

# **Mixed-model QSAR at the Glucocorticoid and Liver X receptors**

**Inauguraldissertation**

zur  
Erlangung der Würde eines Doktors der Philosophie  
vorgelegt der  
Philosophisch-Naturwissenschaftlichen Fakultät  
der Universität Basel

von

**Morena Spreafico**

**aus Galbiate (Italien)**

Basel, April 2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

Auf Antrag von:

Prof. Dr. A. Vedani, Institut für Molekulare Pharmazie, Universität Basel,

Prof. Dr. T. Langer, Prestwick Chemical Inc., Strasbourg, Frankreich, vormals  
Vorsteher Department Pharmazie, Universität Innsbruck, Österreich

Basel, den 28. April 2009

Prof. Dr. Eberhard Parlow

Dekan

## Acknowledgements

My PhD work was carried out at the Institute of Molecular Pharmacy at the University of Basel, and was generously supported by the Margaret and Francis Fleitmann-Foundation (Lucerne, Switzerland).

First of all I would like to thank Prof. Dr. Angelo Vedani for the possibility to work within his group, for his great support, encouragements and challenges that allowed me to grow, from a professional and personal point of view, and Prof. Dr. Beat Ernst, who gave me the opportunity to work in a stimulating and multicultural environment and to improve my competences through international schools and workshops. Thanks to Prof. Dr. Thierry Langer for kindly accepting to be the co-referee for this thesis.

I am extremely grateful to the “modelers” for the support and the friendly environment, and in particular to Dr. Martin Smiesko for the fruitful conversations and friendship, to Prof. Dr. Markus Lill for the enlightening discussions, to Prof. Dr. Max Dobler for the for the various pieces of software he developed and without which this thesis would not have been possible, to Gian for the glimpse of Italy he brought and most of all to have become in such a short time a great friend, and to Rania for sharing working experiences and the everyday life in the lab. Thanks to Anne and to the other students that spent some time in our lab, for the important professional and personal exchange.

My sincere and deep thanks are also going to the friends and colleagues of the IMP: to Mätu first of all, for his patience and strong support in a very difficult moment and everyday since then, to Kiki, because I miss her here very much, to Steffi, because of the activities she organized and because I learnt a lot from our short collaboration, to Jönu for the help with computers and the trip organization, to Meike for the unexpected hugs, to Gabi for the work from the backstage, and to all the others. A special thank goes also to the former members of IMP, in

particular to Karin, Dane, Michele, Alex, Dani, Daniela, Adrian, and all the people that made of this adventure a wonderful experience.

Thanks a lot to the friends I met in Basel: to Anja, Maike, Francesco, and the rest of the crew for the precious friendship that goes far beyond our common linguistic or gastronomical interests. Thanks to the “circolo pane e tulipani” to be so close from far away: you’re a fixed point in my moving life. Thanks to my children in Cuxuryrd, because from you I learnt the value of things.

But the biggest thanks go to my family. Grazie Ste, per questo filo di Arianna che ci lega in qualsiasi parte del mondo ci troviamo. E grazie ai miei genitori, per avermi sempre incoraggiato e supportato anche quando le mie scelte mi hanno portato lontano da voi. Sono davvero fortunata.

The theory is when we know everything and nothing works.  
The practice is when everything works and nobody knows why.  
If the practice and theory are met, nothing works and we do not know why.

Albert Einstein

## Abstract

The presence of hormonally active compounds in the biosphere has become a worldwide environmental concern, and measures such as policy acts and regulations try to address the problem, both in Europe and in the United States. Such compounds, referred to as endocrine disruptors, may alter the functions of the endocrine system and consequently cause adverse health effects in organism, or its progeny, or populations.<sup>1</sup> A safe *in silico* identification of the toxic potential of drugs and chemicals is therefore highly desirable by both regulatory bodies, and the pharmaceutical industry.

Nuclear receptors regulate biological functions such as cell growth and differentiation, metabolic processes, reproduction and development, intracellular signaling and can be involved in carcinogenesis through control of gene expression.<sup>2</sup> Chemicals that disrupt the endocrine system interfere with the function of nuclear receptors, alter their functions and consequently cause adverse health effects.<sup>1</sup>

In this thesis, the development and validation of *in silico* three-dimensional models for the glucocorticoid and the liver X receptors, both belonging to the nuclear receptor superfamily, are presented. These models aim at the screening of drug candidates for glucocorticoid and liver X activity and of environmental chemicals for potential endocrine-disrupting activity.

Different *in silico*-based tools and protocols were used to model receptor-ligand interactions. Molecular dynamics simulations enabled to gain an insight into the dynamical character of the protein-ligand interactions. An appropriate consideration of receptor flexibility (induced fit) was a prerequisite for the identification of realistic binding modes, which was performed with flexible docking. Once a suitable alignment was obtained, QSAR models were built, using two different technologies, and tested by the application to external validation sets, scramble tests and consensus scoring.

The models have been added to the *VirtualToxLab*<sup>TM3, 4</sup> – a technology for the *in silico* identification of the toxic (endocrine-disrupting) potential of drugs and environmental chemicals.

Special consideration was given to the role of hydrophobic effect in ligand binding. An empirical scoring function (*Heidi*: Hydrophobic Effect in Drug Interactions) was developed to quantify the hydrophobic effect for scoring protein–ligand binding energies. The use of *HEidi*, together with electrostatic, van der Waals and hydrogen bond energies, in the ranking of docking poses provided encouraging results when applied to glucocorticoid and liver X receptor complexes, but for a generalized statement more extensive evaluations are needed.

## Abbreviations

CoMFA	Comparative Molecular Field Analysis
ED	Endocrine Disruptor
EDC	Endocrine-Disrupting Chemical
EDSTAC	Endocrine Disruptor Screening and Testing Advisory Committee
e.g.	Exempli Gratia (for Example)
EPA	Environmental Protection Agency
esd	Estimated Standard Deviation
FEP	Free Energy Perturbation
i.e.	Id Est (That is)
IC <sub>50</sub>	Inhibitory Concentration 50%
IPCS	International Programme on Chemical Safety
GR	Glucocorticoid Receptor
K <sub>i</sub>	Inhibition Constant
LIE	Linear Interaction Energy
LXR	Liver X Receptor
MC	Monte Carlo
MD	Molecular Dynamics
mD-QSAR	Multi-Dimensional QSAR
MM	Molecular Mechanics
NR	Nuclear Receptor
NRP50	National Research Programme on Endocrine Disruption
OECD	Organization for Economic Cooperation and Development
$\rho^2$	Predictive Correlation Coefficient
PBSA	Poisson Boltzmann Surface Area
pK <sub>i</sub>	-LOG(K <sub>i</sub> )
$q^2$	Cross-validated Correlation Coefficient
QSAR	Quantitative Structure-Activity Relationship
$r^2$	Correlation Coefficient
REACH	Registration, Evaluation, and Authorization of CHemicals
rms	Root Mean Square
RMSD	Root Mean Square Distance
SAR	Structure-Activity Relationship
SEGRA	Selective Glucocorticoid Receptor Agonist
TI	Thermodynamic Integration
WHO	World Health Organization

# Table of Contents

<b>1 INTRODUCTION.....</b>	<b>1</b>
1.1 ENDOCRINE DISRUPTION.....	1
1.2 NUCLEAR RECEPTORS .....	4
1.2.1 <i>The Glucocorticoid Receptor</i> .....	6
1.2.2 <i>Liver X Receptors</i> .....	8
1.3 CALCULATION OF PROTEIN-LIGAND BINDING AFFINITIES .....	11
1.3.1 <i>Docking and Scoring</i> .....	12
1.3.2 <i>Hydrophobic Effects in Protein-Ligand Binding</i> .....	14
1.3.3 <i>Quantitative Structure–Activity Relationships (QSARs)</i> .....	17
1.3.4 <i>Validation of QSAR Models</i> .....	18
<b>2 AIM OF THE THESIS.....</b>	<b>20</b>
<b>3 MATERIAL AND METHODS.....</b>	<b>21</b>
3.1 MATERIAL.....	21
3.2 SOFTWARE .....	22
3.2.1 <i>AMBER</i> .....	22
3.2.2 <i>AMSOL</i> .....	22
3.2.3 <i>Biograf<sup>x</sup></i> .....	22
3.2.4 <i>Bio<sup>x</sup></i> .....	23
3.2.5 <i>Epik</i> .....	23
3.2.6 <i>Glide</i> .....	24
3.2.7 <i>MacroModel</i> .....	24
3.2.8 <i>PrGen</i> .....	24
3.2.9 <i>Quasar</i> .....	24
3.2.10 <i>Raptor</i> .....	25
3.2.11 <i>Swiss-PdbViewer Deep View</i> .....	26
3.2.12 <i>VMD</i> .....	26
3.2.13 <i>VirtualToxLab<sup>TM</sup></i> .....	26
3.2.14 <i>Xcode Tools</i> .....	27
3.2.15 <i>Yeti</i> .....	27
3.3 METHODS.....	29
3.3.1 <i>Experimental Binding Affinities for GR and LXR Ligands</i> .....	29
3.3.2 <i>Ligand Structure Generation and Conformational Analysis</i> .....	29
3.3.3 <i>Protein Preparation</i> .....	30
3.3.4 <i>The Docking Protocol</i> .....	31



