universität wien

DIPLOMARBEIT

Binding studies on estrogen receptor alpha

Verfasserin

Katharina Maria Ostermann

angestrebter akademischer Grad

Magistra der Naturwissenschaften

Wien, 2009

Studienkennzahl lt. Studienblatt:	A 419
Studienrichtung lt. Studienblatt:	Chemie
Betreuerin / Betreuer:	Ao. Univ.-Prof. Dr. Karl Peter Wolschann

TABLE OF CONTENTS

Danksagung

Mein herzlicher Dank gilt meinen Bertreuern Ao. Univ.-Prof. Dr. Karl Peter Wolschann am Institut für theoretische Chemie und Strahlenchemie an der Universität Wien und Dr. Thomas Stockner am Austrian Institute of Technology, Seibersdorf, für die engagierte Betreuung meiner Diplomarbeit und die zahlreichen wertvollen Anregungen. Ihr fachlicher Beistand war mir beim Erstellen dieser Arbeit eine wesentliche Unterstützung.

Danken möchte ich meinen Eltern für die mentale und finanzielle Unterstützung während meiner gesamten Studienzeit.

Bei meiner ganzen Familie, besonders bei meiner Schwester Bettina, und auch meinen Freunden möchte ich mich bedanken, für alle aufbauenden Worte und guten Ratschläge, die mir auf dem Weg zum Abschluss dieses Studiums sehr geholfen haben.

Die Arbeit ist meinen Eltern gewidmet.

Abstract:

The increasing number of characterized molecular receptors provides the basis for structure based design of active compounds which may be developed to become a new drug. In drug design one starts from a known or hypothetical mode of action or binding mechanism, a lead structure is rationally designed and afterwards tested experimentally.

In contrast to experimental high throughput screening (HTS), which provides hits in terms of chemical compounds, virtual screening (VS) simply suggests computer hits. The compounds are selected on the basis of improved docking algorithms using an approximate energy function to rank them as putative hits.

VS runs are usually validated by comparing the performance of a set of known actives with a large set of “randomly” picked compounds (decoy structures) which are inactive. All structures are submitted to the selected VS protocol, and the performance-ranks of the known actives with respect to the remaining pool are converted into enrichment plots. These plots are accumulation curves that show how the fraction of actives recovered varies with the percent of the database screened.

Estrogens are involved in the growth, development and homeostasis of several tissues. They exert their physiological effects via the estrogen receptor (ER) which is associated with diseases like breast cancer, osteoporosis, neurodegenerative and cardiovascular diseases as well as obesity.

In this work the estrogen receptor was used for a series of docking studies. Agonist and antagonist- bound receptor structures were used, once obtained from the Protein Data Bank (PDB), once from molecular dynamic simulations (MD-simulations).

In the first docking study the receptor´s ligands were docked back into the receptor structures which is referred to as ‘bound docking’. A better docking score was obtained for ligands that docked in a correct position and vice versa.

In the second docking study ligands and decoys were docked into the receptor structures to see whether the performance is able to put potential binders near the top of a score ranked list. The results were validated by enrichment plots.

The study revealed a better enrichment for docking into structures obtained from the PDB than for docking into MD-simulation structures. The final step was ensemble docking which means that for each ligand and decoy docked into all receptors only

the best docking score was used. For both, docking into PDB-crystalstructures and into MD-simulation structures, ensemble docking produced a better enrichment. For MD-simulation structures, ensemble docking gave a result, which was nearly as good as those for the single PDB-structures.

1. Introduction

1.1. Nuclear Receptors

Nuclear receptors are ligand-regulated transcription factors that modulate target gene transcription. These transcription factors, many of which function as receptors for lipophilic hormones, control differentiation, development, homeostasis and behavior. All nuclear receptors are structurally related and belong to the nuclear receptor superfamily. [Gronemeyer2004]

1.1.1. Nuclear Receptors functional domains

A typical nuclear receptor is composed of several functional domains:

A variable NH2-terminal region (A/B-region) which contains the autonomous transcriptional activation function (AF1), which is isoform-specific as well as cell and promoter-specific, suggesting that it is likely to contribute to the specificity of action among different receptor isoforms; a DNA-binding domain (C-region); a hinge region (D-region); a ligand binding domain (COOH-terminal conserved E/F-region) which contains the ligand dependent activation function (AF2) and the dimerization surface. The ligand binding domain (LBD) is the one where coregulator interaction takes place. [Aranda2001]

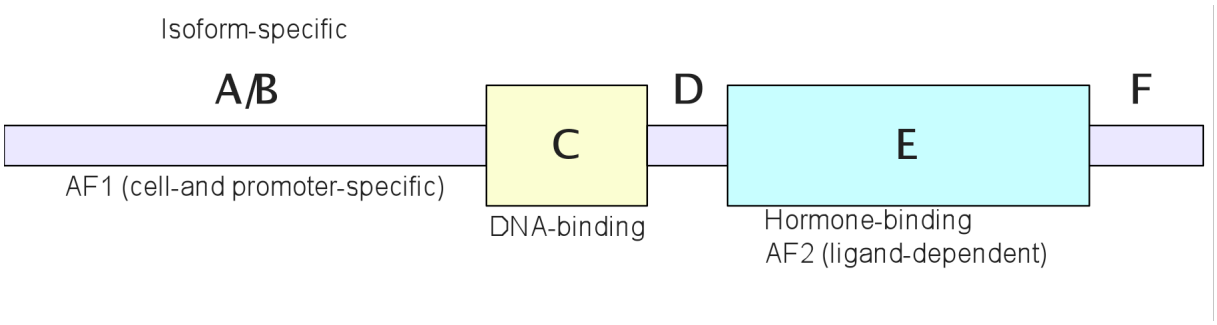


Figure 1: Nuclear receptor domains; AF, activation function.

A/B-region:

The size of the A/B-domain is quite variable and ranges from several hundred amino acids in length for the steroid hormones to only a few amino acids in some of the nonsteroid nuclear receptors. [Burris, ex Nuclear Receptors and Genetic Disease]

Factors which modulate the response of nuclear receptors towards their ligands are alternative splicing and different use of promoters. These factors generate receptors with different A/B-regions and therefore different AF1 functions.
This domain is also the target for phosphorylation and therefore posttranscriptional modifications.
For example nuclear receptors such as the estrogen receptors (ERs) are phosphorylated at serine or threonine residues by the mitogen-activated protein kinase (MAPK) in vitro, and in cells treated with growth factors that stimulate the Ras-MAPK cascade. This phosphorylation enhances transcriptional activity. [Aranda2001]

C-region:

The DNA-binding domain (DBD) is the most conserved domain of nuclear receptors and includes two zinc finger modules. It has the ability to recognize specific target sequences and activate genes.
Amino acid sequences important for dimerization are contained within the DBD (as well as in the LBD). [Olefsky2001]

D-region:

The D-region, often called hinge region, because it is localized between the DBD and the LBD, often harbors nuclear localization signals (NLS) and also residues whose mutation is connected with the loss of interaction with nuclear receptor corepressors. [Aranda2001]

E/F-region:

The LBD is a multifunctional domain that mediates homo- and heterodimarization, interaction with heat shock proteins (HSPs), ligand dependent transcriptional activity and also hormone reversible transcriptional repression.
The LBDs are formed by 12 alpha-helical regions numbered from H1 to H12. [Aranda2001]
11 helices build a compact structure comprising a ligand binding pocket. The entrance to the pocket is guarded by helix 12 which carries the AF2 transactivation function. [Gronemeyer2004]

In the absence of ligand, H12 has been proposed to be exposed to solvent. [Gangloff2001]
The E-domain is the second most conserved region among nuclear receptors. [Burris; ex Nuclear Receptors and Genetic Disease]

The function of the F-domain which is present in the estrogen receptor α and β as well as in retinoic acid receptor is unknown. [Burris; ex Nuclear Receptors and Genetic Disease]

Ligand binding:

Ligand binding initiates a conformational change that results in a different orientation of H12 on the core of the LBD, closing the ligand binding pocket like a lid (mouse trap mechanism). [Gangloff2001]
The flexibility of H12 allows for entry of the ligand into the ligand-binding cavity after displacement of the H12 lid. When the ligand has entered the cavity which is lined with hydrophobic amino acid residues, contacts are made, followed by the conformational shift which also includes closing the H12 lid. [Burris; ex Nuclear Receptors and Genetic Disease]
The orientation of H12 is a consequence of allosteric effects induced by the chemical structure of the specific ligand that is binding. [Gronemeyer2004]

Transactivation:

The effect of a nuclear receptor is to increase the rate of transcription of target genes via interaction with the *cis*-acting DNA response element, a mechanism called transactivation, the activation function (AF1 and AF2) being located in the A/B-region and the E-region, respectively. The two transactivation domains are functionally independent.
AF1 transactivates in a constitutive and hormone-independent manner whereas AF2 acts as a hormone-inducible transactivation domain where the ligand-induced conformational change reveals the activation function. Crystal structure studies indicate, that the AF2 helix is helix 12, the one that reorientates upon ligand binding. In the unliganded state, the AF2 helix projects away from the core of the LBD. Upon

ligand binding the helix folds back on the surface of the LBD and forms a cleft that is able to recruit coactivators which are necessary for transcriptional activation. [Burris; ex Nuclear Receptors and Genetic Disease]

Mechanism of transactivation:

It is believed that nuclear receptors utilize several mechanisms to increase the rate of transcription of target genes. They may directly interact with or recruit general transcription factors that are components of the preinitiation complex (PIC) to increase the rate of initiation of transcription.
The receptors could also act by altering chromatin structure to make the DNA more accessible to various transcription factors and to RNA polymerase II (PolII).
Another suggestion is that receptors interact with components that act to bridge interactions with members of the PIC and proteins that alter chromatin structure.
[Burris; ex Nuclear Receptors and Genetic Disease]

Several steroid receptor coactivators display histone acetyltransferase (HAT) activity which indicates that acetylation of histones is an important mechanism in transactivation. Histone deacetylation is responsible for the chromatin condensation that accounts for the gene-silencing effect of apo-receptors (receptors without ligand). [Gronemeyer2004]

1.1.2. Nuclear receptor subfamilies

Nuclear receptors (NRs) are divided into 7 subfamilies based on amino acid sequence identity. The nomenclature system is based on the evolution of the two conserved domains (DNA- and ligand binding domain) of the NRs.
The nuclear receptor superfamily is divided into 6 subfamilies and 26 groups of receptors. Receptors that contain only one of the two well conserved domains are grouped in subfamily 0. [Laudet1997]

Subfamily 1 which contains eleven groups: TR (thyroid nuclear receptors), RAR (retinoic-acid receptors), PPAR (peroxisome-proliferator-activated receptors), REV-

ERB, E78, RZR/ROR, *Caenorhabditis* CNR14, ECR, VDR, *Drosophila* DHR96 orphan receptor and nematode NHR1 orphan receptor. [Laudet1997]

Subfamily 2 which contains seven groups: HNF4 (hepatocyte nuclear factor 4), RXR (retinoid X receptor), TR2/4 (testicular receptor), *Drosophila* DHR78, TLL, COUP-TF, EAR2. [Laudet1997]

Subfamily 3 which contains three groups: estrogen receptors, ERR (estrogen-related receptors), steroid receptors. [Laudet1997]

Subfamily 4 which to date contains one group: TGFIB. [Laudet1997]

Subfamily 5 which contains two groups: FTZ-F1, DHR39. [Laudet1997]

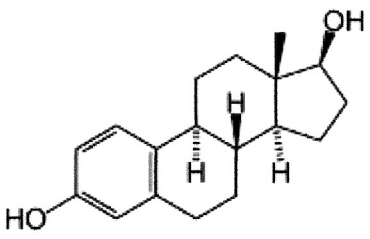
Subfamily 6 which contains one group: GCNF1. [Laudet1997]

Subfamily 0 which contains two groups, one of them lacking the C-domain and one lacking the E-domain. [Laudet1997]

1.1.3. Agonist and Antagonist

The ER's natural ligand 17β-estradiol (E2) acts as a pure agonist in ER α and ER β. [Gangloff2001]

17β-estradiol:



ER- agonist and antagonist bind at the same site within the core of the LBD but show different binding modes. [Avendaño, ex Medicinal Chemistry of Anticancer Drugs]

Ligands can be designed to have different degrees of agonism or antagonism. In the absence of ligand, the LBD of many nuclear receptors is bound to a set of transcriptional corepressors (proteins which recruit histone deacetylases (HDAC)).
An agonistic ligand is responsible for complex allosteric effects that finally lead to corepressor complex dissociation. This holo positioning of H12 (structure with bound ligand) allows coactivators to interact with short LxxLL-like motifs (nuclear-receptor boxes, L=leucine, X=any amino acid) that exist in most coactivators and are common motifs for interaction with nuclear receptor LBDs. [Gronemeyer2004]
It has been observed that in antagonist-bound complexes, H12 is positioned in a structurally conserved cleft where the LxxLL- motif of the coactivator molecule binds. This can be explained by a mechanism for antagonism where H12 and the coactivator compete for a common binding site. [Gangloff2001]

Destabilization of the H12-protein core is the overall mechanism of pure and also partial antagonism. The dominant effect depends on the potency of the ligand to disrupt the active conformation or to prevent the correct binding of coactivators. [Gangloff2001]

Knowledge of the features responsible for inducing and stabilizing a given conformation is very important in order to understand the initial events of nuclear receptor transactivation. [Gangloff2001]

1.1.4. Estrogen receptor

Estrogens, which belong to the steroid hormones, are involved in the growth, development and homeostasis of many tissues. They exert their physiological effects via the estrogen receptor (ER). [Brzozowski1997]

In contrast to G-protein coupled membrane receptors which mediate fast reactions and translate them to long range regulation, nuclear receptors are responsible for slow genomic changes in the nucleus which leads to changes in gene activity.
[Kleine, Rossmanith, ex Hormone und Hormonsystem]

ER and also other steroid nuclear receptors, is less stable in the absence of ligand or protein cofactors like HSP90 (heat shock protein 90). 'The fold stabilization of these proteins is part of the control of gene expression and is ligand-dependent (induced fit mechanism) and controlled by the cellular context (redox potential, nature of the ligand, presence of interacting molecules like coactivators or corepressors).' [Gangloff2001]

Nuclear receptors regulate transcription by binding to specific DNA sequences in target genes. These sequences are called hormone response elements (HRE), and are normally located in regulatory regions of target genes. [Aranda2001]
The symmetry of the HREs suggests that steroid receptors bind to DNA as dimers. [Burris, ex Nuclear Receptors and Genetic Disease]

1.1.5. Nuclear receptor function

Estrogens as well as other steroid hormones act to increase RNA synthesis by stimulating RNA polymerase activity. This stimulatory effect is essential for activity which was proved by treatment of cells with RNA synthesis inhibitor Actinomycin D which resulted in the block of cellular changes induced by estrogens. [Burris, ex Nuclear Receptors and Genetic Disease]

It was demonstrated that transcription of specific, targeted messenger RNAs (mRNAs) was directly stimulated by estrogens and was the rate-limiting step in hormone-dependent induction of protein synthesis. [Burris, ex Nuclear Receptors and Genetic Disease]

The receptors for the classical steroid hormones reside primarily in the cytoplasm associated with HSPs and are unable to bind to DNA. After ligand binding the HSPs dissociate from the receptor which now homodimerizes and binds to HREs. The receptor mediates transcriptional activation of the target genes by interacting with general transcription factors and by recruitment of transcriptional coactivators. [Burris, ex Nuclear Receptors and Genetic Disease]

