

CSC RESEARCH REPORTS R1/2010

COMPUTATIONAL DESIGN OF NON-STEROIDAL MODULATORS OF ANDROGEN RECEPTOR

Annu Söderholm

CSC – IT CENTER FOR SCIENCE LTD.

Department of Biosciences Åbo Akademi University Turku, Finland

Computational Design of Non-Steroidal Modulators of Androgen Receptor

Annu Söderholm

Dissertation for the degree of Doctor of Philosophy to be presented with the permission of the Division for Natural Sciences and Technology, Åbo Akademi University, for public examination in Mauno Koivisto Centre Auditorium, Tykistökatu 6A (BioCity), Turku, on March 12th, 2010 at 12 o'clock noon

Supervisors	Adj. Prof. Tommi Nyrönen	CSC – IT Center for Science Ltd.
-	Dr. Pekka Lehtovuori	CSC – IT Center for Science Ltd.
Reviewers	Prof. C. Gopi Mohan	National Institute of Pharmaceutical
	_	Education and Research, Mohali,
		India
1	Dr. John P. Overington	European Bioinformatics Institute,
		Cambridge, United Kingdom
Opponent	Prof. Antti Poso	University of Eastern Finland
Custos	Prof. Mark S. Johnson	Åbo Akademi University

CSC Research Report R1/2010 CSC – IT Center for Science Ltd., Espoo 2010

ISBN 978-952-5520-38-5 (Print) ISBN 978-952-5520-39-2 (PDF) ISSN 0787-7498 Picaset Oy 2010

Contents

Abstracti
List of Original Papers and Author's Contributioniii
Acknowledgements iv
Abbreviations
1 Introduction 1
2 Review of the Literature
2.1 Ligand-induced modulation of steroid receptor function
2.2 Non-steroidal modulators of AR and PR activity 5
2.2.1 AR modulators
2.2.2 PR modulators
2.2.3 Selective receptor modulators10
2.3 Structural and functional organization of steroid receptors11
2.3.1 N-terminal domain11
2.3.2 DNA-binding and hinge domains12
2.3.3 C-terminal ligand-binding domain13
2.4 Structural basis for activation and inactivation of AR with ligands17
2.4.1 Activation by agonists17
2.4.2 Inhibition by antagonists19
3 Aims of Study21
4 Methods
4.1 Structural and biological data (Papers I-IV)
4.2 Virtual screening (Paper II)
4.3 Molecular docking (Papers I-IV)23
4.3.1 Docking search algorithms24
4.3.2 Docking scoring functions
4.3.3 Docking and scoring in this thesis
4.4 3D QSAR Analysis (Papers I and IV) 29
4.4.1 CoMSIA
4.5 Integration of docking and 3D QSAR for receptor-based virtual screening

4.6 Experimental methods	35
5 Results and Discussion	36
5.1 3D QSAR models (Papers I and IV)	36
5.1.1 CoMSIA of AR ligands	36
5.1.2 CoMSIA of PR ligands	39
5.2 Virtual screening of novel non-steroidal AR ligand scaffolds (Paper II)	40
5.3 Molecular modeling of antagonist mechanism identified for diphenyl- and	
phenylpyridine –based AR ligands (Paper III)	42
5.3.1 Functional characterization in vitro	42
5.3.2 Induced fit docking	43
6 Conclusions	45
7 Bibliography	48

Abstract

Computational methods constitute an integral part of modern drug development. Applications of modeling cover the entire spectrum of drug development process from hit discovery to lead optimization and beyond. A collection of computational methods that are used in order to identify potential hits for a specific target from large compound libraries is termed virtual screening (VS). VS is generally classified into receptor-based and ligand-based VS depending on whether protein structure or ligand structure, respectively, is applied in the computational compound identification. In the present thesis, both types of VS techniques were utilized and combined in the search for new androgen receptor (AR) ligands that could offer new platforms for drug development.

AR belongs to the steroid receptor (SR) family, which is a member of the large nuclear receptor superfamily of ligand-inducible transcription factors. The transcriptional activation of AR target genes is triggered by binding of an androgenic hormone, testosterone or 5α -dihydrotestosterone, to the AR. Androgenic hormones are essential for the development and maintenance of the male sexual phenotype. Disruptions in the gene regulatory function of AR caused, for instance, by receptor mutations and/or altered androgenic hormone concentrations are linked to several disorders, such as prostate cancer (CaP). CaP is the most common cancer in men in Western countries, which is why AR is a major target for drug discovery and attracts the attention of pharmaceutical industry. Design of AR-selective drugs is challenging because the ligand-binding sites of the closely related SRs are very similar and share common ligand-binding features. Progesterone receptor (PR) displays the most similar binding site structure to AR, differing only by six residues. Design of AR-targeted drugs for CaP treatment is additionally complicated by binding site mutations that are frequently found in CaP patients.

A panel of known AR ligands published in the literature was used in this thesis to develop a three-dimensional quantitative structure-activity relationship (3D QSAR) model. Constructed within the AR binding site, the model identified and quantified structural features in ligands that are important for AR binding of the studied compounds. 3D QSAR models are typically used to predict biological activities of compounds structurally related to ones used in model development process. In this thesis however, the 3D QSAR model of AR ligands was integrated as part of our receptorbased VS workflow with the aim to computationally identify new non-steroidal AR hit compounds with structural scaffolds distinct from ones used to derive the model.

Experimental affinity determination confirmed that with our computational approach we indeed found novel AR ligands, yet rather weakly binding. A subset of the new AR ligands was further tested *in vitro* for functional activity in wild-type and T877A mutant AR often associated with CaP. The ligands showed inhibition of AR activity in both AR structures tested. A possible structural mechanism for the inhibitory function was suggested based on computational modeling. Taken together, these AR ligands could serve as potential starting points for development of non-steroidal AR antagonists, for example for CaP treatment.

List of Original Papers and Author's Contribution

This thesis is based on the following papers, which are referred to in the text by Roman numerals (I-IV).

- I Annu A. Söderholm, Pekka T. Lehtovuori, Tommi H. Nyrönen. Three-Dimensional Structure-Activity Relationships of Nonsteroidal Ligands in Complex with Androgen Receptor Ligand-Binding Domain. J. Med. Chem. 48(4): 917-925, 2005.
- II Annu A. Söderholm, Pekka T. Lehtovuori, Tommi H. Nyrönen. Computational Detection of Non-Steroidal Androgen Receptor Ligands. *Manuscript*
- III Annu A. Söderholm, Johanna Viiliäinen, Pekka T. Lehtovuori, Hanna Eskelinen, Daniela Roell, Aria Baniahmad, Tommi H. Nyrönen. Computationally Identified Novel Diphenyl- and Phenylpyridine Androgen Receptor Antagonist Structures. J. Chem. Inf. Model. 48(9): 1882-90, 2008.
- IV Annu A. Söderholm, Pekka T. Lehtovuori, Tommi H. Nyrönen. Docking and Three-Dimensional Quantitative Structure-Activity Relationship (3D QSAR) Analyses of Nonsteroidal Progesterone Receptor Ligands. J. Med. Chem. 49(14): 4261-4268, 2006.

The author's contribution in Papers I – IV is as follows. In Papers I – IV, the author was responsible for performing and analyzing the results of the computational modeling work. The experimental measurements for biological activity described in Papers II and III were not performed by the author, but the author had the main responsibility of the analysis of the results. The author was the primary author of all the Papers and additionally the corresponding author of Paper IV.

Acknowledgements

This work was carried out at CSC – IT Center for Science, where I had the privilege of working during years 2003-2007. CSC is gratefully acknowledged for providing excellent facilities and computational resources for performing this work, which was financially supported by the Academy of Finland, National Graduate School in Informational and Structural Biology (ISB), the Finnish Cultural Foundation and Rector of Åbo Akademi University.

I wish to express my sincere gratitude to my supervisors Adj. Professor Tommi Nyrönen and Dr. Pekka Lehtovuori for their guidance, constructive comments and encouragement during these years. Your positive attitude and trust in my capabilities have been invaluable and greatly appreciated. In addition, I am grateful to you for taking care of the practical and administrative matters during my visit at CSC. I want to extend my gratitude to my supervisory committee members, Professor Kari Keinänen and Adj. Professor Tiina Salminen, for their opinions and advice.

The reviewers of this thesis, Professor C. Gopi Mohan and Dr. John P. Overington, are acknowledged for their careful examination and valuable comments.

Professor Aria Baniahmad and his research groups both at the University of Eastern Finland and at Friedrich Schiller University are warmly thanked for a rewarding collaboration. In particular, Johanna Viiliäinen, Hanna Eskelinen and Daniela Roell are thanked for performing the wet lab work presented in Paper III.

I would especially like to thank the personnel at CSC for contributing to such a pleasant working environment. Technical support from the members of the chemistry group was always close at hand and highly appreciated. Many thanks to Heli Autere, Taavi Hupponen, Aleksi Kallio, Anne Nousiainen, Sami Salonen, Harri Sylvander and Anu Tervo for all the great times shared both in and out of the workplace and for making the long journey to completing this work much more enjoyable. Very special thanks to Susanna Repo for her excellent company to several conferences.

I would also like to thank Professor Mark S. Johnson, the director of the ISB Graduate School that I have been proud to belong to, and the ISB personnel for arranging the inspiring winter and spring meetings each year.

Finally, I wish to thank Thomas for his love and patience and my family for their continuous support and encouragement.

Espoo, December 2009 Annu Söderholm

Abbreviations

3D	Three-dimensional	
AF1, AF2	Activation function 1 and 2	
AR	Androgen receptor	
CaP	Prostate cancer	
СоА	Co-activator	
CoMFA	Comparative Molecular Field Analysis	
CoMSIA	Comparative Molecular Similarity Indices Analysis	
CoR	Co-repressor	
CPA	Cyproterone acetate (steroidal antiandrogen)	
DBD	DNA-binding domain	
DHT	5a-Dihydrotestosterone	
DNA	Deoxyribonucleic acid	
C-terminal	Carboxy-terminal end of polypeptide chain	
ER	Estrogen receptor	
GA	Genetic algorithm	
GOLD	Genetic Optimization for Ligand Docking	
GR	Glucocorticoid receptor	
Н	Helix	
HAT	Histone acetyl transferase	
HDAC	Histone deacetylase	
HF	Hydroxyflutamide	
HRE	Hormone response element	
HTS	High-throughput screening	
IC ₅₀	Half-maximal (50%) inhibitory concentration	
IFD	Induced-fit docking	
LBD	Ligand-binding domain	
LBP	Ligand-binding pocket	
MC	Monte Carlo	
MD	Molecular dynamics	
MR	Mineralocorticoid receptor	
N-terminal	Amino-terminal end of polypeptide chain	
NCoR	Nuclear receptor co-repressor	
NR	Nuclear receptor	

NTD	N-terminal domain
Р	Progesterone
PDB	Protein Data Bank
PLS	Partial least squares
PR	Progesterone receptor
QSAR	Quantitative structure-activity relationship
SAR	Structure-activity relationship
SARM	Selective androgen receptor modulator
SPRM	Selective progesterone receptor modulator
SRM	Selective receptor modulator
SMRT	Silencing mediator of retinoid and thyroid hormone receptor
SR	Steroid receptor
SRC	Steroid receptor co-activator
Т	Testosterone
TFIIF	Transcription factor IIF
VS	Virtual screening
WT	Wild-type

1 Introduction

Steroid hormones are small cholesterol derived lipophilic compounds that through cellular signaling pathways regulate a range of physiological functions. The effects of the hormones are mediated by their respective intracellular steroid receptors (SRs) – androgen (AR), estrogen (ER), progesterone (PR), glucocorticoid (GR), and mineralocorticoid (MR) receptors – which are ligand-dependent transcription factors that in response to hormone binding regulate gene expression in target tissues [1, 2]. SR function is of great interest to the pharmaceutical industry because dysfunction in hormone signaling leads to disorders like cancer, infertility, and inflammatory diseases.

The endogenous androgens, testosterone (T) and its metabolite 5α -dihydrotestosterone (DHT), are essential hormones for male sexual differentiation and maintenance of male characteristics in reproductive and non-reproductive tissues (reviewed in [3]). The androgen-induced effects on the reproductive tissues like prostate are known as androgenic, while the effects on the non-reproductive tissues like muscle and bone are referred to as anabolic. DHT is the primary androgen in the prostate and accounts for the androgenic effect while T is the main androgen in muscle and responsible for the anabolic effects (reviewed in [4]).

The importance of functional AR-mediated signaling in maintaining the male physiology is demonstrated by disorders caused by changes in the endogenous androgen production or receptor activity [3, 4]. Low T levels in hypogonadal men have been associated with e.g. loss of musculoskeletal strength but also with more severe effects such as incomplete sexual differentiation depending on the stage of life at which the T decline occurs. Numerous AR mutations that are characterized as gain-of-function mutations, which confer increased AR activity, have been linked to pathologies like prostate cancer (CaP) [3, 5].

Many disorders related to AR function can be treated with small synthetic compounds (ligands) that either activate or inactivate the AR-mediated signaling pathways. Conditions associated with androgen deficiency (e.g. hypogonadism, muscle wasting, osteoporosis) can be treated with agents that mimic the effects of natural androgens (agonist/androgen), whereas agents able to oppose the effects of natural androgens (antagonist/antiandrogen) are used to treat conditions associated with androgen excess (e.g. hirsutism, acne) and androgen dependent CaP. Currently all the AR agonist and a few AR antagonist drugs on the market are steroidal compounds. Their chronic medical use is however limited due to poor oral bioavailability and side effects such as liver toxicity and other adverse effects caused by cross-reactivity with related SRs (reviewed in [6, 7]). It is well-known that reduced cross-reactivity is easiest achieved with non-steroidal compounds. At present, there are three non-steroidal antiandrogens available clinically but despite improved AR-selectivity, they are not free of side effects. Effort has thus been placed on identification of new non-steroidal AR ligands – several non-steroidal androgens and antiandrogens are in the drug development pipeline [6-10].

The focus of this literature review is on AR function and AR ligands because the main goal of the present thesis was to discover novel AR ligands that could offer new platforms for drug development. AR continues to be a challenging target for drug discovery even if there is a wealth of experimental data available on AR and its ligands. First, only the active structure of AR is currently known, which hampers the design of AR antagonists. Second, the ligand-binding sites of closely related SRs are very similar and share common binding features, which complicate development of drugs with receptor-selective function. Because detection of receptor-selective ligands is of major concern within the SR-field, the most closely related SR to AR, i.e. PR, and its ligands were also studied in this thesis. Judging from literature, a very small chemical change in the structure of a non-steroidal AR ligand can cause a complete change in its activity. One source of this effect arises from binding to similar binding sites in related SRs and other proteins in the signaling pathway.

The goal in the early stage of drug discovery is to identify novel hit compounds for a target protein from large chemical libraries. A hit is defined as a compound that shows activity against a biological target in the primary *in vitro* screening test. Screening hits may be developed into lead compounds if the affinity, selectivity and pharmacokinetic properties of the hits can be improved by manipulating their chemical structure. The hit-to-lead process necessitates synthesis of numerous structural analogues to be tested in both *in vitro* assays and *in vivo* animal models. Further optimization of leads may produce drug candidates that are safe to enter the first clinical trials on human subjects.

Computational methods have been successfully integrated into the drug development process during the last decade. Experimental high-throughput screening (HTS) approaches are now commonly complemented with computer-assisted virtual screening (VS) approaches for hit identification. VS uses various computational methods to explore large chemical libraries to identify new chemical classes of active compounds for a target protein (reviewed in [11-18]). While an entire library is commonly tested for activity in HTS, the objective in VS is to single out only the most potential library compounds for biological testing. The selection of compounds is based on various knowledge-based criteria, which are computationally applied in order to filter out library compounds that are unlikely to interact with the target. Computational methods thus rationally augment plain HTS, which is random in nature.

The computational VS methods are categorized into receptor-based and ligandbased methods. Receptor-based methods are applicable when the three-dimensional (3D) structure of the target protein and its ligand-binding site are known. The most commonly used receptor-based method is molecular docking, which attempts to predict ligand interaction with the target protein and to estimate the strength of the interaction. Ligand-based methods only require that structures of active ligands are known. Information of a few active ligands is usually utilized in similarity searches to identify chemically related ligands. Larger series of ligands displaying differential biological activity can be used to perform quantitative structure-activity relationship (QSAR) analyses. Such analyses attempt to quantify changes in biological activities of ligands to changes in their chemical structures. Both the receptor-based and ligand-based methods are suitable for hit identification as well as lead generation and optimization. In terms of ability to detect structurally novel ligands, the receptor-based methods are more valuable than ligandbased methods because they are not biased by the properties of known ligands [19].

In this thesis, the aim was to uncover new non-steroidal AR ligands with previously unknown molecular scaffolds and with binding selectivity over PR. The binding properties of AR and PR ligands derived from the literature were therefore investigated by means of receptor-based and ligand-based computational drug discovery methods. Even though the emphasis of the thesis was on computational methods, there was a tight collaboration with experimentalists throughout the project for evaluation of computational predictions and generation of additional experimental data for subsequent modeling. In the end, successful drug discovery relies on a concerted effort of a multi-disciplinary team.