3.3.5 Molecular-Dynamics Simulations.....	31
3.3.6 QSAR Studies: Quasar.....	33
3.3.7 Model Validation.....	34
3.3.8 Estimation of Toxic Potential through VirtualToxLab™.....	35
<b>4 RESULTS AND DISCUSSION.....</b>	<b>36</b>
4.1 GLUCOCORTICOID RECEPTOR (GR).....	36
4.1.1 Analysis of the GR Crystal Structures.....	36
4.1.2 Retrieval of the Binding Affinity Data and Preparation of the Compound Structures.....	39
4.1.3 Docking to the GR.....	41
4.1.4 Molecular-Dynamics Simulations on the GR.....	49
4.1.5 Building and Validation of QSAR Models for GR.....	53
4.1.6 Binding of Psychotropic Drugs to the GR.....	61
4.1.7 Conclusions and Applicability of the GR Model.....	66
4.2 LIVER X RECEPTOR (LXR).....	67
4.2.1 Analysis of the LXR Crystal Structures and Receptor Preparation.....	67
4.2.2 Molecular-Dynamics Simulations on the LXR.....	68
4.2.3 Preparation of the Binding Affinity Data and Compound Structures.....	71
4.2.4 Docking to the LXR.....	74
4.2.5 Building and Validation of a QSAR Models for the LXR.....	78
4.2.6 Conclusions and Applicability of the LXR Model.....	85
4.3 HYDROPHOBIC EFFECT ESTIMATION AND EVALUATION.....	86
4.3.1 Development of an Empirical Hydrophobicity Function.....	86
4.3.2 Preliminary Analysis on a Protein–Ligand Complex.....	91
4.3.3 Application to the GR and LXR.....	94
4.3.4 Applicability of the Hydrophobic Effect Function.....	106
<b>5 CONCLUSIONS AND OUTLOOK.....</b>	<b>107</b>
<b>6 APPENDIX A: CHEMICAL STRUCTURES, PK<sub>i</sub> VALUES AND SCRAMBLE TESTS OF GLUCOCORTICOID LIGANDS.....</b>	<b>109</b>
<b>7 APPENDIX B: CHEMICAL STRUCTURES, PK<sub>i</sub> VALUES AND SCRAMBLE TESTS OF LIVER X RECEPTOR LIGANDS.....</b>	<b>124</b>
<b>8 APPENDIX C: LIST OF VAN DER WAALS RADII.....</b>	<b>131</b>
<b>9 REFERENCES.....</b>	<b>132</b>
<b>10 CURRICULUM VITAE.....</b>	<b>150</b>

# 1 Introduction

## 1.1 Endocrine Disruption

The last 50 years, since the publication of Rachel Carson's *Silent Spring*,<sup>5</sup> have witnessed growing scientific concerns and public debate over the potential adverse effects that may result from exposure to chemicals that have the potential to alter the normal functioning of the endocrine system in wildlife and humans.<sup>1</sup> Despite of the disagreement surrounding how to best define endocrine disruptors<sup>6</sup> (EDs), the most frequently applied definition of an endocrine disruptor is that of "*an exogenous agent that interferes with the production, release, transport, metabolism, binding, action or elimination of natural hormones in the body responsible for the maintenance of homeostasis and the regulation of developmental processes*".<sup>7</sup>

Concerns about EDs are primarily due to

- Adverse effects observed in certain wildlife, fish, and ecosystems<sup>8,9</sup>
- The increased incidence of certain endocrine-related human diseases<sup>10-12</sup>
- Endocrine disruption resulting from exposure to certain environmental chemicals observed in laboratory experimental animals, or humans<sup>1</sup>

There are different routes how chemicals can interfere with hormones or disrupt the hormonal system — the most straightforward by directly binding and blocking a hormone receptor. They can also trigger cellular mechanisms by mimicking the action of a natural hormone, or by affecting the synthesis, transport, metabolism and excretion of hormones.<sup>1</sup>

The theory of endocrine disruption postulates that low-dose exposure to chemicals that interact with hormone receptors can interfere with reproduction, development, and other hormonally mediated processes. The timing of exposure is presumed to be critical, since different hormone pathways are active during different stages of development. Whether or not low-level and long-term exposures to such chemicals have adverse effects is the most controversial issue.<sup>1</sup>

Even if many aspects of endocrine disruption still have to be elucidated, abnormalities in laboratory animals and wildlife exposed to endocrine-disrupting chemicals are reported. They include feminization of males, abnormal sexual behaviour, birth defects, altered sex ratios, decreased sperm density, decreased size of testes, breast cancer, testicular cancer, reproductive failure and thyroid dysfunction (Table 1, part of the table from Solomon et al.).<sup>13</sup>

**Table 1.** Examples of endocrine-disrupting chemicals

Chemical	Use	Mechanism	Health effect
Arsenic	Used in herbicides, insecticides, alloys, industrial compounds, mining practice	Glucocorticoid receptor transcription inhibitor	Humans: lung, skin, liver, bladder cancer <sup>14</sup>
DES	Synthetic estrogen	Estrogen receptor agonist	Humans (prenatal exposure): vaginal cancer, reproductive tract abnormalities (females); cryptorchidism, hypospadias, semen abnormalities (males) <sup>15</sup>
Methoxychlor	Insecticide	Metabolite is an estrogen receptor agonist	Rodents: accelerated puberty, abnormal ovarian cycling (females); aggressive behavior (males) <sup>16, 17</sup>
DDT	Insecticide	Metabolite (DDE) is an androgen receptor antagonist	Rodents (males): delayed puberty, reduced sex accessory gland size, altered sex differentiation <sup>18</sup>
Vinclozolin	Fungicide	Androgen receptor antagonist	Rodents (males): feminization, nipple development, hypospadias <sup>19</sup>
PCBs	No longer manufactured; still in electrical transformers, capacitors, toxic waste sites, food chain	Glucocorticoid receptor antagonist Accelerated T <sub>4</sub> metabolism, decreased T <sub>4</sub> levels, elevated TSH levels (high doses: thyromimetic)	Humans ( <i>in utero</i> exposure): delayed neurological development; IQ deficits <sup>20-22</sup>
Atrazine	Herbicide	Reduces gonadotropin-releasing hormone from hypothalamus, reduces pituitary LH levels, interferes with metabolism of estradiol, blocks estrogen receptor binding	Rodents (females): mammary tumors, abnormal ovarian cycling Humans: some evidence of breast and ovarian tumors <sup>23-27</sup>
Dioxin	By-product of industrial processes including waste incineration; food contaminant	Aryl hydrocarbon receptor agonist; increases estrogen metabolism, decreases estrogen-mediated gene transcription, decreases estrogen levels, decreases testosterone levels by interfering with HPG axis	Rodents ( <i>in utero</i> exposure): delayed puberty, increased susceptibility to mammary cancer (females); decreased testosterone, hypospadias, hypospermia, delayed testicular descent, feminized sexual behavior (males) Humans: decreased T <sub>3</sub> and T <sub>4</sub> levels, decreased testosterone levels*, cancer* <sup>28-32</sup>

DES=diethylstilbestrol, DDT=dichlorodiphenyltrichloroethane, PCBs=polychlorinated biphenyls, T<sub>4</sub>=thyroxine, T<sub>4</sub>=triiodothyronine, TSH=thyroid stimulation hormone, IQ=intelligence quotient, LH=luteinizing hormone, HPG axis=hypothalamic-pituitary-gonadal axis.\*Exposures in adults.

Within the last decade, the field of chemical disruption of the endocrine system became an active area of research that captivated the scientific world, and captured the attention of governments, policymakers, and the media.

In 1996, endocrine disruption was identified as one of the six high-priority research topics within the U.S. Environmental Protection Agency (U.S EPA).<sup>7</sup> In the same year, the U.S EPA designated a special task force, the Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC), which was assigned the task of making recommendations for the development of testing and screening programs for endocrine disruptors.<sup>33</sup> Likewise, the Organization for Economic Cooperation and Development (OECD) also established a special activity for endocrine disruptor testing and assessment.<sup>34</sup> Subsequently, the World Health Organization (WHO) tasked the International Programme on Chemical Safety (IPCS) with preparation of a report describing the Global Assessment of the Scientific Literature on Endocrine Disrupting Chemicals.<sup>1</sup>

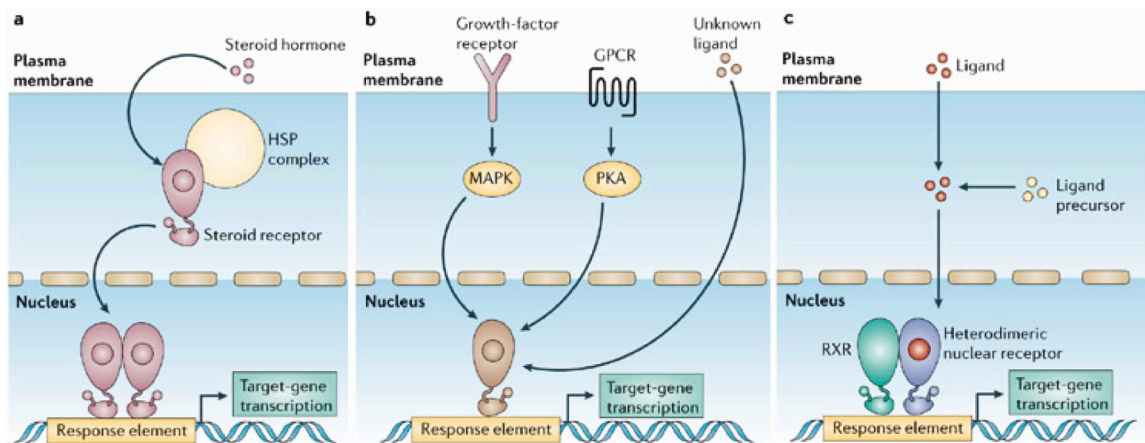
In December 2006, the European Union approved the Registration, Evaluation, and Authorization of CHemicals<sup>35</sup> (REACH, EC 1907/2006), a regulation that covers the production and use of chemical substances. REACH entered into force on 1<sup>st</sup> June 2007. In particular, additional authorization for substances of “very high concern”, such as endocrine disruptors, is required. In Switzerland, the necessity for a coordinated interdisciplinary approach has also been recognized and the National Research Programme on Endocrine Disruption (NRP50) was conducted 2001–2007.<sup>36</sup> According to the REACH regulation, about 30,000 chemicals will have to be registered over 11 years. REACH specifically states that the test methods used are to be revised “in particular to refine, reduce or replace animal testing”, and that “before new tests are carried out to determine the properties listed in this Annex, all available *in vitro* data, *in vivo* data, historical data, data from valid (Q)SARs and data from structurally related substances shall be accessed first”.