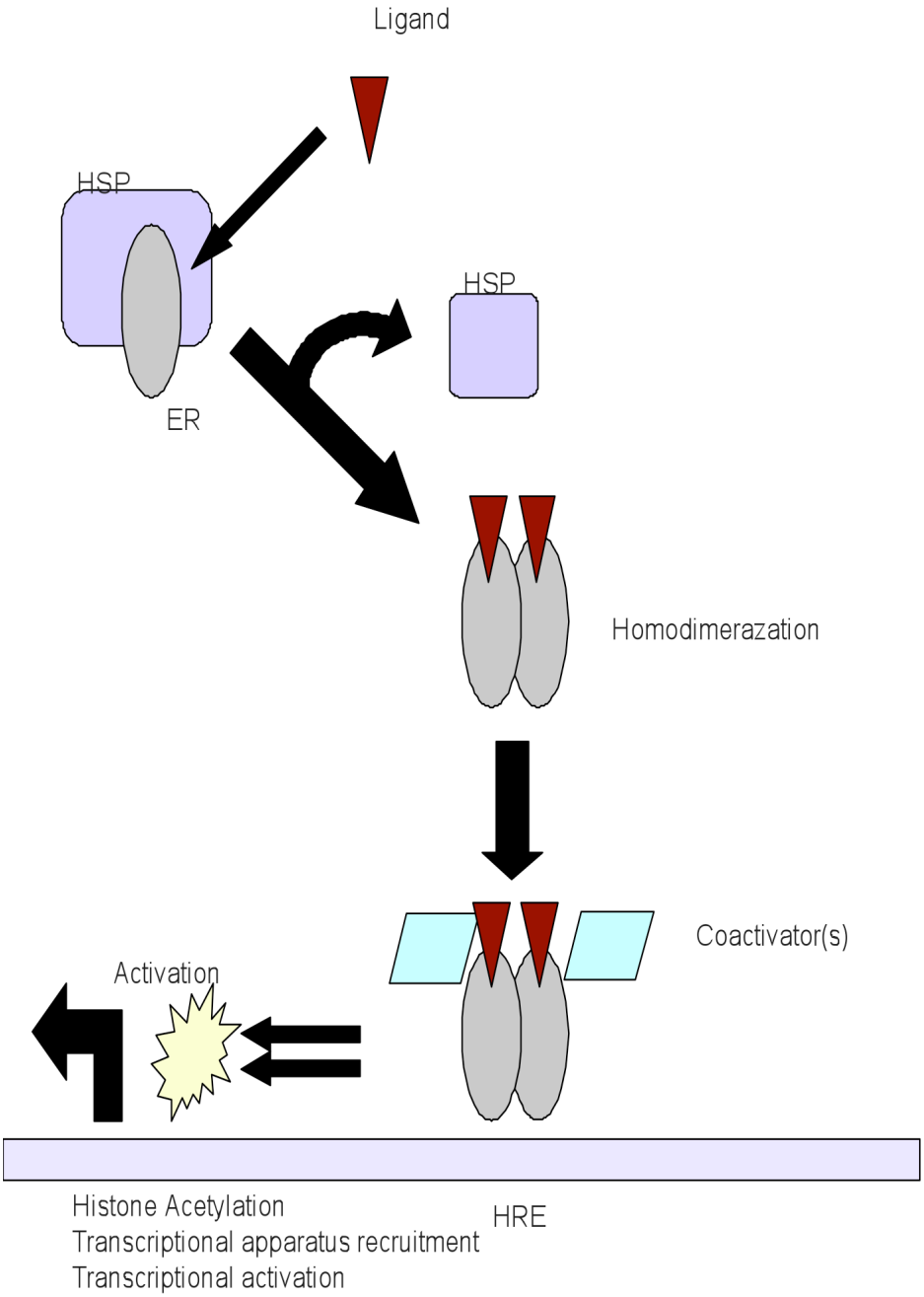


Figure 2: General model of nuclear receptor function. The ER is associated with heat shock proteins (HSPs) and not bound to DNA in the absence of ligand. Upon ligand binding the receptor dissociates from the HSPs, homodimerizes and binds to specific hormone response elements (HREs) in the promoter regions of target genes. Once localized to the promoter, the receptor mediates transcriptional activation by interacting with general transcription factors and also by recruitment of transcriptional coactivators.

1.1.6. DNA-binding

Nuclear receptors bind to DNA by recognizing a hexameric nucleotide sequence known as a core recognition motif or “half-site”. The sequence, arrangement and spacing of the half-sites define the nature and responsiveness of an HRE to various nuclear receptors. The majority of nuclear receptors bind to DNA as dimers with each of the receptors occupying one of the half-sites (Type1). The half site orientation of the ER is an inverted repeat type. Type 2 receptors usually recognize response elements organized into direct repeats. Most of them form heterodimers with retinoid X receptor (RXR) and a few form homodimers or monomers. [Burris; ex Nuclear Receptors and Genetic Disease]

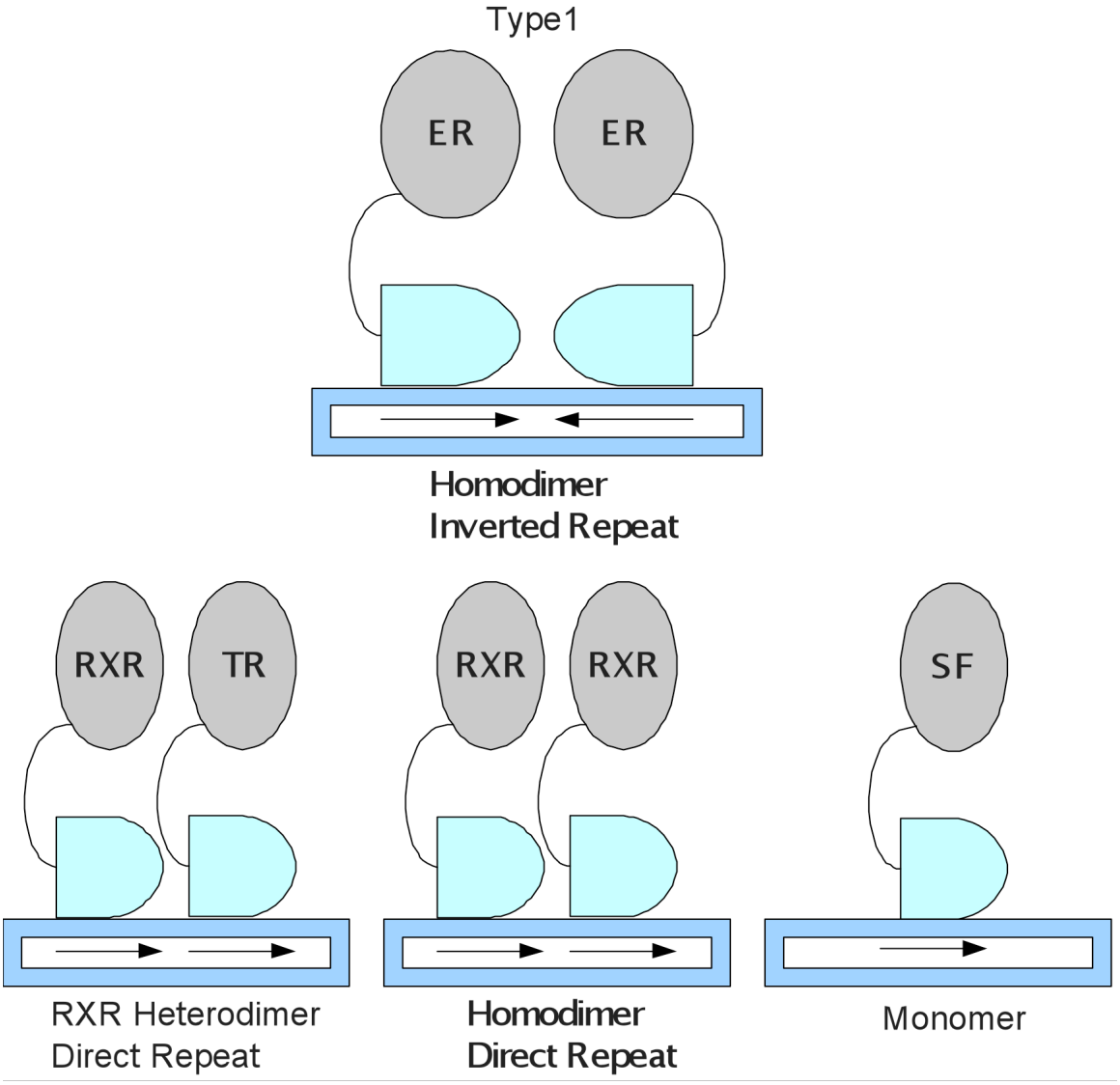


Figure 3: DNA-binding of ER (Type 1); DNA-binding of retinoid X receptor (RXR), thyroid hormone receptor (TR), steroidogenic factor (SF) (Type 2).

1.1.7. Diseases cohesive with ER

Breast cancer:

Estrogen is implicated in the development of breast cancer, based on data from both clinical and animal studies. Risk factors associated with breast cancer reflect cumulative exposure of the breast epithelium to estrogen. [Deroo2006]

There are two hypotheses trying to explain why this causes tumor formation. In the first, binding of estrogen to the ER stimulates proliferation of mammary cells, which increases the target cell number within the tissue. The increase in cell division and DNA synthesis elevates the risk of replication errors, which is supposed to lead to an increasing number of mutations that finally disrupt normal cellular processes such as apoptosis, cellular proliferation and DNA-repair. [Deroo2006]
In the second, estrogen metabolism leads to the production of genotoxic by-products (genotoxic waste) that could damage DNA directly, again resulting in point mutations. There is evidence that estrogen may act through both mechanisms to initiate and/or promote cancer. [Deroo2006]
Selective estrogen receptor modulators (SERMs) and aromatase inhibitors are used in treatment of breast cancer and those patients whose tumors are ER-positive do respond to these therapies. [Deroo2006]

The ER is also involved in several other cancers such as mammary, ovarian, colon, prostate and endometrial cancer. [Deroo2006]

Osteoporosis:

Estrogens regulate skeletal homeostasis. Osteoporosis is due to higher bone resorbtion in both sexes and is associated with estrogen deficiency. Estrogens prevent bone turnover by reducing osteoblast-mediated bone formation. Estrogen and Raloxifen are used as medication for the prevention of bone loss. [Deroo2006]

Neurodegenerative diseases:

Clinical and experimental data support the protective effect of estrogen against neurodegenerative disease in humans such as stroke, Parkinson disease and Alzheimer disease. Experimental evidence from animal and cell culture models show that estrogen treatment protects against neuronal cell death due to insult and both ER α and β are found in several parts of the human brain. [Deroo2006]

Cardiovascular disease:

The incidence of cardiovascular diseases is low in premenopausal women but increases after menopause, suggesting that estrogen protects the female cardiovascular system. [Deroo2006]

Obesity:

Obesity results from excess with adipose tissue, which is considered to be an endocrine organ because of its ability to metabolize steroid hormones. Estrogens regulate the metabolism and the location of white adipose tissue and are involved in adipogenesis, adipose deposition, lipogenesis, lipolysis and adipocyte proliferation. In women the loss of circulating estrogen after menopause is associated with more central body fat. This effect can be attenuated by estrogen treatment. [Deroo2006]

1.1.8. ER alpha and beta

There are 2 subtypes of ER (α and β), which are products of distinct genes, ESR1 and ESR2, that are found at different chromosomal locations. [Deroo2006]
They display structural differences and can mediate overlapping but different sets of biologic functions. The two subtypes can interact with the same ERE (estrogen response elements) and can also form heterodimers, suggesting that in cells that express both ER subtypes, the ratio of the two will effect estrogen action. [Olefsky2001]
ER α is found in endometrium, breast cancer cells, ovarian storma cells and in the hypothalamus. [Yaghmaie2005]

The expression of ER β protein has been documented in kidney, brain, bone and heart. [Babiker2002]

ER α and ER β regulate gene transcription by binding to specific estrogen response elements (EREs) in the promoter of target genes, or by binding to other transcription factors acting at coactivator protein 1 (AP1) and specificity protein 1 (SP1) sites. Pathways in proliferation may be influenced by the interaction of ER with AP1 and SP1. It could be proposed that estrogen regulates differentiation and proliferation through two distinct pathways. [Gronemeyer2004]

The aim in developing selective ER ligands in the future will be to achieve selecitvity for interaction at EREs (estrogen induced differentiation) versus AP1 sites (estrogen induced proliferation functions). [Gronemeyer2004]

1.1.9. Selective ER modulators (SERMS)

Selective estrogen receptor modulators (SERMS) have the ability to antagonize the detrimental effects of estrogen on uterine and breast tissue but producing estrogen-like effects on bone and the cardiovascular system. [Kim2004]

Whether a SERM is an ER agonist or antagonist in a particular tissue depends on several factors:
Binding of a SERM to the ER causes a specific conformational change in the receptor and which coactivators and/or corepressors are recruited to the promoter depends on the resulting 3D-structure. The relative level of corepressors and coactivators in a specific tissue is also a determinant of a SERM's agonistic or antagonistic activity. The coregulators recruited to a particular promoter, depends on the type of ER dependent regulatory sequences that are present in the promoter. As already mentioned the ER α/ER β ratio varies between tissues and which ER form is dominant also effects SERM activity. [Deroo2006]

Tamoxifen, which was the first SERM to be developed, is an ER agonist in bone and uterus but an ER antagonist in the breast and has been used as an effective and save adjuvant endocrine therapy for breast cancer. Tamoxifen recruits a coactivator

complex to estrogen regulated genes in endometrial cells but a corepressor complex to the same gene in breast cancer cells. [Deroo2006]
Raloxifen (RAL) is a highly effective antiestrogen in the reproductive tissue but acts as a partial ER agonist in bone and also lowers blood cholesterol levels. [Pike1999]
Different ligands may differ in their affinity for α and β isoforms of the ER. RAL for example binds preferentially to the α receptor. RAL has the ability to prevent formation of transcriptionally competent AF2 conformation. When RAL is in the binding pocket of ER its bulky site chain prevents H12 forming a lid. So H12 gets in the position of coactivator binding. [Pike1999]

There is a so called “flip-flop” mechanism for H12 positioning, the equilibrium between the H12 agonist and antagonist positions in the coactivator binding site depends on the cellular context (nature and concentration of cofactors). [Gangloff2001]

1.2. Drug design

Structure-based drug design starts from a known or hypothetical mode of action or binding mechanism, a lead structure is rationally designed and subsequently tested experimentally. The results here obtained are fed back into a design cycle as new information. [Gohlke2002]

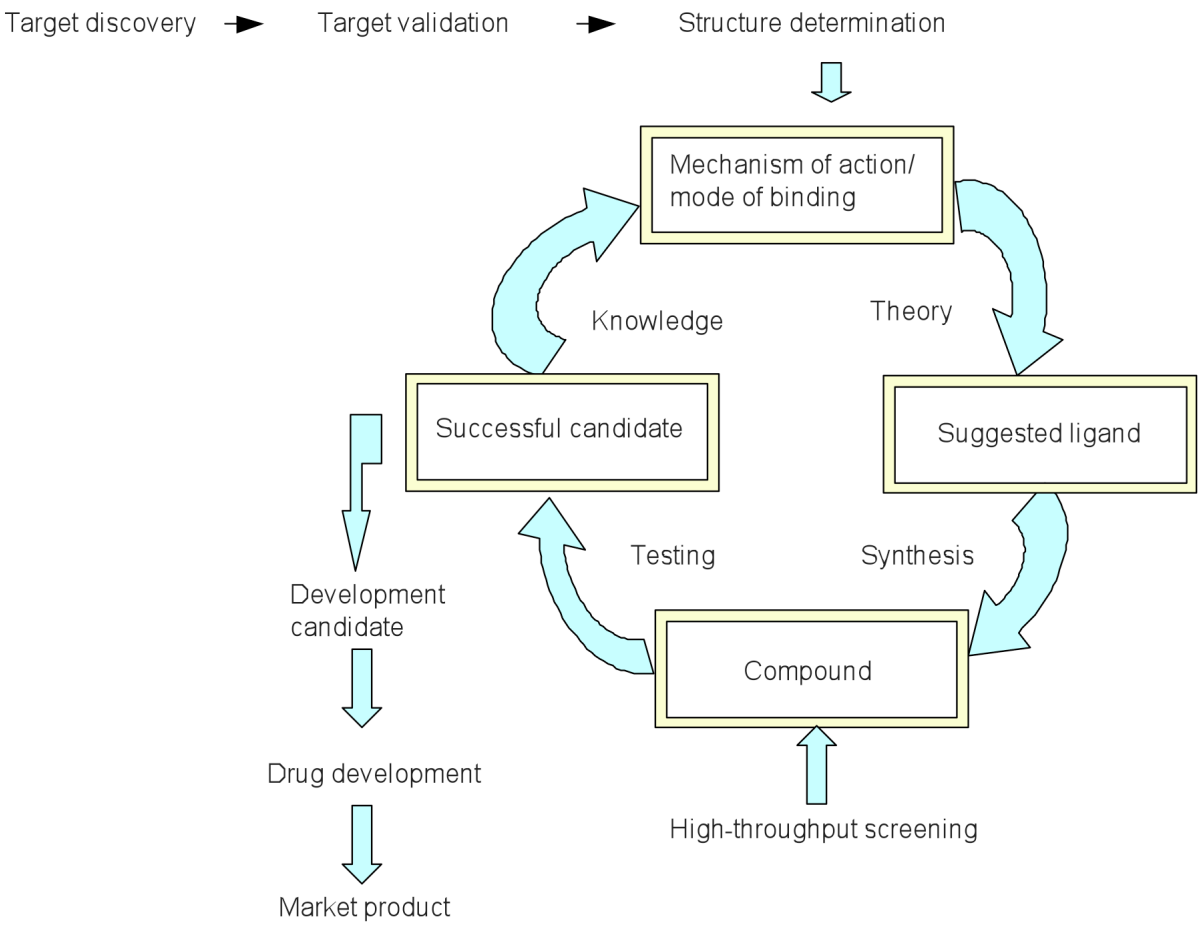


Figure 4: Rational drug design/design cycle. Starting with the discovery of a compound and biological testing, the information about the mechanism of action or binding mode is used for the development of a new drug. In contrast to computational methods the lead structure can also be found by experimental high-troughput-screening (HTS), which involves the testing of large compound libraries.