2 Review of the Literature

2.1 Ligand-induced modulation of steroid receptor function

Ligand binding modulates SR-mediated gene transcription. The ligand may be either the endogenous steroid hormone or a pharmacological substance (agonist, antagonist, or partial agonist/antagonist). Substances having a similar activating effect on target gene expression as the natural hormone(s) are termed agonists. If the synthetic agonist is unable to induce the maximal activation achieved by the endogenous steroid(s), it has a partial agonist activity. Substances that are able to inhibit the effect of natural steroid(s) through competitive binding to the receptor are antagonists. The antagonistic effect may also be partial. Yet another class of substances is the selective receptor modulators (SRMs), which are neither pure agonists nor antagonists but display a mixed agonist/ antagonist activity in a tissue-specific manner.

Upon binding of a ligand, SR undergoes a conformational change that induces its release from a cytoplasmic heat shock protein complex. Receptor-ligand complex translocates into the nucleus and binds as dimers to specific regulatory sequences known as hormone response elements (HREs) located in the promoter regions of its target genes. At the HREs, the liganded SRs recruit a large number of co-regulatory proteins in order to facilitate communication with the basal transcription machinery (for reviews of co-regulators see [20-22]). Co-regulators are structurally and functionally diverse proteins that interact with the SRs either directly or indirectly to activate (co-activators, CoA) or to reduce (co-repressors, CoR) transcriptional activity. During the past ten years it has become evident that the transcriptional regulation of SRs is a highly intricate process: 169 proteins have been classified as potential AR co-regulators that modulate AR activity [22]. Such regulatory complexity, indicated already by the pure number of co-regulators, was unforeseeable when the first CoA (SRC-1: steroid receptor co-activator-1) [23] and CoRs (NCoR: nuclear receptor co-repressor and SMRT: silencing mediator of retinoid and thyroid hormone receptor) [24, 25] were identified over a decade ago. The subject is only briefly covered here to illustrate the modulatory power of ligands on SR interactions with co-regulator complexes.

Agonist binding changes the SR conformation in a way to allow interaction with CoAs, the best-characterized group being the p160 family proteins (referred to as SRC-1, 2, and 3). These CoAs contribute to an increased transcriptional activity by their intrinsic histone acetylase (HAT) activity and by attracting additional CoAs with various histone modifying enzymatic activities. Together the modifications lead to relaxation of chromatin, which enables increased access to the DNA by the basal transcription machinery and hence transcriptional activation. Antagonist binding, on the other hand, induces such conformational changes to the SR structure that facilitate CoR interactions with the receptor. The well-known CoRs, NCoR and SMRT, often form the basal platform for the CoR complexes and serve to recruit other proteins that contain histone deacetylase (HDAC) activity with the consequence of a more condensed chromatin structure and transcriptional repression.

It is obvious that ligands play a crucial role in determining the co-regulator interactions and the transcriptional activity of the SRs. Interestingly, even agonist and partial agonist -bound SRs are able to recruit the CoRs NCoR and SMRT, as shown e.g. for AR [26-29]. The transcriptional effect in response to ligand binding is in fact assumed to be a dynamic balance between the ligand-induced changes on the SR structure and the expression levels of co-regulatory proteins in a given cell type. This assumption underlies the tissue-selective activity of the SRMs (reviewed in [30]). When bound to a SRM the receptor may adopt a conformation that is intermediate or fluctuates between the agonist and antagonist conformations allowing interaction with both CoAs and CoRs, depending on their relative expression [30].

2.2 Non-steroidal modulators of AR and PR activity

Even though steroids are the natural ligands for SRs (Figure 1), the trend within the SR field is to develop drugs with non-steroidal structures. It is thought that non-steroidal ligands enable improved receptor specificity and reduced side effect profiles, in addition to being easier to modify compared to steroids. At present, only three non-steroidal AR-targeted modulators, the antiandrogens flutamide, nilutamide, and bicalutamide, are available for clinical use, whereas no PR-targeted non-steroidal modulators have yet reached the market. Efforts to identify non-steroidal AR and PR modulators in order to overcome problems related to steroidal compounds have resulted in the discovery of many potential chemotypes as modulators of AR and PR activity. The structure-activity relationships (SARs) for the non-steroidal chemotypes of AR and PR modulators have been thoroughly reviewed recently [31-33].



Figure 1. Chemical structures of the endogenous AR and PR ligands.

The following sections concentrate on non-steroidal AR and PR modulators, examples of which are depicted in Figures 2 and 3, respectively. Discussion is based on structural scaffolds rather than functional activities of the chemotypes. It becomes clear from the literature that functional activities may fluctuate from agonism to partial agonism and even to antagonism as a consequence of merely small structural changes in the ligand scaffold. Receptor binding affinity of a ligand is generally affected by the minor changes made to the ligand structure too. Overall, the biologically measured outcome of a chemical modification is many times ambiguous and hard to understand only by visual comparison of ligand structures. In this work, advanced computational methods have therefore been applied to study the features of non-steroidal AR and PR modulators to be able to identify and quantify the features that are important for high-affinity binding to their respective receptors (Paper I and II).

2.2.1 AR modulators

Flutamide was the first non-steroidal antiandrogen that was approved in the 1980's for treatment of CaP [34]. Its primary active metabolite, hydroxyflutamide (HF) (1 in Figure 2), served as a lead compound in the search for sex organ -selective antiandrogens. This resulted in the discovery of bicalutamide (2) [35], which is currently the preferred antiandrogen for CaP treatment.

SAR research around the aryl propionamide chemotype (bicalutamide derivatives) has been very active within the past decade and resulted in a number of analogues with various functional activities [36-40]. Aryl propionamide derivatives also constituted a major group of compounds for our 3D QSAR analysis (Paper I). It is well-known that good binding affinity and functional activity require electron attracting substituents on the anilide ring (A-ring) and a tertiary carbinol separated by an amide linkage from the A-ring [35-41]. Optimal activity is generally achieved by either nitro or cyano substitution in the *para* position together with a chloro or trifluoromethyl group in the *meta* position of the A-ring. High affinity and activity is also dependent on the correct stereochemistry. Analogues from the bicalutamide series with a sulfonyl (-SO₂-) linkage usually demonstrate antagonist activity, while replacing the sulfonyl with a thio (-S-) linkage generally

introduces agonist activity [40]. Changing the thio linkage in the agonist derivatives into metabolically more stable ether (-O-) linkage, like in S-1 (4) and S-4 (5), preserves the binding affinity and produces higher agonist activity [39]. B-ring substitutions, with a preference for *para* and/or *meta* substitutions, are also important for the functional activity.



Figure 2. Chemical structures of clinical antiandrogens (1-3) and investigational AR modulators (4-14). The *in vitro* function of the ligands is given.

The third clinically available antiandrogen, nilutamide (3), is a hydantoin derivative of flutamide. The SAR of nilutamide template has been studied [42, 43] and many derivatives resembling the agonist DTIB (6) are included in our 3D QSAR analysis (Paper I). Related succinimide (7) and hydantoin (8) derivatives of nilutamide incorporating a bridged bicycle have recently been identified as AR antagonists [44-46]. Reduction of the bulky bicyclic system of the hydantoin analogues and incorporation of a hydroxyl group to appropriate position reverses the functional activity from antagonism to agonism [47]. A highly potent agonist BMS-564929 (9) resulted from optimization of the anilide ring [48].

Linear tricyclic quinolinone derivatives have been identified as AR modulators [49] and many SAR studies have been performed around the quinolinone template (10-13) [50-56]. Compound LG120907 (10) is a potent antagonist with *in vivo* efficacy superior to flutamide [49]. Bisalkylation of the carbon C8 adjacent to the piperidine nitrogen was shown to be responsible for the antagonist activity [51, 53]. Agonist activity was observed for quinolinone derivatives with small alkyl substitutions on C6 or C7, or combinations on C6/C7 or C7/C8. An ethyl group on C6 gives a full agonist LG121071 (11), which was the first reported orally active non-steroidal AR agonist [51]. It has become clear that the substitution pattern including the size and positioning of the substituents on the C-ring plays an important role in determining the functional activity of the quinolinone derivatives [52, 56]. C-ring substitutions converting the piperidine to other heterocycles like pyrrolidine and oxazino (12) rings have been explored and provide additional SAR data for different quinolinone derivatives [31, 57]. Various A-ring substitutions have also been studied [32, 50, 53]. Replacement of the A-ring lactam (LG120907) with e.g. a lactone (13) switches the activity profile from antagonism to agonism. SAR studies of bicyclic quinolinones, obtained by breaking of the C-ring, indicated that monoalkylation of the amino group gives good antagonists whereas bisalkylation generates highly potent agonists like LGD2226 (14) [54, 55].

2.2.2 PR modulators

There are several applications for PR modulators in female health care due to the central role of progesterone (P) and PR in the regulation of female reproductive function. PR agonists are used for e.g. contraception and hormone-replacement therapy. The use of PR antagonists is currently more limited and primarily focus on medical termination of pregnancy although new clinical applications are emerging (reviewed in [58]). All clinically used PR modulators have a steroidal structure.

During the past decade quite a few non-steroidal chemotypes have been discovered as PR modulators [31, 33], some of them similar to the AR chemotypes. The resemblance between the AR and PR modulators is not surprising because of the structural similarity of the ligand-binding sites of active AR and PR structures (see next chapter, Figure 7, and Table 1). Tetrahydropyridazine-based chemotype, exemplified by agonist RWJ 60130 (1 in Figure 3), was among the early non-steroidal chemotypes for PR modulators [59-61]. Different substitution patterns on the aromatic moieties (2) affect both binding affinity and functional activity [31]. Compounds of a related pyrazolinebased chemotype (3) function as antagonists [62].

One of the most actively studied chemotypes has been the quinoline family of compounds [63-73], a series of which was studied in this work with 3D QSAR analysis (Paper IV). The 6-aryl-dihydroquinoline LG001447 (4) was demonstrated to possess antagonist activity, which was increased by electron-withdrawing substituents at the *meta* positions of the 6-aryl moiety [65]. Conformationally constrained analogs, like the dihydrochromenoquinolines (5), show antagonist activity as well [71]. Nevertheless, lipophilic substituents on the 5-position generate potent agonists (6) (e.g. [63, 73]).



Figure 3. Chemical structures of investigational PR modulators. The *in vitro* function of the ligands is given.

Modifications to the structural elements at either end of the dihydroquinoline template influence the functional activity of the compounds. Bioisostere replacement of the 6-phenyl for a 6-thienyl group results in antagonistic analogues (7) [70]. Replacing the quinoline template itself with a benzoxazine template (8) generally produces potent agonists when the substituents at the 2-position are smaller alkyl moieties [74]. Larger 2-substituents switch the activity to antagonism. 2-Thiocarbonyl derivatives (9) of the 6-aryl benzoxazine class of compounds exhibit agonist activity [75] whereas 2-carbonyl analogues (10) are potent antagonists [76]. The antagonist function by introducing a

5-cyano-2-pyrrole group to replace the 6-aryl group (11) [77]. An even more potent agonist, Tanaproget (12), is obtained by combining the 5-cyano-2-pyrrole with the 2-thiocarbonyl substituent on the benzoxazine template [78]. Together the above-described examples clearly show that subtle structural modifications to a specific chemotype of PR modulators may cause a complete reversal of functional activity.

Tanaproget is the only non-steroidal PR modulator that has reached clinical trials for contraception and hormone-replacement therapy so far, but was pulled from all the trials in 2006 due to its side effect profile [33].

2.2.3 Selective receptor modulators

With SRMs it is possible to achieve tissue-selective activity profiles and such modulators are the ultimate goal in the development of SR ligands (potential therapeutic applications of selective AR modulators (SARMs) are reviewed in [9, 79]). A few SARMs with nonsteroidal structures have recently been discovered [48, 54, 80, 81]. The above introduced compounds S-4 (5), BMS-564929 (9), and LGD2226 (14) are orally available SARMs that have similar *in vivo* pharmacological profiles. They exhibit full agonist activity in the skeletal muscle and bone but only partial agonist activity in the prostate in rodent models [48, 80, 82]. In the prostate they basically function as competitive antagonists in the presence of full agonists like the endogenous DHT [83]. These SARMs thus behave as strong anabolic agents with weak androgenic activity. SARMs with such a profile could be suitable for treatment or prevention of age- and disease-related muscle wasting and osteoporosis. Additional applications that could benefit from the low androgen activity on prostate include hormone-replacement therapy in aging males, but maybe even CaP [7, 32]. S-4, BMS-564929, and LGD2226 have advanced into clinical trials but thus far none of them or other clinical SARM candidates have entered the market [7, 9, 10].

Selective PR modulators (SPRMs) have also been identified [33, 58]. All the currently reported potential SPRMs are steroidal compounds however and there are no non-steroidal SPRMs in clinical trials.

2.3 Structural and functional organization of steroid receptors

The SRs share a conserved modular domain organization comprising of an N-terminal domain (NTD), a DNA-binding domain (DBD), a hinge domain, and a C-terminal ligand-binding domain (LBD). Despite variability in sequence conservation of the domains among the SR family members, their functional activities are preserved (Figure 4). In the following sections, the domain structures and their functions are introduced in more detail. The focus is on the LBD that via ligand binding regulates the SR activity.



Figure 4. Structural and functional organization of the SRs. Sequence identities (adapted from [84]) between the human SRs are shown in relation to AR (percentage values) and the length of each sequence is indicated. The general transactivation functions 1 and 2 (AF1 and AF2), including the helices (H) forming the AF2, are indicated within their respective structural domains. The Figure was adapted with permission from [6]. Copyright 2005 American Chemical Society.

2.3.1 N-terminal domain

For the least conserved NTD there is no high-resolution structure available for any member of the SR family. Neither is the function of this highly divergent domain fully understood. Nevertheless, the NTD is known to be important for transcriptional activation and it harbors a ligand-independent transactivation function termed activation function 1 (AF1). A large number of co-regulatory proteins, both CoAs and CoRs, interact with the NTD via AF1 sequences (reviewed in [22, 85]). Distinguished from other SRs, AR exhibits strong AF1 activity as revealed by the weak transcriptional activity of an NTD-deletion mutant of the receptor [86, 87]. This is explained by the weak recruitment of CoAs by the ligand-dependent activation function 2 (AF2) on the AR LBD compared to other SRs (see sections 2.3.3 and 2.4.1) [88].

Structural characterization of AF1 polypeptides by means of biochemical, biophysical and bioinformatics methods has revealed that the region is flexible and mostly unstructured in aqueous solution, e.g. in AR only 16% of AF1 is α -helical and 24% β structure [89, 90]. The AF1 has been shown to adopt a more folded structure upon binding to target proteins as is shown for AR AF1 when interacting with the transcription factor IIF (TFIIF) from the basal transcription machinery [89, 90]. Based on the observations, an induced folding model for transforming the unstructured NTD into more folded conformation upon interaction with different factors has been proposed [85, 90]. The model suggests that additional binding surfaces on NTD are introduced for further protein-protein interactions upon induced NTD folding. Creation of new binding surfaces following TFIIF-induced folding of AR AF1 has been supported by enhanced binding of a CoA protein from the p160 family to AR AF1 [89].

2.3.2 DNA-binding and hinge domains

The transcriptional regulation of the SRs is dependent on their interaction with the regulatory DNA sequences of their target genes. The DBD located at the center of the protein sequence is responsible for the recognition and binding to the appropriate HREs. Reflecting the common need for SRs to interact with the DNA, the sequence and structure of the DBD is well conserved (Figure 4).

Currently, the crystal structure of DBD as a dimer in complex with DNA (Figure 5) has been solved for all SRs except for MR [91-94]. The core of the DBD consists of two zinc-finger motifs, where each of the zinc atoms is tetrahedrally coordinated by four highly conserved cysteine residues [95]. The zinc-fingers fold into a globular domain comprising two α -helices. The first α -helix, which contributes to the N-terminal zinc-finger, regulates the recognition of specific HREs via amino acids within a region termed P-box [91]. The P-box residues make nucleotide-specific contacts when the first helix, also known as the recognition helix, is positioned within the major groove of DNA. The second α -helix that contributes to the C-terminal zinc-finger packs against the recognition helix in a perpendicular fashion and forms hydrophobic interactions to stabilize the complex. DNA-binding induces receptor dimerization, which in the DBD is mediated by residues within the C-terminal zinc-finger [91, 96].

Following the highly conserved core of the DBD there is a variable C-terminal extension, which is suggested to contribute to DNA-binding affinity and specificity of SRs through interactions with DNA regions outside the HREs [92]. The C-terminal extension is part of the flexible hinge domain, which serves as a linker region between DBD and LBD. The hinge domain houses a nuclear localization signal making it important for nuclear targeting of SRs [87, 97, 98].



Figure 5. Crystal structure of AR DBD dimer co-crystallized with a stretch of DNA (PDB 1R4I) [94]. (A) Side view. (B) Top view. The highly conserved DBD consists of two zinc-finger motifs, which fold together to form a globular structure comprising two perpendicularly oriented α -helices. Zinc (Zn) atoms are shown as red spheres.