A reliable *in silico* identification of the endocrine-disrupting (or, more general, the toxic) potential of drugs and chemicals is therefore regarded as highly desirable by both regulatory bodies and the pharmaceutical, chemical, and food industry.

Endocrine-disrupting chemicals (EDCs) can act at multiple sites through multiple mechanisms of action. Using receptor-binding assays and receptor-based functional assays, researchers have shown that some environmental chemicals interact with nuclear receptors and cause adverse effects in humans and rodents (Table 1).

## 1.2 Nuclear Receptors

Nuclear receptors (NRs) are one of the most abundant classes of transcriptional regulators in animals and probably the most studied class associated with endocrine disruption. They comprise a superfamily of structurally conserved, ligand-dependent transcription factors that regulate diverse aspects of development, metabolism, reproduction and homeostasis<sup>37</sup>. NR can be subdivided into three classes (Figure 1), based on their ligand-binding and DNA-binding properties.<sup>38-41</sup>



**Figure 1.** The three classes of NRs. **a:** Steroid receptors (e.g. glucocorticoid receptors) are synthesized in inactive forms that are associated with heat-shock protein (HSP). Hormone binding causes dissociation of steroid receptors from HSP complexes and allows binding to specific response elements in target genes. **b:** Heterodimeric NRs (e.g. liver X receptors) bind constitutively to DNA with retinoid X receptors as obligate partners. **c:** A subset of NRs bind DNA as monomers. In most cases, these receptors are designated as 'orphans'. They might mediate transcription through changes in their expression or post-translational modifications (Figure from Glass et al.).<sup>42</sup>

The classical steroid- and thyroid-hormone receptors (e.g. the glucocorticoid and estrogen receptors) are the first and most extensively studied class of nuclear receptors. In the unbound state, they are generally associated in the nucleus or in the cytoplasm, to other proteins, such as heat-shock proteins. Following binding, they dissociate from heat-shock proteins and initiate transcriptional activity. The ligand binding domain mediates the recruitment of coactivator or co-repressor protein in a ligand-dependent manner and the transcription of target genes is activated or repressed.

Orphan receptors, a second class of nuclear receptors, show the structurally conserved features of the nuclear-receptor superfamily, but they have not been linked to naturally occurring ligands and, in some cases, function in a ligand-independent manner.

A third class of nuclear receptors consists of the so-called 'adopted' orphan receptors (e.g. the liver X receptors). Those receptors were initially classified as

orphan receptors, but subsequent studies identified naturally occurring ligands and determined their physiological roles.<sup>42</sup> They exist already bound to DNA, together with the retinoid X receptor to form heterodimers. Following ligand binding, the heterodimeric receptors switch their conformation and activate or repress transcription through binding to various coactivator or corepressor proteins.

Three structural and functional domains, variable in length, form the general structure of NRs.<sup>2</sup> The variability involves also some functions, such as activators binding, which do not always correspond to a simple structural domain. The modulatory N-terminal domain is the least conserved domain across the superfamily, and it houses activation function 1 (AF-1), one of the two regions involved in transactivation functions. Adjacent to the N-terminal domain is the DNA binding domain (DBD). This is the most conserved region of the receptor, and interacts with DNA, by allowing a specific recognition of short nucleotide sequences. At the C terminus, the ligand binding domain (LBD) is responsible for hormone binding, and contains the second activation function (AF-2), tightly regulated by hormone binding, that accounts for coactivators and corepressors recruitment. The interactions with coactivator and corepressor proteins are due to conformational changes in the  $\alpha$ -helical region of AF2.

The presence of ligands that can bind to the NRs is one of the most important determinants in NRs activity. Several recent studies have reported microarray analysis of the impact of NRs ligands on inflammatory programs of gene expression.<sup>43-45</sup> Steroid receptors, such the glucocorticoid receptor (GR), regulate inflammatory gene expression in response to circulating hormones that are mainly produced in tissues under the control of the hypothalamic-pituitary-adrenal axis. By contrast, liver X receptors (LXRs) are heterodimeric NRs that are mainly regulated by ligands produced in a paracrine or autocrine manner. From these studies emerges a specific gene regulation for each NR, suggesting a distinct, but cooperative (systemic and local) role in inflammation processes.<sup>42</sup>

Xenobiotics that bind to NRs can disturb the cooperation of signaling pathways, giving rise to adverse health effects that are addressed by numerous studies.

Among NRs, GR and LXRs were studied in this thesis with molecular modeling techniques, in order to build QSAR models able to predict the binding affinity of compounds for the two receptors. An *in silico* prediction of the binding affinity towards GR and LXRs would be desirable both in pharmaceutical industry, in the design of new drugs, and in environmental toxicology, for the investigation of endocrine-disrupting potential of environmental compounds.

## 1.2.1 The Glucocorticoid Receptor

The GR is a ligand-activated transcription factor controlling a wide variety of biological processes including development, metabolism, and the immune response of the organism.<sup>46</sup>

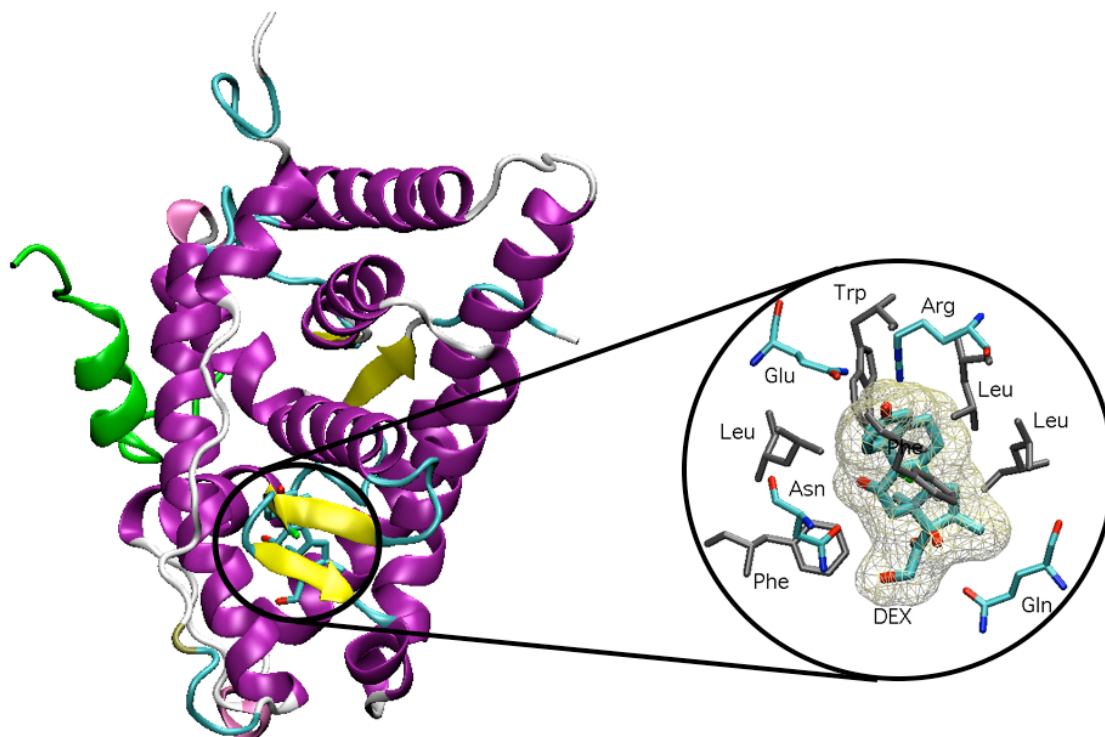
Ligands of the GR, the glucocorticoids, are therapeutically widely used for their anti-inflammatory and immunosuppressive activities, to treat numerous pathological conditions such as asthma, allergic rhinitis, rheumatoid arthritis, or acute transplant rejection.<sup>47</sup> However, a range of side effects including osteoporosis, metabolic syndrome, impaired development, and blunted growth, limits their clinical use.<sup>48</sup> Therefore, the identification of new glucocorticoids is still an endeavor in pharmaceutical R&D, and *in silico* tools assisting the rational design of glucocorticoids, — particularly by quantifying their binding affinity — are, consequently, much needed.

Two isoforms of GR, due to alternative splicing, have been identified:<sup>49</sup> GR $\alpha$  and GR $\beta$ , which differ in the structural composition of the ligand binding domain and therefore in their ability to bind glucocorticoid ligands. In GR $\beta$  helix 12 is missing in the C terminus, resulting in a receptor apparently unable to bind ligands and to activate responsive promoters. Recent studies however, show some evidence of ligand binding and of transcriptional activation.<sup>50</sup> Additionally, GR $\beta$  may act as a dominant negative to repress the transcriptional activity of GR $\alpha$ .<sup>51, 52</sup> Because of the ability of GR $\beta$  to antagonize the action of GR $\alpha$ , it has been hypothesized that changes in the expression of GR $\beta$  may underlie the development of glucocorticoid resistance.<sup>53</sup>

In contrast, GR $\alpha$  is the classically functional GR. It can both activate and repress the transcription of target genes via binding to glucocorticoid responsive elements or cross-talk with other transcription factors, such as activator protein-1 or nuclear factor- $\kappa$ B,<sup>54, 55</sup> to repress their gene activation activities. This GR-mediated repression of pro-inflammation genes, has been postulated to be a molecular basis for the anti-inflammatory and immunosuppressive activities of glucocorticoids.<sup>56-59</sup> The side effects of glucocorticoids seem to be associated with both repression and activation of specific genes:<sup>60</sup> the bone-related side effects appear related with repression of genes involved in osteoblast function and bone formation,<sup>61-63</sup> while the metabolic side effects seem due to transcriptional activation of enzymes involved in gluconeogenesis, lipid metabolism, and enzymes involved in muscle metabolism.<sup>64, 65</sup>

A common feature of the GR ligand binding domains is a helical sandwich fold that nests a ligand binding pocket within the bottom half of the domain (Figure 2). The ligand binding pocket of GR is extremely adaptable and is able to accommodate a diverse set of ligands. The volume of the binding pocket is ranging from 540 Å<sup>3</sup> (when bound to dexamethasone) to 1,070 Å<sup>3</sup> (when bound to deacylcortivazol).<sup>66</sup> The cavity is mostly lined with hydrophobic residues, but it is clear that both polar and non-polar residues play specific roles in ligand

recognition.<sup>67</sup> When bound to an agonist, like dexamethasone, the AF-2 function at the C terminus of the ligand binding domain is positioned in the classical agonistic conformation, allowing for coactivator recognition. In contrast, when GR is bound to an antagonist, like mifepristone, AF-2 is displaced such that there is no direct interaction between the residues of AF-2 and the ligand, and the receptor loses the ability to bind to the coactivator.