The following strategy in rational design depends on whether the three-dimensional structure of the biological target is known or not. If the structure is not available, “quantitative structure-activity relationship” (QSAR-methods) can be used to find a relationship between molecular structure and biological activity in order to allow an affinity prediction for unknown compounds. [Gohlke2002]

Another approach is the generation of a pharmacophore model from a series of active compounds whose properties are represented in geometric terms. In the following step potentially active candidate molecules are obtained from a compound library that obey this pharmacophore hypothesis. [Gohlke2002]

The growing number of structurally characterized macromolecular receptors provides the possibilities for any structure-based design of active compounds. [Gohlke2002]
The three-dimensional structure of a target can either be determined by crystallography, nuclear magnetic resonance (NMR) or construction on the basis of homologous proteins. [Klebe; ex Virtual Screening in Drug Discovery]

There are two possible strategies for computer-aided complex-generation:

-*De novo* docking where novel leads are generated in the binding pocket starting from prepositioned seed atoms that are subsequently grown into entire molecules. [Gohlke2002]
-Another method would be screening a compound library for ligands in agreement with the binding site requirements. Several thousands of compounds from an in-silico database are often docked into a receptor binding site and afterwards ranked according to their fit. [Gohlke2002]

The success of computer-aided drug design depends on the generation of reasonable ligand-binding modes (configuration-generation problem) and the recognition of those binding modes that suit best to the given situation based on a reasonable assessment of the expected binding affinity (affinity prediction problem). [Gohlke2002]

1.2.1. Molecular Docking

Docking is a tool often used to predict the binding orientation of small molecule drug candidates to their protein targets to predict the affinity and activity of the small molecule towards this target. Hence, docking plays an important role in drug design. [Kitchen2004]

Molecular docking can be thought of as a problem of “lock and key”, where one is interested in finding the correct relative orientation of the right key which will open up the lock. Here the protein can be though of as the lock and the ligand as the key. The molecular docking problem can be defined as predicting the correct bound association of two molecules with the given atomic coordinates. [Halperin2002]

Docking is computationally difficult because there are many ways of putting two molecules together (three translational and rotational degrees of conformational freedom). With the size of the components, the number of possibilities grows exponentially. Furthermore protein flexibility and screening large compound databases makes the computational problem even more difficult. [Halperin2002]

The interactions between ligand molecules and their receptors are dynamic and complex. Techniques in computer aided drug design should therefore account for the conformational flexibility of the ligand and the receptor which changes the “lock and key” problem into a “hand in glove” problem (induced fit). [Gohlke2002]
Ligand flexibility exploring is already well established and new methods accounting for receptor flexibility are already available. [Gohlke2002]
This is also important for discovering ligands that bind to one set of receptors but not to another potentially similar one and to design selective nuclear receptor modulators which means ligands that selectively modulate different receptor subtypes and/or act in a cell-type or tissue specific manner. [Fernandes2004]

1.2.2. Methods accounting for protein flexibility

Ensemble docking:

Ensemble docking means docking into multiple conformations of the receptor where the ligands are treated as rigid bodies to focus on the effect of protein flexibility. The Ensemble docking algorithm can dock a ligand simultaneously into multiple protein structures and automatically select an optimal protein conformation. [Huang2007]

Serial docking:

This method docks a candidate ligand serially to a set of different receptors and the results obtained for the structure determine whether to continue on to the next receptor or not. [Fernandes2004]
The purpose of serial docking is to find homology models for virtual screening and to find ligands that bind to one set of receptors but not to another potentially similar one (selective targeting). [Fernandes2004]

Relaxed complex method (RCM):

The RCM also incorporates receptor flexibility, starting from molecular dynamic (MD) simulations followed by docking of special libraries of candidate ligands to a large ensemble of the receptor's MD conformations. [Amaro2008]

Structures from the protein data bank (PDB) are mostly static and docking against them does often not reflect the true dynamical nature of most protein-ligand interactions. [Amaro2008]
RCS combines the advantage of docking algorithms with dynamic structural information provided by MD simulations, which are carried out for the target of interest. [Amaro2008]

Four-Dimensional Docking: using conformers as fourth dimension.

The purpose here is to account for conformational changes induced by ligand binding. [Bottegoni2009]
Receptor flexibility can be defined as the fourth discrete dimension of the small molecule conformational space, with multiple recomputed 3-D grids from optimally superimposed conformers merged into a 4-D object. The receptor conformations are represented by a single set of 4-D grids. In this method, no postprocess step is needed. [Bottegoni2009]

Accounting for side chain rotations:

Today several docking programs like GLIDE, DOCK, GOLD and AUTODOCK are using rotamer libraries to represent side chain flexibility. [Meiler2006]

1.2.3. Virtual Screening (VS)

For virtual screening, a knowledge about the spatial and energetic criteria responsible for the binding process needs to be available, which means either the

3-D structure of the target molecule or a rigid reference ligand with a known bioactive conformation in the assumed receptor bindingsite. [Klebe2006]
In contrast to experimental high throughput screening (HTS), which provides the medicinal chemist with hits in terms of chemical compounds that actually bind to the target of interest, VS simply suggests computer hits. [Klebe2006]
In VS, compounds are selected on the basis of improved docking algorithms using an approximate energy function to mutually rank them as putative hits. [Klebe2006]
The advantage of using computer programs for prediction of binding a compound to a target is that the compounds do not necessarily need to exist and that experimental deficiencies such as limited solubility, aggregate formation or any sort of influence that could possibly interfere with experimental assay conditions, do not need to be considered. [Klebe2006]
Moreover, hits discovered by HTS, even through they represent real molecules, do not help us to understand why and how they act upon the target. Any gain in knowledge is only obtained once structural biology or molecular modeling come into play and detect structural similarities or possible binding modes among the discovered hits. [Klebe2006]

1.2.4. Enrichment rates used to control the achievements of virtual screening

Enrichment is the ability of the docking program to put ligands that are known to bind to the target near the top of a score ranked list. [Moustakas, Pegg, Kuntz; ex Virtual Screening in Drug Discovery]
VS of compounds for possible drug leads requires identifying the few candidates, out of perhaps millions, which can bind with significant affinity (100 μM or better) to a target of a known structure. [Moustakas, Pegg, Kuntz; ex Virtual Screening in Drug Discovery]

VS runs are usually monitored and validated by comparing the performance of a set of known actives with a large set of “randomly” picked compounds, which are inactive (decoy structures). All structures are submitted to the selected VS protocol, and the performance-ranks of the known actives with respect to the remaining pool are converted into enrichment plots. These plots are accumulation curves that show how the fraction of actives recovered varies with the percent of the database screened.

An enrichment factor (EF) can be defined as:

$$EF = (a/n)/(A/N)$$

Where a is the number of active compounds in the n top-ranked compounds of a total database of N compounds of which A are active. [Moustakas, Pegg, Kuntz; ex Virtual Screening in Drug Discovery]

Successful screening implies EF>>1. This requires the identification of the best ligand conformation/position/orientation (pose) in the target binding site, that is, the solution of the docking problem. This in turn requires the ability to accurately calculate the binding affinity of a given pose (at least relative to another pose), which is the solution of the binding problem. [Moustakas, Pegg, Kuntz; ex Virtual Screening in Drug Discovery]

Basically an enrichment ratio shows how much work one saves by performing a virtual screen followed by testing those compounds in the hit list, compared with random screening of the entire collection. [Halgren, Murphy, Friesner; ex Virtual Screening in Drug Discovery]

1.2.5. Database preparation

Compound libraries used in lead finding programs should generally be filtered first to remove unsuitable compounds that would not reach and pass clinical trails anyway because of undesired properties. A good method to evaluate drug-likeness is the Lipinski “Rule-of-Five” which is suggesting that poor absorbtion or permeation are more likely when the molecular weight is over 500, the calculated octanol/water partition coefficient (logP) is higher than 5, when there are more than 10 hydrogen bond acceptors and more then 5 hydrogen bond donors. Compounds that fulfil two or more of these conditions are likely to show poor permeability and should be removed from the database. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

Meanwhile, filters for specific pharmacokinetic properties (absorbtion, distribution, metabolism, excretion=ADME) for prediction of aqueous solubility, membrane permeation and metabolic clearance are being developed. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]
Another possibility for filtering is to create a universal filter from databases that automatically distinguishes between drugs and chemicals. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

1.2.6. Stages for optimization / Scoring function

The success of a docking program depends on two important components: the search algorithm and the scoring function. [Gohlke2002]
After a ligand with a detectable affinity for a given receptor is found several stages for optimization are followed. [Klebe2006]
To characterize the binding affinity of putative lead candidates experimentally, the binding constant or its inverse, the dissociation constant (or inhibition constant), is determined. [Klebe2006]

$$K_A=K_D^{-1}=K_i^{-1}=[R`L`]/[R][L]$$

The scoring function takes a pose as input and returns a number indicating the likelihood that the pose represents a favourable binding interaction. [Klebe2006]

Some scoring functions are physics-based molecular mechanics force fields that estimate the energy of the pose. A low (negative) energy indicates a stable system and a likely binding interaction. [Klebe2006]
Alternatively one can derive a statistical potential for interactions from a database of protein-ligand complexes, such as the protein databank (PDB) and evaluate the fit of the pose according to this inferred potential. [Klebe2006]

If one assumes that the basic rules of equilibrium thermodynamics can be applied, an equilibrium constant that describes the formation of a protein-ligand complex can be defined. This equilibrium constant is logarithmically related to the Gibbs free energy,

which contains both an enthalpic and an entropic contribution. Whereas the former relates to energetic features the latter is related to configurational and ordering aspects. The entropic term estimates how the energy content of the system is distributed over internal and external molecular degrees of freedom. [Klebe2006]

$$\Delta G° = \Delta H° - T\Delta S° = -RT \ln K_A$$

It is known that electrostatic interactions determine noncovalent ligand-receptor binding. They include salt bridges, hydrogen bonds, dipole-dipole interactions, and interactions with metal ions. [Gohlke2002]
Solvation and desolvation contributions and the mutual, spatial complementary in the van der Waals (vdW) interactions are also of great importance. [Gohlke2002]

1.2.7. Different scoring functions

There are energy-based scoring functions that are related to energy terms in force fields, empirical scoring functions whose parameters are more dependent on empirical data, and knowledge-based scoring functions that are derived from databases. The improvements in scoring functions account for partial charges, desolvation effects and balance of different score terms as used in consensus scoring which means combining multiple scoring functions. [Lee2008]

1.2.8. Water-Treatment

Water is a frequently ignored binding partner because it is very difficult to treat properly. [Klebe; ex Virtual Screening in Drug Discovery]
The importance of water molecules as additional partners in protein-ligand interactions has to be considered in docking. Neglecting tightly bound water molecules in the binding site can result in a high desolvation penalty and unfavourable contributions to binding affinity. [Klebe; ex Virtual Screening in Drug Discovery]
Taking water molecules into account is usually performed by considering them as integral parts of the binding pocket. This requires reliable criteria whether to classify a water molecule as tightly or loosely bound and even the loosely bound waters can

mediate important interactions between ligand and protein. Furthermore the displacement of water to another position after ligand binding has to be considered. One possibility to incorporate water molecules is the sound analysis of solvation patterns as observed in crystal structures of the target protein with a lot of structurally diverse ligands. [Klebe; ex Virtual Screening in Drug Discovery]

1.2.9. Glide: Grid-based ligand docking with energetics

The Glide algorithm approximates a systematic search for positions, orientations and configurations of the ligand in the receptor-binding site using hierarchical filters that allow for respectable computational speed. The shape and properties of the receptor are represented on a grid of several sets of fields that provide progressively more accurate scoring of the ligand pose. The binding site is defined by a rectangular box confining the translations of the mass center of the ligand. [Klebe; ex Virtual Screening in Drug Discovery]

A set of initial ligand conformations is generated through search of the torsional minima, and the conformers are clustered in a combinatorial fashion. In the first stage, each cluster is docked as a single object. The search begins with a rough positioning and scoring phase that narrows the search space and reduces the number of poses to be further considered. The selected poses are minimized on precomputed OPLS-AA (molecular-mechanics force field) vdW and electrostatic grids for the receptor. [Klebe; ex Virtual Screening in Drug Discovery]

The 5 to 10 lowest-energy poses obtained are subjected to a MC (multiple copy) procedure in which nearby torsional minima are examined and the orientation of peripheral groups of the ligand is refined. The minimized poses are then rescored using the GlideScore function. It has been shown that rescoring docked poses with a secondary function can help to improve the selection of poses to be used at the rank ordering stage in database screening. The final choice of the best pose is made using a model energy score (EModel) that combines the energy grid score, Glide score, and the internal strain of the ligand. [Klebe; ex Virtual Screening in Drug Discovery]

1.2.10. CHEMScore/GLIDEScore

The ChemScore function was developed as a method to predict binding affinity. It consists of linear combination of four terms: lipophilic, H-bonding, metal binding, and an entropic penalty based on the number of frozen rotatable bonds. GlideScore has been designed to maximize enrichment in database screening. The ChemScore function has been modified with additional terms that differentiate charged and neutral H-bonds and terms that account for intermolecular Coulomb- and vdW-interactions and desolvation. [Perola, Walters, Charifson, ex Virtual Screening in Drug Discovery]

1.2.11. Glide XP

A new scoring function to estimate protein-ligand binding affinities has been developed named Glide 4.0 XP (extra precision) scoring function. Additional features characterizing XP Glide scoring are the application of large desolvation penalties to ligand and protein polar and charged groups in appropriate cases and the identification of specific structural motifs that provide large contributions to enhanced binding affinity. [Friesner2006]
Beside the unique water desolvation energy terms, protein-ligand structural motifs leading to enhanced binding affinity are included:

-**Hydrophobic enclosure which is surrounding of ligand lipophilic atoms or groups by lipophilic protein atoms.**

-**Neutral-neutral single or correlated hydrogen bonds in a hydrophobically enclosed environment.**

-**Five contributions of charged-charged hydrogen bonds.**

The aim is to semiquantitatively rank the ability of candidate ligands to bind to a specified conformation of the protein receptor and to exclude false positives. [Friesner2006]

In order to make the docking protocols effective within the receptor approximation, some ability to deviate from the restriction of the hard wall vdW-potential of the receptor conformation used in docking must be built into the potential energy function to predict ligand binding. [Friesner2006]
In XP and SP (standard precision) Glide, this is accomplished by scaling the vdW-radii of nonpolar protein and/or ligand atoms which when done effectively introduces the modest “induced fit” effect. [Friesner2006]

The appropriate fitting is judged by two factors: the ability to make key hydrogen bonding and hydrophobic contacts and the ability to achieve an appropriate root-mean-square deviation (RMSD) as compared to the native complex. [Friesner2006]

Contributions to protein-ligand binding affinity:

-**Displacement of waters by the ligand from “hydrophobic regions” of the protein active site.** Displacement of these waters by a suitable designed ligand will lower the overall free energy of the system. Considering entropic effects, if a water molecule is restricted in mobility in the protein cavity the release into solvent through the ligand will result in an entropy gain. Furthermore transfer of a hydrophobic moiety on the ligand from solvent exposure to a hydrophobic pocket can also contribute favourably to binding. [Friesner2006]
Hydrophobic bonding or entropic effects play a very important role in each drug-receptor interaction. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

-**Protein-ligand hydrogen bonding interactions, as well as other strong electrostatic interactions such as salt bridges.** Again displacement of waters by the ligand in the protein cavity leads to favourable entropic effects. Contributions to binding affinity also depends on the quality and type of hydrogen bonds formed, net electrostatic interaction energies and specialized features of hydrogen-bonding geometries. [Friesner2006]
Interactions of the displaced waters with the protein environment near the hydrogen bond can also have effects on binding affinity. [Friesner2006]

Electrostatic interactions are particularly important due to their long-range character for the attraction between ligand and receptor. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