2.3.3 C-terminal ligand-binding domain

The LBD is a multifunctional domain, which is responsible for ligand recognition but contributes also to dimerization and co-regulator interactions. Despite the relatively low sequence conservation among the LBDs of different SRs (Figure 4), numerous 3D structures of SR LBDs show that they adopt a similar α -helical sandwich structure that generally consists of 12 α -helices (H) and 1-2 small anti-parallel β -sheets arranged into three layers (Figure 6) [99, 100]. Conventionally the helices are numbered from H1-H12, even if some helices are unidentifiable in some SRs, such as H2 in AR and PR LBD structures [101, 102]. The central layer, formed by helices H4, H5, H8, and H9 and the first β -sheet, is flanked from one side by helices H1-H3, and from the other side by helices H6, H7, H10, and H11 and the second β -sheet. The central layer is missing from the lower part of the domain structure, generating a cavity that serves as the LBP (Figure 7). The bottom half of the domain has shown to be intrinsically flexible to support binding of ligands with different sizes. Dynamic behavior is observed especially for H12, which is able to adopt various positions depending on the bound ligand. The mobility of H12 has been connected to the activity of the SRs and will be discussed in more detail in sections 2.4.1 and 2.4.2. In the active, agonist-bound LBD conformation, as shown for AR in Figure 6, H12 is placed over the LBP like a lid and it spans all three helical layers and stabilizes the active conformation. Additional stabilization is provided in AR and PR LBDs by a second β -sheet formed by the most C-terminal region following H12.



Figure 6. Crystal structure of the agonist-bound AR LBD representing the conserved threelayered α -helical sandwich structure of SRs (PDB 1137) [103]. A) Front view. B) Side view. The layers are shown with different colors (blue, green, and yellow). LBP resides within the lower half of the central layer and is enclosed by the helix H12 (red). DHT agonist is shown in stick representation. Helices H1-H12 are marked; H2 is missing from the AR LBD.

Activation function 2

On the surface of the active SR LBD structure there is a hydrophobic cleft formed by residues from helices H3, H4, and H12 (Figure 6 and 8A) that mediates interactions to CoAs containing a common LXXLL motif called NR box [104], where X represents any amino acid. This surface cleft is known as the ligand-dependent AF2 because the activity of AF2 is regulated by the binding ligand. The AF2 surface forms a complementary binding interface for the LXXLL motif, which adopts an amphipathic α -helical structure with the leucines interacting with the hydrophobic floor of the AF2 cleft [105]. The hydrophobic center of the AF2 surface is flanked by patches of opposite charge at the ends of the cleft [106]. Among these charged residues there is a particularly important charge clamp of conserved lysine and glutamate residues [100] from helices H3 and H12, respectively, that stabilize CoA binding by hydrogen bonding to the NR box residues [105].

Ligand-binding pocket

The LBP resides in the interior of the LBD just underneath the central helical layer. Structurally the LBPs of different SRs are closely related owing to the similar chemistries of the natural steroidal ligands. This buried cavity is lined by residues protruding from six α -helices (H3, H4, H5, H8, H11 and H12) and the β -strand located between H5 and H6. Most of the LBP residues that envelop the hormone are hydrophobic in character and interact with the steroid body through van der Waals interactions. Hydrophobic interactions have consequently an essential role in ligand binding. Only a few polar residues are presented to the binding site surface at opposite ends of the LBP. In AR, the two polar patches are formed by residues Q711 and R752 at one end and by residues N705 and T877 at the other end of the cavity, while the corresponding residues in PR are Q725 and R766, and N719 and C891, respectively (Figure 7A-B). They anchor the binding ligands via hydrogen bonding networks and are thus important for both binding affinity and specificity.

All steroidal and non-steroidal AR and PR ligands co-crystallized with their respective receptor LBDs have been shown to form hydrogen bonds to the polar patch of residues Q and R either directly or indirectly via a structural H₂O molecule [47, 102, 103, 107-110]. Hydrogen bonding to the other polar area is more varied. The 17-hydroxyl group of the steroids T and DHT hydrogen bonds to both N705 and T877 of AR [103, 108], while no comparable polar interactions occur between the 17-acetyl group of P and the residues N719 and C891 of PR [102]. The non-steroidal AR or PR ligands have been shown to form a single hydrogen bond to the N residue [47, 107, 109] or to make no hydrogen bond contacts at this end of the LBP [110] when co-crystallized with their respective receptors.

The crystal structures determined to date for the wild-type (WT) AR and PR LBDs in complex with various synthetic agonists, including steroidal and non-steroidal molecules, have demonstrated that the LBP is remarkably flexible (e.g. [107, 111, 112]). The LBP volume can be adapted to accommodate ligands with various structures and sizes much larger than the natural steroids without disrupting the active AF2 surface. This is primarily achieved by altering the side chain conformations of only a few LBP residues and/or by small changes in the protein main chain [107, 111, 112]. In AR, the mobility of W741 and M895 is important for accommodation of larger ligands (Figure 7C). The volume of the AR LBP has been shown to increase from about 600 Å³ (DHT-bound) to over 800 Å³ upon binding of a bulky steroid agonist EM5744 mainly by reorienting the side chains of W741 and M895 [111]. A similar increase in volume is observed when the non-steroidal agonist S-1 (4 in Figure 2) [39, 80] is bound to the AR LBP [107]. Besides affecting the volume of the LBP, such ligand-induced adjustments in the LBP structure generate additional non-polar and polar interaction sites that are important for the ligand binding affinity.



Figure 7. The LBP structures of AR LBD (A and C) and PR LBD (B) depicted in comparable orientations. The residues shown are the six residues that differentiate AR LBP from PR LBP (AR residues M749, M780, Q783, F876, T877, L880), the polar residues forming hydrogen bond interactions with the ligands (N705, Q711, R752, T877), and the two residues important for accommodation of larger ligands to the AR LBP (W741 and M895). Comparison of figures A and C reveals the flexibility of the AR LBP: reorientation of W741 and M895 increases the LBP volume drastically and enables binding of larger ligands to AR LBP. A) (PDB 1137) [103], B) (PDB 1A28) [102], C) (PDB 2AXA) [107].

The LBP serves also as the binding site for antagonists. From a mechanistic point of view, ligands become antagonistic when the LBP is unable to adapt its volume to the binding ligand. This push introduces changes in the tertiary structure of the LBD that inhibit the receptor from functioning. However, the mechanistic viewpoint is too simple to represent the whole truth. For example, the AR antagonist HF is comparable in size to the natural androgens, indicating a different structural basis for antagonism. Currently, there are no crystal structures available for AR antagonists in complex with the WT AR. The structural modifications upon antagonist binding are discussed in more detail in section 2.4.2.

Understanding receptor-ligand interactions and receptor flexibility is important when one attempts to identify new target-selective hit compounds for drug development. The design of new AR-selective modulators is tricky because the LBPs of SRs are very similar and share common ligand-binding features. PR demonstrates the most similar LBP structure to AR: the LBP of PR differs only by six residues from the LBP of AR (Table 1). The design of AR antagonists for CaP treatment is additionally complicated by LBP mutations that are frequently found in CaP patients (Table 1) [5]. The mutated ARs, T877A and W741C variants in particular, have the ability to bind and be activated by a broadened spectrum of ligands including a variety of hormones but also antiandrogenic drugs [107, 113-115]. T877A mutant confers antagonist-to-agonist conversion of the non-steroidal antiandrogen HF and the steroidal antiandrogen cyproterone acetate (CPA) [107, 115-118]. W741C/L mutations result in agonist activity of bicalutamide [107, 113, 117]. Design of AR antagonists that retain their activity in several mutated AR variants besides the WT AR is a major goal for CaP treatment [8].

AR LBP - PR LBP	CaP (AR LBP)
residues	mutations
M749 - L763	L701H
M780 - F794	W741C
Q783 - L797	M749I
F876 - Y890	T877A/S
T877 - C891	L880Q
L880 - T894	M895V

Table 1. Selective AR-targeted ligands are structurally difficult to design because 1) there are only six residues that differentiate AR LBP from PR LBP in the hormone-bound active form (left column) and 2) AR LBP mutations that broaden ligand specificity are frequently detected in CaP patients (right column).

2.4 Structural basis for activation and inactivation of AR with ligands

2.4.1 Activation by agonists

A functional AF2 surface on the LBD that enables CoA recognition is essential for the transcriptional activation of SRs (Figure 8A-B). AR appears to differ from the other SRs in terms of CoA recruitment. Compared to other SRs the AF2 in AR LBD interacts poorly with the LXXLL motif containing CoAs and displays therefore weak AF2 activity [88, 119, 120]. Instead, the AF2 of AR prefers binding of phenylalanine-rich FXXLF

motifs that are found within the AR NTD and in a subset of AR interacting CoAs [121-124]. The N-terminal ²³FQNLF²⁷ sequence has been shown to mediate direct, interdomain interaction between the NTD and LBD (N/C interaction) in an androgendependent manner [121] and this interaction is proposed to have functional importance in the regulation of some, but not all, AR target genes [125]. Due to the N/C interaction the AF2 surface becomes unavailable for interaction with LXXLL-containing CoAs [119], thus making the NTD AF1 the main activation region responsible for CoA recruitment in AR [126, 127].



Figure 8. Crystal structures of SR LBDs illustrating the location of the functional AF2 surface (A) and the dynamic behavior of helix H12 (green) (B-D) upon ligand binding (see text for details). Ligands are shown as spheres. A) **AR-agonist complex** (PDB 1137) [103]. The helices H3, H4, and H12 forming the AF2 surface are highlighted in green. The charge clamp residues are shown as blue (lysine) and red (glutamate) sticks. B) **AR-agonist-CoA complex** (PDB 1XOW) [120]. Agonist-bound AR LBD structure with a CoA peptide (light blue) interacting with the AF2. C) **ER-antagonist complex** (PDB 3ERT) [105]. Antagonist-bound structure of ER LBD where H12 has swung onto the surface of AF2 thereby inhibiting CoA binding to AF2. D) **PR-SPRM-CoR complex** (PDB 2OVM) [128]. SPRM-bound PR LBD structure with H12 displaced from the active position and CoR peptide (magenta) bound to AF2.

The structural basis for the distinct preference of AR AF2 for FXXLF motifs compared to other SRs has recently been elucidated by analyzing the contacts formed between the agonist-induced functional AF2 surface and various peptide fragments conforming to either FXXLF or LXXLL motifs [120, 129, 130]. The experimentally determined structures show that the general mechanisms for binding the FXXLF motifs are similar to those used in binding LXXLL motifs by other SRs [105, 131]. The common hydrophobic and electrostatic interactions are formed when the α -helical peptide fragment docks into the AF2 cleft. Nevertheless, the AR AF2 is unique in being able to rearrange the AF2 surface into a deeper cleft that can accommodate the bulky aromatic side chains of the FXXLF motif, thus providing enhanced hydrophobic interactions. Hydrogen bonds are formed to both charge clamp residues (K720 and E897) by the FXXLF-containing peptides, while a shift in the position of LXXLL-containing peptides enables interactions with only one charge clamp residue (K720). Together, these differences may explain why the AR LBD binds the FXXLF-containing peptides with higher affinity than the LXXLL-containing peptides [106, 120, 130].

2.4.2 Inhibition by antagonists

The first antagonist-bound SR structures were determined for ER and provided insight into the structural basis of antagonism [105, 132]. The structures revealed that antagonist ligands with a bulky extension sterically interfered with positioning of H12 into its active orientation as a consequence of not being able to fit the extension within the binding cavity. Instead of adopting the agonist conformation, H12 swings onto the surface of AF2 (Figure 8C). This relocation is feasible because H12 of ER harbors a recognition sequence (⁵⁴⁰LLEML⁵⁴⁴) that resembles the LXXLL motif of CoAs and is able to mediate interactions with the AF2 cleft. Because the AF2 becomes engaged in binding of the repositioned H12, there is no functional AF2 surface available for CoA recruitment. The AF2 antagonist model, as it is called, was proposed as the general mechanism for SR antagonism [99] and has subsequently been observed also in GR [133].

In AR, the sequence corresponding to ⁵⁴⁰LLEML⁵⁴⁴ of H12 in ER is ⁸⁹⁵MAEII⁸⁹⁹. Whether this H12 sequence interacts with the AF2 surface to inhibit AR activity by blocking N/C interaction and/or CoA recruitment is unknown at the moment. The structural basis for AR antagonism is unclear, because no antagonist-bound structures of WT AR LBD are available. In the lack of proper antagonist AR structure, insight into the structural mechanism for AR antagonism has been sought by co-crystallizing AR antagonists with AR LBD mutants that are known to induce agonist activity for the antagonists, thus permitting the complex to be crystallized in the agonist conformation [45, 107, 118, 134]. AR antagonists have been crystallized in complex with either T877A (HF, CPA, and isoindoledione-based compound) or W741L (bicalutamide) variants of AR LBD [45, 107, 118, 134]. Comparison of the CPA-complexed T877A structure with the DHT-bound WT structure demonstrated a steric overlap between T877 and the bulky antagonist. The clash is expected to induce unfolding of the C-terminal end of H11 in the WT AR, ultimately leading to displacement of H12 and disruption of a functional AF2 surface. Indeed, dislocation of the loop linking H11 and H12 was demonstrated in the CPA-T877A AR structure [118]. The isoindoledione-based antagonist is assumed to inhibit AR activity in a similar fashion [44-46]. Interestingly, binding of the smaller HF antagonist seems not to be hindered by T877, suggesting a different structural origin for HF-based antagonism. Comparison of the bicalutamidebound W741L AR with the WT AR structure bound to the structurally similar agonist

S-1 (Figure 2) [107] revealed that the sulfonyl linker of bicalutamide would sterically prevent positioning of M895 at the N-terminal end of H12 [107, 134]. The hindrance to accommodate M895 is likely to induce unfolding of H12, which offers a structural explanation for bicalutamide-based AR antagonism. Mutating the methionine residue into a smaller threonine (M895T) restored agonist activity for bicalutamide *in vitro* confirming the role of M895 in the bicalutamide-initiated antagonism [107]. Taken together, these structures provide a lot of data on the primary structural source of AR antagonism but do not reveal the whole picture. Whether the predicted AR unfolding only results in disruption of the functional AF2 surface or also in movement of H12 to the CoA-binding cleft cannot be deciphered based on the available structural data.

Other mechanisms for AR antagonism have been proposed including recruitment of CoR proteins NCoR and SMRT [26-29]. A recent structural study on PR confirmed that when bound to a steroidal SPRM, H12 was blocked from assuming the active orientation by a CoR peptide binding to a region of the LBD that partly overlaps with the AF2 surface (Figure 8D) [128]. Due to dislocation of H12, a longer groove is formed between helices H3 and H5. The interaction of CoR peptides with this groove is mediated by a consensus motif LXX(I/H)IXXX(I/L) termed the CoRNR box [21, 135]. The CoRNR box resembles the NR box of CoAs but forms a longer helix thereby preventing the active positioning of H12. Comparison of H12 sequences in various NRs/SRs indicates that those NRs that are known to display poor CoR binding possess a H12 sequence resembling the CoRNR box sequence, whereas NRs that exhibit good CoR binding have a shorter H12 with less similarity to the CoRNR box sequence [136]. As shown for ER, there is an internal CoRNR box substitute within H12 that interacts with the AF2 surface itself and acts as a barrier to CoR binding (AF2 antagonist model) [105, 132, 136]. Other SRs have a shorter H12 with an incomplete CoRNR box-like sequence, which is unable to oppose CoR binding [136]. Given that CoR peptide interaction was demonstrated for PR and that AR and PR possess H12 of equal lengths, it is possible that AR antagonism is also mediated by CoR binding to AR LBD. Nevertheless, the structural basis for AR antagonism remains unclear.

3 Aims of Study

The ultimate goal of the study was to discover non-steroidal AR modulators with novel structural scaffolds and selectivity over PR by means of computational modeling, which relied on publicly available data. To reach the goal both ligand-based and receptor-based methods were used in the studies reported in Papers I-IV with the following aims:

Paper I – The aim was to combine molecular docking and 3D QSAR analysis to develop a computational approach that identifies, by means of molecular field data, structurally important features for binding affinity of a diverse set of non-steroidal AR ligands. Docking was used to predict the bioactive conformations and to generate the ligand alignment needed for 3D QSAR analysis.

Paper II – The aim was to find novel non-steroidal scaffolds as hit compounds for ARtargeted lead development by receptor-based VS. Methods developed in the 3D QSAR analysis of AR ligands were integrated into the VS protocol because an additional aim was to find out whether the derived 3D QSAR model was able to contribute to detection of such AR hits.

Paper III – The aim was to characterize the functional activities of compounds from novel AR hit classes and to use advanced docking that incorporates protein flexibility to explain the experimentally determined activity profiles.

Paper IV – The aim was identical to that in Paper I, but now the study was carried out on a set of non-steroidal PR ligands. The underlying idea is to integrate the results of the two 3D QSAR analyses as part of a future VS protocol, whereby their ability to detect ligands with selectivity to AR over PR and vice versa could be assessed.

4 Methods

Computational methods that can be used in drug development processes are at the heart of this work and they are the focus of the following chapter. The concept of VS is introduced shortly, while a more comprehensive presentation is given on molecular docking and 3D QSAR analysis, which were the main computational tools applied in this thesis and also integrated in the receptor-based VS for AR hits. *In vitro* experiments, including binding affinity determination and functional profiling, were combined with computational methods to create a coherent picture of the ligand-induced modulation of receptor function from both computational and experimental viewpoints.