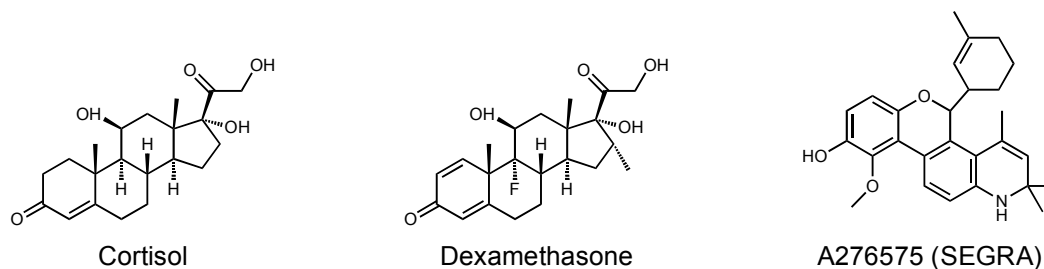


**Figure 2.** Left: side view of the crystal structure of the GR ligand binding domain, when bound to dexamethasone (PDB 1M2Z).<sup>68</sup> The structure is a helical sandwich where the ligand occupies the lower part of the domain. In the agonist form, GR can bind a coactivator peptide. The receptor is represented as ribbons and colored by secondary structure. The coactivator peptide is colored in green, and the ligand dexamethasone represented as licorice. Right: zoom into the binding pocket. Dexamethasone and amino-acid residues lining the binding pocket are represented as licorice. Amino-acid residues involved in hydrogen bonds are colored by atom type, and the ones involved in hydrophobic interaction are colored in gray.

The name *glucocorticoid* derives from early observations that these hormones are involved in glucose metabolism. The vast majority of glucocorticoid activity in most mammals is from cortisol, also known as hydrocortisone. It is essential for life, and regulates or supports a variety of important cardiovascular, metabolic, immunologic, and homeostatic functions. Various synthetic glucocorticoids are also available, such as dexamethasone. They are used either as replacement therapy in glucocorticoid deficiency or to suppress the immune system. Unfortunately the anti-inflammatory and immunosuppressive effects of classical glucocorticoids are frequently accompanied by undesired side effects. Recent efforts in pharmaceutical industry is then devoted to discover selective glucocorticoid receptor agonist (SEGRA) compounds,<sup>69</sup> able to retain the positive anti-



inflammatory activity but deprived from the side-effects.<sup>69, 70</sup> Currently, they are being investigated in cellular<sup>69</sup> and animal models,<sup>71</sup> but none of them has reached yet clinical trials. Cortisol, dexamethasone and an example of a SEGRA compound are shown in Figure 3.



**Figure 3.** Chemical structure of cortisol, dexamethasone and a non steroidal selective glucocorticoid receptor agonist.

### 1.2.2 Liver X Receptors

Liver X receptors (LXRs) were initially described as orphan receptors. Later, oxidized cholesterol derivatives or oxysterols were identified as specific ligands of LXRs,<sup>72</sup> which are therefore also named “oxysterols receptor”. Studies performed during the last decade suggest that LXRs have an important role in sensing the intracellular sterol level by regulating genes for controlling the absorption, storage, transport and metabolism of cholesterol.<sup>73, 74</sup> In particular, LXRs protect from cholesterol overload by

- Inhibiting intestinal cholesterol absorption
- Stimulating cholesterol efflux from cells to high-density lipoproteins
- Activating the conversion of cholesterol to bile acids in the liver
- Activating biliary cholesterol and bile acid excretion

In addition, LXR agonists activate *de novo* fatty acid synthesis by stimulating the expression of a lipogenic transcription factor leading to the elevation of plasma triglycerides and liver steatosis.<sup>75</sup> Finally, recent studies demonstrate that LXRs are implicated in negative regulation of macrophages inflammatory gene expression,<sup>44</sup> inhibiting inflammation and autoimmune reactions in several *in vivo* models.<sup>44, 76-78</sup>

Despite the many biological effects of LXRs agonists, little is known about changes in endogenous LXR in pathological conditions. Inherited disorders of cholesterol metabolism, like the Smith-Lemli-Opitz syndrome or the Niemann-Pick type C disease, are associated with defective cholesterol synthesis or

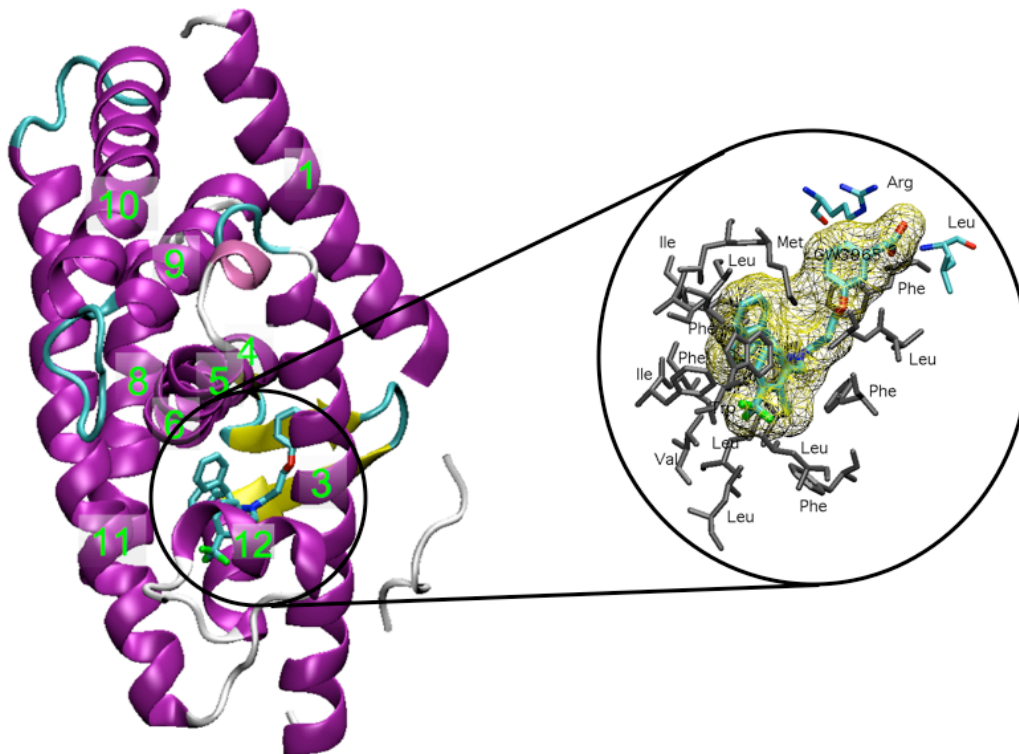
storage,<sup>79, 80</sup> but it remains to be established the precise relationship between such disorders and LXRs.<sup>81</sup>

Given the potent effect of LXR agonists on cholesterol balance and inflammation, these compounds should be very helpful in the prevention and/or treatment of many disorders, including diabetes, inflammatory diseases, atherosclerosis, Alzheimer's disease, and hypogonadism.<sup>81</sup> However, possible side effects must be taken into account, including enhancement of lipogenesis, hypertriglyceridemia, and liver steatosis. So far no studies have addressed the effect of LXRs agonists in humans.

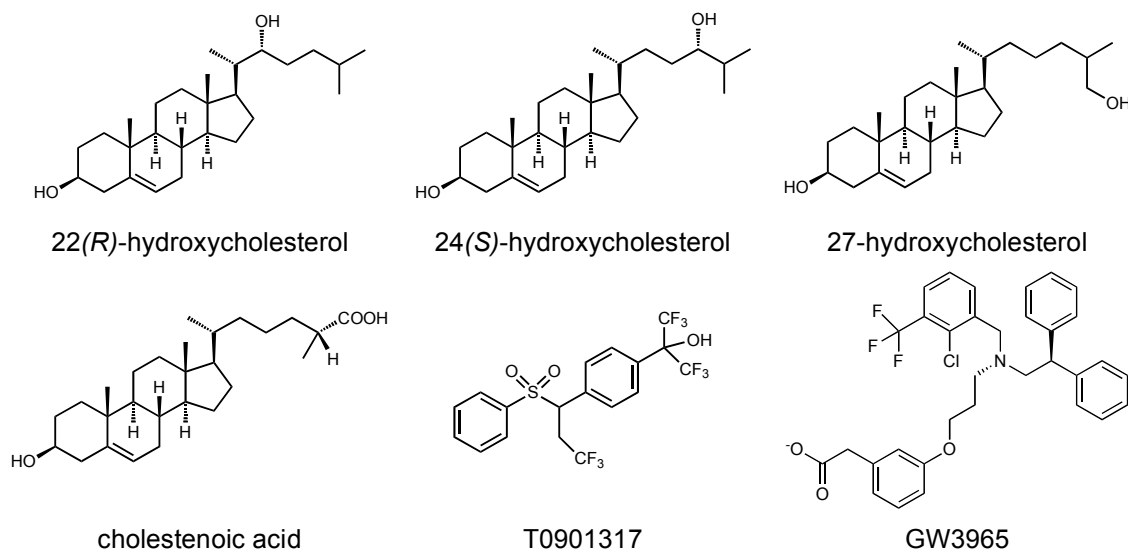
Two isoforms of LXR have been identified in mammals (non mammals have only one isoform) and are referred to as LXR $\alpha$  and LXR $\beta$ . They possess an amino acid sequence similarity of 77% and both bind at DNA at response elements after forming permissive heterodimers with RXR. Both subtypes are expressed in the enterohepatic system, but each has a distinct pattern of expression in other tissues: whereas LXR $\beta$  is ubiquitously expressed, LXR $\alpha$  expression is restricted to tissues rich in lipid metabolism (e.g liver),<sup>82</sup> suggesting different roles in regulating physiological functions.