-**Desolvation effects.** Polar and charged groups of the ligand or the protein that were exposed to solvent may become desolvated because they get in contact with groups to which they cannot hydrogen bond effectively. Those effects can only reduce binding affinity. [Friesner2006]

-**Entropic effects due to restriction on binding of the motion of flexible protein or ligand groups.** The major contributions are due to restriction of ligand translational/orientational motion and protein and ligand torsions but also contributions of vibrational entropies. These effects also reduce binding affinity. [Friesner2006]

-**Metal-ligand interactions.** Special terms are necessary concerning metal-ligand interactions. Metal-specific parameterization is a very complex contribution which needs large effort to be treated correctly. [Friesner2006]

XP Glide sampling methodology:

XP Glide sampling begins with SP Glide docking but using a wider “docking funnel” to obtain a greater diversity of docked structures. For XP docking to succeed, SP docking must provide one structure in which a key fragment of the molecule is properly docked. [Friesner2006]
The following step in XP sampling is to use various fragments of the molecule as “anchors” and starting from these anchors, to attempt to build a better scoring pose for the ligand. Typical anchors are rings but can also be other rigid fragments. Afterwards various positions of the anchors are clustered, representative members of each cluster are chosen, and the growing of the side chain from appropriate positions on the anchor is initiated. Glide “rough scoring” function is used to screen the initial side chain conformations which can be grown at extremely high resolutions because the total number of configurations considered is always pruned through screening and clustering algorithms. [Friesner2006]

After individual side chains are grown, a set of complete molecules is chosen by combining high scoring individual conformations at each position and rejecting structures with steric clashes between side chains. Candidate structures are minimized with the standard Glide total energy function. Afterwards, the grid based water addition technology is applied to a set of top structures, penalties are assessed and the full XP-scoring function is computed. [Friesner2006]

Parameterization:

Because the terms are calculated through fast empirical functions a lot of parameterization is required to obtain results in good agreement with experiment. These parameters are required to convert different geometrical criteria into specific scores. The number of parameters in XP scoring function is on the order of 80. [Friesner2006]

2. Materials and Methods

2.1. Protein Data Bank (PDB)

The success of a drug design study is strongly dependent on already available experimental data which is used as the basis for modeling procedures. A complete 3D-structure of a receptor obtained from x-ray crystallography or NMR measurements would provide a good basis of information. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

The most important database for structural information is the Protein Data Bank (PDB) which is available via the World Wide Web (http://www.rcsb.org) and contains a collection of protein and DNA structures. The PDB gives information about the resolution of a crystal structure which should be between 2.5 and 1.5 Å, or better. The structures do not include hydrogen atoms, so they have to be added especially when studying protein-ligand interactions. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

Two different estrogen-receptor conformations have been used in this study:

First, the agonist bound complexes where ligand binding leads to the rotation of Helix 12 which is then placed against the ligand binding cavity (mouse trap mechanism). This activation step is illustrated in Figure 5. [Avendaño, ex Medicinal Chemistry of Anticancer Drugs]

Second, the antagonist bound complexes where the ligand prevents the rotation of H12. Figure 6 shows the activation state conformation and Figure 7 the repression state conformation. The conserved amino acid clamp which is supposed to accomodate the ligand in both conformational states is illustrated in Figure 8. The different binding modes are shown in Figure 9. In case of the agonist Estradiol, binding involves both polar and nonpolar interactions. The phenolic A-ring and the A/B interface interact with amino acids: Ala-350, Leu-387 and Phe-404. The D-ring interacts with amino acids: Ile-424, Gly-521 and Leu-525. [Avendaño, ex Medicinal Chemistry of Anticancer Drugs]

The hydroxyl group of the A-ring establishes hydrogen bonds with the carboxylate group of Glu-353 as well as with the guanidinium group of Arg-394 and a water molecule. [Avendaño, ex Medicinal Chemistry of Anticancer Drugs]

The hydroxyl-group in the D-ring establishes a hydrogen bond with the imidazole ring of His-524. In case of the antagonist Raloxifene, this imidazole ring rotates to be able to bind to the hydroxyl group in Raloxifene. The other difference is that there are additional hydrophobic interactions because of Raloxifene´ s side chain. Furthermore a hydrogen bond is established between the basic group of the side chain and the carboxylate group of Asp-351. [Avendaño, ex Medicinal Chemistry of Anticancer Drugs]

The alignment of H12 over the binding cavity is prevented, because the side chain is too long to fit in the ER binding pocket and hence protrudes from the cavity between helix3 and helix11 (see Figure 5 for the position). [Avendaño, ex Medicinal Chemistry of Anticancer Drugs]

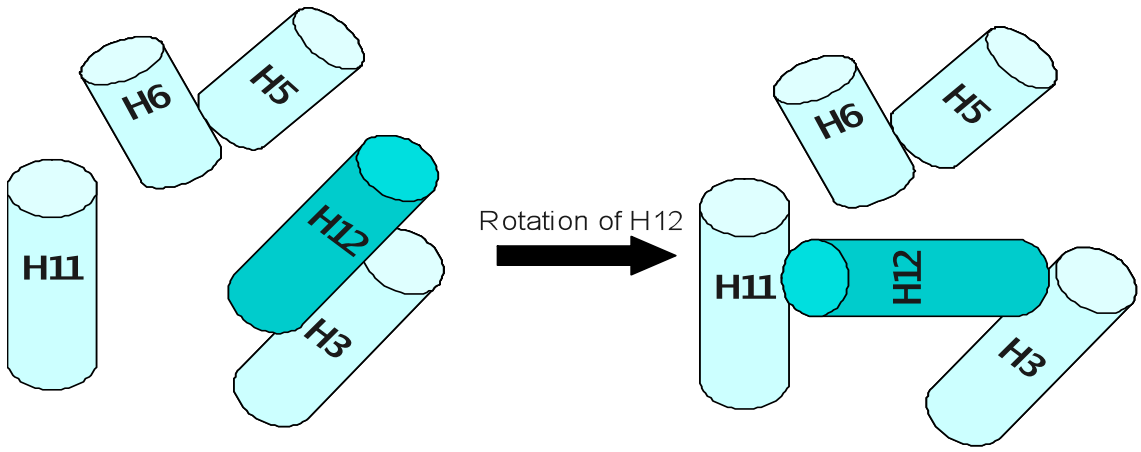


Figure 5: Activation of the Estrogen receptor. Symbolization of the rotation of Helix 12 after binding of an ER-agonist (Estradiol): H12 projects its inner, hydrophobic surface towards the ligand, its outer, charged surface is able to make interactions with coactivators. H3, H5, H6, H11 are the helices surrounding the binding pocket.

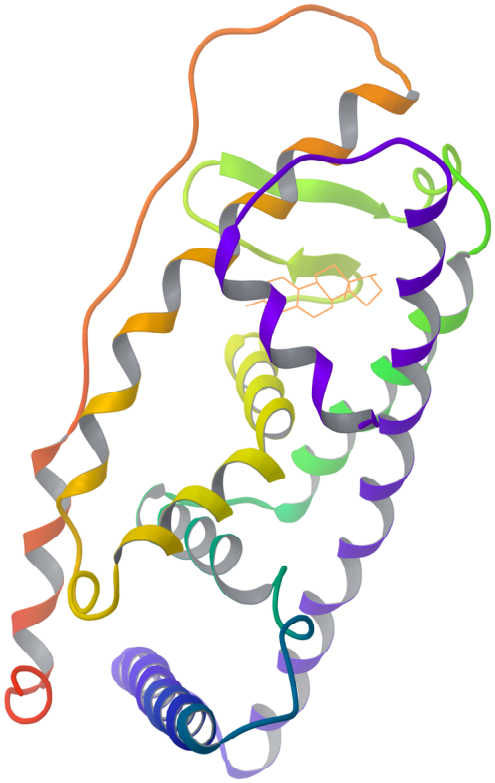


Figure 6: Estrogen Receptor and ligand Estradiol: **Activation state**, Helix 12 closes the ligand binding pocket.

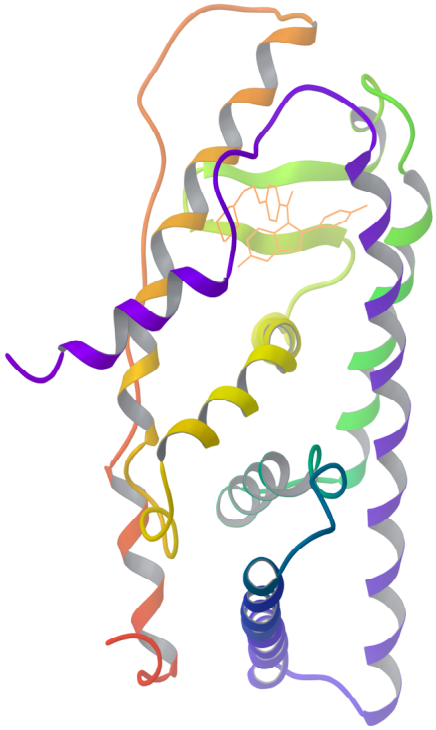


Figure 7: Estrogen Receptor and ligand Raloxifene: **Repression state**, Helix 12 protrudes from the cavity.

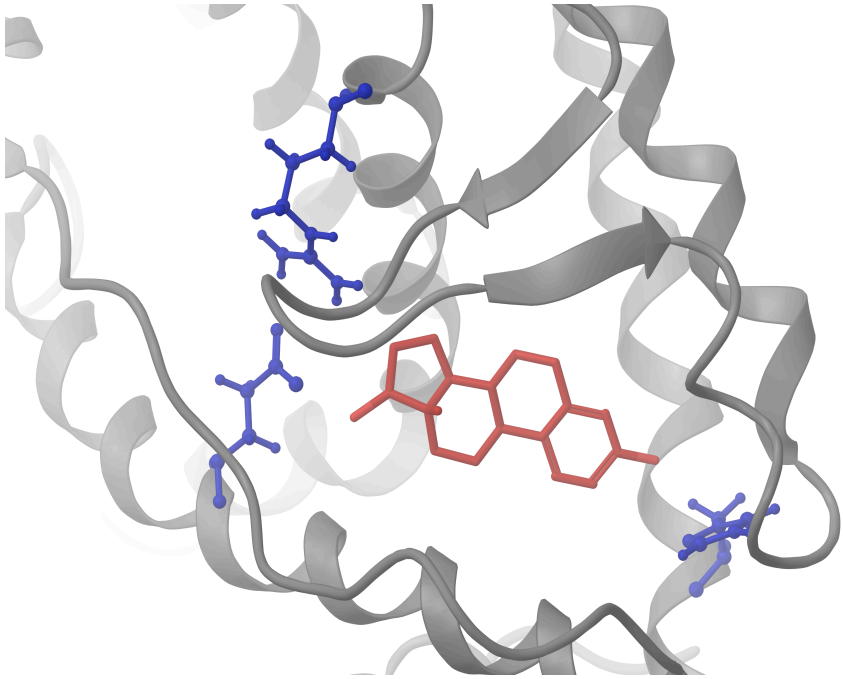


Figure 8: Amino acids **Arginine 353**, **Glutamate 394** and **Histidine 524** (blue) and ligand Estradiol (red)**.** These 3 residues build a sort of clamp with Arginine and Glutamate on one side and Histidine on the other side. This clamp is supposed to accommodate the ligand in the binding pocket. The clamp is conserved for all agonist and antagonist structures.

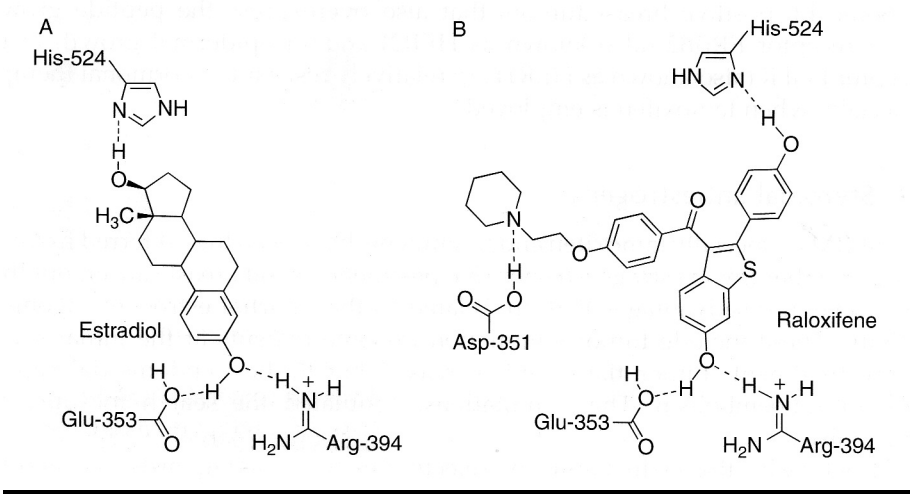


Figure 9: Binding modes of ER-agonist Estradiol (A) and ER-antagonist Raloxifene (B). See text for details. [Avendaño, ex Medicinal Chemistry of Anticancer Drugs]

2.2. ZINC- a free database of commercially available compounds for virtual screening

ZINC-database is a free available library of compounds with three dimensional structure, which can be used for docking (http://zinc.docking.org). The structures have biologically relevant protonation states and are annotated with information for molecular weight, calculated logP and number of rotatable bonds. The structures are available with multiple protonation states and tautomeric forms. [Irwin2005]

2.3. Conformational analysis

As the motional energy at room temperature is large enough to let atoms in a molecule move permanently, their absolute position is far from being fixed. Compounds which contain one or more single bonds exist in many different so called *conformers* or *rotamers*. The transformations are mostly related to changes in torsion angle about single bonds. The changes in molecular conformations can be regarded as movements on a multi-dimensional surface which describes the relationship between the potential energy and the conformation of the molecule, where stable conformations are the local minima on this surface. Conformational energies can be calculated either by quantum mechanical methods or molecular mechanical methods, the latter being used for large and flexible molecules because they are less time consuming. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.4. Molecular Mechanics and force fields

Molecular Mechanics is a well established computational method to calculate molecular geometries and energies. The simplification in molecular mechanics is based on considering the atoms in a molecule to be a collection of masses interacting with each other via harmonic forces. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.4.1. Molecular mechanics force fields

Atoms in molecular mechanics force fields are treated as rubber balls (different atom types), joined together by springs of varying length (bonds).