4.1 Structural and biological data (Papers I-IV)

Publicly available structural and biological data for proteins and ligands were used throughout this work. The crystal structures of AR LBD and PR LBD were obtained from the Protein Data Bank (PDB) [137]. The PDB codes for the used AR and PR structures are 1GS4 and 1I37, and 1ZUC, respectively. The structural and biological data for AR [36, 38, 40, 42, 43] and PR ligands [64, 65, 70-72] were collected from several publications, which report experimentally comparable binding data for the respective ligand series. A commercial collection of over 200 000 compounds provided in the Gold Collection of Asinex Europe (http://www.asinex.com) was used as the VS database.

4.2 Virtual screening (Paper II)

VS encompasses a variety of computational methods that are used to evaluate large chemical databases in order to discover new ligands for a drug target of interest [11-18]. Typically, the objective is to identify hits with structurally novel scaffolds that can be used as starting points for development of new, patentable lead molecules. VS processes usually apply sequential computational phases to reduce the search space for potential drug candidates by filtering out compounds unlikely to interact with the target. In the

initial phases there is a need for computationally inexpensive filters because databases may comprise millions of compounds, most of which have little chance of being hits for a specific drug target [11, 138]. Such initial filters are rather unspecific for a particular target and consider simple compound properties like molecular weight, clogP, and number of hydrogen bonding groups and rotatable bonds [138, 139]. A common practice is also to use substructure queries to filter out compounds that contain moieties known to be problematic, like highly reactive functional groups [138]. When the database decreases in size, more sophisticated and computationally demanding VS methods with target-specificity can be applied. The applicable methods take advantage of information on ligands known to interact with the target protein (ligand-based VS) or 3D structure of the target itself (receptor-based VS). Prioritization of database compounds requires a means to measure the likelihood of compounds being active towards the given target. Computed similarities (e.g. using the Tanimoto coefficient [140] as a measure of similarity) between the structures of the database compounds and the known bioactive structure(s) are used to prioritize compounds in ligand-based similarity methods [15], while estimates of binding affinities computed by so called scoring functions are employed in receptor-based docking methods [141-143]. There is often a need for additional data analysis together with visual inspection of the top-ranked compounds before the final decision on compounds to be selected for experimental testing is made.

In this thesis, a receptor-based VS protocol was developed to discover chemically novel hit compounds for AR. Initial filtering protocol, which is described in more detail in the Results and Discussion chapter, was applied to reduce the screening database into a more manageable size prior to docking into the AR LBP. The pre-generated 3D QSAR model of AR ligands (Paper I) together with three common scoring functions were then applied in rank ordering the docked compounds according to their predicted binding preferences for AR and in selection of compounds for *in vitro* assessment of binding affinity.

4.3 Molecular docking (Papers I-IV)

It is well-known that the activity of a drug is attained through its interaction with the active site of the target protein. In drug discovery, one therefore attempts to find compounds that match the geometry and chemistry of the protein binding site. The best way to study intermolecular interactions is to solve the structure of a protein-ligand complex experimentally, e.g. by X-ray crystallography. Experimental structure determination is however an impractical approach in the initial phases of drug discovery where large amount of compounds need to be screened. To this end, computational approaches have become very useful tools in rational drug design. Molecular docking is a broadly used computational method for predicting ligand conformation and orientation in the protein

binding site. It is therefore a useful strategy for 1) studying the binding mode of a single ligand, 2) finding new hit compounds with good complementarities for the target in virtual database screening, and 3) aligning (superimposing) ligands in 3D space for e.g. 3D QSAR analyses [144-147]. During this work, docking was used for all these three purposes.

There is a large number of different docking programs available (reviewed in [141, 142, 148]). A common feature for all of them is a combined use of a search algorithm for sampling ligand binding modes and generating ensembles of docking solutions called poses, and a scoring function for evaluation of the different poses to determine the most likely biological binding mode for each ligand.

4.3.1 Docking search algorithms

Early search algorithms treated both the protein and the ligand as rigid bodies [149]. Today, the most commonly used docking programs introduce ligand flexibility into the docking protocol [148]. Because ligand binding often induces conformational changes in protein binding site (e.g. [107, 111, 128]) docking programs that consider protein flexibility as well are increasing in number.

The search algorithms used to explore the conformational space of the ligand are classified into systematic, stochastic, and simulation methods [141, 142]. Systematic search algorithms seek to explore all conformations of a ligand. The search space increases rapidly as the degrees of freedom increase for a ligand. In order to avoid combinatorial explosion, many search algorithms employ an incremental construction approach where the ligand is initially fragmented into rigid and flexible components, which are then sequentially linked back together to grow the ligand within the binding site. The ligand growth is started from the rigid anchor fragments that are placed within the binding site by rigid body docking. The flexible fragments are then linked to the anchor one at a time while systematically scanning the conformational space of each added fragment. The success of this approach depends whether the anchor fragments form good binding interactions or not. FlexX [150] and DOCK 4.0 [151] are examples of docking programs that use incremental construction algorithms for docking.

Stochastic search algorithms explore ligand flexibility by introducing random changes in an iterative fashion to a single ligand or a population of ligands. Monte Carlo (MC) methods (e.g. AutoDock [152]) and genetic algorithms (GAs) (e.g. GOLD [153]) are two widely used random search approaches. At each iteration in a standard MC search the ligand conformation in the binding site is changed by randomly assigning new torsion angles or cartesian coordinates. Following energy minimization Metropolis criterion is used to determine whether the resulting ligand structure is retained [141, 142]. A large number of iterations are needed to ensure detection of low energy conformations.

To enhance the exploration of the conformational space and to increase the chance of reaching the global minimum energy conformation, simulated annealing MC performs repeated MC cycles at gradually reduced temperatures with the first cycle simulated at a high temperature. In GA methods a population of initial ligand conformations, or possible solutions, are generated. Possible solutions are encoded into data structures called chromosomes, which undergo genetic operations such as mutations and crossing-over between two parent chromosomes in order to create a new generation of solutions. The GA thus mimics the process of evolution and introduces even evolutionary pressure by biasing the selection of parent chromosomes towards the fittest individuals of the population based on the assigned fitness scores. Due to the evolutionary origin of GAs the results are influenced by the initial population size, mutation and cross-over rates as well as number of evolutionary cycles that are ran [141].

Ligand flexibility may also be explored by simulation methods such as molecular dynamics (MD) and energy minimization. The quality of the results in simulation methods depends heavily on the starting conformation of the system under study because these algorithms tend to get stuck in local minima due to their inability to cross highenergy barriers [141, 142, 148]. Use of multiple starting conformations and increased temperatures in the simulations are examples of approaches used to circumvent the problem [141, 142, 148]. Contrary to MD, energy minimization is rarely used as an independent search strategy but rather as a complement to other search algorithms such as incremental search methods [151].

Inclusion of protein flexibility into docking adds to the complexity of the system as the degrees of freedom increase significantly. MD simulations can be used to introduce protein flexibility into docking but such methods are computationally expensive and are therefore restricted to studies on a small set of ligands. Thus, various approaches requiring less computational power have been developed to deal with protein flexibility in docking [154]. Treatment of protein flexibility is usually restricted to binding site residues. The simplest method is soft docking to rigid protein, which in an indirect way accounts for small adjustments in protein structure by allowing some steric overlap between the ligand and protein. Other methods allow for side chain flexibility e.g. by sampling the positions of terminal hydrogen bonding groups to optimize hydrogen bonding networks or by sampling side chain conformations using rotamer libraries. For more advanced approaches available for treating protein mobility in docking simulations the reader is advised to the review of Alonso *et al.* [154].

4.3.2 Docking scoring functions

Scoring functions have a dual role in molecular docking. First, when docking a single ligand a scoring function is needed to evaluate the fitness of the generated poses to be

able to distinguish between the correct and incorrect binding conformations. Second, when docking a series of ligands, as in VS, a scoring function is additionally required to rank order the ligands so that active compounds can be distinguished from inactives. Scoring functions try to accomplish these tasks by estimating the total free energy change upon binding of a ligand to its target [141, 143].

Molecular recognition is a highly complex phenomenon that is driven by a number of factors – hydrophobic effect, van der Waals interactions, electrostatic interactions, hydrogen bonding interactions, desolvation and entropy changes among others – each of which influence the total free energy of ligand binding to its target (i.e. the binding affinity) [141, 143]. Estimating the binding free energies accurately requires consideration of all contributing factors, which makes the calculations computationally complex and timeconsuming. As a consequence, all scoring functions used in docking simulations make simplifying approximations in order to speed up computational estimation of binding free energies, thereby enabling screening of vast amount of ligands against a target protein.

A wide range of scoring functions has been developed owing to their impact on the quality of docking results. They are commonly categorized into force field -based, empirical, or knowledge-based scoring functions [141-143]. Binding energy estimates using the scoring functions based on molecular mechanics force fields are obtained by summing up contributions from electrostatic and van der Waals interaction energies for the protein-ligand complex. A weakness of standard force field -based functions is that they ignore most of solvent and entropic effects, because the functions were originally formulated to model binding energies or enthalpies in the gas phase. Empirical scoring functions estimate binding energies as a sum of several uncorrelated terms, which are functions of the ligand and protein coordinates and describe specific interactions contributing to the binding event such as hydrogen bonding, hydrophobic interactions, and entropic contributions. The weights of each term in the function are derived from a regression analysis using a training set of protein-ligand complexes and associated experimental binding energies. Consequently, the major limitations of empirical scoring functions are related to the experimental data used to train the function: successful predictions of binding free energy are only expected for compounds making similar interactions as ones used to train the scoring function. Knowledge-based scoring functions are founded on statistical analyses of interatomic contact frequencies observed in a database of experimental protein-ligand complex structures. The premise of such scoring functions is that interactions occurring with high frequency in the experimental structures are assumed to be energetically favorable while low occurrence frequency reflects energetically unfavorable interactions. Summing up the many interatomic contact contributions gives the binding estimates of knowledge-based scoring functions. One limitation on deriving such scoring functions is their dependence on information encoded in limited sets of experimental protein-ligand complex structures.

The performance of a given docking program is influenced by the search algorithm and scoring function being used as well as by the chemical nature of the ligand and the protein active site. Comparative evaluations have shown that no single docking program performs well on all targets and that success is largely case dependent [155, 156]. A general conclusion is that many currently available docking search algorithms are capable of thorough sampling of ligand conformations/orientations and reproducing the crystallographically determined ligand binding mode with adequate accuracy [155, 156]. On the other hand, there is room for improvement for scoring functions in the detection of the experimentally determined, "true", binding conformation as the top-scoring pose from a pool of decoy conformations [155, 156]. An additional problem of scoring functions entails binding affinity prediction: there is at best only a moderate correlation between the predicted and the experimentally determined ligand binding affinities [156-158]. Together these weaknesses in scoring cause a variety of challenges in hit identification and lead optimization by computational methods. Enhanced performance has been achieved by using different scoring functions for fitness evaluation during docking and for final ranking of the generated ligand poses [159]. Consensus scoring schemes, which combine the predictions of two or more scoring functions, have on several occasions suggested to augment the identification of true binding conformations as well as true actives from databases as a result of improved scoring accuracy [158, 160-162]. The rationale for consensus scoring is that the reliability of a prediction improves if the same prediction is made by a number of scoring functions. Each of the scoring functions incorporated should, however, have relatively high performance individually and have quite different scoring characteristics in order to improve the probability of identifying actives [163]. Nevertheless, automated consensus scoring is not yet possible in our opinion; an intelligent filter, a scientist, is needed to interpret the computational results for the best outcome.

4.3.3 Docking and scoring in this thesis

In Papers I, II, and IV and Paper III, the docking programs GOLD [153] and Glide [164], respectively, were used. Both of the docking algorithms treat ligands as flexible structures. Protein flexibility is disregarded in Glide, while GOLD introduces partial protein flexibility by allowing optimization of -OH group of serine, threonine, and tyrosine, as well as -NH₃⁺ group of lysine. GOLD was used to predict the bioactive binding conformations and to derive an alignment for series of known non-steroidal AR (Paper I) and PR ligands (Paper IV) that were subjected to 3D QSAR analysis. Additionally, GOLD was used as the docking tool in VS for the identification of new hit compounds for AR (Paper II). The Induced Fit Docking (IFD) protocol of Schrödinger [165, 166], which samples both ligand and protein flexibility by combining Glide docking with pro-

tein side chain predictions of binding site residues with the program Prime, was used for a more detailed analysis of ligand-induced structural changes in AR (Paper III).

GOLD applies a GA to optimize ligand conformations in complex with the target binding site. The GA of GOLD utilizes three types of genetic operators: mutations for introducing random changes, cross-over for combining features from two parent chromosomes, and migration for exchanging entire chromosomes between subpopulations, which are used in GOLD instead of a single large population. Each operator is assigned a weight that determines its probability of selection. Initially, random subpopulations of chromosomes are generated and each member is assigned a fitness score. An operator is then randomly selected based on operator weights and applied to parent chromosome(s) chosen with a slight selection pressure towards the best members of the population. The fitness of the child chromosome(s) is evaluated and if not already present in the population, it replaces the least-fit member of the population. The GA run terminates after a predefined number of genetic operations are performed or when no improvement to the docking solution is achieved. The fitness of the docking solutions is evaluated using a force field -based scoring function, Gold-Score, which sums up energy terms that account for the hydrogen bonding energy of the protein-ligand complex, the interaction energy between the ligand and protein, and the internal steric energy of the ligand conformation. GOLD has been shown to reproduce the experimental binding modes of hydrophilic ligands with high success rates, whereas with hydrophobic ligands it may run into problems because the algorithm does not properly account for hydrophobic interactions [153]. The binding free energy predictions are also unlikely to be accurate as the scoring function underestimates the hydrophobic effect, as it does not account for desolvation or entropic effects. GOLD is thus likely to encounter problems when rankordering a set of ligands. In Papers II and IV alternative scoring functions were used to re-rank the docking solutions generated with GOLD. The empirical scoring function F-Score, which is drawn from the scoring function used by FlexX [150] and which is integrated within the CScore [162] module in the molecular modeling package SYBYL [167], was used in Paper IV. F-Score estimates protein-ligand binding free energies as a sum of terms accounting for hydrogen bonding interactions (neutral and ionic), interactions between aromatic groups, lipophilic contacts, and torsional entropy losses of the ligand upon protein-ligand complexation. In Paper II, the docked database compounds were reranked with our 3D QSAR model of AR ligands, F-Score, and another empirical scoring function X-Score [168].

Glide, implemented in the IFD protocol [165, 166], uses a systematic search method for exploring ligand flexibility [164]. To narrow the search space a series of hierarchical filters is applied to a pre-generated library of ligand conformations. In the IFD, an initial ensemble of ligand poses is produced using rigid protein docking with Glide together with a soft docking approach. Soft docking applies reduced van der Waals radii of both the ligand and protein atoms, allowing close contacts to be formed between the atoms of the ligand and protein. The initial poses are ranked according to the GlideScore, which combines both empirical and force field -based terms. The function for GlideScore includes terms for lipophilic and metal-ligation contacts, three types of hydrogen bonding interactions (neutral-neutral, neutral-charged, charged-charged) and interactions of polar non-hydrogen atoms in hydrophobic environments. The function also comprises terms for estimates of entropy losses, solvation effects, and contributions from electrostatic and van der Waals interaction energies. In the following step, the side chains of the protein binding site residues of the top-scoring initial poses are subjected to conformational sampling and energy minimization using the Refinement module of Prime. The ligand and protein backbone of the low-energy according to the total energy estimate of the system, defined as Prime energy, are subsequently used to re-dock the ligand with Glide, now with default parameters. Final ranking of the re-docked ligand poses is based on IFDScore, which is a composite score of GlideScore (95 %) and Prime energy (5%).

4.4 3D QSAR Analysis (Papers I and IV)

The basic idea underlying quantitative structure-activity relationship (QSAR) analysis is that the biological activity of a molecule is dictated by its chemical structure. As a consequence, molecules with similar structure will have similar bioactivities arising from interactions with the same target proteins and changes in structure will be reflected as changes in bioactivity. The objective of QSAR analyses is to establish a quantitative correlation of molecular structure and biological activity for a series of biologically graded compounds using statistical methods. QSAR techniques are in regular use in modern drug design because the generated QSAR models, if successful, can be used to predict activities for compounds whose bioactivity is unknown, thereby assisting the optimization of existing lead series or even enabling discovery of completely new hit compounds as exemplified in this work (Paper II).

QSAR techniques utilize molecular descriptors of diverse nature to characterize the structures and physicochemical properties of the molecules. Depending on the type of descriptors used, the QSAR techniques can be classified into classical (2D) QSAR and nD QSAR (n = 3,4,5,...) methods [169-171]. 3D QSAR methods differ from classical QSAR in that they incorporate the ligands' 3D structures and inherent properties affecting their interaction with the target protein into building the QSAR model [170]. Conformational flexibility of each ligand is incorporated as the fourth dimension in 4D QSAR [172], while an ensemble of induced fit receptor models constitute the fifth dimension in 5D QSAR [173].