The overall structure of the LXRs binding domain comprises a core layer of three helices (H5, H6, H9 and H10) enclosed in two additional layers of helices (H1–H4, H7, H8, and H11, respectively) and represents a typical nuclear receptor LBD fold (Figure 4). This arrangement is organized in a wider upper part, which shows the highest degree of sequence conservation between different nuclear receptors, and a lower, narrower, part that is folded to form a hydrophobic cavity into which the ligand can bind. The remaining secondary elements, an antiparallel  $\beta$ -sheet comprising three strands and H12 (that includes the AF-2 motif), reside on either side of the ligand-binding cavity. The volume of the LXR binding pocket is of average size, ranging from 560 (when bound to the compound T0901317) to 680 Å<sup>3</sup> (when bound to GW3965). A substantial fraction of the binding pocket is hydrophobic. The remaining (upper part in Figure 4) is polar.

Oxysterols, the oxygenated derivatives of cholesterol, such as 22(R)-hydroxycholesterol, 24(S)-hydroxycholesterol, 27-hydroxycholesterol, and cholestenic acid, are the natural ligands for LXRs. Some synthetic LXRs agonists have been developed, including non-steroidal T0901317 and GW3965. Structures of natural and synthetic LXRs ligands are shown in Figure 5.



**Figure 4.** Left: crystal structure of the LXR $\beta$  ligand binding domain (PDB 1PQ6),<sup>83</sup> when bound to GW3965. The structure is a helical sandwich (helices numbered based on that of the thyroid hormone receptor). The receptor is represented as ribbons and colored by secondary structure and the ligand GW3965 is represented as licorice. Right: zoom into the binding pocket. GW3965 and amino-acid residues lining the binding pocket are represented as licorice. Amino-acid residues involved in hydrogen bonds are colored by atom type, and the ones involved in hydrophobic interaction are colored in gray.



**Figure 5.** Chemical structure of some natural and synthetic LXR ligand.

### 1.3 Calculation of Protein-Ligand Binding Affinities

Interactions between small molecule ligands and protein receptors are the basis of the mechanism of the great majority of pharmaceutically active compounds. The ability to determine the structures and free energy of binding of protein-ligand complexes is, therefore, a key objective of computational structure-based drug design.

In principle, simulations of the protein and ligand can solve this problem. In practice, obstacles such as the computation time and the limitations of force field accuracy make a direct physical chemistry approach problematic. These considerations have led to the development of approximate methods that, while still based on physical chemistry principles, make use of empirically optimized models (the scoring functions) and determine structures via specially designed conformational search algorithms (the docking algorithms).

Docking and scoring technology is applied at different stages of the drug discovery process for three main purposes:

- Predicting the binding mode of a known active ligand;
- Identifying new ligands using virtual screening;
- Predicting the binding affinities of related compounds from a known active series.

Currently, of these three challenges, the successful prediction of a ligand binding mode in a protein active site is perhaps the most straightforward and is the area where most success has been achieved<sup>84</sup>. There are many published examples of successful virtual screens to identify new hit molecules, but the correlation between the ability of a program to produce a correct binding mode and its success in a virtual screen, remains to be demonstrated.<sup>84</sup> In predicting the binding mode and identifying new ligands, other techniques, such as pharmacophore models, prove to be fast and accurate<sup>85, 86</sup> as well. In principle, the functions used in the docking to calculate the scores predict also the free energies of binding of every molecule being screened. In practice, however, the best that can be obtained is a correct ranking of the molecules, and even this is typically beyond current methods.<sup>84</sup>

Accurate prediction of binding affinities for a diverse set of molecules is so difficult because, at the simplest level, is a problem of small differences between large numbers, inaccurately calculated, to arrive at a small number.<sup>84, 87</sup> The large numbers are on one hand the protein–ligand interaction energy and on the other hand the cost of bringing the two molecules out of solvent and into a complex. The result of this subtraction is the small number that gives the free energy of binding.

In order to predict accurately binding affinities, other techniques have to be applied in alternative, or in addition, to docking techniques. Methods such as

thermodynamic integration (TI) and free energy perturbation (FEP) allow for precise quantification of the binding affinities, but they are computationally intensive and are limited to small structural changes of the investigated molecules.<sup>88</sup> Approximate methods based on the sampling of several conformations, such as Molecular Mechanics Poisson Boltzmann Surface Area (MM-PBSA) or Linear Interaction Energy method (LIE) are faster but still require for each putative binding mode an ensemble of conformations from molecular dynamics or Monte Carlo simulations. Quantitative Structure–activity relationships (QSARs) are the most frequently used approach as they allow for a fast and quantitative determination of the binding affinity based on linear or multiple regression techniques.<sup>89-91</sup>

In this thesis, a mixed-modeling approach, by docking the ligands to the X-ray crystal structure and quantifying their binding affinity using a quasi-atomistic receptor-surface model, was employed.

### 1.3.1 Docking and Scoring

Methods to calculate molecular interactions can be divided in docking, that is the prediction of a ligand conformation and orientation (or pose) within a targeted binding site, and scoring, the quality assessment of docked ligands. As several recent reviews have made clear,<sup>92-94</sup> the technology has been productive for both finding and elaborating bioactive molecules.

There are three principal algorithmic approaches to docking small molecules into macromolecular binding sites.<sup>95</sup> A first class of algorithms aims at simultaneously optimizing the conformation and orientation of the molecule in the binding site. Because of the tremendous complexity of this combined optimization problem, systematic solutions are out of reach, and stochastic algorithms such as genetic algorithms (e.g. *GOLD*,<sup>96, 97</sup> *AutoDock*<sup>98</sup>) or Monte Carlo simulations (*Yeti/Autodock*<sup>99-101</sup>) are usually employed. Docking programs based on such stochastic algorithms, in particular, can give very accurate docking solutions even for very large and flexible ligands,<sup>102, 103</sup> but they require more computation time than the other two methods. A second class of algorithms (i.e. *Glide*,<sup>104</sup> *FRED*,<sup>105, 106</sup> *LigandFit*<sup>107</sup>) separates the conformational search of the small molecule from its placement in the binding site. A conformational analysis is carried out first, and all relevant low-energy conformations are then rigidly placed in the binding site, whereby only the remaining six rotational and translational degrees of freedom of the rigid conformer must be considered. This approach can be referred to as “multiconformer docking”. Finally, a third class of docking algorithms (e.g. *FlexX*,<sup>108</sup> *eHiTS*,<sup>109</sup> *DOCK*<sup>110</sup>) exploits the fact that most molecules contain at least one small, rigid fragment that is able to form specific, directed interactions with a receptor. Such so-called base-fragments are docked rigidly at various favorable positions of the binding site. Docking solutions are

then built starting from these various initial base fragment positions in an incremental construction process, thereby exploring the (torsional) conformational space of the newly added fragments.

Scoring functions have two tasks. First, they serve as an objective function to differentiate between diverse poses of a single ligand in the receptor binding site. Second, after docking a series of compounds, a scoring function is needed to give a rough estimation of binding affinities for different receptor–ligand complexes and to rank order the compounds. Due to the crucial role of scoring, a large number of functions have been developed. They can be classified in three categories. The most widely used class is constituted by the empirical scoring functions (e.g. as in *Glide*<sup>104</sup>).<sup>111</sup> They approximate the free energy of binding as a weighted sum of terms, each term being a function of the ligand and protein coordinates and describing a different type of interaction such as lipophilic contacts and hydrogen bonds between receptor and ligand. The second class of scoring functions is based on molecular mechanics force fields (e.g. *Yeti/Autodock*,<sup>99-101</sup> *MedusaScore*<sup>112</sup>). The binding affinity is estimated by summing up the electrostatic and van der Waals interaction energies between receptor and ligand. Contributions as hydrogen-bond or solvation energies might also be included. Finally, so-called knowledge-based scoring functions<sup>113</sup> (e.g. *eHiTS*,<sup>109</sup> *PMF*<sup>114, 115</sup>) are derived from statistical analyses of experimentally determined protein–ligand X-ray structures. The underlying assumption is that interatomic contacts occurring more frequently than average are energetically favorable. Knowledge-based functions are sums of many atom-pair contact contributions for combinations of protein and ligand atom types.

Furthermore, scoring functions can be divided into *soft* and *hard*. The scoring functions that contain no directional (angular) terms and that have large distance cutoffs can be regarded as soft functions, because their values do not change abruptly with slight changes of ligand orientation and emphasize lipophilic contacts and general steric fit. Soft scoring functions are knowledge-based ones like *PMF*<sup>114, 115</sup> and *DrugScore*,<sup>113</sup> but also the “piecewise linear potential” (PLP)<sup>116</sup> and the Gaussian shape fitting procedure by *OpenEye*.<sup>105</sup> The empirical functions *ChemScore*<sup>117, 118</sup> and the closely related *FlexX* scoring function<sup>119</sup> are “hard”, because they contain angular terms for hydrogen bond interactions and emphasize these directed interactions more strongly. Force fields also belong to the category of hard functions, because they naturally include not only attractive, but also repulsive interactions that lead to steeper potential surfaces.