The total potential energy is: $\mathbf{E_{tot} = E_{str}+E_{bend}+E_{tors}+E_{vdW}+E_{elec}}$

$\mathbf{E_{tot}}$ = total energy of the molecule
$\mathbf{E_{str}}$ = bond-streching energy term
$\mathbf{E_{bend}}$ = angle bending energy term
$\mathbf{E_{tors}}$ = torsional energy term
$\mathbf{E_{vdW}}$ = Van der Waals energy term
$\mathbf{E_{elec}}$ = electrostatic energy term

Molecular mechanics makes it possible to calculate the total steric energy of a molecule in terms of deviations from reference “unstrained” bond lenghs, angles and torsions. Together with empirically derived fit parameters (force constants), these unstrained values are denoted the *force field*. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

Equation terms:

1. Bond stretching term

$$\mathbf{E_{str}=1/2\ k_b(b-b_0)^2}$$

k_b=bond stretching force constant
b_0=unstrained bond length
b=actual bond length

2. Angle bending term

$$\mathbf{E_{bend}=1/2\ k_\theta(\theta-\theta_0)^2}$$

k_θ=angle-bending force constant
θ_0=equilibrium value of θ
θ=actual value of θ

3. Dihedral potential energy term

$$\mathbf{E_{tors}=1/2\ k_\varphi(1+\cos(n\varphi-\varphi_0))}$$

k_φ=torsional barrier
φ=actual torsional angle
n=number of energy minima within one full cycle)
φ_0=reference torsional angle

4. Van der Waals interaction term: Lennard-Jones potential

$$\mathbf{E_{vdw}=\Sigma\ 4\varepsilon\ \{(\sigma_{ij}/r_{ij}^{12})-(\sigma_{ij}/r_{ij}^{6})\}}$$

ε= dielectric constant
σ_{ij}/r_{ij}^{12}=repulsive term
σ_{ij}/r_{ij}^{6}=attractive term
r_{ij}=distance between atom i and j

5. Coulomb interaction term for electrostatic forces

$$\mathbf{E_{elec}=(1/\varepsilon)\ [(Q_1Q_2)/r]}$$

ε=dielectric constant
Q_1Q_2=atomic charges of interacting atoms
r=interatomic distance

The underlying idea in molecular mechanics is that bonds and angles have a "natural" length. The equilibrium values of these bond lengths and angels together with the force constants used in the energy function are termed the *force field parameters* and are defined in the force field. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

A deviation from these equilibrium values will result in an increased total energy. So the total energy can be considered as a measure of intramolecular strain relative to a hypothetical "ideal" molecule. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.4.2. Force fields for bigger molecules

Force fields for Protein modeling differ from small molecule force fields in that specific parameters and simplifications are being introduced. Such simplifications are used in the *united atom model* where only polar hydrogens that may be partners in hydrogen bonding are treated explicitly while non-polar hydrogens are only included in the treatment of the heavy atoms to which they are bonded. Force fields like GROMOS

are examples which make use of this model. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]
Other simplifications make use of cut-off radii to ignore non-bonded interactions between atoms with larger distances than the defined cut-off and use continuum electrostatic models beyond the cutoff. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

What should be noted is that simplifications can always lead to a loss in accuracy and which force field to use, strongly depends on the given situation. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.5. Energy minimizing procedures/algorithms

2.5.1. Steepest Descent Minimizer

The steepest descent procedure is usually used for a rough minimization of little refined crystallographic data which is then followed by another minimization like the conjugate gradient method. The energy minimum is obtained by calculating the first derivative of the energy function. The energy-calculation starts with the initial geometry and is repeated for all atoms when they move to new positions until they finally reach the minimum on the energy surface. The procedure stops when a predetermined minimum condition is achieved. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.5.2. Conjugate Gradient Method

The computational effort for conjugate gradients is greater than for steepest descent. Here, the information obtained after each iteration is concentrated, the gradient is calculated over and over, and is used for computing the new direction vector. As much better convergence to the minimum can be achieved with this method it is often used for larger systems like proteins. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.5.3. Newton-Raphson Minimizer

The Newton-Raphson Minimizer uses the gradient and supplementary the second derivative to calculate the curvature of the function in order to find the search direction. The method is usually used for problems where fast convergence form an already partially optimized geometry to a precise minimum is needed. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.6. Conformational analysis using Monte Carlo Methods

The Monte Carlo Method is a random search or statistical technique. Starting from an optimized structure, each stage in a Monte Carlo procedure generates a new conformation by a random change of the former one. The new conformation is minimized by molecular mechanics and is only stored if it is unique. In principle molecules of any size can be correctly treated with this method. However, to cover all regions of conformational space the process may have to run extremely long depending on the number of conformers which have already been discovered. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.7. Conformational analysis using Molecular Dynamics

Molecular dynamics are based on molecular mechanics. Here, the aim is to reproduce the time-dependent motional behavior of a molecule. According to the rules of the force field, the atoms in the molecule interact with each other. At normal time intervals the classical equation of motion is solved:

$\mathbf{F_i(t)=m_i a_i(t)}$ (Newton's second law)

where F is the force on atom i at time t, m_i is the mass of atom i and a_i is the acceleration of atom i at time t. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

The forces on the atoms are calculated with the use of the gradient of the potential energy function and the initial velocities on the atoms are generated randomly at the beginning of the dynamics run. Positions and velocities can be calculated based on

the initial atom coordinates of the system and the atoms will then be moved to these new positions. The collection of conformations produced is called an *ensemble*. An important advantage of molecular dynamics is the ability to overcome energy barriers between different conformations which offers the possibility to find local minima other than the nearest in the potential energy surface. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.7.1. Model refinement of large molecules by molecular dynamics simulations

Molecular Dynamics Simulations are used to find the energetically most realistic three dimensional structure of a large molecule with hundreds of rotatable bonds. A simulation for a molecular system is performed by integrating the classical equations of motion over a period of time. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.7.2. GROMACS: GROningen MAchine for Chemical Simulations

GROMACS is a software for molecular dynamic simulations that was developed at the University of Groningen, The Netherlands. GROMACS works very fast, by carefully optimizing neighbor searching and inner loop performance. It doesn't have a force field of its own, but is usable with several previously mentioned force fields like GROMOS, OPLS and AMBER. The program was especially designed for the versatile simulation of biological (macro)molecules in liquid and membrane environments. [Van_Der_Spoel2005]
For further versatility the software package is provided with quantum mechanical packages like for example MOPAC, GAMES and GAUSSIAN to perform mixed MM/QM simulations. [Van_Der_Spoel2005]

2.8. Simulated Annealing and Monte Carlo Simulations

A special molecular dynamic simulation is simulated annealing, where the simulation temperature is cooled down at regular time intervals which leads to the trap of the system in the nearest local minimum conformation. In Monte Carlo simulated annealing (MCSA), random changes are made during each constant temperature

cycle. A new conformation is accepted when the energy is lower than the energy of the state before. A probability expression (Boltzmann equation) provides the basis to the decision of accepting or rejecting a compound. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

$P=\exp(-\Delta E/kT)$

ΔE=difference in energy from the previous step
T=absolute temperature in Kelvin
k=Boltzmann constant

Full force field based MCSA minimization is used in Glide for high level calculations. Glide uses a new algorithm for fast conformational generation. Computational costs are minimized by clustering the core regions of the generated 3D ligand conformations and treating the positions of the rotamer groups at the ends independently. [Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling]

2.9. Glide-docking tool

As already said in the introduction, Glide uses a series of hierarchical filters to flexibly dock a ligand in the active site of a protein-receptor. The search for ligand conformations starts with computational inexpensive initial screens to locate favorable ligand poses. After the initial poses have been selected the ligand is minimized in the receptor field using OPLS-aa (Optimized Potentials for Liquid Simulations) in association with a dielectric model. Three to five low-energy poses are obtained and are submitted to a Monte Carlo procedure. [Perola, Walters, Charifson, ex Virtual Screening in Drug Discovery]

OPLS: Optimized Potentials for Liquid Simulations

The OPLS force field was developed by Prof. William L. Jorgensen. Its functional form is similar to that of AMBER. [Perola, Walters, Charifson, ex Virtual Screening in Drug Discovery]

Different OPLS parameters are used in OPLS-ua (united atom) and OPLS-aa (all atom). OPLS-ua, where hydrogens next to carbon atoms are included in the carbon parameters, is used to safe simulation time. In OPLS-aa every atom is explicitly included. OPLS parameters were specifically optimized to fit experimental properties of liquids like density and heat of vaporization. OPLS makes use of the single point charge (SPC) or TIP3P water model for aqueous solution simulations. [Perola, Walters, Charifson, ex Virtual Screening in Drug Discovery]

2.9.1. Protein and Ligand preparation

For accurate docking with Glide a proper protein preparation is necessary. In order to yield favorable vdW-interactions for the receptor-ligand complex, steric clashes which are often found in crystallographically determined protein structures, have to be adjusted. Furthermore it has to be taken care that hydrogen bonding patterns and protonation states are correct. To adjust protonation states in the structures especially histidines, asparagines and glutamines have to be analyzed.

The adjustments have been carried out with *Protein Preparation Wizard* in Maestro. The procedure adjusts protonation states and performs a series of restrained minimizations to improve hydrogen orientations and relax unphysical steric clashes in the protein-ligand complex. Although the preparation-procedure is mostly able to make the right preparation choices, protonation states have been checked to see whether there is an incorrect H-bonding.

2.9.2. Glide-scoring function

Glide Score is an extended version of the ChemScore function:

Score:

$$\Delta G_{bind} = C_{lipo\text{-}lipo} \Sigma f(r_{lr}) + C_{hbond\text{-}neut\text{-}neut} \Sigma g(\Delta r) h(\Delta\alpha) + C_{hbond\text{-}neut\text{-}charged} \Sigma g(\Delta r) h(\Delta\alpha)$$

$$+ C_{hbond\text{-}charged\text{-}charged} \Sigma g(\Delta r) h(\Delta\alpha) + C_{max\text{-}metal\text{-}ion} \Sigma f(r_{lm}) + C_{rotb} H_{rotb}$$

$$+ C_{polar\text{-}phob} V_{polar\text{-}phob} + C_{coul} + E_{coul} + C_{vdW} + E_{vdW} + \text{Solvation Terms}$$

Glide Score has a unique solvation term to account for solvation of solvent-exposed moieties and water molecules in hydrophobic protein pockets. [Englebienne2007]

3. Results and Discussion

3.1. Bound Docking

The aim in bound docking is to reconstruct a complex with the use of the bound structures of a receptor and its ligand. [Halperin2002]

33 crystallographic agonist bound receptor structures and 20 crystallographic antagonist bound receptor structures were obtained from protein databank (PDB) and have been used to perform bound docking. Each of the receptors had its ligand included. [Halgren, Murphy, Friesner, ex Virtual Screening in Drug Discovery]

Optimization of the vdW-Scale Factors

A factor that makes Glide such a successful docking program is its ability to recognize detailed interactions with the use of hard interaction energetics on a Coulomb-vdW grid. Glide provides a way to adjust the interaction with the protein site relative to what the full vdW potential would produce. This mechanism works by scaling down the vdW-radii of nonpolar protein or ligand atoms. It has been shown that this is important to allow some room for the ligands that are a little bit larger than the native ligand. [Halgren, Murphy, Friesner, ex Virtual Screening in Drug Discovery]

Settings for Ligand docking

Glide does not scale the protein radii, it scales the radii of non-polar ligand atoms by **0.7** which corresponds to a **1.0/0.7** scaling. [Halgren, Murphy, Friesner, ex Virtual Screening in Drug Discovery]

Scaling factor: 0.7
Partial charge cutoff: 0.15
Number of poses per ligand to include: 5
Poses with Coulomb-vdW energy greater that 10 kcal/mol, have been rejected, because global strain energies of 10 kcal/mol have been proposed to be common for receptor-ligand complexes. [Perola2007]

Choosing the Enclosing Box

The enclosing box should be chosen smaller than the default-sized box, which allows ligands with up to 100 atoms to dock. A small box will make the calculation faster and will eliminate those ligands that cannot fit because they will either find no viable position within the box or will be given unfavorable docking scores due to unresolved steric clashes. It also keeps available 'slots' in the rough scoring stage of the docking for ligand poses that could be of interest in the more detailed stages. A smaller box will also prevent from finding positions outside the active site like for example positions on the protein surface. [Halgren, Murphy, Friesner, ex Virtual Screening in Drug Discovery]

Settings for Receptor grid generation

Scaling factor: 0.7

Partial charge cutoff: 0.25

Enclosing Box:

Supplied x, y, z coordinates: x=-4.78; y=-4.87; z=20.24 (used for agonist receptorstructures)

Supplied x, y, z coordinates: x=-7.78; y=-2.87; z=20.24 (used for antagonist receptorstructures)

Dock ligands with length ≤ 12 Å

Maestro: all structure-calculations and manipulations have been carried out in Maestro, a graphical user interface.

3.1.1. Results for Agonist structures

All agonist ligands have been docked in each agonist receptor, including the docking of the receptor's own ligand back into its structure which is referred to as self-docking. [Rao2008]

The question here is how reproducible the structure is or whether the receptor `finds` its own ligand, meaning that the own ligand has the most negative docking score. Protein receptors that have the best docking score for their own ligand are: **2B1ZB**, **1L2IA** and **1L2IB**. Table 1 shows the docking scores for each ligand docked into each receptor structure. Table 1 also shows at which position the receptors own ligand lies.

Table 1

Table 1: Docking of each ligand (shown in orange) into each receptor-structure (shown in red). Docking Scores for docking of receptors own ligand back into its structure (self-docking) is shown in grey; the bold numbers at the bottom show the positions at which the own ligand of the receptor lies (compared to the other docking scores for the other ligands). For receptors 2B1ZB, 1L2IA and 1L2IB the receptors own ligand has the most negative docking score (position 1, shown in light blue). Empty fields indicate cases, where the docking program produced no ligand pose.

To see how a ligand differs in orientation in the different receptor structures, the receptor's own ligand has been fixed in the project table in Maestro and the *root mean square deviation* (RMSD) between this ligand and the same ligand docked into all the other receptor structures has been calculated. The procedure has been carried out with the *Superposition tool* in Maestro. To compare the ligand orientations the option *calculate 'in place' (no transformation)* has been chosen which calculates the RMSD without moving the structures.

Table 2 shows the RMSD for the agonist structures. Ligands that have a RMSD of 6 or higher lie side inverted in the binding pocket. The reason for ligands lying side inverted is due to a different geometry of amino acids Methionine 343, Methionine 421 and Isoleucine 424. Figure 10 and 11 show the different orientation of Met 343 and Met 421 together with the ligand 17β-Estradiol in the correct and side inverted orientation respectively.

To find out more about the interactions in this particular case, 17β-Estradiol has been docked in 2 receptors which differ in orientation of Met 421, with Glide Extra Precision (see Introduction). Receptor 1, which orients the ligand correctly, has a more negative value for the lipophilic pair term and fraction of the total protein-ligand van der Waals energy and also a more negative value for the hydrophobic enclosure energy reward. Additionally it has a negative value for the electrostatic reward term, which is missing in Receptor 2, which orients the ligand side inverted:

	Receptor 1	**Receptor 2**
Glide Score	-7.93	-6.88
Lipophilic pair term	-5.07	-4.98
Hydrophobic encosure	-0.72	-0.28
Electrostatic reward	-0.57	0

Table 2

Table 2: RMSD of agonist structures. A RMSD of 6 or higher means that the ligand lies side inverted in the binding pocket. It has been shown that this is due to a different orientation of amino acids Met 343, Met 421 and Ile 424.

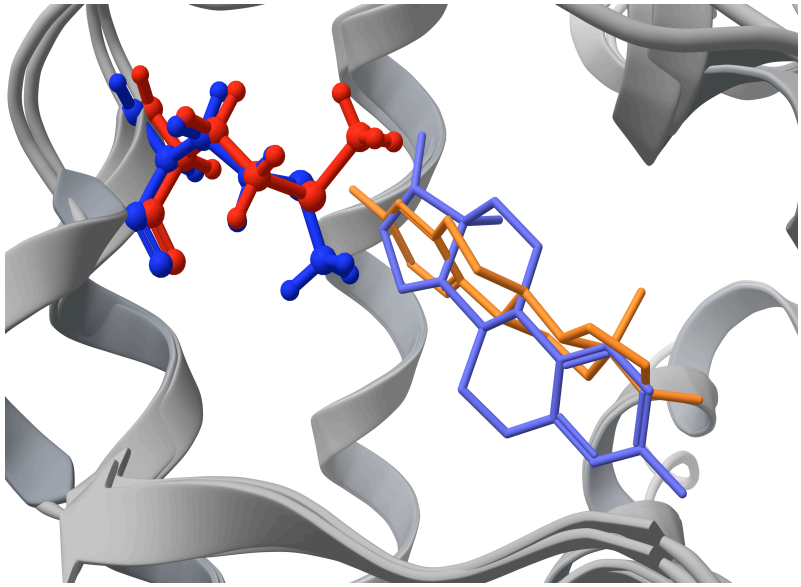


Figure 10

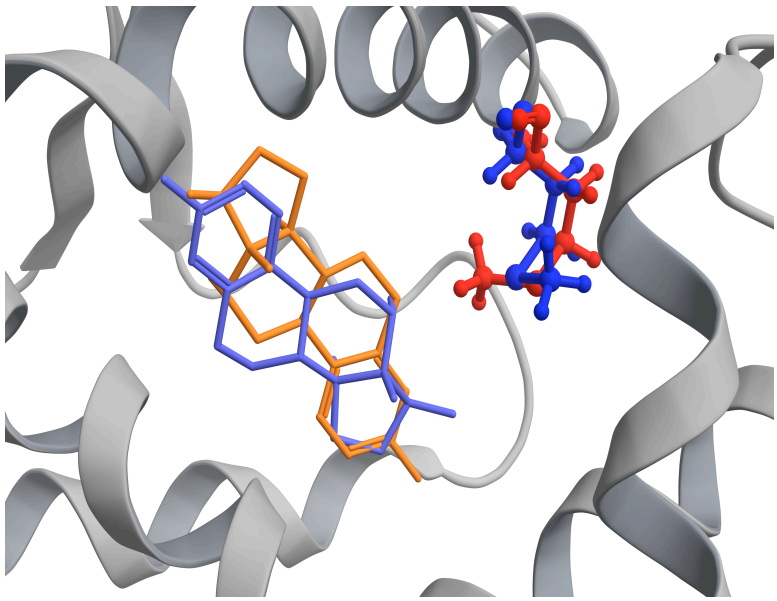


Figure 11

Figure 10 and 11: Methionine 421 (Figure 10) and Methionine 343 (Figure 11) and 17β-Estradiol. In both cases, the orientation of Met in blue orients the ligand correctly (ligand is shown in light blue), the orientation of Met in red docks the ligand side inverted (ligand is shown in orange).

