogues of α_1-adrenoceptor antagonist WB4101

L. Fumagalli, C. Bolchi, S. Colleoni, M. Gobbi, B. Moroni, M. Pallavicini, A. Pedretti, L. Villa, G. Vistoli, E. Valoti

Bioorg. Med. Chem. 13 (2005) 2547-2559

62   Binding site analysis of full-length α_{1a} adrenergic receptor using homology modelling and molecular docking

A. Pedretti, M.E. Silva, L. villa, G. Vistoli

Biochem. Biophys. Res. Commun. 319 (2004) 493-500

63   Synthesis, screening and molecular modelling of new potent and selective
      antagonists at the a_{1d} adrenergic receptor
      A. Leonardi, D. Barlocco, F. Montesano, G. Cignarella, G. Motta, R. Testa, E.
      Poggesi, M. Seeber, P.G. de Benedetti, F. Fanelli,
      J. Med. Chem. 47 (2004) 1900-1918

64   Structure-based drug discovery using GPCR homology modelling: successful virtual
      screening for antagonists of the alpha1A adrenergic receptor
      A. Evers, T. Klabunde
      J. Med. Chem. 48 (2005) 1088-1097

65   Modelling the interaction of catecholamines with the a_{1A} adrenoceptor towards a
      ligand-induced receptor structure
      G.K. Kinsella, I. Rozas, G.W. Watson
      J. Comput. Aided Mol. Des. 19 (2005) 357-367

66   Comparative molecular dynamics simulations of uncomplexed, 'agonist-bound' and
      'antagonist-bound' a_{1A} adrenoceptor models
      G.K. Kinsella, I. Rozas, G.W. Watson
      Biochem. Biophys. Res. Commun. 333 (2005) 737-741

67   Computational study of antagonists/ a_{1A} adrenoceptor complexes – observations of
      conformations variations on the formation of ligand/receptor complexes.
      G.K. Kinsella, I. Rozas, G.W. Watson
      J. Med. Chem 49 (2006) 501-510

68   A two model receptor system of the $alpha_{1D}$ adrenergic receptor to describe
      interactions with epinephrine and BMY7378
      D.L. Bautista, D.H. Morris, L. Stein, W. asher, T. Hammitt
      J. Chem. Inf. Model. 46 (2006) 334-344

69   Prediction of transmembrane alpha-helices in procariotic membrane proteins: the
      Dense Alignment Surface method
      M. Cserzo, E. Wallin, I. Simon, G. von Heijne, A. Elofsson
      Protein Eng. 10 , 6 (1997) 673-676

70   A hidden Markov model for predicting transmembrane helices in protein sequences
      E.L. Sonnhammer, G. Von Heijne, A. Krogh
      Proc. Int. Conf. Intell. Syst. Mol. Biol. 6 (1998) 175-182

71   Tmbase – A database of membrane spanning proteins segments
      K. Hofmann, W. Stoffel
      Biol. Chem. Hoppe-Seyler, 374 (1993) 166

72   Homology modelling

      E. Krieger, SB. Nabuurs, G. Vriend

      Methods Biochem. Anal. 44 (2003) 509-23

73   Multiple alignment using hidden Markov models

      S.R. Eddy

      Proc. Int. Conf. INtell. Syst. Mol. Biol. 3 (1995) 114-120

74   T-coffee: A novel method for fast and accurate multiple sequence alignment

      C. Notredame, D.G. Higgins, J. Heringa

      J. Mol. Biol. 8, 302, 1 (2000) 205-217

75   Comparative protein modelling by satisfaction of spatial restraints

      A. Sali, T. blundell

      J. Mol Biol. 324 (1993) 779-815

76   Procheck: a program to check the stereochemical quality of protein structure

      R.A. Laskowski, M.W. MacArthur. D.S. Moss, J.M. Thornton

      J. Appl. Cryst. 26 (1993) 183-291

77   Assessment of protein models with three-dimensional profiles

      R. Luthy, J.U. Bowie, D. Eisenberg

      Nature, 356 (1992) 83-85

78   Verification of protein structures: patterns of nonbonded atomic interactions

      C. Coloyos, T.O. Yeates

      Prot. Sci. 2 (1993) 1511-1519

79   Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein
      structures

      R.W.W. Hooft, C. Sander, G. Vriend

      Proteins 26 (1996) 363-376

80   What if: a molecular modeling and drug design program

      G. Vriend

      J. Mol. Graph 8, 1, 29 (1990) 52-56

81   Molecular Modelling principles and applications. Chapter 2.1 Introduction

      A.R. Leach

      Pearson Education Ltd 1996, second edition 2001

      ISBN 0-582-38210-6

82   Essentials of computational chemistry: Theories and models. Chapter 4.1 Quantum
        mechanics and the wave function
        C.J. Cramer
        John Wiley & Sons, second edition 2004
        ISBN 0-470-09182-7