From this short summary on available methods, it is clear that many options for combinations of docking algorithms and scoring functions are available, and the performance of the chosen method strongly depends on characteristic of the target structure.<sup>120</sup>

The treatment of protein flexibility during docking is nowadays one of the current challenges, being less advanced than that of ligand flexibility: most docking methodologies treat the ligand as flexible, whereas the protein is kept rigid.<sup>84</sup> Within the methodologies that flexibly model at least part of the protein, various approaches have been attempted. Local induced fit can be addressed by

allowing side-chains flexibility (e.g. *Yeti*<sup>99, 100</sup>). A simple implicit approach to account for side-chain flexibility is to use a *soft* interface (by scaling the sterical interactions), which allows partial penetration of the partners.<sup>121, 122</sup> Other methods adjust side chain conformations explicitly during a refinement stage following the rigid-body search.<sup>123, 124</sup> This is typically performed only for a selected set of protein side chains close to the putative binding site and side chain conformations are represented as a discrete set of rotamers.<sup>97, 124</sup> Part of the protein backbone can also be treated by discrete sets of backbone structures compatible with the protein 3D-fold.<sup>125</sup> While several methods have been developed to tackle the problem of side chain and to some degree also local backbone conformational changes, computational efficient treatment of global deformations during docking remains a challenge. One possibility is to approximately account for receptor flexibility by representing the receptor target as an ensemble of structures.<sup>110, 126</sup> During docking the ligand interacts with a mean-field due to the ensemble of receptor structures.<sup>127</sup> In case of a limited number of ligands, it is also possible to combine docking with molecular dynamics (MD) simulations. In the relaxed complex method, an ensemble of protein structures is generated using molecular dynamics simulations prior to docking.<sup>128, 129</sup> Subsequently, docking approaches that assume a rigid receptor are applied to dock putative ligands to the individual conformational snapshots of the simulation.<sup>128, 129</sup> This approach has shown promising results on test cases,<sup>129</sup> however, it can become computationally expensive, since depending on the size of the conformational ensemble, docking to many target receptor structures (possibly several hundred) needs to be performed.

### 1.3.2 Hydrophobic Effects in Protein-Ligand Binding

Solvation effects, especially hydrophobic effect in aqueous solutions, play a key role in the protein-ligand binding process. Due to the complexity of solute-solvent interactions, solvation free energy is considered as one of the most difficult energy term to be calculated.<sup>130</sup>

The hydrophobic effect is often accounted for by an additional solvation energy term that is proportional to molecular surface area, with a positive coefficient.<sup>131</sup> The effect is to add a positive (unfavorable) solvation energy to conformations with more surface area and thus to favor binding, which reduces surface area. Combining the Poisson-Boltzmann<sup>132, 133</sup> (PB) or Generalized Born<sup>134</sup> (GB) electrostatics models with such a surface area (SA) term yields the PBSA<sup>135</sup> and GBSA<sup>136</sup> solvation models, respectively. These are called implicit solvent models because they do not treat any water molecules explicitly. Once parameterized, the PBSA and GBSA models provide rather good agreement with experimental solvation energies of small model compounds,<sup>135, 137</sup> but they may be less accurate for more complex molecules, such as proteins, that can bind or comple-

tely sequester individual water molecules. Unfortunately, it is not straightforward to generate experimental data that directly address this issue, but computational studies are beginning to be applied to it.<sup>138-141</sup>

The influence of solvent on binding can also be treated with molecular dynamics (MD) or Monte Carlo (MC) simulations that include explicit water molecules modeled with an empirical force field.<sup>142-145</sup> Dielectric screening, the solvation of polar groups, and the hydrophobic effect all emerge automatically with this approach. In addition, it should provide a better treatment of bound and sequestered water molecules, at least in principle. However, an explicit treatment of solvent is substantially more costly computationally than an implicit model (by perhaps an order of magnitude, depending on the specifics of the comparison).

Several methods address the calculation of hydrophobic effect contribution in an empirical way. Empirical approaches such as the atomic solvation parameters (ASP<sup>146</sup>) method can provide simple and quick ways to evaluate solvation energy in accuracy comparable to theoretical methods. Alternatively, algorithms can be developed to recognize regions of hydrophobic enclosure in protein active sites (e.g. *Glide*<sup>104</sup>): when groups of lipophilic ligand atoms occupy such sites, the predicted free energy is adjusted to reflect the additional free energy gained beyond the standard scoring function representation of the hydrophobic effect. A similar example is given by the program *LUDI*:<sup>111</sup> small molecules are positioned into clefts of protein structures in such a way that hydrogen bonds can be formed with the enzyme and hydrophobic pockets are filled with hydrophobic groups. The program works in three steps. First it calculates interaction sites, which are discrete positions in space suitable to form hydrogen bonds or to fill a hydrophobic pocket. The second step is the fit of molecular fragments onto the interaction sites. The final step in the present program is the connection of some or all of the fitted fragments to a single molecule, by the use of bridge fragments. *X-Score*<sup>147</sup> (formerly known as *X-CCScore*) is an empirical scoring function used for molecular docking that includes, among terms that account for van der Waals, hydrogen bonding and deformation energies, three different algorithms for the quantification of the hydrophobic effect: one depending on the polar surface area of the ligands, a second one that consist of an hydrophobic matching algorithm and a third term that accounts for hydrophobic contacts between protein and ligand. This last term in particular was the inspiration for the development of the hydrophobic effect function described in the section 4.3. The term in *X-Score* is calculated by summing up the hydrophobic atom pairs formed between the ligand and the protein, according to a distance function:

$$X\text{-Score: Hydrophobic contacts term} = \sum_i^{\text{ligand}} \sum_j^{\text{protein}} f(d_{ij}) \quad (1)$$

if $d_{ij} > d_{ij0} + 2 \text{ \AA}$	then	$f(d_{ij}) = 0$
if $d_{ij} \leq d_{ij0} + 0.5 \text{ \AA}$	then	$f(d_{ij}) = 1$
if $d_{ij0} + 0.5 \text{ \AA} < d_{ij} \leq d_{ij0} + 2 \text{ \AA}$	then	$f(d_{ij}) = (1/1.5) \cdot (d_{ij0} + 2.0 - d_{ij})$



where  $d_{ij}$  is the distance between an atom  $i$  of the ligand and an atom  $j$  of the protein and  $d_{ij0}$  is the sum of their van der Waals radii

This distance function reflects the intuition that the strength of ‘hydrophobic interaction’ will reach the maximum when two hydrophobic atoms form van der Waals contact and diminish gradually with the increase in the inter-atomic distance.

Other approaches make use of empirical indices or scoring functions accounting for the hydrophobic interaction to be used in the search of quantitative structure–activity relationship (QSAR). The logarithm of the octanol–water partition coefficient ( $\log P$ ) is one of the first and most used chemical descriptors in QSAR.<sup>148–150</sup> It describes the intrinsic lipophilicity of a molecule and therefore the ability or inability to cross membranes, but is limited to the ligand and doesn’t consider specific hydrophobic interactions with a protein. More recently, in the approach of Akahane,<sup>151</sup> two indices have been proposed: the first is a hydrophobic field-effect (Hf) index, which indicates the hydrophobic nature of the binding site of a host molecule such as a protein, and the second is a hydrophobic correlation (Hc) index, which indicates the hydrophobic correspondence between the host molecule and its guest molecule such as a ligand. In QSAR development, hydrophobic fields have also been proposed (e.g. *HINT*,<sup>152</sup> *Raptor*,<sup>153</sup> *MLP*<sup>154, 155</sup>). In *HINT*,  $\log P$  is used as the only quantity and applied it successfully for structure-based design<sup>156</sup> and receptor modeling.<sup>157</sup> In *Raptor*<sup>153</sup> physico-chemical fields (hydrophobic and hydrogen bonding), along with a cost for a topological adaptation and for changes in entropy during ligand binding, are contributing to the scoring function that leads the QSAR search. The binding site in *Raptor*<sup>153</sup> is represented by two three-dimensional surfaces, populated with quasi-atomistic properties. The two shells account for different induced-fit mechanisms, observed, for example, with agonists and antagonists or ligands differing substantially in size. In *Raptor*, the induced fit is not only determined by steric aspects, but also by the adaptation of both fields (hydrophobic and hydrogen bonding) on the receptor surrogate. The molecular lipophilicity potential (*MLP*) has been developed as a field expressing in three-dimensions and in a conformation–dependent manner the intermolecular forces encoded in lipophilicity. *MLP* can be introduced as an additional field in three-dimensional QSAR (3D-QSAR) computations, leading to successful predictions of binding constants and biological activities.<sup>158–162</sup>

The physical or empirical calculation of the contribution given by the hydrophobic effect in ligand binding aims at providing a rapid, reasonably accurate way to recognize hydrophobic interactions and quantitatively evaluate how various ligands, characterized by different features, capture the free energy gains available due to their ability to fit hydrophobic cavities in the protein.

### 1.3.3 Quantitative Structure–Activity Relationships (QSARs)

Quantitative structure–activity relationships (QSARs), including quantitative structure–property relationships (QSPRs), are based on the assumption that the structure of a molecule (i.e. its geometric, steric and electronic properties) contains the features responsible for its physical, chemical, and biological properties (e.g. the binding affinity towards a receptor), and rely on the possibility to correlate such properties with a numerical representation of the molecule.

A reliable *in silico* prediction of the binding affinity of a series of compounds towards a protein receptor is extremely desirable by both regulatory bodies and pharmaceutical industry not only because economical benefits would be provided, being *in silico* calculations cheaper than experimental assays, but also because time saving and ecological benefits would be allowed, due to the possibility of a more rational use of resources.

Since the first QSAR models developed by Hammett,<sup>163, 164</sup> Taft,<sup>165</sup> and Hansch,<sup>89, 148, 166, 167</sup> where electronic, steric, and lipophilic parameters were correlated with a biological activity, QSAR techniques have been extensively used and evolved during the years. The Free-Wilson approach<sup>168</sup> was one of the first developments, and addressed for the first time structure–activity studies in a congeneric series by using substituent constants which related biological activity to the presence of a specific functional group at a specific location on the parent molecule.