Interestingly protein receptors **2B1ZA** and **2B1ZB** only differ in the side chain geometry of Isoleucine 424 (Figure 12). Receptor 2B1ZA docks ligands 1erea, 1ereb, 1erec and 1ered with a RMSD above 6 (side inverted) and receptor 2B1ZB docks the same ligands with a RMSD under 1. In this particular case the docking scores for 2B1ZB, where the ligands have the right orientation, have been better (below -7 kcal/mol) than for 2B1ZA (above -7 kcal/mol). The comparison of these 2 receptors is shown in Table 3.

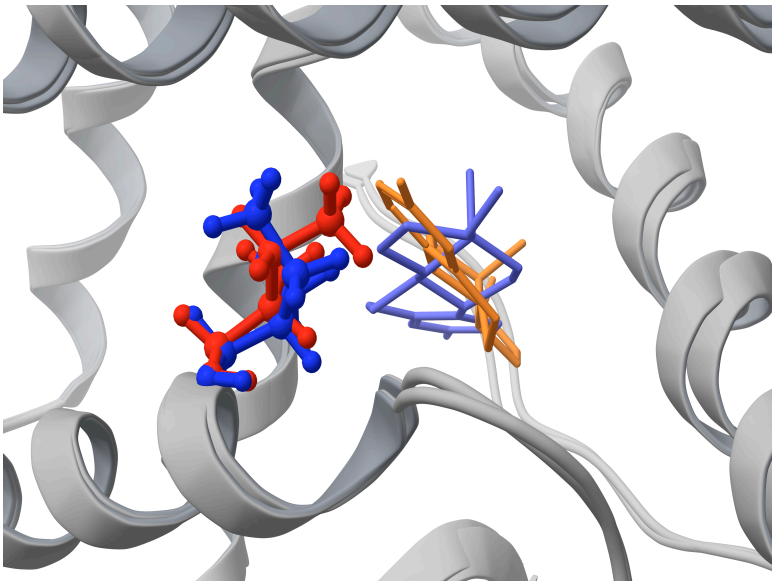


Figure 12: Isoleucine 424 and 17β-Estradiol. Ile 424 in blue orients the ligand correctly (ligand is shown in light blue) and Ile 424 in red leads to a side inverted orientation of the ligand (ligand is shown in orange).

	Docking Score	Standard deviation
2B1zA	-6.67 kcal/mol	0.11
2B1zB	-7.82 kcal/mol	0.07
	RMSD	
2B1zA	6.57	0.11
2B1zB	0.7	0.04

Table 3: Mean value and standard deviation of docking score and RMSD for receptors 2B1ZA and 2B1ZB.

17β-Estradiol has been docked into receptors 2B1ZA and 2B1ZB with Glide- Extra Precision. 2B1ZB has a more negative value for the lipophilic pair term and fraction of the total protein-ligand van der Waals energy and a more negative value for the hydrophobic enclosure energy reward:

	2B1ZA	**2B1ZB**
Glide Score	-6.88	-7.37
Lipophilic pair term	-3.66	-4.1
Hydrophobic encosure	-0.58	-1.03

It has also been shown that for receptors 1G50A, 1G50B and 1G50C, receptor 1G50A is the only one which orients all ligands correctly and this receptor has the overall docking score minimum (-9.72 kcal/mol). Receptor 1ERED has the second lowest minimum value (-9.51 kcal/mol) and compared to receptors 1EREA, 1EREB, 1EREC, 1EREE and 1EREF; 1ERED again is the only one which orients all ligands correctly. All receptors and their overall minimum docking score are shown in Table 5.

Receptor	Minimum value
1G50A	-9.72
1ERED	-9.51
3ERDB	-9.36
1X7EB	-9.27
2G44B	-9.24
1GWQB	-9.18
1GWRB	-9.14
1X7EA	-9.13
1PCGA	-9.1
3ERDA	-9.06
1GWQA	-9.01
2B1ZA	-8.91
2B1ZB	-8.89
1G50B	-8.86
1G50C	-8.86
2G44A	-8.84
1L2IA	-8.8
1QKUB	-8.78
2FAIB	-8.76
1EREE	-8.71
1QKUA	-8.71
1QKUC	-8.63
2B1VB	-8.62
1EREA	-8.58
1PCGB	-8.56
1EREB	-8.52
1GWRA	-8.46
1EREC	-8.41
2FAIA	-8.38
1L2IB	-8.31
1ZKYB	-8.31
1EREF	-8.28
2B1VA	-8.09

Table 4: Receptors and their overall minimum score.

3.1.2. Results for Antagonist structures

The same procedure as for the agonist structures has been carried out with the antagonist structures and here there are four receptors that have the best docking score for their own ligand: **1ERRA**, **1SJ0A**, **1XP9A** and **2AYRA** (Table 5). Table 6 shows the RMSD for the antagonist structures. Once again there are ligands that lie side inverted in the binding pocket.

	1erra	1errb	1qkta	1r5ka	1r5kb	1r5kc	1sj0a	1uoma	1xp1a	1xp6a	1xp9a	1xqca	1xpca	1xqcb	1xqcc	1xqcd	1yina	1yima	2ayra	3erta
1ERRA	-9.2	-8.5	-8.1	-5.1	-5.6	-5.2	-8.3	-8.6	-7.6	-7.3	-7.8	-8.1	-8.6	-8.7	-8.3	-8.8	-8.5	-8.3		-6.5
1ERRB	-8.5	-8.6	-7.7	-5.1	-4.6	-4.9	-8.9	-7.9	-7.6	-8.3	-8.8	-8.2	-8.4	-8.9	-8.1	-8.4	-9.1	-7.7		-6.5
1QKTA	-7.8	-7.3	-7.5	-6.0	-6.0	-5.9	-6.3	-7.5	-5.3	-7.3	-6.0	-6.9	-6.8	-6.9	-7.1	-7.5	-6.9	-6.7	-5.2	-7.0
1R5KA	-7.7	-7.9	-7.5	-5.2	-5.9	-5.8	-7.4	-8.1	-4.2	-7.1	-7.4	-8.8	-7.5	-8.4	-9.3	-8.3	-8.0	-7.7	-8.1	-7.2
1R5KB	-8.3	-8.1	-7.8	-6.6	-7.1	-6.8	-8.9	-7.5	-7.1	-4.9	-8.5	-9.4	-8.5	-9.7	-9.3	-9.1	-7.7	-8.3	-9.5	-7.8
1R5KC	-7.6	-8.0	-7.6	-6.3	-6.2	-6.7	-8.3	-7.4	-5.2	-7.9	-8.6	-8.7	-8.0	-8.8	-8.8	-8.8	-8.9	-7.7	-7.8	-7.9
1SJ0A	-7.8	-7.6	-7.4	-6.7	-7.2	-6.6	-9.3	-8.8	-9.2	-8.4	-9.2	-8.6	-8.7	-8.9	-9.2	-8.8	-8.0	-8.6	-5.1	-8.0
1UOMA	-8.0	-8.3	-6.8	-6.4	-5.9	-6.0	-8.5	-8.3	-6.8	-7.8	-7.2	-8.5	-7.4	-8.4	-8.8	-8.7	-9.0	-7.7	-4.9	-7.7
1XP1A	-9.0	-8.7	-7.1	-5.8	-5.3	-6.1	-8.2	-8.0	-7.7	-7.1	-7.9	-9.6	-8.5	-8.1	-9.7	-8.9	-8.9	-8.0	-6.8	-7.0
1XP6A	-9.2	-8.4	-7.2	-6.9	-6.4	-6.7	-9.0	-8.3	-7.8	-9.8	-9.0	-9.1	-9.9	-8.4	-9.1	-8.3	-9.9	-9.2		-6.9
1XP9A	-8.1	-8.0	-7.2	-5.6	-5.6	-4.6	-8.5	-8.1	-6.1	-8.8	-9.1	-8.5	-8.9	-8.6	-8.8	-8.4	-8.4	-8.6	-6.0	-7.7
1XQCA	-7.9	-8.6	-7.9	-5.6	-4.7	-5.9	-8.0	-7.1	-7.2	-7.7	-7.3	-8.1	-7.9	-8.8	-8.4	-8.6	-7.3	-7.0	-5.9	-7.0
1XPCA	-8.3	-8.4	-7.5	-4.5	-5.5	-4.6	-9.0	-8.3	-6.0	-8.5	-8.1	-8.8	-8.3	-8.8	-8.8	-8.8	-8.6	-7.9	-5.2	-8.0
1XQCB	-7.8	-7.1	-6.7	-4.7	-5.1	-3.0	-6.9	-8.0	-6.8	-7.3	-7.5	-7.9	-8.0	-7.9	-7.8	-7.8	-7.6	-7.8	-4.9	-6.3
1XQCC	-7.6	-8.2	-7.4	-5.2	-5.0	-5.7	-8.4	-8.0	-7.5	-8.0	-7.7	-8.5	-8.5	-8.1	-8.3	-8.3	-7.5	-8.1	-3.2	-7.2
1XQCD	-7.2	-7.4	-7.2	-4.3	-5.1	-5.3	-7.9	-7.8	-7.7	-7.5	-7.7	-8.2	-8.0	-8.4	-8.5	-8.3	-7.9	-7.8	-7.1	-6.6
1YINA	-8.9	-9.1	-7.7	-6.2	-7.1	-7.1	-9.1	-8.3	-6.6	-9.3	-9.2	-9.6	-8.8	-9.2	-8.7	-9.1	-9.1	-9.0	-6.7	-6.4
1YIMA	-8.0	-7.3	-7.5	-6.4	-7.1	-6.5	-8.7	-7.6	-5.7	-5.9	-8.3	-8.2	-8.4	-7.4	-7.2	-7.6	-9.1	-8.2	-6.5	-7.6
2AYRA	-8.5	-7.8	-7.7	-6.6	-6.7	-6.6	-8.0	-7.4	-5.8	-5.2	-7.8	-8.2	-8.1	-8.1	-8.4	-9.3	-8.6	-7.4	-9.8	-7.3
3ERTA	-8.3	-8.2	-7.4	-5.4	-3.6	-5.0	-7.6	-7.5	-7.6	-7.7	-7.9	-8.4	-7.5	-8.7	-8.5	-8.4	-7.6	-7.5	-3.7	-7.9
	1	**5**	**4**	**19**	**17**	**17**	**1**	**7**	**13**	**3**	**1**	**5**	**10**	**3**	**4**	**3**	**5**	**5**	**1**	**8**

Table 5: Docking scores for antagonist structures.

	1erra	1errb	1qkta	1r5ka	1r5kb	1r5kc	1sj0a	1uoma	1xp1a	1xp6a	1xp9a	1xqca	1xpca	1xqcb	1xqcc	1xqcd	1yina	1yima	2ayra	3erta
1ERRA	1.6	2.2	2.9	5.3	1.5	1.5	2.4	1.4	4.8	4.8	2.7	1.3	2.3	1.1	1.4	1.5	2.3	1.7		2.1
1ERRB	1.3	2.1	6.3	7.0	1.9	1.9	5.0	2.5	10.7	1.4	1.3	1.3	1.8	0.9	1.2	1.2	1.8	2.5		1.3
1QKTA	2.6	5.9	0.6	1.4	1.3	1.3	6.0	1.5	7.6	7.2	1.5	1.6	1.4	1.2	1.4	1.3	1.7	2.6	8.8	1.6
1R5KA	3.3	3.7	3.0	1.9	1.1	1.2	4.8	3.1	10.6	3.6	3.5	3.2	3.7	2.8	2.7	2.8	4.9	4.5	3.0	1.7
1R5KB	3.6	3.3	6.7	1.3	0.9	1.2	4.4	3.2	6.2	6.0	3.4	2.6	3.6	2.7	2.6	2.6	2.7	3.2	3.6	1.8
1R5KC	3.1	3.8	0.7	6.5	1.0	6.5	3.8	2.7	10.7	3.9	3.5	2.7	3.5	2.8	2.7	2.7	3.8	3.0	2.9	1.4
1SJ0A	5.4	5.6	2.0	5.2	1.2	1.7	5.9	1.5	3.0	1.2	2.0	1.8	1.7	1.8	1.6	1.8	5.4	1.4	8.2	1.3
1UOMA	2.0	5.5	1.1	1.8	1.9	1.7	1.9	1.7	5.5	2.0	1.4	1.5	2.3	1.3	1.5	1.4	5.5	1.4	9.3	1.3
1XP1A	1.2	4.5	6.4	1.8	1.2	1.9	1.4	1.5	10.7	10.8	2.1	0.8	3.2	0.8	1.0	1.6	1.8	1.5	11.7	2.0
1XP6A	2.3	3.5	6.5	1.6	1.5	1.9	2.4	5.1	3.5	2.8	2.9	1.3	2.1	1.2	1.2	5.4	1.5	1.2		2.1
1XP9A	1.6	3.7	1.3	1.5	1.5	2.0	1.9	1.5	5.1	1.9	1.6	1.7	2.4	1.0	1.4	1.7	2.2	1.8	8.3	2.3
1XQCA	4.4	2.5	6.6	1.7	2.1	1.2	5.8	2.6	3.6	3.0	3.0	1.0	2.6	2.9	2.7	2.5	2.2	5.6	6.4	1.8
1XPCA	2.1	2.9	1.3	2.1	1.4	2.1	2.0	1.4	4.9	2.1	1.3	1.8	2.7	1.8	1.8	1.8	1.8	1.3	9.4	1.6
1XQCB	4.4	5.4	1.8	1.7	1.4	8.4	2.6	3.1	3.1	3.2	3.0	2.8	2.6	3.1	2.9	2.9	1.7	1.6	9.8	1.4
1XQCC	3.2	3.4	1.1	1.2	1.8	1.3	2.7	2.5	3.0	3.2	3.6	2.5	2.6	2.9	2.7	2.5	1.2	1.4	9.8	2.0
1XQCD	2.4	2.8	0.3	1.6	1.8	1.7	1.8	1.5	1.6	2.3	2.4	2.7	2.5	2.5	2.8	1.0	2.8	1.5	6.4	2.2
1YINA	5.0	2.7	6.6	5.3	1.8	1.8	1.4	2.2	10.5	1.3	1.6	5.0	1.6	4.8	5.4	5.0	1.3	3.4	8.8	2.2
1YIMA	2.0	5.5	1.9	1.6	1.6	1.6	1.8	2.1	10.4	10.7	3.1	1.8	3.0	1.2	1.5	1.4	3.2	1.5	9.0	1.6
2AYRA	4.6	2.2	3.3	1.6	1.1	1.1	2.0	1.7	4.9	10.5	3.2	3.3	2.5	3.6	3.3	3.1	3.2	3.8	1.5	1.7
3ERTA	6.3	7.3	1.3	1.4	8.4	1.8	6.1	2.9	1.3	3.8	3.4	2.7	3.8	2.9	2.7	2.6	5.5	5.5	10.1	1.9

Table 6: RMSD for antagonist structures.

The results for the antagonist structures also revealed a correlation between docking score and ligands lying correct or side inverted. Moreover there is evidence that the results are associated with the overall docking score minimum, as having observed for the agonist structures before:

Ligand 1r5kc docked in receptors 1XQCA, 1XQCC and 1XQCD resulted in a docking score of under -5 kcal/mol and for receptor 1XQCB in a docking score of -3 kcal/mol. 1XQCB is the only one which docks this ligand side inverted with a RMSD of 8.4, while the three others have a RMSD of 1.2, 1.3 and 1.7. Furthermore receptor

1XQCB has the highest overall minimum value and the highest mean value compared to the other ones. (Table 7 and Table 8)

The structural analysis showed that 1XQCB has a different orientation of Met 343 compared to the other three which orient Met 343 in a similar way.

Receptor	Minimum value
1XP6A	-9.89
2AYRA	-9.77
1XP1A	-9.69
1R5KB	-9.65
1YINA	-9.57
1SJ0A	-9.27
1R5KA	-9.25
1ERRA	-9.18
1ERRB	-9.14
1YIMA	-9.12
1XP9A	-9.12
1XPCA	-9.04
1UOMA	-8.96
1R5KC	-8.87
1XQCA	-8.75
3ERTA	-8.67
1XQCD	-8.52
1XQCC	-8.51
1XQCB	-8
1QKTA	-7.83

Table 7

Receptor	Mean value
1YINA	-8.26
1SJ0A	-8.1
1R5KB	-8.05
1XP6A	-7.98
1XP1A	-7.83
1R5KC	-7.77
1XP9A	-7.68
2AYRA	-7.66
1XPCA	-7.59
1UOMA	-7.54
1YIMA	-7.44
1R5KA	-7.37
1ERRA	-7.35
1XQCA	-7.34
1XQCC	-7.31
1ERRB	-7.3
1XQCD	-7.3
3ERTA	-7.21
1XQCB	-6.85
1QKTA	-6.7

Table 8

Table 7: Receptors and their overall docking score minimum value.