83   A guide to molecular mechanic and quantum chemical calculations. Chapter 2
        Quantum chemical models
        W.J. Hehre
        Wavefunction Inc. 2003
        ISBN 1-890661-18-X

84   Molecular Modelling principles and applications. Chapter 2.5 The Hartree-Fock
        equations
        A.R. Leach
        Pearson Education Ltd 1996, second edition 2001
        ISBN 0-582-38210-6

85   Essentials of computational chemistry: Theories and models. Chapter 6 Ab Initio
        implementation of hartree-Fock molecular orbital theory
        C.J. Cramer
        John Wiley & Sons, second edition 2004
        ISBN 0-470-09182-7

86   Molecular Modelling principles and applications. Chapter 3.7 Density functional
        theory
        A.R. Leach
        Pearson Education Ltd 1996, second edition 2001
        ISBN 0-582-38210-6

87   Essentials of computational chemistry: Theories and models. Chapter 8 Density
        functional theory
        C.J. Cramer
        John Wiley & Sons, second edition 2004

88   Essentials of computational chemistry: Theories and models. Chapter 8.4
        Exchange-correlation functionals
        C.J. Cramer
        John Wiley & Sons, second edition 2004
        ISBN 0-470-09182-7

89   Molecular Modelling principles and applications. Chapter 2.6 Basis sets

A.R. Leach

Pearson Education Ltd 1996, second edition 2001

ISBN 0-582-38210-6

90   Essentials of computational chemistry: Theories and models. Chapter 6.2 Basis sets

C.J. Cramer

John Wiley & Sons, second edition 2004

ISBN 0-470-09182-7

91   Molecular Modelling principles and applications. Chapter 2.9 Semi-empirical

methods

A.R. Leach

Pearson Education Ltd 1996, second edition 2001

ISBN 0-582-38210-6

92   Essentials of computational chemistry: Theories and models. Chapter 5

Semiempirical implementations of molecular orbital theory

C.J. Cramer

John Wiley & Sons, second edition 2004

ISBN 0-470-09182-7

93   Molecular interactions in solution: an overview of methods based on continuous

distributions of the solvent

J. Tomasi, M. Persico

Chem. Rev. 94 (1994) 2027-2094

94   Theoretical proton affinities of α1 adrenoceptor ligands

G.K. Kinsella, G.W. Watson, I. Rozas

Bioorg. Med. Chem. 14 (2006) 1580-1587

95   Theoretical calculation of pK_a using the cluster-continuum model

J. Pliego, J.M. Riveros

J. Phys. Chem. A. 106 (2002) 7434-7439

96   Adding explicit solvent molecules to continuum solvent calculations for the

calculation of aqueous acid dissociation constants

C.P. Kelly, C.J. Cramer, D.G. Truhlar

J. Phys. Chem. A. 110 (2006) 2493-2499

97   Computational determination of aqueous pK_a values of protonated benzimidazoles

(part 1)