In 1988 Cramer et al.<sup>169</sup> published the QSAR method known as comparative molecular field analysis (*CoMFA*). For the first time, such structure–activity relationships were based on the three-dimensional structure of the ligand molecules (3D-QSAR). In 3D-QSAR the ligands' interaction with chemical probes is mapped onto a surface or grid surrounding a series of compounds (superimposed in 3D space). This surface or grid represents a surrogate of the binding site of the true biological receptor. The quality of the QSAR model here depends critically on the correct superposition of the ligands, the identification of which is almost impossible in the absence of structural information for the target protein.<sup>170</sup> A possible solution to this problem could be provided by 3D approaches that do not depend on a mutual alignment of the molecules, like the grid-independent descriptors (*GRIND*<sup>171</sup>). These are autocorrelation transforms that are independent of the relative orientation of the molecules in 3D space. Another possibility is provided by 4D methods that are able to deal simultaneously with different conformations, orientations, and protonation states.<sup>172-175</sup> Some of these approaches provide also the possibility of simulating induced fit in an explicit manner, by means of a topological adaptation of the model of the binding-site surface to the individual ligand molecules.<sup>153, 176</sup> Different induced-fit protocols and solvation models constitute the fifth and sixth QSAR dimension (5D-QSAR,<sup>176</sup> 6D-QSAR<sup>177</sup>), respectively. More details concerning such methodologies are given in Chapter 3.2.9 and 3.2.10.

Different approaches describe the molecules not through field calculation, but through parameters, referred to as molecular descriptors. A molecular descriptor can be defined as the result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number (theoretical descriptor) or the result of some standardized experiment<sup>178</sup> (experimental descriptor). Different statistical methods are used to build QSAR models that correlate a set of (theoretical or experimental) descriptors with the studied activity (i.e. multiple regression analysis,<sup>179, 180</sup> genetic algorithms,<sup>181, 182</sup> artificial neural networks<sup>183, 184</sup>). While at an early stage of QSAR history only few experimental descriptors (e.g. logP, ionization constant, molar refractivity) were used as QSAR parameters, later hundreds of quantum chemical, steric, topological, connectivity, and other theoretical descriptors were generated.<sup>178</sup>

### 1.3.4 Validation of QSAR Models

With the increasing number of descriptors, but also with the proliferation of QSAR models and technologies, many authors addressed the importance of a critical assessment of model predictivity by means of extensive validation.<sup>90, 185-189</sup> There's general agreement that a good QSAR model should be characterized by a good fit and a good predictivity. Whereas fit can easily be checked by  $r^2$  (coefficient of determination) or by the F-test of the regression, corresponding measures and criteria for predictivity are not so well defined. By convention, *leave-one-out* cross-validation or cross-validation in groups (*leave-many-out*) is accepted as the method of choice to test predictivity. However, many statisticians agree in recognizing that the *leave-one-out* procedure is asymptotically inconsistent,<sup>190</sup> and too conservative in the sense that it tends to select large models unnecessarily.<sup>191</sup> Moreover, internal predictivity as assessed by the *leave-one-out* method, has been demonstrated to be not correlated with test set predictivity.<sup>187, 192, 193</sup> A high value of internal predictivity is a necessary condition for high predictive power, but it is not a sufficient condition. In other words, although a low value of  $q^2_{LOO}$  ( $r^2$  measured on the training set with the *leave-one-out* method) may well indicate low test set predictivity in a model, high  $q^2_{LOO}$  does not necessarily imply high predictivity.<sup>187</sup> This effect has been called, from the author that first raised the problem,<sup>192</sup> "Kubinyi paradox".<sup>194, 195</sup> Another risk of very fit models is overfitting, that occurs when the training of the model has been performed too long or when the size of the training set is too small. In such cases the model may adjust to very specific random features of the training data, that have no causal relation to the modeled property: the model in this case is trained to reproduce not only the relevant information included in the training set but also the noise, and fails to predict test set compounds.<sup>196</sup> Moreover, for models characterized by a high number of variables, the risk of chance correlation should

be considered. Already in 1972, Topliss pointed out that a large number of variables increase the risk of chance correlation.<sup>185, 186</sup>

To address these problems, a number of criteria have been proposed.<sup>187-189, 191, 197-201</sup> At a workshop held in Setubal, Portugal in 2002, a set of principles was proposed to define the validity and applicability domain of QSAR models. These then evolved into the Organization for Economic Co-operation and Development (OECD) principles in 2004.<sup>202</sup> Reliable and predictive QSAR models should be statistically significant and robust, should be validated by predictions for external data sets that were not used in the model development, and should have their application boundaries defined.<sup>188</sup> In particular, many authors recommend the use of scramble tests, *leave-many-out* cross validation as well as predictions for an external set of data.<sup>90, 187, 188, 191</sup> Furthermore, Todeschini et al.<sup>191</sup> asserted that, in order to transform statistical models into real scientific knowledge, a model should be built on a sufficient large information about the modeled response, should contain only relevant variables, should have an acceptable predictive ability, and should avoid some pathological characteristics, for example models with a marked difference between fitting and prediction power should be avoided.<sup>191</sup> Consensus between different methodologies is an approach often used to ensure robustness of predictions.<sup>203-205</sup>

QSAR models that satisfy qualitative and predictive requirements are a useful tool for pharmaceutical research, in the field of drug design and virtual screening, but probably even more for regulatory bodies, where an increasing number of toxicity tests is required, and alternatives to reduce or replace animal testing are highly desired.

## 2 Aim of the Thesis

The presence of compounds with hormonal activity in the biosphere has become a worldwide environmental concern, and it has been addressed by various regulations, both in the U.S.<sup>7</sup> and Europe.<sup>1, 34, 35</sup> Besides natural and synthetic hormones, a broad variety of chemicals, referred to as endocrine disruptors, are able to disrupt the hormonal system, and cause adverse health effects.<sup>7</sup> There are different routes how chemicals can trigger adverse effects; the most straightforward is via nuclear-receptors binding. A reliable *in silico* identification of the endocrine-disrupting (or, more general, the toxic) potential of drugs and chemicals is therefore regarded as highly desirable by both regulatory bodies and the pharmaceutical, chemical, and food industry.

This thesis focuses on two members of the nuclear receptor superfamily: the glucocorticoid and liver X receptors, which through their role in metabolism and inflammation, cooperate in maintaining homeostasis.<sup>37, 41, 43-45</sup> Molecular modeling techniques, applied to study the interactions between the receptors and their ligands, and to build predictive models aiming at the quantification of the binding affinity of small molecules towards the receptors, are presented and discussed.

First, the identification of the binding mode by flexible docking is described. In particular, the results obtained from manual and automated protocols is compared. The generation of multi-dimensional QSAR models (software *Quasar*), able to quantify the binding affinity of small molecules to the glucocorticoid and liver X receptors, are then reported. Special attention is paid to the model validation, by evaluating the predictive ability, when applied to external sets of compounds. Robustness is verified by the consensus with a second QSAR methodology (*Raptor*), sensitivity with a series of scramble tests. Predictive QSAR models provide a tool for an *in silico* estimation of the binding affinity for new ligands or environmental compounds.

Finally, the implementation of empirical scoring function in a small C program (*HEidi*: Hydrophobic Effect in Drug Interactions), aiming at the quantification of the hydrophobic effect for scoring protein–ligand binding energies, is described. The use of *HEidi*, together with electrostatic, van der Waals, and hydrogen bond energies, in the ranking of docking poses is evaluated.

### 3 Material and Methods

#### 3.1 Material

In Table 2 is listed the hardware that was used during this thesis, together with the corresponding software.

**Table 2:** Overview of the computer, operating systems and software used.

<b>Computer Hardware</b>	<b>Processors</b>	<b>RAM</b>	<b>Operating System</b>	<b>Software</b>
Macintosh Power Mac G5	PowerPC G5 2×1.8GHz	1GB	OSX	<i>Yeti</i> <i>AMBER</i> <i>VMD</i>
Macintosh Mac Pro	Intel® Xeon™ 8×3.0GHz	8GB	OSX	<i>Biograf<sup>x</sup></i> <i>Quasar</i> <i>Raptor</i>
Dell Precision™ Workstation 650	Intel® Xeon™ 2×2.8GHz	1GB	Kubuntu	<i>MacroModel</i> <i>VirtualToxLab™</i>
Dell Precision™ Workstation 530	Intel® Xeon™ 2×2.4GHz	1GB	Kubuntu	<i>MacroModel</i> <i>VirtualToxLab™</i>
Cronos Linux Cluster (for more details see Pic XX)	Intel® Xeon™ 76×2.4GHz	2GB/node	Red Hat	<i>AMBER</i>
Silicon Graphics Workstation Octane	2×250 MHz MIPS R10000	1.2GB	IRIX	<i>MacroModel</i> <i>PrGen</i> <i>AMSOL</i> <i>AMBER</i>
Silicon Graphics Workstation O <sub>2</sub>	250 MHz MIPS R10000	320MB	IRIX	<i>MacroModel</i> <i>PrGen</i> <i>AMSOL</i>

## 3.2 Software

The main software used throughout this thesis will be described in short. The procedures will be described in the paragraph 3.3.

### 3.2.1 *AMBER*

*AMBER 7.0*<sup>206</sup> is a package of molecular simulation programs for biomolecules (including energy minimization, molecular dynamics, NMR refinement, free energy calculations, energies and trajectories analysis, etc.). In this work, *AMBER* was exclusively used to perform MD simulations in explicit solvent. The modules used in this thesis are the following: *Antechamber* for the ligand preparation, *Leap* for the ligand and target preparation, *Sander* for the minimization and the molecular dynamic simulations, and *Ptraaj* for the analysis of trajectories, energies and temperatures.