Table 8: Receptors and their docking score mean value.

3.1.3. MD-Simulations

Macromolecular Systems

Five different simulation systems were established, each one representing one of the distinct receptor conformations found among ER α structures available in the Protein Data Bank (PDB).

The systems are simulated as monomer (**M**) and dimer (**D**) respectively, once with the ligand bound in the active site (apo (**A**) conformation), once without the ligand (holo (**H**) conformation), and where adequately, with the bound coregulator peptide (**C**); no coregulator peptide (**N**).

Coordinates were taken from the best crystalstructure with the respective backbone conformation. Furthermore, coordinates from two PDB files have been combined in some cases, so that unresolved residues in one structure are resolved in the other one.
Among the five systems, two were activation state conformations (**act**), meaning the protein bound to ER agonists, and two were repression state conformations (**rep**), meaning the protein bound to ER antagonists. One system was in the domain switched tetramer configuration (PDB code 1A52).

MD Runs

The simulations where carried out with Gromacs version 3.3 (simulation length 20ns).
Force field: **AMBER**
Protein solvatation: **TIP3P water**

For an accurate comparison of MD-simulations and docking into PDB-crystal structures, for the simulations, the same settings as for the previous docking studies have been used. The structures have all been fitted so that their place and orientation is the same as for the crystal structures. This procedure has been carried out with the *protein structure alignment tool* in Maestro. Afterwards the same supplied x,y,z-coordinates for the respective enclosing box (agonist or antagonist) could be used for the simulations.
As before the dockings have been performed with Glide SP. The settings for ligand docking and receptor grid generation have been the same as for the previous dockings.
For comparing the ligand orientations in the MD-simulation docking study the ligand **17beta-Estradiol** from **1GWR** PDB-structure has been fixed in the workspace of Maestro. This PDB structure was used to generate the **act2** simulation system. The RMSD between 17beta-Estradiol from 1GWR and this ligand docked in the act2 simulation structures has been calculated and is shown in Table 9. Again there are ligands that lie side inverted due to a different orientation of Isoleucine 424.

Receptors	RMSD	Receptors	RMSD	Receptors	RMSD	Receptors	RMSD
act2DNA2	1.2	**act2MNA12**	6.6	**act2DNH25**	7.0	**act2MNH16**	7.5
act2DNA8	1.2	**act2MNA2**	1.6	**act2DNH30**	6.6	**act2MNH2**	7.5
act2DNA5	3.3	**act2MNA4**	3.7	**act2DNH38**	7.1	**act2MNH23**	7.3
		act2MNA6	2.0	**act2DNH40**	2.4	**act2MNH39**	4.3
		act2MNA8	6.9	**act2DNH44**	1.1	**act2MNH51**	4.0
				act2DNH9	7.0	**act2MNH7**	7.2

Table 9: RMSD of 17β-Estradiol from 1GWR and this ligand docked in the act2 simulation structures. Ligands with a RMSD of 6 or higher lie side inverted in the binding pocket due to a different orientation of Ile 424.

3.1.4. Conclusion

Although it has been shown, that a receptor 'finds' its own ligand in only some cases, there is a correlation between docking score and the orientation of the ligand. A lower docking score is connected with lower RMSD and vice versa. Furthermore there is evidence that the overall docking score minimum (in case of the agonist and antagonist structures) and also the mean value (in case of the antagonist structures) can reveal specific receptors as being able to orient ligands correctly.
In case of the agonist structures a side inverted orientation of a ligand comes from a different orientation of a few amino acids: Met 343, Met 421 and Ile 424. A different orientation of Ile 424 also leads to a side inverted orientation of 17beta-Estradiol docked into the MD-simulation structures.

3.2. Cross-Docking

Cross Docking refers to docking a ligand into each of the superimposed protein structures originally bound with other ligands in the ensemble, in other words it employs a protein structure with a bound ligand, but where the ligands to be predicted are different. [Huang2006]

Cross docking has been performed with Glide SP, the settings have been the same as those used in the bound docking studies.

Molecular structures

37 crystallographic agonist bound receptor structures and 20 crystallographic antagonist bound receptor structures from protein databank (PDB) have been used to perform cross docking.

67 agonist ligands and 2570 agonist decoys (drug-like molecules- low molecular weight, high solubility) as well as 39 antagonist ligands and 1448 antagonist decoys form ZINC database (a free available database for virtual screening) have been docked in all agonist- and antagonist bound receptors.

3.2.1. Enrichment Results: Docking of agonist- and decoy- ligands in agonist- receptor structures

As a key objective is to find active compounds as early as possible in the ranked database, the enrichment plots have been visually compared and classified in three groups: **good, bad and quite good enrichment**.

The enrichment plots show the percentage of known actives found (y-axis) versus percentage of the ranked database screened (x-axis). Table 10 gives the list of the classified receptors. Figure 13 shows the enrichment plots for all protein receptors.

For seven receptors good enrichment has been achieved, the plots are steep meaning that the active compounds enrich very soon in the database. For 19 receptors quite good enrichment has been obtained and 11 receptors show bad enrichment, the active compounds enrich very late in the database.

The reason for choosing this classification, is that for receptors with good enrichment there might probably be one compound that binds significantly well (one top compound), while for the less steep enrichment plots (with quite good enrichment) there may be more compounds that have good affinities for these receptors.

List of receptors with good enrichment: **1g50A, 1gwqA, 1gwqB, 1gwrB, 1x7eA, 2b1zA, 2g44B**
List of receptors with bad enrichment: **1ereB, 1ereD, 1ereE, 1ereF, 1gwrA, 1x7rA, 2b23A, 3erdA, 3erdB, 1l2iA, 2b1vB**
List of receptors with quite good enrichment: **1ereA, 1ereC, 1g50B, 1g50C, 1pcgA, 1pcgB, 1qkuA, 1qkuB, 1qkuC, 1x7eB, 1zkyA, 1zkyB, 2b1vA, 2b1zB, 2b23B, 2faiA, 2faiB, 2g44A, 1l2iB**

Table 10

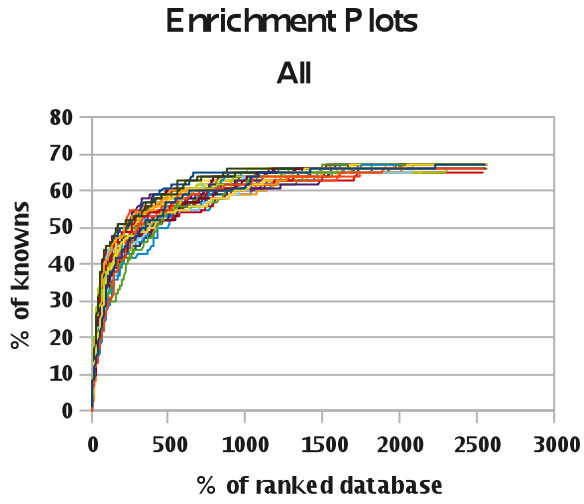


Figure 13: All enrichment plots are shown. Percent of known actives found (y-axis) versus percent of the ranked database screened (x-axis).

Figure 14 and 15 show the ZINC-Number versus docking score plots for the ligands (potential binders) and the decoys (nonbinders) respectively. With this representation the difference in docking score between ligands and decoys can be seen. Quite a lot of ligands have a docking score between -10 and -8 kcal/mol whereas the lowest docking score for the decoys is -8 kcal/mol. This observation suggests that for the ligands with docking scores between -10 and -8 kcal/mol there are no false positive hits, meaning decoys with a lower docking score than the ligands.

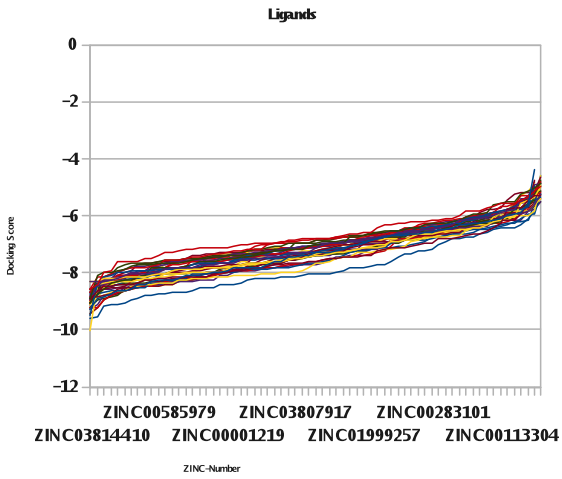


Figure 14: ZINC-number (x-axis) versus docking score (y-axis) for the ligands.

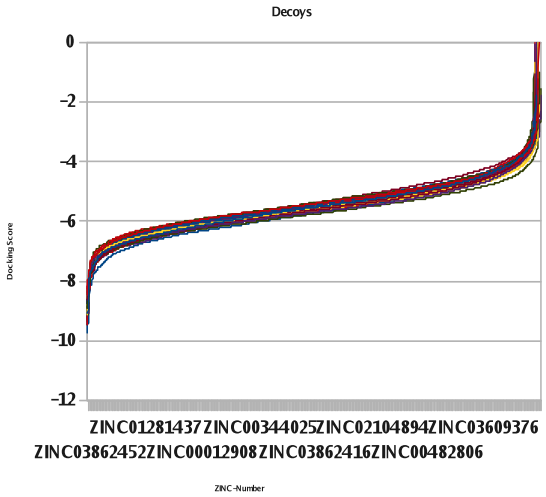


Figure 15: ZINC-number (x-axis) versus docking score (y-axis) for the decoys.

3.2.2. Enrichment Results: Docking of antagonist- and decoy- ligands in agonist- receptor structures

This docking study has been performed because especially competitive antagonists (antagonists which bind reversible and can be displaced by agonists) often bind partially in the agonist receptor binding site. [Höltje]

However, for docking of antagonist- ligands and decoys in agonist bound receptor structures a bad performance was expected, because the ligands are often too big for the binding pocket of agonist receptors. Figure 16 shows the enrichment plots for all receptorstructures. In fact almost all structures show bad enrichment and most plots are very short because a lot of ligands could not dock at all. A better enrichment is only obtained for one structure (2B23A).

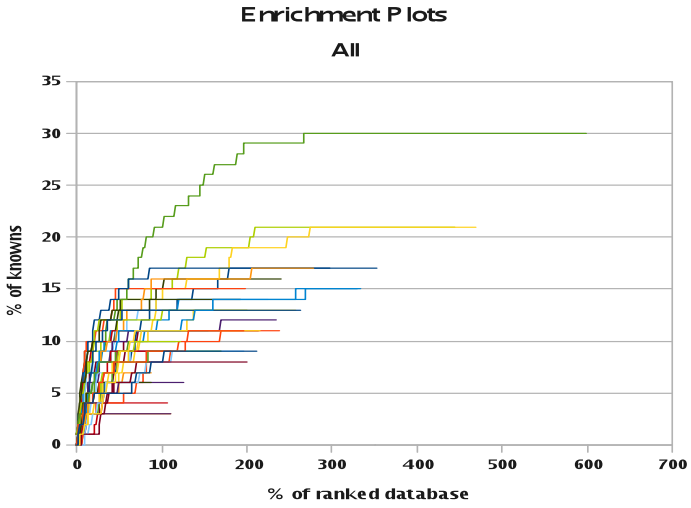


Figure 15: Bad enrichment for all receptorstructures; except for 2B23A (light green) a better enrichment was obtained. Some plots are very short because ligands could not dock at all.

3.2.3. Enrichment Results: Docking of agonist- and decoy- ligands in antagonist- receptor structures

Good enrichment has been obtained for 4 receptorstructures, 13 structures gave quite good enrichment and 3 gave bad enrichment (Table 11, Figure 16).

List of receptors with good enrichment: **1qktA, 1r5kB, 1xp6A, 2ayrA** List of receptors with bad enrichment: **1r5kC, 1xp1A, 1yinA** List of receptors with quite good enrichment: **1errA, 1errB, 1r5kA, 1sj0A, 1uomA, 1xp9A, 1xqcA, 1xpcA, 1xqcB, 1xqcC, 1xqcD, 1yimA, 3ertA**

Table 11

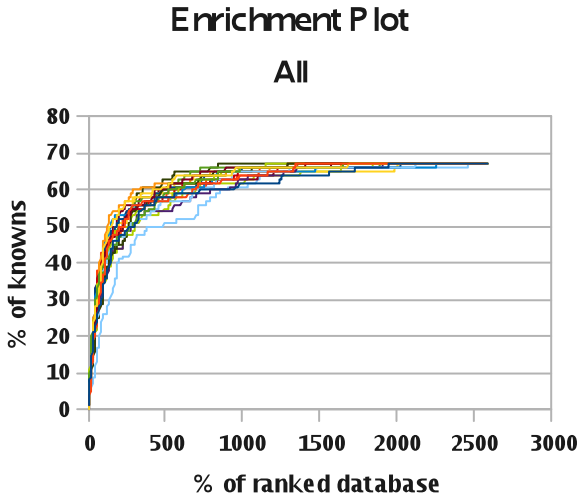


Figure 16: All enrichment plots for docking of agonist- and decoy- ligands in antagonist receptor structures are shown.

3.2.4. Enrichment Results: Docking of agonist- and decoy- ligands in antagonist- receptor structures

5 receptors gave good enrichment, 6 receptors gave bad enrichment and 9 receptors gave quite good enrichment (Table 12, Figure 17).

List of receptors with good enrichment: **1r5kB, 1xp9A, 1xqcB, 1xqcD, 3ertA** List of receptors with bad enrichment: **1errB, 1sj0A, 1xp6A, 1xqcA, 1xpcA, 1yinA** List of receptors with quite good enrichment: **1errA, 1qktA, 1r5kA, 1r5kC, 1uomA, 1xp1A, 1xqcC, 1yimA, 2ayrA**

Table 12

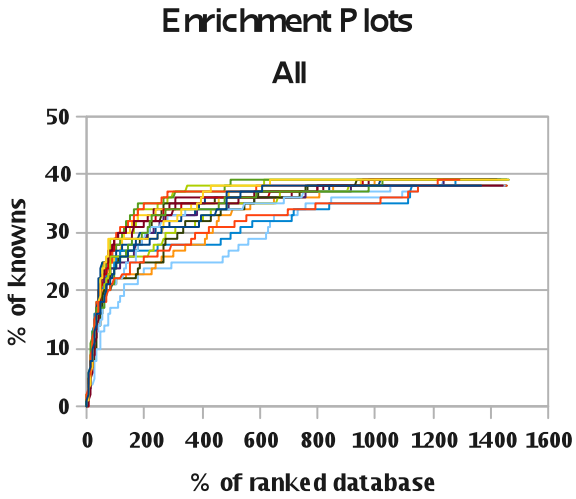


Figure 17: All enrichment plots for docking of agonist- and decoy- ligands in antagonist receptor structures are shown.

3.2.5. Docking Sensitivity

To see whether Glide can distinguish between receptor agonist and antagonist the ZINC-number (x-axis) of ligand agonists versus docking score of receptor agonist and antagonist (y-axis) has been plotted.

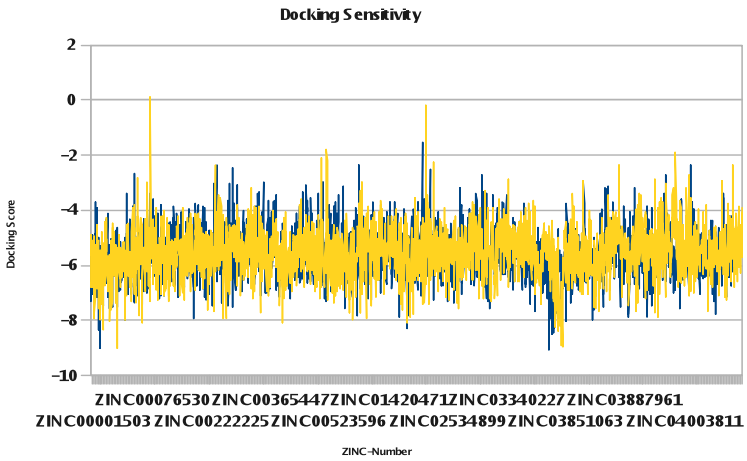


Figure 18: ZINC-number (x-axis) versus docking scores for agonist and antagonist structures (y-axis).

The Plot shows that no significant difference can be observed meaning that agonist and antagonist receptors cannot really be distinguished by the docking program.

The main structural difference between the two receptor types is that in contrast to agonist receptorstructures the antagonist receptorstructures have an opening which makes their bindingpockets slightly broader. However since this opening is only a small part of the bindingpocket-surface there will not be a big difference in the number and type of interactions between ligand- and receptoratoms.