T.N. Brown, N. Mora-Diez

J. Phys. Chem. B 110 (2006) 9270-9279

98   Absolute pK$_a$ determinations for substituted phenols

        M.D. Liptak, K.C. Gross, P.G. Seybold, S. Feldgus, G.C. Shields

        J. Am. Chem. Soc 124 (2002) 6421-6427

99   Computational approach to the basicity of a series of α1-adenoceptor ligands in
        aqueous solution

        G.K. Kinsella, F. Rodriquesz, G.W. Watson, I. Rozas

        Bioorg. Med. Chem. 15, 8 (2007) 2850-2855

100   The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from
        Cluster-Ion Solvation Data

        M.D. Tissandier, K.A. Cowen, W.Y. Feng, E. Gundlach,

        M.H. Cohen, A.D. Earhart, J.V. Coe

        J. Phys. Chem. A 102 (1998) 7787-7794

101   Comment on "accurate experimental values for the free energies of hydration of
        H$^+$, OH$^-$ and H$_3$O"

        D.M. Camaioni, C.A. Schwerdtfeger

        J. Phys. Chem. A 109 (2005) 10795-10797

102   Theoretical evaluation of pK(a) in phosphoranes: implication sfor phosphate ester
        hydrolysis

        X. Lopez, M. Schaefer, A. Dejaegere, M. Kaplus

        J. Am. Chem. Soc. 1, 124, 18 (2002) 5010-5018

103   Structure-based virtual screening: an overview

        P.D. Lyne

        DDT, 7, 20 (2002) 1047-1055

104   Virtual screening and fast automated docking methods

        G.S. Schneider, H.J. Bohm

        DDT, 7, 1 (2002) 64-70

105   Docking and scoring in virtual screening for drug discovery: methods and
        applications

        D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath

        Nat. Rev. 3 (2004) 935-949

106   Automated docking using a Lamarckian genetic algorithm and empirical binding
        free energy function

        G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J.
        Olson

        J. Comp. Chem. 19 (1998) 1639-1662

107   Development and validation of a modular, extensible docking program: DOCK 5

        D.T. Moustakas, P.T. Lang, S. Pegg, I.D. Kuntz, N. Brooijmans, R.C. Rizzo

        J. Comput. Aided. Mol. Des (2006)

108   Further development and validation of empirical scoring functions for structure-
        based binding affinity prediction

        R. Wang, L. Lai, S. Wang

        J. Comput.-Aided Mol. Des. 16 (2002) 11-26

109   Consensus scoring for ligand/protein interactions

        R.D. Clark, A. Strizhev, J.M. Leonard, J.B. Matthew

        J. Mol. Graph. Model. 20, 4 (2002) 281-295

110   Development and validation of a genetic algorithm for flexible docking

        G. Jone, P. Willett, R.C. Glen, A.R. Leach, R. Taylor

        J. Mol. Biol. 3, 267 (1997) 727-748

111   A geometric approach to macromolecule-ligand interactions

        I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin

        J. Mol. Biol. 161, 2 (1982) 269-288

112   Prediction of binding constants of protein ligands: a fast method for the
        prioritization of hits obtained from de novo design or 3D database search
        programs.

        H.J. Böhm

        J. Comput. Aided. Mol. Des. 12 (1998) 309-323

113   Empirical scoring functions: I. The development of empirical scoring function to
        estimate the binding affinity of ligands in receptor complexes.

        M.D. Eldrige, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee

        J. Comput-aided. Mol. Design 11 (1997) 425-445

114   A fast flexible docking method using an incremental construction algorithm

        M. Rary, B. Kramer, T. Lengauer, G. Klebe

        J. Mol. Biol. 261 (1996) 470-489

115   A general and fast scoring function for protein-ligand interactions: a simplified
        potential approach

        I. Muegge, Y.C. Martin

        J. Med. Chem. 42, 5 (1999) 791-804

116   Knowledge-based scoring function to predict protein-ligand interactions

        H. Gohlke, M. Hendlich, G. Klebe

        J. Mol. Bio. 14, 295, 2 (2000) 337-356

117   Consensus scoring for protein-ligand interactions
       M. Feher
       DDT. 11, 9/10 (2006) 421-428

118   LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions
       A.C. Wallace, R.A. Laskowski, J.M. Thornton
       Protein Eng. 8, 2 (1995) 127-134

119   Molecular modelling, Principles and applications. Chapter 4 Empirical force field
       models: molecular mechanics, chapter 7 molecular dynamics simulation methods
       A.R. Leach
       Prentice Hall, second edition 1996, 2001
       ISBN 0-582-38210-6

120   Molecular modelling, Principles and applications. Chapter 5 Energy minimisation
       and related methods for exploring the energy surface
       A.R. Leach
       Prentice Hall, second edition 1996, 2001
       ISBN 0-582-38210-6

121   Molecular modelling, Principles and applications. Chapter 4.3 Bond strechting,
       chapter 4.4 Angle bending, chapter 4.5 Torsional terms, chapter 4.6 Improper
       torsions and out-of-plane bendig motions, chapter 4.8 Introduction to non-bonded
       interactions, chapter 4.9 Electrostatic interactions, chapter 4.10 Van der Waals
       interactions
       A.R. Leach
       Prentice Hall, second edition 1996, 2001
       ISBN 0-582-38210-6