The earliest classical QSAR approach (Hansch analysis) was introduced already in the 1960's by Hansch and Fujita [174], while the first applicable 3D QSAR approach

(comparative molecular field analysis, CoMFA) was published in the 1980's by Cramer *et al.* [175]. The CoMFA methodology is one of the most broadly known and used 3D QSAR techniques. It has laid the basis for the development of many other currently used 3D QSAR methodologies. One CoMFA-modification, which was developed to improve the limitations of CoMFA, is the comparative molecular similarity indices analysis (CoMSIA) [176]. In this work, CoMSIA was applied to two distinct ligand sets, namely non-steroidal AR and PR ligands (Papers I and IV), to identify structural features important for their activity. Additionally, the CoMSIA model of AR ligands was applied in the VS as a "tailored scoring function" for AR and used to rank-order the database compounds. In the following sections, the CoMSIA method is described in more detail.

4.4.1 CoMSIA

Data preparation

CoMSIA, like other alignment dependent 3D QSAR methods, requires that the molecules are aligned in space in their putative bioactive conformation. The generation of a rational/reliable alignment is the most critical and challenging step in 3D QSAR analysis. At many instances, this task is facilitated by the knowledge of an experimentally determined bioactive conformation for a molecule included in the series under investigation. There were no co-crystal structures of non-steroidal AR ligands available at the time of the 3D QSAR analysis of AR ligands. Because the 3D structure of AR was known, the prediction and alignment of bioactive conformations was performed with molecular docking. The same alignment approach was utilized also for the non-steroidal PR ligands for their 3D QSAR analysis. Combining docking as an alignment generation tool to 3D QSAR analysis has previously been shown to yield predictive models [144-147].

For calculation of CoMSIA descriptors, the aligned compounds for which atomic point charges have been calculated are enclosed in a sufficiently large box (4 Å beyond the compound set as default) and a regularly spaced grid (default grid spacing 2 Å) is placed over the molecules. In CoMSIA each molecule is assessed for the steric, electrostatic, hydrophobic, and hydrogen bond donor and acceptor property fields [176, 177]. These molecular descriptor fields are calculated as similarity indices between each molecule of the dataset and a common probe atom at every grid point. The similarity indices $A_{F,k}$ for the five physicochemical properties *k* at each grid point *q* are calculated as a sum over all atoms *i* of the molecule *j* using the following equation

$$A_{F,k}^{q}(j) = -\sum_{i=1}^{n} w_{probe,k} w_{ik} e^{-\alpha r_{iq}^{2}}$$

where $w_{probe,k}$ is the property of the probe atom and w_{ik} is the actual value of property k of atom *i*. The properties of the probe atom comprise a 1Å radius, +1 charge, +1 hydrophobicity, and +1 for the hydrogen bond donor and acceptor properties [176, 177]. For each atom *i*, the steric property is evaluated by the third power of atomic radii, the electrostatic property by pre-calculated point charges, hydrophobic property by atom-based parameters developed by Viswanadhan *et al.* [178], and hydrogen bonding properties by representative spatial sites for donors and acceptors defined from distributions of small molecule crystal structures [177].

In the similarity determination, the distance dependence r_{iq} between the probe atom and the atoms of the molecule is considered using a Gaussian-type function and adjusted with the attenuation factor α (default value being 0.3). Use of this functional form allows calculation of descriptors at every grid point, including those within the space occupied by a molecule.

Data analysis

In order to establish a relationship between the vast amount of calculated similarity indices (X-variables) and the biological activity values (Y-variables), multivariate statistical technique of partial least squares (PLS) [179] is used. PLS is an algorithm that extracts linear combinations of principal component -like vectors termed latent variables from the X-variables in a manner that best models the Y-variables. The latent variables are orthogonal, meaning that they only describe variance within the X- and Y-variables not explained by previously introduced latent variables. Large amount of the X-variables are useless for explaining the Y-variables and are detrimental for the robustness of prediction, since they can start to model the noise of the measured values [170]. To avoid inclusion of irrelevant X-variables in the PLS analysis, variable selection approaches are applied, such as a minimum sigma cut-off value, which is used to eliminate those grid points that show a variance below a user defined cut-off value (2 kcal/mol by default). More sophisticated techniques have been developed but these procedures are not within the scope of this thesis and thus not covered.

It is important to determine the optimum number of latent variables to be included in the PLS analysis in order to avoid over-fitting of the data. Generally five to six latent variables are sufficient for generating a realistic QSAR model [180]. The optimum number in each case is usually determined by cross-validation, which is an approach to evaluate the internal predictivity or goodness-of-prediction of the QSAR model. In leave-oneout cross-validation one compound at a time is excluded from the model building and the biological activity of the excluded compound is predicted by the model built from the residual compounds. The same procedure is repeated for every compound within the analyzed series of compounds, after which the statistical values of q^2 (cross-validated correlation coefficient) and *SDEP* (standard deviation of the error of prediction) are calculated from the experimental and predicted activities according to following equations:

$$q^{2} = 1 - \frac{\sum_{i=1}^{N} (Y_{exp} - Y_{pred})^{2}}{\sum_{i=1}^{N} (Y_{exp} - Y_{mean})^{2}}$$

$$SDEP = \sqrt{\frac{\sum_{i=1}^{N} (Y_{exp} - Y_{pred})^2}{N}}$$

Here Y_{exp} is the experimental activity value for a compound *i*, Y_{pred} is the predicted activity value, Y_{mean} is the mean value of the experimental activities, and N is the number of compounds. The q^2 measures the predictive ability of the model. Extraction of statistically significant latent variables in the PLS analysis results in an increase in q^2 and a decrease in *SDEP* values, whereas extraction of too many variables decreases q^2 and increases *SDEP* [170]. Hence, the lowest *SDEP* value is recommended as the criterion for the optimal number of latent variables to be used in the final model building. Models with a q^2 value greater than 0.5 are usually regarded as robust and predictive [180] but we generally apply 0.6 as the threshold value.

PLS analysis calculates also the conventional r^2 (non-cross-validated correlation coefficient) of the final model and the corresponding standard deviation. These values are calculated by replacing Y_{pred} with Y_{calc} (activity value calculated by the model) in equations used to compute q^2 and *SDEP*. The r^2 is a measure of how well the input data is fitted by the model. Goodness-of-fit improves as more latent variables are extracted, ultimately leading to perfect correlations. Therefore, the conventional r^2 is not a good criterion for the validity of the model without other statistical parameters [170].

The result of the PLS analysis is a regression equation consisting of a large number of terms and coefficients [170]. This equation can be used to predict the biological activity for an unknown molecule, providing that the similarity indices have been calculated for this molecule in its' aligned conformation within the grid box. A more informative way of presenting the PLS results is to visualize them as contour plots, which represent spatial areas of physicochemical properties that significantly contribute to the model. Both favorable and unfavorable areas of these properties are contoured at userdefined levels to denote locations having a positive or negative effect on activity.

Model validation

Cross-validation provides a way to define the number of statistically significant latent variables but is also used to estimate the predictive power of the model. The above described leave-one-out cross-validation procedure often gives too optimistic results for predictivity, particularly when analyzing redundant data sets [170]. More rigorous cross-validation approaches, which exclude several compounds from the data set at a time, are regularly applied for larger data sets for obtaining more reliable q^2 values. The group of compounds to be eliminated in a cross-validation run is assigned randomly, thus being different in each run and affecting the statistical outcome. As a consequence, evaluation of the statistical significance of the model using cross-validation in random groups needs to be repeated several times. In Papers I and IV, cross-validation to leave-one-out cross-validation.

It has become evident that high q^2 values do not necessarily confer high predictive ability for the model [181]. Therefore, external validation with a sufficiently large test set of compounds that are completely excluded from the model building is proposed as the only way to estimate the true predictive power of the model [181]. The external test set must provide both structure and activity wise a good representation of the compounds used in model building to obtain reliable statistics for the comparison of experimental and predicted activities for these compounds (predictive r^2). A reliable model is characterized by both high internal (q^2) and external predictivity (predictive r^2). The 3D QSAR models reported in Papers I and IV were therefore validated also with external test sets.

4.5 Integration of docking and 3D QSAR for receptor-based virtual screening

Docking and 3D QSAR models constructed for our work aim to describe the same biological phenomenon – binding interactions of two molecules resulting from weak forces between them. These two independent descriptions of the phenomenon can be logically combined. Independency of the descriptions means that the models can be constructed separately from one another using different initial data and that construction of one model does not require preliminary information of the other.

Formation of intermolecular receptor-ligand complexes is a fundamental requirement for proper function of receptors and in consequence of entire cells. In docking simulations mathematical and statistical functions, which model the physical and chemical forces of molecular interactions, are applied together with diverse algorithms for conformational optimization to guide the binding ligand into the receptor binding site. Based on these functions each ligand is assigned a score that represents the strength of the binding interaction. Comparison of the interaction strength between structurally dissimilar ligands using the assigned computed scores as the only reference is not by itself reliable prediction. It is therefore a good idea to look for complementary information to support the docking models.

3D QSAR models are constructed typically from a set of few dozens of ligands that are known to interact with the target receptor as a result of ligand binding experiments. Statistical methods are used to explain data derived from spatially overlaid ligand structures in order to uncover structural features that have a positive or a negative effect on ligand binding. An essential precondition for the construction of a 3D QSAR model is that the ligand set comprises structurally sufficiently diverse, but yet comparable, molecules. Additionally, model building requires expertise to identify chemically equivalent structural moieties, so called bioisosteres, within the diverse ligands. Carboxylic acid (-C(O)OH) and a five-membered pyrazole ring with two adjacent nitrogen atoms, one of which is in a protonated state (Smiles string c1cn[nH]c1), are examples of bioisosteric groups. They are comparable in size and capable of both donating and accepting hydrogen. Before statistical analysis can be applied in 3D QSAR modeling, the bioisosteric groups of ligands need to be overlaid in 3D space. This is a stage where docking becomes useful. Docking provides an automated and objective idea of positions of the different bioisosteres of ligands within the receptor binding site. A docking simulation performed for each of the molecules gives as a result a proposal of the best binding pose for each compound. These poses represent a natural alignment of the binding ligands from the receptor viewpoint where positioning is guided by the receptor binding site structure. Docking simulation thus produces a suggestion how, and if, bioisosteric moieties of the structurally dissimilar molecules become naturally overlaid, which is a requirement for the construction of 3D QSAR model. In this manner, the two separate computational methods of the binding phenomenon can be logically combined.

Performing the statistical analysis is straightforward after the bioisosteric moieties of the ligands are aligned in the 3D space. Properties such as the ability to donate or accept hydrogen bonds are computed for the docked ligands and then used in the statistical 3D QSAR analysis. If bioisosteres such as carboxylic acid and pyrazole are aligned in the docking simulations and if the hydrogen bond interactions formed by them turn out as favorable for ligand binding in the statistical analysis, the spatial volume where such a bioisostere should be positioned for increased biological activity is delineated in the 3D space of the receptor binding site. The volume can be visualized within the receptor binding site because docking was used for alignment generation. The presence of complementary interacting chemical groups in the receptor structure in the proximity of such a volume, which is created solely from information derived from ligands, validates the quality of the 3D QSAR model.

The volumes that are revealed in the 3D QSAR analysis and that represent favored and disfavored interactions for ligand binding are useful in many ways. They can be used for interpreting the docking results, which include the final pose in the binding site and the assigned score for the pose. As mentioned above, scoring with current technology is not by itself reliable enough for predicting ligand binding. Yet, if a molecule is assigned a good score, and its bioisosteric groups are positioned in the binding site regions that are favorable for its intermolecular interaction types according to the receptor structure and 3D QSAR model, it means that the two independent computational methods have converged into a same prediction. Experimentally observed ligand binding is explained with data originating from receptor and ligand structures. Furthermore, the favored and disfavored volumes can be used to examine that a binding ligand does not have chemically incompatible or structurally clashing moieties according to the 3D QSAR model. If a ligand is able to fulfill all the conditions set by the computational methods, it is more likely to display true binding affinity for the receptor when tested for activity.

This ensemble of computational models is applicable in VS of molecular databases. During the last decade, all the major chemical vendors have provided their compound collections in an electronic format. A VS workflow that is capable of searching new bioisosteric moieties, as those computationally described in 3D QSAR models, can be set up to screen molecular databases for novel bioactive scaffold structures for a target receptor. It is likely that the databases encompass molecules that are able to fulfill the conditions set by docking and 3D QSAR models, but whose interactions with biological systems have not yet been determined. Such molecules are important starting points for drug discovery and thus intensely sought after by means of computations.

4.6 Experimental methods

Modern drug discovery and development integrates computational and experimental approaches in an iterative fashion. Computational techniques can be used to explain preexisting experimental knowledge and to introduce new ideas for the next generation of compounds to be tested. Experimentalists –chemists, biologists, pharmacologists– then carry out series of biological analyses, which in turn generate new raw data to be used for the next round of computational modeling. Discovery and optimization phases of drug development require repetitive experimental and computational cycles, thus making drug development a lengthy collaborative effort of a multidisciplinary team. In this thesis, all the experimental measurements performed for compounds selected with computational methods were either outsourced to a biopharmaceutical company (Paper II) or performed by our collaborators (Paper III). Detailed descriptions of the experimental analyses are given in the original Papers II and III.

5 **Results and Discussion**

This thesis work aimed at the discovery of structurally novel non-steroidal AR ligands from a commercial compound library that was subjected to receptor-based VS. The success of receptor-based VS relies on how well the selected scoring function manages to rank the active database compounds in the top positions of the rank list of docked compounds. It is well-known that current scoring functions have difficulties in performing this task. Due to this limitation of scoring functions, we were eager to test an alternative scoring approach, namely scoring with a 3D QSAR model. Such models are typically used in the optimization of a particular ligand scaffold structure for a target protein of interest, not as a scoring function in receptor-based VS. Nevertheless, we were interested in challenging the conventional use of a 3D QSAR model and assessing its potential as a tailored scoring function for AR in receptor-based VS.

In this chapter the results from our receptor-based VS experiment are presented step-by-step, starting from the generation of the 3D QSAR model of AR ligands to be used in rank-ordering the docked database compounds in VS, continuing with the VS procedure itself, and ending with the functional profiling of a subset of identified AR hit compounds. In addition, the 3D QSAR model of PR ligands is introduced. When used in concert, these two 3D QSAR models could be valuable tools for detection of AR- and PR-selective ligands in receptor-based VS.

5.1 3D QSAR models (Papers I and IV)

5.1.1 CoMSIA of AR ligands

3D QSAR analysis with CoMSIA [176] was performed on a series of 61 AR ligands (Tables 1-6 in Paper I) [36, 38, 40, 42, 43] to determine the structural features that contribute to the binding affinity of these compounds. The analyzed ligands were diverse in terms of both structure and function. Six structural classes, the majority of which were derivatives of the non-steroids HF, nilutamide, and bicalutamide, and the entire spectrum of functionalities ranging from agonists to partial agonists and antagonists were

represented in the ligand set. Generating an alignment, which is needed to quantify binding characteristics with 3D QSAR statistical methods, of such a diverse set of ligands within the AR binding site posed a significant challenge. Moreover, at the start of the work we lacked experimentally confirmed knowledge about the binding mode for the non-steroidal ligands and structural changes induced by antagonists on AR LBD. The latter still remains unknown as discussed in the literature review.

The ligand alignment was derived within the active, agonist structure of the AR ligand-binding site by generating possible binding poses with the docking program GOLD [153] and using Gold-Score in the selection of a representative docking pose for each ligand (Figure 1 in Paper I). Hydrogen bonding to the few polar residues within the otherwise hydrophobic binding site (Figure 7) is proven to be important for binding of both steroidal and non-steroidal ligands [103, 182]. Formation of receptor-ligand hydrogen bond interactions was emphasized by using the DHT structure to guide pose generation in docking simulations. To enable the large non-steroidal AR ligands to form these key interactions there was a need for a modification in the AR LBP structure. A F876A mutation that we expected to partly account for the structural changes taking place upon binding of the large non-steroidal ligands was introduced. This mutation opened up the binding cavity, providing an entrance to the receptor surface. As a result, the search space for ligand conformations in docking increased significantly and therefore DHT guided docking served also the purpose of limiting the number of alternative poses and ensuring that the non-steroids were positioned completely within the binding cavity.

PLS analysis and various cross-validation methods indicated that hydrophobic and hydrogen bond acceptor descriptors calculated for the aligned ligands yield a statistically significant and robust model for binding affinity predictions. Leave-one-out crossvalidation yielded a q^2 value of 0.656, while more rigorous cross-validation of ten and five groups yielded average q^2 values of 0.612 and 0.571 per 25 repetitions, respectively. Despite the somewhat lower q^2 values for the random group cross-validations all the values reflect an internally consistent, or predictive, model. The final model was derived using the optimum of five latent variables as determined by leave-one-out crossvalidation and yielded a non-cross-validated r^2 value of 0.911. The predictivity of the model was further evaluated with an external test set of nine compounds representing structural features and a wide range of activities included in the compound set used to train the model. The model managed to clearly separate the high- and low-affinity compounds. The test set yielded a predictive r^2 value of 0.800 indicating a good external predictive power for the model. Taken together, the internal and external validation methods suggest that the model is robust and predictive (Figure 2 in Paper I).

The statistically relevant results of the CoMSIA model were visualized as 3D contour plots within the receptor binding site (Figure 3 in Paper I) and the contours explaining the variation of the binding affinity of the ligands were interpreted in terms of the binding site structure. The favored and disfavored volumes for placing hydrophobic, hydrophilic or hydrogen bond acceptor groups within the binding cavity correlate with the amino acid residues of the AR binding site in proximity to these volumes. Possible interactions formed between these binding site residues and the ligands in the training and external test sets in the docking simulations. Altogether, this consensus creates confidence towards the generated ensemble of models because the discovered, statistically significant correlation is only dependent on the ligand structures and their alignment, and not on the protein binding cavity.