3.2.6. Ensemble Docking

The lowest docking score of each agonist ligand and each agonist decoy has been used to make a further enrichment plot and to see whether this procedure gives a better accumulation curve or not. Figure 19 shows the enrichment plot of ensemble docking together with the best original ones.

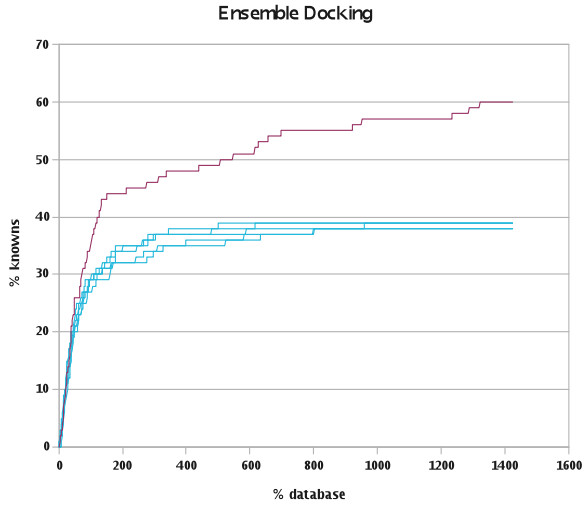


Figure 19: Enrichment plot for ensemble docking is shown in purple, the single enrichment plots are shown in blue.

As can be seen in this result, to identify the receptors where a ligand or decoy gives the best value and to use only these docking scores clearly makes sense.
Ensemble docking incorporates protein flexibility which is often neglected in molecular docking. Local rearrangements of side chains but also domain movements are common in receptor conformational changes induced by ligand binding (induced fit). Sometimes even small changes in protein conformation can effect ligand binding affinity. [Huang2007]

3.2.7. MD-Simulations

The same agonist and antagonist ligands and decoys used before have been docked in the MD-simulation structures and the results have been compared with those from the crystalstructure docking studies. Figure 20 shows the enrichment plots of both studies together in one plot and this revealed, that the dockings for the PDB-structures gave better enrichment than those from the MD-simulations. This trend has been obtained for all simulation systems.

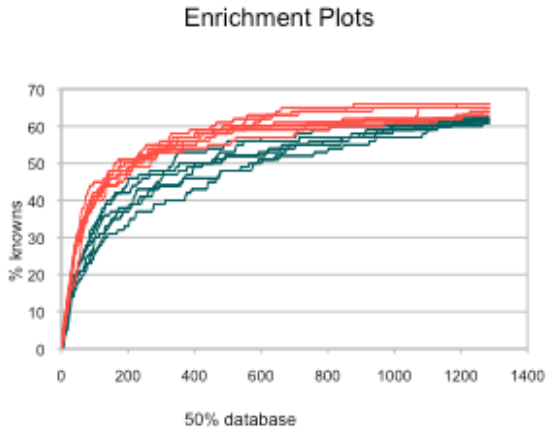


Figure 20: Enrichment plots for structures from the PDB are shown in red, enrichment plots for structures from MD-simulations (simulation system 2DNH) are shown in dark green.

3.2.8. Comparison with previous experimental evidence

One problem in simulation techniques is that they are often too computationally expensive to be used for extensive sets of compounds which are common for biological targets. [Amaro2008]

Previous studies mentioned the importance of correct treatment of solvent contributions. Water molecules can inhibit the flexibility of a bound ligand or they may even occlude potential areas of binding. Relaxed complex scheme (RCS) dockings have been carried out with and without cavity water molecules and the latter identified the best ligands. [Amaro2008]
What has been deduced is that the consequence of introducing water molecules is a significant reduction of the configurational space available to the ligand. Correct

sampling of receptor, ligand and solvent phase space is primarily reached by more expensive free energy calculations. [Amaro2008]
An important improvement would be first of all the refinements of the physical models describing ligand-binding thermodynamics. Secondly a more accurate description of the solvent contributions and the role of ligand entropy should be further investigated. Thirdly as already mentioned, the role of receptor flexibility is generally underestimated. [Amaro2008]
One way to overcome these problems would be using a more generally parameterized MD-type force field to evaluate the docked complexes. This may also allow for increased transferability of the method to other sets of systems. [Amaro2008]
A further development would also be an improved treatment of enthalpy-entropy compensation which should be especially considered when discussing the thermodynamics of proteins, ligands and nucleic acids.
The computation of absolute entropies would be another challenge for investigation. [Amaro2008]

A former ensemble-docking study mentioned the problem of improper consideration of receptor conformational changes. To overcome this problem Huang et al. proposed that an empirical energy correction term would have to be added. Another limitation is the optimization method. The problem arises from wrong predictions which originate from the optimization method, meaning that ligands may be trapped in a local minimum. Here a possible solution would be the use of a relatively global minimization method like genetic algorithm which means using operations similar to mutations and crosses. Here, the quality of the results is a function of the starting genes, mutations and crosses (evolutionary events) and the scoring function to pick favorable conformers. [Halperin2002] [Huang2007]

3.2.9. Ensemble Docking

Ensemble docking for the MD-simulation structures showed, that in the beginning of the accumulation curve, the plot is nearly as good as those for the single PDB-structures. Figure 21 shows both ensemble docking plots and the single PDB-structure plots.

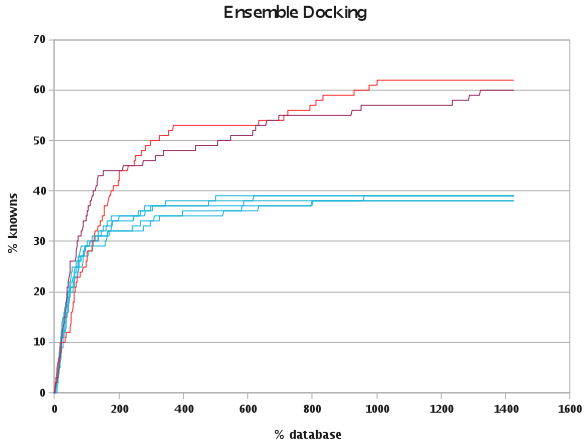


Figure 21: MD-ensemble docking is shown in red, ensemble docking for PDB-structures is shown in purple and the single PDB-structure plots are shown in blue.

4. Conclusion

What these studies show is that one can not only identify protein receptors which give a better performance for docking, furthermore, with this information one can look for structural reasons why certain receptors give better results.

Even small structural differences around the binding pocket of the estrogen receptor resulted in differences which could be measured with the refined docking programm Glide-XP.

In my study it has been shown, that the structures obtained from the PDB have good ligand binding conformations for ligand ranking.

Using ensembles of structures clearly gave an improved accuracy of compound ranking. Comparing ligand ranks with ensemble docking provides a method to pick a specific receptor for a virtual screen study.

The information obtained from ensemble docking suggests that with the use of more MD-simulation structures it should definitely be possible to improve this method to get results as good as those for the crystal structures.

Acknowledgements:

This work was supported by the Austrian Institute of Technology, Seibersdorf. Special thanks to Dr. Thomas Stockner for providing the structures obtained from MD- simulations.

Zusammenfassung:

Die wachsende Anzahl an charakterisierten molekularen Rezeptoren liefert die Basis für das struktur-basierte Design von aktiven Verbindungen zur Entwicklung eines neuen Medikamentes. Ausgehend von einer bekannten oder hypothetischen Wirkungsweise oder einem Bindungsmechanismus wird eine vernünftige Vorlage entworfen, die später experimentell getestet werden kann.
Im Gegensatz zum experimentellen High Throughput Screening (HTS), mit dem man potentielle Medikament-Kandidaten in Form von chemischen Verbindungen erhält, liefert das sogenannte Virtual Screening (VS) einfach Computer-Treffer.
Die Verbindungen werden auf der Basis eines verfeinerten Docking Algorithmus mithilfe einer approximativen Energiefunktion selektiert um sie als mutmaßliche Treffer zu reihen.

Ein solcher virtueller Screen wird üblicherweise validiert, indem man die Durchführung mit bekanntlich aktiven Verbindungen, mit der gleichen Prozedur mit einer Serie inaktiver Verbindungen, sogenannter Köder, vergleicht. Hierbei werden alle Strukturen einem selektierten VS-Protokoll unterbreitet und der Rang der aktiven Verbindungen bezüglich der übrigen wird in Anreicherungskurven konvertiert.
Diese Kurven zeigen, wie die Fraktion der aktiven Komponenten mit dem Prozentsatz der gescreenten Datenbank variiert.

Estrogene sind in Wachstum, Entwicklung und Homöostase von diversen Geweben involviert. Sie betätigen diese physiologischen Effekte über den Estrogen Rezeptor der mit Krankheiten wie Brustkrebs, Osteoporose, neurodegenerativen und kardiovaskulären Krankheiten sowie Fettleibigkeit assoziiert ist.

In dieser Studie wurde der Estrogen Rezeptor für eine Serie von Docking Studien verwendet. Es wurden Agonist- und Antagonist Strukturen benutzt sowohl aus der Protein Daten Bank (PDB) als auch aus Moleküldynamik-Simulationen.
In der ersten Studie wurden die einzelnen Liganden der Rezeptoren in die Rezeptoren zurückgedockt, eine Methode die man als ‚Bound Docking' bezeichnet.
Ein besserer Dockingwert wurde für jene Liganden erhalten die in einer korrekten Orientierung gedockt haben und vice versa.

In der zweiten Studie wurden Liganden und Köder in die Rezeptoren gedockt um festzustellen ob die Durchführung dazu fähig ist, potentielle Binder an den Beginn einer Rangliste zu setzen. Die Resultate wurden mit Anreicherungskurven validiert. Das Resultat ergab bessere Ergebnisse für die PDB-Strukturen als für die Moleküldynamik-Simulationsstrukturen.

Im letzten Schritt wurde Ensemble-docking durchgeführt was bedeutet, dass für jeden Ligand und jeden Köder gedockt in alle Rezeptoren nur der beste Dockingwert verwendet wurde. Für beide Systeme, PDB-Strukturen und Moleküldynamik-Simulationsstrukturen wurden bessere Resultate mit Ensemble-docking erhalten.

Für letztere Strukturen ergab Ensemble-docking ein Resultat das fast so gut war wie jenes für die einzelnen PDB-Strukturen.

Bibliography:

Amaro2008: Amaro RE, Baron R, McCammon JA, An improved relaxed complex scheme for receptor flexibility in computer-aided drug design, 2008

Aranda2001: Aranda A, Pascual A, Nuclear hormone receptors and gene expression, 2001

Avendaño, ex Medicinal Chemistry of Anticancer Drugs: Carmen Avendaño and J. Carlos Menéndez, Medicinal Chemistry of Anticancer Drugs, Elsevier B.V., 2008

Babiker2002: Babiker FA, De Windt LJ, van Eickels M, Grohe C, Meyer R et al., Estrogenic hormone action in the heart: regulatory network and function, 2002

Bottegoni2009: Bottegoni G, Kufareva I, Totrov M, Abagyan R, Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking, 2009

Brzozowski1997: Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T et al., Molecular basis of agonism and antagonism in the oestrogen receptor, 1997

Burris; ex Nuclear Receptors and Genetic Disease: Burris T.P., Nuclear Receptors and Genetic Disease, Burris T.P., McCabe E.R.B., ACADEMIC PRESS, 2001

Deroo2006: Deroo BJ, Korach KS, Estrogen receptors and human disease, 2006

Englebienne2007: Englebienne P, Fiaux H, Kuntz DA, Corbeil CR, Gerber-Lemaire S et al., Evaluation of docking programs for predicting binding of Golgi alpha-mannosidase II inhibitors: a comparison with crystallography, 2007

Fernandes2004: Fernandes MX, Kairys V, Gilson MK, Comparing ligand interactions with multiple receptors via serial docking, 2004

Friesner2006: Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR et al., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes, 2006

Gangloff2001: Gangloff M, Ruff M, Eiler S, Duclaud S, Wurtz JM et al., Crystal structure of a mutant hERalpha ligand-binding domain reveals key structural features for the mechanism of partial agonism, 2001

Gohlke2002: Gohlke H, Klebe G, Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, 2002

Gronemeyer2004: Gronemeyer H, Gustafsson J, Laudet V, Principles for modulation of the nuclear receptor superfamily, 2004

Halgren, Murphy, Friesner; ex Virtual Screening in Drug Discovery: Halgren, Murphy, Friesner, Virtual Screening in Drug Discovery, Juan Alvarez, Brain Shoichet, CRC Press, 2005

Halperin2002: Halperin I, Ma B, Wolfson H, Nussinov R, Principles of docking: An overview of search algorithms and a guide to scoring functions, 2002

Höltje, Sippl, Rognan, Folkers; ex Molecular Modeling: Höltje H.-D., Sippl W., Rognan D., Folkers G., Molecular Modeling Basic Principles and Applications, Second Edition, WILEY-VCH, 2003

Huang2007: Huang S, Zou X, Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking, 2007

Irwin2005: Irwin JJ, Shoichet BK, ZINC--a free database of commercially available compounds for virtual screening, 2005

Kim2004: Kim S, Wu JY, Birzin ET, Frisch K, Chan W et al., Estrogen receptor ligands. II. Discovery of benzoxathiins as potent, selective estrogen receptor alpha modulators, 2004

Kitchen2004: Kitchen DB, Decornez H, Furr JR, Bajorath J, Docking and scoring in virtual screening for drug discovery: methods and applications, 2004

Klebe; ex Virtual Screening in Drug Discovery: Klebe G., Virtual Screening in Drug Discovery, Juan Alvarez, Brain Shoichet, CRC Press, 2005

Klebe2006: Klebe G, Virtual ligand screening: strategies, perspectives and limitations, 2006

Kleine, Rossmanith, ex Hormone und Hormonsystem: Kleine B., Rossmanith W. G., Hormone und Hormonsystem Eine Endokrinologie fuer Biowissenschaftler, Springer, 2007

Laudet1997: Laudet V, Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor, 1997

Lee2008: Lee J, Seok C, A statistical rescoring scheme for protein-ligand docking: Consideration of entropic effect, 2008

Meiler2006: Meiler J, Baker D, ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility, 2006

Moustakas, Pegg, Kuntz; ex Virtual Screening in Drug Discovery: Moustakas D.T., Pegg S.C.H., Kuntz I.D., Virtual Screening in Drug Discovery, Juan Alvarez, Brain Shoichet, CRC Press, 2005

Olefsky2001: Olefsky JM, Nuclear receptor minireview series, 2001

Perola, Walters, Charifson, ex Virtual Screening in Drug Discovery: Perola E., Walters W.P., Charifson P.S., Virtual Screening in Drug Discovery, Juan Alvarez, Brain Shoichet, CRC Press, 2005

Perola2007: Perola E, Walters WP, Charifson P, Comments on the article "On

evaluating molecular-docking methods for pose prediction and enrichment factors", 2007

Pike1999: Pike AC, Brzozowski AM, Hubbard RE, Bonn T, Thorsell AG et al., Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist, 1999

Rao2008: Rao S, Sanschagrin PC, Greenwood JR, Repasky MP, Sherman W et al., Improving database enrichment through ensemble docking, 2008

Van_Der_Spoel2005: Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE et al., GROMACS: fast, flexible, and free, 2005

Yaghmaie2005: Yaghmaie F, Saeed O, Garan SA, Freitag W, Timiras PS et al., Caloric restriction reduces cell loss and maintains estrogen receptor-alpha immunoreactivity in the pre-optic hypothalamus of female B6D2F1 mice, 2005

Curriculum Vitae

Personal information

First names and Surname	Katharina Maria Ostermann
Date of birth	02.01.1981
Nationality	Austrian
Residence	1140 Vienna, Austria

Education and training

Dates	*1999 – present*
Name and type of organisation providing education and training	Vienna University, General Chemistry
Specializations	Biochemistry, Analytical Chemistry
Level in national or international classification	Universitiy studies

Dates	*03.09 2004 – 10.12.2004*
Studies abroad	University of Sussex, Brighton, UK
Practices	Research Methods in Molucular Genetics, Bio-organic Chemistry, Chemistry of Hormones, Medicinal Chemistry

Dates	*01.08.2003 – 14.09.2003*
Work experience	Böhringer Ingelheim Austria GmbH, 1121 Vienna
Practices	Laboratory for organic synthesis

Dates	*1991 – 1999*
Name and type of organisation providing education and training	Wirtschaftskundliches Realgymnasium der Dominikanerinnen, 1130 Vienna
Major subjects	Chemistry, English language, Psychology, Music
Level in national or international classification	High School studies / Matura