122   The Amber biomolecular simulation programs
       D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C.
       Simmerling, B. Wang, R. Woods
       J. Comp. Chem 26 (2005) 1668-1688

123   CHARMM: A program for macromolecular energy, minimization and dynamics
       calculations
       B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus
       J. Comp. Chem. 4 (2005) 1701-1718

124   Gromacs: fast flexible and free
       D. van der Spoek, E. Lindah, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen
       J. Comp. Chem. 16 (2005) 1701-1718

125   Molecular modelling, Principles and applications. Chapter 7.3 Molecular dynamics
       with continuous potentials
       A.R. Leach
       Prentice Hall, second edition 1996, 2001
       ISBN 0-582-38210-6

126   Molecular modelling, Principles and applications. Chapter 6.5 Boundaries
       A.R. Leach
       Prentice Hall, second edition 1996, 2001
       ISBN 0-582-38210-6

127   Molecular modelling, Principles and applications. Chapter 6.7 Truncating the
       potential the minimum image convention
       A.R. Leach
       Prentice Hall, second edition 1996, 2001
       ISBN 0-582-38210-6

128   Molecular modelling, Principles and applications. Chapter 6.8 Long-range forces
       A.R. Leach
       Prentice Hall, second edition 1996, 2001
       ISBN 0-582-38210-6

129   Towards an accurate representation of electrostatics in classical force fields:
       efficient implementation of multipolar interaction sin biomolecular simulations
       C. Saqui, L.G. Pedersen, T.A. Darden
       J. Chem. Phys, 1, 120, 1 (2004) 73-87

130   CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment
       through sequence weighting, position-specific gap penalties and weight matrix
       choice.
       D. Higgins, J. Thompson, T. Gibson. J.D. Thompson, D.G. Higgins, T.J. Gibson
       Nucleic Acids Res. 22 (1994) 4673-4680

131   Construction of a sequence motif characteristic of aminergic G protein-coupled
       receptors
       E. Huang
       Protein Sci. 12 (2003) 1360-1367

132   Computational development of an α_{1A}-adrenoceptor model in a membrane mimic.
       G.K. Kinsella, I. Rozas, G.W. Watson
       Biochem. Biophys. Res. Commun. 12,324 (2004) 916-912

133   Force fields for protein simulations.

J.W. Ponder and D.A. Case.

Adv. Prot. Chem. 66, 27-85 (2003)

134   Comparison of multiple Amber force fields and development of improved protein backbone parameters.

V. Hornak, R. Abel, A. Okur, B. Roitberg, C. Simmerling

Proteins, 15, 65 (2006) 712-725

135   Characteristics for a salt-bridge switch mutation for the α1b adrenergic receptor

J. E. Porter, D. M. Perez

J. Biol. Chem. 274 (1999) 34535-34538

136   High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor

V. Cherezov. D. M. Rosenbaum, M. A. Hanson, S. G. F. Rasmussen, F. S> thian, T. S. Kobilka, H-J. Choi, P. Kuhn. W.I. Weis, B. K. Kobilka. R. C. Stavens

Science 328 (2007) 1258-1265

137   Solvent effect on the tautomerism of 4-aminopyrazino[2,3-c]-[1,2,6]thiadiazine 2,2-dioxides

N. Campillo, I. Alkorta, J. Antonia Paez, P. Goya

J. Chem. Soc. Perkin Trans. 2 (1998) 1889-1892

138   Sybyl7.2, Molecular Modelling system, Tripos Associates: St. Louis, MO.

139   Synthesis and pharmacological characterization of 3-[2-((3aR,9bR)-cis-6-methoxy-2,3,3a,4,5,9b-hexahydro-1H-benz[e] isoindol-2-yl)ethyl]pyrido-[3',4':4,5]thieno[3,2-d]pyrimidine-2,4 (1H,3H)-dione (A-131701): a uroselective alpha 1A adrenoceptor antagonist for the symptomatic treatment of benign prostatic hyperplasia

M.D. Meyer, R.J. Altenbach, F.Z. Basha, W.A. Carroll, I. Drizin, S.W. Elmore, P.P. Ehrlich, S.A. Lebold, K. Tietje, K.B. Sippy, M.D. Wendt, D.J. Plata, F. Plagge, S.A. Buckner, M.E. Brune, A.A. Hancock, J.F. Kerwin