The predictive potential of the model was assessed one step further with a small set of bicalutamide-related high-affinity compounds reported by Marhefka *et al.* incorporating structural features not included in the model building [39]. The docking poses generated for each compound were rank-ordered according to the binding affinities predicted with the model. The pose with the highest prediction was picked as the representative binding conformation for each compound. These docking poses were well aligned with the ligands used to build the model. With this approach, the model was able to recognize these compounds as high-affinity AR ligands, although with a larger standard error of prediction compared to the nine compound external test set. Based on the results, the model was suggested to be applicable for identification of active compounds in virtual database screening.

Good statistics and consistency of the model with the AR LBP structure indicated that the alignment used to build the CoMSIA model consisted of biologically active conformations of the non-steroidal ligands. Prior to the publication of our CoMSIA model Bohl *et al.* reported their CoMFA model of AR ligands [144]. They also generated the ligand alignment with docking but used a PR LBD -based homology model of AR LBD as the target structure instead. Their alignment differed from ours but produced a statistically significant model as well. Later, Bohl *et al.* solved the binding mode of the bicalutamide-like agonist S-1 (Figure 2) by crystallography [107]. It turned out to be different from both our and their predictions made by docking simulations.

Despite the inconsistency in the experimental and predicted binding modes, the CoMSIA model was considered suitable for receptor-based VS purposes because the contour plots were in agreement with the AR LBP structure. Incorporation of various functional profiles into the model permits the model to be used for binding affinity prediction of putative AR ligands without consideration to the possible modulation of receptor activity. New AR ligand scaffolds were indeed found and the model predictions contributed to their detection as indicated in Papers II and III.

5.1.2 CoMSIA of PR ligands

3D QSAR analysis with CoMSIA [176] was performed to study the structure-activity relationships of 64 PR ligands (Tables 1-3 in Paper IV) [64, 65, 70-72]. An additional goal was to derive the model in a comparable manner to the model of AR ligands and thereby enable their parallel use in e.g. VS for receptor-selective non-steroidal ligands for AR and PR. Therefore, the modeling procedure was analogous to the previous analysis. As in Paper I the ligands represented several different scaffold structures and various pharmacological activities, majority being antagonists however. Similarly to AR, the antagonist-induced changes to the receptor structure were unknown at the time of starting the work as it was only very recently when the antagonist structure of PR LBD was reported [128]. Nevertheless, the co-crystal structure of the non-steroidal agonist Tanaproget with the PR LBD became available while the work was initiated [109]. Tanaproget has a closely related structure to many of the studied PR agonists and antagonists. The co-crystal structure thereby provided insight into the binding interactions for such compounds and assisted in the alignment generation for the 3D QSAR analysis.

The alignment was constructed in the agonist structure of PR LBD from ligand poses generated with the docking program GOLD [153] (Figure 1 in Paper IV). Selection of representative docking poses for the alignment was based on the FlexX-derived [150] scoring function F-Score from the CScore [162] module in SYBYL [167] instead of Gold-Score as used for the AR ligands in Paper I, because the F-Score-based alignment resulted in a higher statistical quality for the model. F-Score was also able to identify the best Tanaproget docking solution with the lowest RMSD from its crystallographically determined conformation and was therefore considered the most suitable scoring function for this work.

The structure-activity relationships for the analyzed set of ligands were best explained with the electrostatic, hydrogen bond donor, and hydrophobic properties of the ligands. PLS analysis of the three descriptor fields and binding affinities for the ligands yielded a statistically significant and robust model determined with internal crossvalidation methods. Leave-one-out cross-validation gave a q^2 value of 0.637, and crossvalidation with ten and five randomly selected groups yielded average q^2 values of 0.601 and 0.563 per 25 repetitions, respectively. The optimum of six latent variables was used to derive the final model yielding a non-cross-validated r^2 value of 0.878 (Figure 2 in Paper IV). External validation was performed with a test set of ten compounds, which produced a predictive r^2 value of 0.833, indicating strong predictive ability for the model. Regarding the statistical quality, the CoMSIA model of PR ligands was equivalent to the model of AR ligands.

Significant statistical data of the final model was represented as 3D contour plots within the ligand-binding site (Figure 3 in Paper IV) and interpreted with respect to the receptor structure. A good agreement exists between the chemical environment of the

binding site and the favorable and unfavorable electrostatic, hydrogen bond donor, and hydrophobic contours. Many of the contour plots are in line with experimental SAR data and the crystallized Tanaproget structure, which further supports the reliability of the model.

The model was considered to be of high quality and to be derived from an alignment that comprised biologically active conformations of the PR ligands. The binding conformations predicted by docking are corroborated by the bioactive conformation determined for the structural analogue, Tanaproget [109].

5.2 Virtual screening of novel non-steroidal AR ligand scaffolds (Paper II)

A workflow for receptor-based VS was developed and experimentally tested to find nonsteroidal AR ligand scaffolds with structural novelty. Based on the previous study (Paper I), 3D hydrophobic and hydrogen bond acceptor interactions (3D QSAR model) can explain most of the variation in ligand binding affinity to AR. Hence, the 3D QSAR model of AR ligands was applied as part of the workflow as a tailored scoring function for AR to assist in the selection of candidate compounds for experimental assessment of AR binding. QSAR methods that are based on an overlay of molecules (such as CoMSIA) are often criticized due to lack of credibility to identify novel actives for a specific target by VS and are therefore rarely applied in VS. This VS study served thus as a further test for the predictive quality of the model because the goal was to seek compounds structurally unrelated to those used to build the 3D QSAR model.

The use of a successful 3D QSAR model as part of an ensemble of computational methods in receptor-based VS is well-founded and covered already in Methods section 4.5. In short, the statistical correlation in 3D QSAR is dependent only on the ligand structures and their alignment. Docking in turn uses the structural information within the binding cavity to create superimposed ligand conformations. Docking can thus extend the 3D QSAR method by supplying a superimposed set of ligands for the 3D QSAR model building while incorporating information from the receptor binding site. Further, the regions or volumes of space around the ligand that are found in a successful 3D QSAR analysis to affect binding can be visualized within the binding cavity. In our model, the volumes are consistent with the residues providing interactions to the ligands within the AR binding cavity in proximity to these volumes. For that reason docking and 3D QSAR models can be used in concert to predict binding affinities for database compounds in search for new AR ligands. First, different poses of a compound to be evaluated are created with docking inside the binding cavity. Second, different poses are scored using the 3D QSAR model of AR ligands. When the top-scoring pose is chosen, it is typically automatically aligned within the AR binding cavity in a biologically sensible

conformation. Third, since structurally and chemically different scaffolds can create similar molecular interactions that are generalized in the 3D QSAR model, the model was expected to detect new compound scaffolds as AR ligands. For these reasons, created 3D QSAR model of AR ligands was considered suitable for compound screening purposes.

A commercial compound library provided by Asinex was used as the screening database in the VS. The database, which comprised over 200 000 compounds, was subjected to a chain of filters (Table 1 in Paper II) in order to reduce the size of the database prior to docking. Due to the hydrophobic nature of the AR binding site, some bias toward compounds with higher lipophilicity was allowed. Filtering was started with simple physicochemical filters and continued with a more target-specified substructure query derived from known AR ligands deposited in the MDDR (MDL Drug Data Report) database. The search query was not made too stringent so as to retain a large number of distinct scaffold structures in the following stages of VS. After removal of chiral and highly reactive compounds, the database had reduced to a size of less than 4000 compounds. These were docked into the AR LBP with the program GOLD to predict their binding conformations within the binding site. The docking procedure was similar to that reported in Paper I for generation of ligand alignment for the 3D QSAR analysis and Gold-Score was used in the docking simulations to evaluate the fitness of the generated poses. Subsequently four scoring approaches, i.e. the 3D QSAR model and the scoring functions F-Score [150, 167], Gold-Score [153], and X-Score [168], were used in ranking the docked database compounds. Compound selection for experimental testing was guided by these rankings with most weight put on the 3D QSAR model predictions. Three hundred top-ranked compounds from the rank list based on 3D QSAR model predictions were evaluated visually and a few non-steroidal structures with distinct scaffold structures were selected with support from the predictions of the other scoring functions. Selection of distinct scaffolds was done because the objective was to form small clusters of analogous compounds rather than picking out single representatives of non-steroidal structures. For each scaffold, a few chemically similar compounds were then chosen (Table 2 and Figure 1 in Paper II). Chemical similarity is one of the most important concepts in pharmaceutical development and chemical toxicology based on the hypothesis that similar compounds have similar properties. There are numerous examples supporting this hypothesis as well as examples where this hypothesis is not valid. In the latter case, a small change in chemical structure can paradoxically lead to a significant change in biological activity. Extension of the chosen scaffolds from the ranked sets of compounds was done to avoid this paradox.

The *in vitro* binding of the non-steroidal compounds for AR LBD was analyzed in a competitive binding assay. All the compounds were able to displace the reference ligand from the binding site, thus showing affinity for the AR LBD (Figure 2 in Paper II). The IC₅₀ values were below 10 μ M for most of the compounds, below 3 μ M for some of the

compounds and in nM range for the highest affinity compound (Table 2 in Paper II). Experimental screening thereby verified that the selected scaffolds interact with the target receptor, yet with rather low binding affinities. A few compounds were additionally tested for PR binding using a corresponding assay system for PR as was used for AR binding. Only one of the tested compounds showed better binding to PR than AR, while others did not bind or had a lower PR affinity (Table 3 in Paper II). Taken together, the binding experiments confirmed detection of structurally novel AR ligands and even ligands with some binding preference for AR over PR. Nevertheless, it must be noted that the AR binding affinity predictions by the 3D QSAR model did not correlate with the experimentally measured affinities. This was no surprise however, since the inability to accurately estimate binding affinities is a well-known limitation of all generally used scoring functions [156-158].

An additional aim of this study, besides the discovery of previously unknown nonsteroidal ligand scaffolds as AR hits, was to find out whether the 3D QSAR model could contribute to identification of such hits and serve as a tailored scoring function for AR. Examination of the rankings assigned to the experimentally tested compounds by the 3D QSAR model and the common scoring functions F-Score, Gold-Score, and X-Score clearly displayed lack of correlation. In fact, the rankings of the model usually opposed the rankings of the other scoring functions, while somewhat better correlations exist between the three scoring functions (Figure 3 in Paper II). Despite the discrepancies in the rankings, the VS results indicated that the model was able to contribute to detection of several reported AR binding ligand scaffolds either independently or together with other scoring functions. The model enabled even detection of scaffold classes that would probably remain unidentified by the common scoring functions due to low rankings. According to the results of this work, it can be concluded that the 3D QSAR model can be integrated in receptor-based VS as a tailored scoring function for AR to uncover weakly binding, structurally novel AR hits with non-steroidal scaffolds distinct from ones used in model building.

5.3 Molecular modeling of antagonist mechanism identified for diphenyl- and phenylpyridine –based AR ligands (Paper III)

5.3.1 Functional characterization in vitro

The 3D QSAR model used in our receptor-based VS predicts only the likelihood of a compound to bind to AR. It is unable to predict the functional activities of AR binding compounds because it is derived from compounds representing a mixture of functionalities. We selected a set of six µM AR ligands, which represented two structural scaffolds

(Figure 1 in Paper III), from the hits identified in the receptor-based VS and subjected them to in vitro functional profiling. Biological effects induced by the diphenyl- and phenylpyridine-based compounds (termed E5 and G1-G5, respectively, according to internal nomenclature) upon interaction with WT and T877A variant of AR were studied. The assays were performed in two cell lines: 1) CV1 cell line, which is devoid of endogenously expressed SRs and thus allows responses to particular transfected receptor to be measured, and 2) LNCaP cell line, which expresses AR with the binding site mutation T877A often associated with CaP [183]. Agonist activity on AR transactivation was not observed for any of the compounds in neither WT nor mutant AR as demonstrated by reporter gene assays in CV1 cells. On the contrary, a range of antagonist effects was displayed by the compounds at the tested concentration of 10 µM. The degree of antagonism was found to be different for the compounds on the T877A mutant compared to the WT AR. This observation indicated that the compounds affect the AR transactivation through binding to the ligand-binding site of AR LBD since the activity is affected after introduction of the mutation in the binding site. Two of the newly identified AR antagonists (G1 and G4) were able to repress the androgen-dependent cell proliferation of LNCaP cells with the mutant receptor. This finding was in line with their antagonist character on AR transactivation in CV1 cells transfected with mutant AR. Together, the *in vitro* results suggest discovery of two structurally new non-steroidal AR antagonists with inhibitory activity for both WT and T877A mutant AR.

Detection of compounds that inhibit the activity of mutated AR variants found in CaP patients besides WT AR is of great value when designing drugs for CaP treatment. Several binding site mutations are known to cause resistance to the current antiandrogens that are used in the treatment of CaP. T877A mutation, for example, overcomes the antagonist activity of HF and converts it to an agonist instead. Bicalutamide, on the other hand, retains its antagonistic behavior in the T877A variant of AR and can therefore be used as a second-line treatment when resistance to HF following T877A mutation has occurred. Bicalutamide is in fact the most attractive antiandrogen currently available for CaP treatment. AR mutations that confer antagonist-to-agonist conversion of bicalutamide are also known however and they may occur during the progression of CaP too. There is therefore a need for new sex organ -selective AR antagonists that are able to inhibit the activity of several AR variants that are observed in CaP [8]. Future will show whether such drugs can be designed and developed as safe treatments.

5.3.2 Induced fit docking

Inspired by the newly characterized AR antagonists we performed additional computational modeling with the aim of gaining insight into the mechanism behind the antagonism of the phenylpyridine-based compounds. Since the docking program GOLD that was used in receptor-based VS ignores protein flexibility apart from optimization of hydrogen bonds, we decided to explore the antagonist-induced conformational changes in the AR structure using a more advanced computational docking method. The IFD protocol of Schrödinger [165, 166] addresses protein flexibility more thoroughly by means of side chain predictions and allowing small movements in the protein backbone. Modeling of large movements, like repositioning of H12, as is shown to take place in the antagonism of the related ER [105, 132], are out of scope with current computational methodologies including this docking method. Whether AR antagonism involves such structural movement, is in any case uncertain. For this reason, induced fit docking seemed to be a sufficient tool to extend the modeling one step further from what was achieved with GOLD in VS. The aim was to obtain suggestions for interaction points between the antagonists and protein, which in turn could induce more extensive structural changes affecting the function of the protein.

Induced fit dockings were performed for three antagonists (G1, G2, and G5) to both WT and T877A mutant AR in order to find a possible structural explanation for the observed experimental results (Figure 4 in Paper III). In the WT AR, all the docked antagonists were shown to induce a noticeable reorientation of residues W741 and M895. This result is in line with the residue movements in the published crystal structures of AR LBD. W741 and M895 are among the most flexible LBP residues and enable accommodation of various larger-sized ligands to AR LBP [107, 111]. The antagonistinduced movement of the indole ring of W741, which also necessitates repositioning of M895 side chain, was proposed to be needed to enhance hydrogen bonding to the receptor. In the mutant AR, similar movement of W741 and M895 side chains was observed only for one of the antagonists (G1). In the in vitro assays, this compound inhibited the T877A mutant form of AR in the transactivation assay and also repressed the androgen-dependent cell growth of LNCaP cells. On the other hand, the other two compounds (G2 and G5) exhibited reduced or no inhibition of mutant AR transactivation and completely lacked inhibitory effect on LNCaP cell proliferation. Because there are no other major structural modifications predicted to take place according to the representative induced fit docking poses for G1, G2, and G5, it was hypothesized that the movement of W741 and M895 could be mechanistically involved in the initiation of the antagonism for the studied phenylpyridine-based antagonists. The movement of these residues could in turn lead to more extensive structural reorganization required for antagonism.

6 Conclusions

It has become a regular practice in modern drug discovery to combine computational methods with experimental methods throughout the drug development process from hit discovery to lead optimization and beyond. Computational modeling is regularly used in drug development to explain the existing experimental data, e.g. data on ligand binding affinity, and to use the gained knowledge when designing the next generation of compounds to be tested. In the present thesis, the discovery of new structural scaffolds for AR-targeted drug development was attempted using computational methods and information from public structural and biological data resources. The wealth of experimental data stored in scientific literature and in public databases offers an excellent opportunity to initiate drug discovery as academic projects, even without an own wet-lab as was the situation in this project. In such cases collaboration with experimentalists is of utmost importance in order to prove the modeling results right or wrong, but also to generate together new research hypotheses for subsequent collaborative work.

The major computational tools employed during the course of the work were molecular docking and 3D QSAR analysis with CoMSIA method. In Paper I, the binding properties of a diverse set of AR ligands were investigated by combining the two methods. The results demonstrated the usability of molecular docking as the alignment generation tool for deriving a statistically significant 3D QSAR model that is applicable in the prediction of AR binding affinities of docked compounds. The reliability and predictivity of the 3D QSAR model was extensively validated; the model was statistically of high quality and it reflected the structure and possible interactions of the AR ligand-binding site. Overall, the results indicated that the 3D QSAR model could be combined into receptor-based VS for detection of new AR ligands from compound databases. The future plans for the project include updating the model because lots of new information on AR ligands and their interactions with the receptor structure has become available since the initial model was generated.

Arranging the docked database compounds according to their binding preference for the target receptor is a major problem in receptor-based VS. Scoring functions are typically used for this purpose but receptor-based VS is still far from being automated because scoring functions, although improved, still have limited and often case-

dependent success in identifying active compounds from inactive compounds. In Paper II, receptor-based VS of AR ligands from a commercial database was performed. In addition to uncovering novel non-steroidal ligand scaffolds as AR hits, the work aimed at exploring whether our previously generated 3D QSAR model of AR ligands could serve as a tailored scoring function for AR and contribute to detection of such AR hits. The 3D QSAR model was hence applied in the VS workflow as a scoring function together with three common scoring functions. The VS led to the discovery of AR hits exhibiting previously unknown non-steroidal scaffold structures but rather weak binding affinities. Nevertheless, the results indicated that the 3D QSAR model, as part of the developed workflow, contributed to detection of these new AR ligand scaffolds, even ones that are likely to remain unidentified by the common scoring functions due to low rank positions. Furthermore, many of the identified AR ligands displayed a binding preference for AR over PR. This indicated that a 3D QSAR model that is built within the ligand-binding site of a target protein and therefore reflects the chemical environment of a particular binding site could be used as a tailored scoring function for identification of ligands with binding preference for the particular target. The future goal of the project is to focus on a few of the identified scaffold structures and try to increase their binding affinity for AR. A second round of VS should thus be carried out.

In Paper III, the functional profile of a few AR ligands discovered in the VS study was studied *in vitro*. Interestingly, AR ligands with antagonist activity in both WT AR and T877A mutant AR frequently associated with CaP were uncovered. These antagonists, although they clearly have too low binding affinities at this stage of development, could potentially be used in the future as a starting point for design of more potent AR antagonists for CaP treatment. Furthermore, in this study, the structural basis for the observed AR antagonism was modeled with induced fit docking methodology. Docking results suggested that movement of residues W741 and M895 could mechanistically initiate the structural reorganization ultimately giving rise to their antagonist activity. More detailed study of the structural alterations triggered by the analyzed antagonism at the structural level. It would be very important to solve the AR LBD structure in complex with an antagonist ligand so that the mechanism for AR antagonism could be verified. Moreover, the antagonist structure could also make the design of AR antagonists easier.

In Paper IV, the binding properties of diverse PR ligands were investigated in a similar fashion with docking and 3D QSAR analysis as the AR ligands were studied in Paper I. The resulting 3D QSAR model of PR ligands could also be implemented in receptor-based VS of new PR ligands, but this was not done in this thesis. In the future, further investigations on the suitability of 3D QSAR models as custom scoring functions in receptor-based VS are definitely required and whether the 3D QSAR model of PR

ligands is able to identify new PR ligands with binding selectivity over AR and perhaps even other SRs should be tested.

In summary, the case studies presented in this work show that computational methods are able to provide added value to existing experimental data in the drug discovery process, especially in finding new molecular scaffolds and structural ideas for development. Computations are not independent, however, and require collaboration with chemists and biologists who can carry out experimental investigations based on the modeled theories. It is worth remembering that the quality of the experimental data sets the limits for the quality of theoretical modeling throughout the drug development process: the quality of modeling results can never exceed the quality of experimental data on which the modeling is based on, making good quality modeling feasible only from high quality experimental data. Hence, successful drug design projects always rely on a concerted effort of a multidisciplinary team.

7 Bibliography

- [1] Robinson-Rechavi M, Escriva Garcia H, and Laudet V. J. Cell. Sci., **116** (2003) 585-586.
- [2] Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schutz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P, and Evans RM. *Cell*, 83 (1995) 835-839.
- [3] Quigley CA, De Bellis A, Marschke KB, el-Awady MK, Wilson EM, and French FS. Endocr. Rev., 16 (1995) 271-321.
- [4] Snyder PJ. Androgens, In Goodman & Gilman's The Pharmacological Basis of Therapeutics, 11th ed., Brunton LL, Lazo JS, Parker KL, Eds., The McGraw-Hill Companies, Inc., New York, 2006, pp.1573-1586.
- [5] Gottlieb B, Beitel LK, Wu JH, and Trifiro M. Hum. Mutat., 23 (2004) 527-533.
- [6] Gao W, Bohl CE, and Dalton JT. Chem. Rev., **105** (2005) 3352-3370.
- [7] Gao W, Kim J, and Dalton JT. Pharm. Res., 23 (2006) 1641-1658.
- [8] Tran C, Ouk S, Clegg NJ, Chen Y, Watson PA, Arora V, Wongvipat J, Smith-Jones PM, Yoo D, Kwon A, Wasielewska T, Welsbie D, Chen CD, Higano CS, Beer TM, Hung DT, Scher HI, Jung ME, and Sawyers CL. *Science*, **324** (2009) 787-790.
- [9] Narayanan R, Mohler ML, Bohl CE, Miller DD, and Dalton JT. Nucl. Recept. Signal, 6 (2008) e010.
- [10] Mohler ML, Bohl CE, Jones A, Coss CC, Narayanan R, He Y, Hwang DJ, Dalton JT, and Miller DD. *J. Med. Chem.*, **52** (2009) 3597-3617.
- [11] Walters WP, Stahl MT, and Murcko MA. Drug Discov. Today, 3 (1998) 160-178.
- [12] Ekins S, Mestres J, and Testa B. Br. J. Pharmacol., 152 (2007) 9-20.
- [13] Lyne PD. Drug Discov. Today, 7 (2002) 1047-1055.
- [14] Miller MA. Nat. Rev. Drug Discov., 1 (2002) 220-227.
- [15] Willett P. Drug Discov. Today, **11** (2006) 1046-1053.
- [16] Lengauer T, Lemmen C, Rarey M, and Zimmermann M. Drug Discov. Today, 9 (2004) 27-34.
- [17] Good AC, Krystek SR, and Mason JS. Drug Discov. Today, 5 (2000) 61-69.
- [18] Ekins S, Mestres J, and Testa B. Br. J. Pharmacol, 152 (2007) 21-37.

- [19] Ghosh S, Nie A, An J, and Huang Z. Curr. Opin. Chem. Biol., 10 (2006) 194-202.
- [20] Lonard DM and O'Malley B W. Mol. Cell, 27 (2007) 691-700.
- [21] Perissi V and Rosenfeld MG. Nat. Rev. Mol. Cell. Biol., 6 (2005) 542-554.
- [22] Heemers HV and Tindall DJ. Endocr. Rev., 28 (2007) 778-808.
- [23] Onate SA, Tsai SY, Tsai MJ, and O'Malley BW. Science, 270 (1995) 1354-1357.
- [24] Chen JD and Evans RM. Nature, 377 (1995) 454-457.
- [25] Horlein AJ, Naar AM, Heinzel T, Torchia J, Gloss B, Kurokawa R, Ryan A, Kamei Y, Soderstrom M, Glass CK, and et al. *Nature*, **377** (1995) 397-404.
- [26] Cheng S, Brzostek S, Lee SR, Hollenberg AN, and Balk SP. Mol. Endocrinol., 16 (2002) 1492-1501.
- [27] Hodgson MC, Astapova I, Cheng S, Lee LJ, Verhoeven MC, Choi E, Balk SP, and Hollenberg AN. J. Biol. Chem., **280** (2005) 6511-6519.
- [28] Dotzlaw H, Papaioannou M, Moehren U, Claessens F, and Baniahmad A. *Mol Cell Endocrinol*, **213** (2003) 79-85.
- [29] Berrevoets CA, Umar A, Trapman J, and Brinkmann AO. *Biochem. J.*, **379** (2004) 731-738.
- [30] Smith CL and O'Malley BW. Endocr. Rev., 25 (2004) 45-71.
- [31] Buijsman RC, Hermkens PH, van Rijn RD, Stock HT, and Teerhuis NM. Curr. Med. Chem., 12 (2005) 1017-1075.
- [32] Cadilla R and Turnbull P. Curr. Top. Med. Chem., 6 (2006) 245-270.
- [33] Madauss KP, Stewart EL, and Williams SP. Med. Res. Rev., 27 (2007) 374-400.
- [34] Anonymous. Med. Lett. Drugs Ther., **31** (1989) 72.
- [35] Tucker H, Crook JW, and Chesterson GJ. J. Med. Chem., 31 (1988) 954-959.
- [36] Dalton JT, Mukherjee A, Zhu Z, Kirkovsky L, and Miller DD. *Biochem. Biophys. Res. Commun.*, **244** (1998) 1-4.
- [37] He Y, Yin D, Perera M, Kirkovsky L, Stourman N, Li W, Dalton JT, and Miller DD. Eur. J. Med. Chem., 37 (2002) 619-634.
- [38] Kirkovsky L, Mukherjee A, Yin D, Dalton JT, and Miller DD. J. Med. Chem., 43 (2000) 581-590.
- [39] Marhefka CA, Gao W, Chung K, Kim J, He Y, Yin D, Bohl C, Dalton JT, and Miller DD. J. Med. Chem., 47 (2004) 993-998.
- [40] Yin D, He Y, Perera MA, Hong SS, Marhefka C, Stourman N, Kirkovsky L, Miller DD, and Dalton JT. *Mol. Pharmacol.*, 63 (2003) 211-223.
- [41] Morris JJ, Hughes LR, Glen AT, and Taylor PJ. J. Med. Chem., 34 (1991) 447-455.
- [42] Van Dort ME and Jung YW. Bioorg. Med. Chem. Lett., 11 (2001) 1045-1047.
- [43] Van Dort ME, Robins DM, and Wayburn B. J. Med. Chem., 43 (2000) 3344-3347.
- [44] Balog A, Salvati ME, Shan W, Mathur A, Leith LW, Wei DD, Attar RM, Geng J, Rizzo CA, Wang C, Krystek SR, Tokarski JS, Hunt JT, Gottardis M, and Weinmann R. *Bioorg. Med. Chem. Lett.*, 14 (2004) 6107-6111.

- [45] Salvati ME, Balog A, Shan W, Wei DD, Pickering D, Attar RM, Geng J, Rizzo CA, Gottardis MM, Weinmann R, Krystek SR, Sack J, An Y, and Kish K. *Bioorg. Med. Chem. Lett.*, **15** (2005) 271-276.
- [46] Salvati ME, Balog A, Wei DD, Pickering D, Attar RM, Geng J, Rizzo CA, Hunt JT, Gottardis MM, Weinmann R, and Martinez R. *Bioorg. Med. Chem. Lett.*, 15 (2005) 389-393.
- [47] Sun C, Robl JA, Wang TC, Huang Y, Kuhns JE, Lupisella JA, Beehler BC, Golla R, Sleph PG, Seethala R, Fura A, Krystek SR, Jr., An Y, Malley MF, Sack JS, Salvati ME, Grover GJ, Ostrowski J, and Hamann LG. J. Med. Chem., 49 (2006) 7596-7599.
- [48] Ostrowski J, Kuhns JE, Lupisella JA, Manfredi MC, Beehler BC, Krystek SR, Jr., Bi Y, Sun C, Seethala R, Golla R, Sleph PG, Fura A, An Y, Kish KF, Sack JS, Mookhtiar KA, Grover GJ, and Hamann LG. *Endocrinology*, **148** (2007) 4-12.
- [49] Hamann LG, Higuchi RI, Zhi L, Edwards JP, Wang XN, Marschke KB, Kong JW, Farmer LJ, and Jones TK. J. Med. Chem., 41 (1998) 623-639.
- [50] Edwards JP, Higuchi RI, Winn DT, Pooley CL, Caferro TR, Hamann LG, Zhi L, Marschke KB, Goldman ME, and Jones TK. *Bioorg. Med. Chem. Lett.*, 9 (1999) 1003-1008.
- [51] Hamann LG, Mani NS, Davis RL, Wang XN, Marschke KB, and Jones TK. J. Med. Chem., 42 (1999) 210-212.
- [52] Higuchi RI, Edwards JP, Caferro TR, Ringgenberg JD, Kong JW, Hamann LG, Arienti KL, Marschke KB, Davis RL, Farmer LJ, and Jones TK. *Bioorg. Med. Chem. Lett.*, 9 (1999) 1335-1340.
- [53] Kong JW, Hamann LG, Ruppar DA, Edwards JP, Marschke KB, and Jones TK. Bioorg. Med. Chem. Lett., 10 (2000) 411-414.
- [54] van Oeveren A, Motamedi M, Mani NS, Marschke KB, Lopez FJ, Schrader WT, Negro-Vilar A, and Zhi L. J. Med. Chem., 49 (2006) 6143-6146.
- [55] van Oeveren A, Pio BA, Tegley CM, Higuchi RI, Wu M, Jones TK, Marschke KB, Negro-Vilar A, and Zhi L. *Bioorg. Med. Chem. Lett.*, **17** (2007) 1523-1526.
- [56] Zhi L, Tegley CM, Marschke KB, and Jones TK. Bioorg. Med. Chem. Lett., 9 (1999) 1009-1012.
- [57] Edwards JP, West SJ, Pooley CL, Marschke KB, Farmer LJ, and Jones TK. Bioorg. Med. Chem. Lett., 8 (1998) 745-750.
- [58] Chabbert-Buffet N, Meduri G, Bouchard P, and Spitz IM. Hum. Reprod. Update, 11 (2005) 293-307.
- [59] Combs DW, Reese K, Cornelius LA, Gunnet JW, Cryan EV, Granger KS, Jordan JJ, and Demarest KT. J. Med. Chem., 38 (1995) 4880-4884.
- [60] Combs DW, Reese K, and Phillips A. J. Med. Chem., 38 (1995) 4878-4879.

- [61] Palmer S, Campen CA, Allan GF, Rybczynski P, Haynes-Johnson D, Hutchins A, Kraft P, Kiddoe M, Lai M, Lombardi E, Pedersen P, Hodgen G, and Combs DW. J. Steroid Biochem. Mol. Biol., 75 (2000) 33-42.
- [62] Jones DG, Liang X, Stewart EL, Noe RA, Kallander LS, Madauss KP, Williams SP, Thompson SK, Gray DW, and Hoekstra WJ. *Bioorg. Med. Chem. Lett.*, 15 (2005) 3203-3206.
- [63] Edwards JP, West SJ, Marschke KB, Mais DE, Gottardis MM, and Jones TK. J. Med. Chem., 41 (1998) 303-310.
- [64] Hamann LG, Winn DT, Pooley CL, Tegley CM, West SJ, Farmer LJ, Zhi L, Edwards JP, Marschke KB, Mais DE, Goldman ME, and Jones TK. *Bioorg. Med. Chem. Lett.*, 8 (1998) 2731-2736.
- [65] Pooley CL, Edwards JP, Goldman ME, Wang MW, Marschke KB, Crombie DL, and Jones TK. *J. Med. Chem.*, **41** (1998) 3461-3466.
- [66] Tegley CM, Zhi L, Marschke KB, Gottardis MM, Yang Q, and Jones TK. J. Med. Chem., 41 (1998) 4354-4359.
- [67] Zhi L, Tegley CM, Kallel EA, Marschke KB, Mais DE, Gottardis MM, and Jones TK. J. Med. Chem., 41 (1998) 291-302.
- [68] Zhi L, Tegley CM, Edwards JP, West SJ, Marschke KB, Gottardis MM, Mais DE, and Jones TK. *Bioorg. Med. Chem. Lett.*, 8 (1998) 3365-3370.
- [69] Zhi L, Tegley CM, Marschke KB, Mais DE, and Jones TK. J. Med. Chem., 42 (1999) 1466-1472.
- [70] Zhi L, Tegley CM, Pio B, West SJ, Marschke KB, Mais DE, and Jones TK. Bioorg. Med. Chem. Lett., 10 (2000) 415-418.
- [71] Zhi L, Ringgenberg JD, Edwards JP, Tegley CM, West SJ, Pio B, Motamedi M, Jones TK, Marschke KB, Mais DE, and Schrader WT. *Bioorg. Med. Chem. Lett.*, 13 (2003) 2075-2078.
- [72] Zhi L, Tegley CM, Pio B, Edwards JP, Jones TK, Marschke KB, Mais DE, Risek B, and Schrader WT. *Bioorg. Med. Chem. Lett.*, **13** (2003) 2071-2074.
- [73] Zhi L, Tegley CM, Pio B, Edwards JP, Motamedi M, Jones TK, Marschke KB, Mais DE, Risek B, and Schrader WT. J. Med. Chem., 46 (2003) 4104-4112.
- [74] Zhang P, Terefenko EA, Fensome A, Zhang Z, Zhu Y, Cohen J, Winneker R, Wrobel J, and Yardley J. *Bioorg. Med. Chem. Lett.*, **12** (2002) 787-790.
- [75] Zhang P, Terefenko EA, Fensome A, Wrobel J, Winneker R, and Zhang Z. Bioorg. Med. Chem. Lett., 13 (2003) 1313-1316.
- [76] Zhang P, Terefenko EA, Fensome A, Wrobel J, Winneker R, Lundeen S, Marschke KB, and Zhang Z. J. Med. Chem., 45 (2002) 4379-4382.
- [77] Collins MA, Hudak V, Bender R, Fensome A, Zhang P, Miller L, Winneker RC, Zhang Z, Zhu Y, Cohen J, Unwalla RJ, and Wrobel J. *Bioorg. Med. Chem. Lett.*, 14 (2004) 2185-2189.