J. Med. Chem 40,20 (1997) 3141-3143

140   Receptor binding profile of cyclazosin, a new alpha 1B-adrenoceptor antagonist

D. Giardina, M. Crucianelli, C. Melchiorre, C. Taddei, R. Testa

Eur. J. Pharmacol. 287,1 (1995) 13-16

141 The alpha 1-adrenergic receptor that mediates smooth muscle contraction in human prostate has the pharmacological properties of the cloned human alpha 1c subtype
C. Forray, J.A. Bard, J.M. Wetzel, G. chiu, E. Shapiro, R. Tang, H. Lepor, P.R. Hartig, R.L. Weinshank, T.A. Branchek
Mol. Pharmacol. 45,4 (1994) 703-708

142 Pharmacological characterization of the uroselective alpha-1 antagonist Rec 15/2739 (SB 216469): role of the alpha-1L adrenoceptor in tissue selectivity, part II.
R. Testa, L. Guarneri, P. Angelico, E. Poggesi, C. Taddei, G. Sironi, D. Colombo, A.C. Sulpizio, D.P. Naselsky, J.P. Hieble, A. Leonardi
J. Pharmacol. Exp. Ther, 281,3 (1997) 1284-1293

143 Pharmacological options in the treatment of benign prostatic hyperplasia
B. Kenny, S. Ballard, J. Blagg, D. Fox
J. Med. Chem. 25, 40 (1997) 1293-1315

144 M.J. Konkel, J.M. Wetzel, M. Cahir, D. Craig, S.A. Noble. C. Gluchowski,
The 216th National Meeting, Boston 23-27 August, 1998 MEDI#129

145 Alpha 1-adrenoceptor subtype selectivity: molecular modelling and theoretical quantitative structure-affinity relationships
P.G. De Benedetti, F. Fanelli, M.C. Menziani, M. Cocchi, R. Testa, A. Leonardi
Bioorg. Med. Chem. 5, 5 (1997) 809-816

146 Use of recombinant alpha 1-adrenoceptors to characterize subtype selectivity of drugs for the treatment of prostatic hypertrophy
R. Foglar, K. Shibata, K. Horie, A. Hirasawa, G. Tsujimoto
Eur. J. Pharmacol. 16, 288 (1995) 201-207

147 (-)-Discretamine, a selective alpha 1D-adrenoceptor antagonist, isolated from Fissistigma glaucescens.
F.N. Ko, J.H. Guh, S.M. Yu, Y.S. Hou, Y.C. Wu, C.M. Teng
Br. J. Pharmacol. 112, 4 (1994) 1174-1180

148 Alpha- and beta-adrenoceptors: from the gene to the clinic. 2. Structure-activity relationships and therapeutic applications.
R.R. Ruffolo, W. Bondinell, J.P. Hieble
J. Med. Chem. 15, 38, 19 (1995) 3681-3716

149   Alpha 1-adrenoceptor classification: sharpening Occam's razor.

A.P. Ford, D.R. Blue, D.E. Clarke

Trends Pharmacol. Sci. 15, 6 (1994) 167-170

150   Sybyl 7.02. www.tripos.com

151   Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004

152   Molecular dynamics simulations of membrane proteins

P.C. Biggin, P.J. Bond

Methods Mol. Bio. 443 (2008) 147-160

153   Automatic generation of 3D atomic coordinates for organic molecules

J. Gasteiger, C. Rudolph, J. Sadowski

Tetrahedron Com. Method. 3 (1990) 537-547

154   Benchmarking sets for molecular docking

N. Huang. B.K. Shoichet, J.J. Irwin

J. Med. Chem. 49 (2006) 6789-6801

155   ZINC – a free database of commercially available compounds for virtual screening

J.J. Irwin, B.K. Shoichet

J. Chem. Inf. Model. 45 (2005) 177-182

156   Structural interaction fingerprint (SIFt): a novel methods for analyzing three-dimensional protein-ligand binding interactions

J. Med. Chem. 15 (2004) 337-344

157   The use of protein-ligand interaction fingerprints in docking

S.C. Brewerton

Curr. Opin. Drug. Discov. Devel. 3 (2008) 356-364

References