- [78] Fensome A, Bender R, Chopra R, Cohen J, Collins MA, Hudak V, Malakian K, Lockhead S, Olland A, Svenson K, Terefenko EA, Unwalla RJ, Wilhelm JM, Wolfrom S, Zhu Y, Zhang Z, Zhang P, Winneker RC, and Wrobel J. J. Med. Chem., 48 (2005) 5092-5095.
- [79] Chen J, Kim J, and Dalton JT. Mol Interv, 5 (2005) 173-188.
- [80] Yin D, Gao W, Kearbey JD, Xu H, Chung K, He Y, Marhefka CA, Veverka KA, Miller DD, and Dalton JT. *J. Pharmacol. Exp. Ther.*, **304** (2003) 1334-1340.
- [81] Zhang X, Li X, Allan GF, Sbriscia T, Linton O, Lundeen SG, and Sui Z. Bioorg. Med. Chem. Lett., 17 (2007) 439-443.
- [82] Miner JN, Chang W, Chapman MS, Finn PD, Hong MH, Lopez FJ, Marschke KB, Rosen J, Schrader W, Turner R, van Oeveren A, Viveros H, Zhi L, and Negro-Vilar A. *Endocrinology*, 148 (2007) 363-373.
- [83] Gao W and Dalton JT. Drug Discov. Today, **12** (2007) 241-248.
- [84] Zhou ZX, Wong CI, Sar M, and Wilson EM. Recent Prog. Horm. Res., 49 (1994) 249-274.
- [85] Lavery DN and McEwan IJ. Biochem. J., 391 (2005) 449-464.
- [86] Jenster G, van der Korput HA, Trapman J, and Brinkmann AO. J. Biol. Chem., 270 (1995) 7341-7346.
- [87] Simental JA, Sar M, Lane MV, French FS, and Wilson EM. J. Biol. Chem., 266 (1991) 510-518.
- [88] He B, Kemppainen JA, Voegel JJ, Gronemeyer H, and Wilson EM. J. Biol. Chem., 274 (1999) 37219-37225.
- [89] Kumar R, Betney R, Li J, Thompson EB, and McEwan IJ. *Biochemistry*, **43** (2004) 3008-3013.
- [90] Reid J, Kelly SM, Watt K, Price NC, and McEwan IJ. J. Biol. Chem., 277 (2002) 20079-20086.
- [91] Luisi BF, Xu WX, Otwinowski Z, Freedman LP, Yamamoto KR, and Sigler PB. Nature, 352 (1991) 497-505.
- [92] Roemer SC, Donham DC, Sherman L, Pon VH, Edwards DP, and Churchill ME. Mol. Endocrinol., 20 (2006) 3042-3052.
- [93] Schwabe JW, Chapman L, Finch JT, and Rhodes D. Cell, 75 (1993) 567-578.
- [94] Shaffer PL, Jivan A, Dollins DE, Claessens F, and Gewirth DT. Proc. Natl. Acad. Sci. USA, 101 (2004) 4758-4763.
- [95] Freedman LP, Luisi BF, Korszun ZR, Basavappa R, Sigler PB, and Yamamoto KR. Nature, 334 (1988) 543-546.
- [96] Dahlman-Wright K, Wright A, Gustafsson JA, and Carlstedt-Duke J. J. Biol. Chem., 266 (1991) 3107-3112.
- [97] Jenster G, Trapman J, and Brinkmann AO. *Biochem. J.*, **293** (1993) 761-768.
- [98] Zhou ZX, Sar M, Simental JA, Lane MV, and Wilson EM. J. Biol. Chem., 269 (1994) 13115-13123.

- [99] Bourguet W, Germain P, and Gronemeyer H. Trends Pharmacol. Sci., 21 (2000) 381-388.
- [100] Wurtz JM, Bourguet W, Renaud JP, Vivat V, Chambon P, Moras D, and Gronemeyer H. Nat. Struct. Biol., 3 (1996) 87-94.
- [101] Matias PM, Donner P, Coelho R, Thomaz M, Peixoto C, Macedo S, Otto N, Joschko S, Scholz P, Wegg A, Basler S, Schafer M, Egner U, and Carrondo MA. *J. Biol. Chem.*, 275 (2000) 26164-26171.
- [102] Williams SP and Sigler PB. Nature, **393** (1998) 392-396.
- [103] Sack JS, Kish KF, Wang C, Attar RM, Kiefer SE, An Y, Wu GY, Scheffler JE, Salvati ME, Krystek SR, Jr., Weinmann R, and Einspahr HM. Proc. Natl. Acad. Sci. USA, 98 (2001) 4904-4909.
- [104] Heery DM, Kalkhoven E, Hoare S, and Parker MG. Nature, 387 (1997) 733-736.
- [105] Shiau AK, Barstad D, Loria PM, Cheng L, Kushner PJ, Agard DA, and Greene GL. Cell, 95 (1998) 927-937.
- [106] He B and Wilson EM. Mol. Cell. Biol., 23 (2003) 2135-2150.
- [107] Bohl CE, Miller DD, Chen J, Bell CE, and Dalton JT. J. Biol. Chem., 280 (2005) 37747-37754.
- [108] Pereira de Jesus-Tran K, Cote PL, Cantin L, Blanchet J, Labrie F, and Breton R. *Protein Sci.*, **15** (2006) 987-999.
- [109] Zhang Z, Olland AM, Zhu Y, Cohen J, Berrodin T, Chippari S, Appavu C, Li S, Wilhem J, Chopra R, Fensome A, Zhang P, Wrobel J, Unwalla RJ, Lyttle CR, and Winneker RC. J. Biol. Chem., 280 (2005) 28468-28475.
- [110] Wang F, Liu XQ, Li H, Liang KN, Miner JN, Hong M, Kallel EA, van Oeveren A, Zhi L, and Jiang T. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun., 62 (2006) 1067-1071.
- [111] Cantin L, Faucher F, Couture JF, de Jesus-Tran KP, Legrand P, Ciobanu LC, Frechette Y, Labrecque R, Singh SM, Labrie F, and Breton R. J. Biol. Chem., 282 (2007) 30910-30919.
- [112] Madauss KP, Deng SJ, Austin RJ, Lambert MH, McLay I, Pritchard J, Short SA, Stewart EL, Uings IJ, and Williams SP. J. Med. Chem., 47 (2004) 3381-3387.
- [113] Hara T, Miyazaki J, Araki H, Yamaoka M, Kanzaki N, Kusaka M, and Miyamoto M. Cancer Res., 63 (2003) 149-153.
- [114] Suzuki H, Akakura K, Komiya A, Aida S, Akimoto S, and Shimazaki J. Prostate, 29 (1996) 153-158.
- [115] Veldscholte J, Berrevoets CA, Ris-Stalpers C, Kuiper GG, Jenster G, Trapman J, Brinkmann AO, and Mulder E. J. Steroid. Biochem. Mol. Biol., 41 (1992) 665-669.
- [116] Culig Z, Hoffmann J, Erdel M, Eder IE, Hobisch A, Hittmair A, Bartsch G, Utermann G, Schneider MR, Parczyk K, and Klocker H. Br. J. Cancer, 81 (1999) 242-251.

- [117] Taplin ME, Rajeshkumar B, Halabi S, Werner CP, Woda BA, Picus J, Stadler W, Hayes DF, Kantoff PW, Vogelzang NJ, and Small EJ. J. Clin. Oncol., 21 (2003) 2673-2678.
- [118] Bohl CE, Wu Z, Miller DD, Bell CE, and Dalton JT. J. Biol. Chem., 282 (2007) 13648-13655.
- [119] He B, Bowen NT, Minges JT, and Wilson EM. J. Biol. Chem., 276 (2001) 42293-42301.
- [120] He B, Gampe RT, Jr., Kole AJ, Hnat AT, Stanley TB, An G, Stewart EL, Kalman RI, Minges JT, and Wilson EM. *Mol. Cell*, 16 (2004) 425-438.
- [121] He B, Kemppainen JA, and Wilson EM. J. Biol. Chem., 275 (2000) 22986-22994.
- [122] He B, Minges JT, Lee LW, and Wilson EM. J. Biol. Chem., 277 (2002) 10226-10235.
- [123] Zhou ZX, He B, Hall SH, Wilson EM, and French FS. *Mol. Endocrinol.*, 16 (2002) 287-300.
- [124] Hsu CL, Chen YL, Yeh S, Ting HJ, Hu YC, Lin H, Wang X, and Chang C. J. Biol. Chem., 278 (2003) 23691-23698.
- [125] He B, Lee LW, Minges JT, and Wilson EM. J. Biol. Chem., 277 (2002) 25631-25639.
- [126] Alen P, Claessens F, Verhoeven G, Rombauts W, and Peeters B. Mol. Cell. Biol., 19 (1999) 6085-6097.
- [127] Bevan CL, Hoare S, Claessens F, Heery DM, and Parker MG. Mol. Cell. Biol., 19 (1999) 8383-8392.
- [128] Madauss KP, Grygielko ET, Deng SJ, Sulpizio AC, Stanley TB, Wu C, Short SA, Thompson SK, Stewart EL, Laping NJ, Williams SP, and Bray JD. Mol. Endocrinol., 21 (2007) 1066-1081.
- [129] Estebanez-Perpina E, Moore JM, Mar E, Delgado-Rodrigues E, Nguyen P, Baxter JD, Buehrer BM, Webb P, Fletterick RJ, and Guy RK. J. Biol. Chem., 280 (2005) 8060-8068.
- [130] Hur E, Pfaff SJ, Payne ES, Gron H, Buehrer BM, and Fletterick RJ. *PLoS Biol.*, 2 (2004) 1303-1312.
- [131] Bledsoe RK, Montana VG, Stanley TB, Delves CJ, Apolito CJ, McKee DD, Consler TG, Parks DJ, Stewart EL, Willson TM, Lambert MH, Moore JT, Pearce KH, and Xu HE. *Cell*, **110** (2002) 93-105.
- [132] Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson JA, and Carlquist M. *Nature*, **389** (1997) 753-758.
- [133] Kauppi B, Jakob C, Farnegardh M, Yang J, Ahola H, Alarcon M, Calles K, Engstrom O, Harlan J, Muchmore S, Ramqvist AK, Thorell S, Ohman L, Greer J, Gustafsson JA, Carlstedt-Duke J, and Carlquist M. J. Biol. Chem., 278 (2003) 22748-22754.

- [134] Bohl CE, Gao W, Miller DD, Bell CE, and Dalton JT. Proc. Natl. Acad. Sci. USA, 102 (2005) 6201-6206.
- [135] Perissi V, Staszewski LM, McInerney EM, Kurokawa R, Krones A, Rose DW, Lambert MH, Milburn MV, Glass CK, and Rosenfeld MG. *Genes Dev.*, 13 (1999) 3198-3208.
- [136] Heldring N, Pawson T, McDonnell D, Treuter E, Gustafsson JA, and Pike AC. J. Biol. Chem., 282 (2007) 10449-10455.
- [137] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, and Bourne PE. Nucleic Acid Res., 28 (2000) 235-242.
- [138] Charifson PS and Walters WP. Mol. Divers., 5 (2002) 185-197.
- [139] Lipinski CA, Lombardo F, Dominy BW, and Feeney PJ. Adv. Drug Del. Rev., 23 (1997) 3-25.
- [140] Tanimoto TT. IBM Internal Report 17th Nov., (1957)
- [141] Brooijmans N and Kuntz ID. Annu. Rev. Biophys. Biomol. Struct., 32 (2003) 335-373.
- [142] Kitchen DB, Decornez H, Furr JR, and Bajorath J. Nat. Rev. Drug Discov., 3 (2004) 935-949.
- [143] Jain AN. Curr. Protein Pept. Sci., 7 (2006) 407-420.
- [144] Bohl CE, Chang C, Mohler ML, Chen J, Miller DD, Swaan PW, and Dalton JT. J. Med. Chem., 47 (2004) 3765-3776.
- [145] Sippl W. Bioorg. Med. Chem., 10 (2002) 3741-3755.
- [146] Tervo AJ, Nyrönen TH, Rönkko T, and Poso A. J. Chem. Inf. Comput. Sci., 44 (2004) 807-816.
- [147] Wolohan P and Reichert DE. J. Comput. Aided Mol. Des., 17 (2003) 313-328.
- [148] Taylor RD, Jewsbury PJ, and Essex JW. J. Comput. Aided Mol. Des., 16 (2002) 151-166.
- [149] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, and Ferrin TE. J. Mol. Biol., 161 (1982) 269-288.
- [150] Rarey M, Kramer B, Lengauer T, and Klebe G. J. Mol. Biol., 261 (1996) 470-489.
- [151] Ewing TJ, Makino S, Skillman AG, and Kuntz ID. J. Comput. Aided Mol. Des., 15 (2001) 411-428.
- [152] Goodsell DS and Olson AJ. Proteins, 8 (1990) 195-202.
- [153] Jones G, Willett P, Glen RC, Leach AR, and Taylor R. J. Mol. Biol., 267 (1997) 727-748.
- [154] Alonso H, Bliznyuk AA, and Gready JE. Med. Res. Rev., 26 (2006) 531-568.
- [155] Kontoyianni M, McClellan LM, and Sokol GS. J. Med. Chem., 47 (2004) 558-565.
- [156] Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, and Head MS. J. Med. Chem., 49 (2006) 5912-5931.

- [157] Wang R, Lu Y, Fang X, and Wang S. J. Chem. Inf. Comput. Sci., 44 (2004) 2114-2125.
- [158] Wang R, Lu Y, and Wang S. J. Med. Chem., 46 (2003) 2287-2303.
- [159] Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, and Taylor RD. Proteins, 52 (2003) 609-623.
- [160] Bissantz C, Folkers G, and Rognan D. J. Med. Chem., 43 (2000) 4759-4767.
- [161] Charifson PS, Corkery JJ, Murcko MA, and Walters WP. J. Med. Chem., 42 (1999) 5100-5109.
- [162] Clark RD, Strizhev A, Leonard JM, Blake JF, and Matthew JB. J. Mol. Graph. Model., 20 (2002) 281-295.
- [163] Yang JM, Chen YF, Shen TW, Kristal BS, and Hsu DF. J. Chem. Inf. Model., 45 (2005) 1134-1146.
- [164] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, and Shenkin PS. J. Med. Chem., 47 (2004) 1739-1749.
- [165] Schrödinger Suite 2007 Induced Fit Docking protocol; *Glide*, version 4.5, LLC, New York, NY, 2005; *Prime*, version 1.6, Schrödinger, LLC, New York, NY, 2005.
- [166] Sherman W, Day T, Jacobson MP, Friesner RA, and Farid R. J. Med. Chem., 49 (2006) 534-553.
- [167] SYBYL, Tripos Inc., St. Louis, MO, USA.
- [168] Wang R, Lai L, and Wang S. J. Comput. Aided Mol. Des., 16 (2002) 11-26.
- [169] Esposito EX, Hopfinger AJ, and Madura JD. 3D- and nD-QSAR methods, In Handbook of Chemoinformatics: From Data to Knowledge, Gasteiger J, Ed., Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2003, pp. 1576-1603.
- [170] Kubinyi H. Comparative Molecular Field Analysis (CoMFA), In Handbook of Chemoinformatics: From Data to Knowledge, Gasteiger J, Ed., Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2003, pp. 1555-1574.
- [171] Kubinyi H. QSAR in drug design, In Handbook of Chemoinformatics: From Data to Knowledge, Gasteiger J, Ed., Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2003, pp. 1532-1554.
- [172] Hopfinger AJ, Wang S, Tokarski JS, Jin B, Albuquerque M, Madhav PJ, and Duraiswami C. J. Am. Chem. Soc., 119 (1997) 10509-10524.
- [173] Vedani A and Dobler M. J. Med. Chem., 45 (2002) 2139-2149.
- [174] Hansch C and Fujita T. J. Am. Chem. Soc., 86 (1964) 1616-1626.
- [175] Cramer III RD, Patterson DE, and Bunce JD. J. Am. Chem. Soc., **110** (1988) 5959-5967.
- [176] Klebe G, Abraham U, and Mietzner T. J. Med. Chem., 37 (1994) 4130-4146.
- [177] Klebe G and Abraham U. J. Comput. Aided Mol. Des., 13 (1999) 1-10.

- [178] Viswanadhan VN, Ghose AK, Revankar GR, and Robins RK. J. Chem. Inf. Comput. Sci., 29 (1989) 163-172.
- [179] Wold S, Ruhe A, Wold H, and Dunn WJ. SLAM J. Sci. Stat. Comput., 5 (1984) 735-743.
- [180] Höltje HD, Sippl W, Rognan D, and Folkers G. Molecular modeling: Basic Principles and Applications. 2nd ed., Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2003.
- [181] Golbraikh A and Tropsha A. J. Mol. Graph. Model., 20 (2002) 269-276.
- [182] Poujol N, Wurtz JM, Tahiri B, Lumbroso S, Nicolas JC, Moras D, and Sultan C. J. Biol. Chem., 275 (2000) 24022-24031.
- [183] Taplin ME, Bubley GJ, Ko YJ, Small EJ, Upton M, Rajeshkumar B, and Balk SP. *Cancer Res.*, **59** (1999) 2511-2515.