### 3.2.2 *AMSOL*

*AMSOL 5.4.1*<sup>207</sup> is a semiempirical quantum-chemistry program that allows to compute partial CM-1 atom charges and solvation energies.<sup>208</sup>

### 3.2.3 *Biograf*<sup>X</sup>

*Biograf*<sup>X</sup> 1.1<sup>209</sup> is a mQSAR tool running on Macintosh and Linux platform combining the multi-dimensional QSAR tools *Quasar*<sup>177</sup> and *Raptor*,<sup>153</sup> together with the pharmacophore alignment module *Symposar* and tools for preparing the ligands to a QSAR study.

The *Biograf*<sup>X</sup> concept is illustrated in Figure 6.

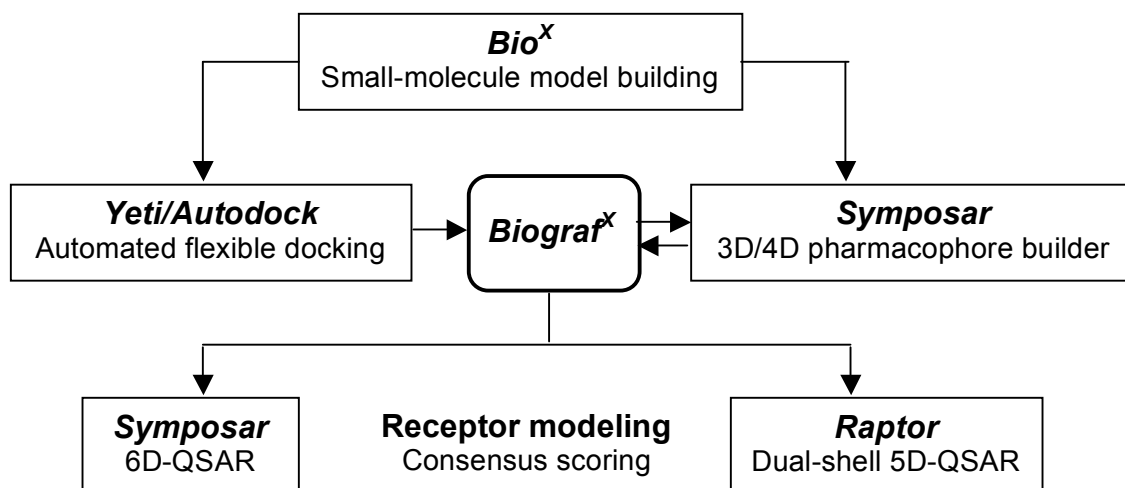


Figure 6. The *Biograft<sup>X</sup>* concept

### 3.2.4 *Bio<sup>X</sup>*

*Bio<sup>X</sup> 3.1.8*<sup>209</sup> is a general molecular modeling program able to build molecules from scratch and optimize with a molecular-mechanics minimizer, using the *Yeti* force field<sup>99</sup> featuring directional terms for hydrogen bonds. In this thesis *Bio<sup>X</sup>* was used to inspect structures, analyze ligand alignments and to build molecules submitted to *VirtualToxLab<sup>TM</sup>*.

### 3.2.5 *Epik*

*Epik*<sup>210</sup> is a module which is part of Schrödinger Inc. software suite for pK<sub>a</sub> prediction. It was used to predict the protonation state of the ligands studied in this thesis at physiological pH.



### 3.2.6 *Glide*

*Glide*<sup>104</sup> is a ligand docking program within the Schrödinger suite for predicting protein-ligand binding modes and ranking ligands via high-throughput virtual screening. *Glide* enables to perform flexible ligand docking into a rigid protein structure.

### 3.2.7 *MacroModel*

*MacroModel 8.0*<sup>211</sup> is a molecular-modeling software package from Schrödinger that was used for ligand generation, optimization in GB/SA solvation<sup>212</sup> model and for conformational searches. For the preparation of the ligands binding to the GR, *Macromodel 6.5* was used, as version 8.0 was not available at the time.

### 3.2.8 *PrGen*

*PrGen 2.1*<sup>213</sup> (Pseudoreceptor Generator) was used to prepare the GR and LXR ligands for charges and solvation calculation in preparation to the docking studies.

### 3.2.9 *Quasar*

The software was used to generate receptor models for the GR and LXR. *Quasar 5.2*<sup>176, 177, 214</sup> is a receptor-modeling concepts that allows for multidimensional QSAR. In *Quasar*, the binding site of the protein is represented by a surrogate, which consists of a three-dimensional surface, surrounding the ligands superimposed in their bioactive conformation (as obtained, for example, from docking studies at the true biological receptor) at van der Waals distance. The topology of this surface mimics the shape of the binding site. This surface is then populated with quasi-atomistic properties, corresponding to those of the amino-acid residues: positively and negatively charged salt bridges; hydrogen-bond donors and acceptors; neutral, positively and negatively charged hydrophobic properties; hydrogen-bond flip-flop, as well as solvent. Apart from accepting 4D compound input (conformations, poses, protonation states, tautomers),<sup>174</sup> *Quasar* allows specifically for the simulation of induced fit (corresponding to side-

chain flexibility and moderate backbone motion at the true biological receptor) whereby six different protocols are evaluated simultaneously (5D-QSAR).<sup>176</sup> The model family, typically consisting of 200–500 models, is evolved using a genetic algorithm and provides an averaged prediction for each compound along with the variation over the model family. *Quasar* employs the following scoring function derived from the directional *Yeti*<sup>99</sup> force field:

$$E_{\text{binding}} = E_{\text{ligand-receptor}} - E_{\text{ligand desolvation}} - E_{\text{ligand internal strain}} - T\Delta S - E_{\text{induced fit}} \quad (2)$$

where  $E_{\text{ligand-receptor}} = E_{\text{electrostatic}} + E_{\text{van der Waals}} + E_{\text{hydrogen bonding}} + E_{\text{polarization}}$

$$\Delta G_{\text{exp}} = |a| \cdot E_{\text{binding}} + b \quad (3)$$

Using the ligands of the training set, a linear regression of the experimental ( $\Delta G_{\text{exp}}$ ) and calculated ( $E_{\text{binding}}$ ) binding affinity is then obtained (eqn. 3). The coefficients  $a$  and  $b$  are derived from the correlation of the training set in cross-validation mode and, later on, applied to molecules of the test set or new compounds for which binding affinity should be predicted. In *Quasar*, the solvent contribution can be calculated explicitly, allowing the presence of solvent properties on the surrogate surface, or implicitly, where the solvation terms (ligand desolvation and solvent stripping) are independently scaled for each model within the surrogate family. Each different scaling reflects a different solvent accessibility of the binding site. Solvation terms are associated with weights that evolve throughout the simulation (6D-QSAR).<sup>177</sup>

### 3.2.10 *Raptor*

*Raptor 3.2.10*<sup>153</sup> is a receptor-modeling concepts that allow for multi-dimensional QSAR. It was used to generate models in consensus mode for the prediction of the free binding energy of ligands. In contrast to *Quasar*, the binding site in *Raptor* is represented by two three-dimensional surfaces, populated with quasi-atomistic properties. The two shells allow to account for different induced-fit mechanisms, observed, for example, with agonists and antagonists or ligands differing substantially in size. Induced fit may, of course, exert different shapes and magnitudes for the two shells. The model development employs a multi-step optimization protocol<sup>153</sup> including domain assignment, *tabu* search,<sup>153</sup> and local search. In *Raptor*, the scoring function includes directional terms for hydrogen bonding and hydrophobicity as well as terms for the cost of the topological adaptation and the changes in entropy upon ligand binding:

$$E_{\text{binding}} = E_{\text{ligand-receptor}} - T\Delta S - E_{\text{induced fit}} \quad (4)$$

where  $E_{\text{ligand-receptor}} = w (E_{\text{hydrogen bonding (shell-1)}} + E_{\text{hydrophobic (shell-1)}}) + (1.0-w) \cdot (E_{\text{hydrogen bonding (shell-2)}} + E_{\text{hydrophobic (shell-2)}})$

with  $w$  being the interpolation weight between the two shells.

### 3.2.11 Swiss-PdbViewer Deep View

*Swiss-PdbViewer 4.0*<sup>215</sup> (aka *DeepView*) is an application that provides a graphical interface allowing to analyze several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino-acid mutations, H-bonds, angles and distances between atoms can be obtained thanks to the graphic and menu interface.

*Swiss-PdbViewer* was used to align protein and compare different crystal structures of the same protein or different isoforms of the same receptor.

### 3.2.12 VMD

*VMD 1.8.6*<sup>216</sup> (Visual Molecular Dynamics) 1.8.6 is a molecular visualization program for displaying, animating, and analyzing large biomolecular systems.

In this thesis it was used to display trajectories from the molecular dynamics and to render static pictures of ligand binding modes or to produce movies from molecular dynamics simulations.

### 3.2.13 VirtualToxLab™

The *VirtualToxLab*<sup>TM3, 4</sup> is an *in silico* tool for predicting the toxic (endocrine-disrupting) potential of drugs, chemicals and natural products. Its fully automated protocol — accessible through the Internet — calculates the binding affinity of any molecule of interest towards a series of 12 proteins, known or suspected to trigger adverse effects and estimates the underlying toxic potential. In contrast to other approaches in the field, the technology does not only provide a binding affinity but allows to verify a prediction at the molecular level by interactively

inspecting the binding mode of the tested compound with all target proteins in 3D.

The toxic potential of existing and hypothetical compounds (drugs and environmental chemicals) is estimated by simulating and quantifying their interactions towards a series of macromolecular targets at the molecular level using automated flexible docking combined with multi-dimensional QSAR (mQSAR). Currently, those targets comprise 12 proteins: the androgen, aryl hydrocarbon, estrogen  $\alpha/\beta$ , glucocorticoid, liver X, mineralocorticoid, thyroid  $\alpha/\beta$  and the peroxisome proliferator-activated receptor  $\gamma$  as well as the enzymes cytochrome P450 3A4 and P450 2A13.

In the present work the *VirtualToxLab*<sup>TM</sup> was used to check the predictions by the automatic procedure for the compounds used in this thesis as well as to predict the toxic potential of new compounds to the GR and LXR, respectively.

### 3.2.14 Xcode Tools

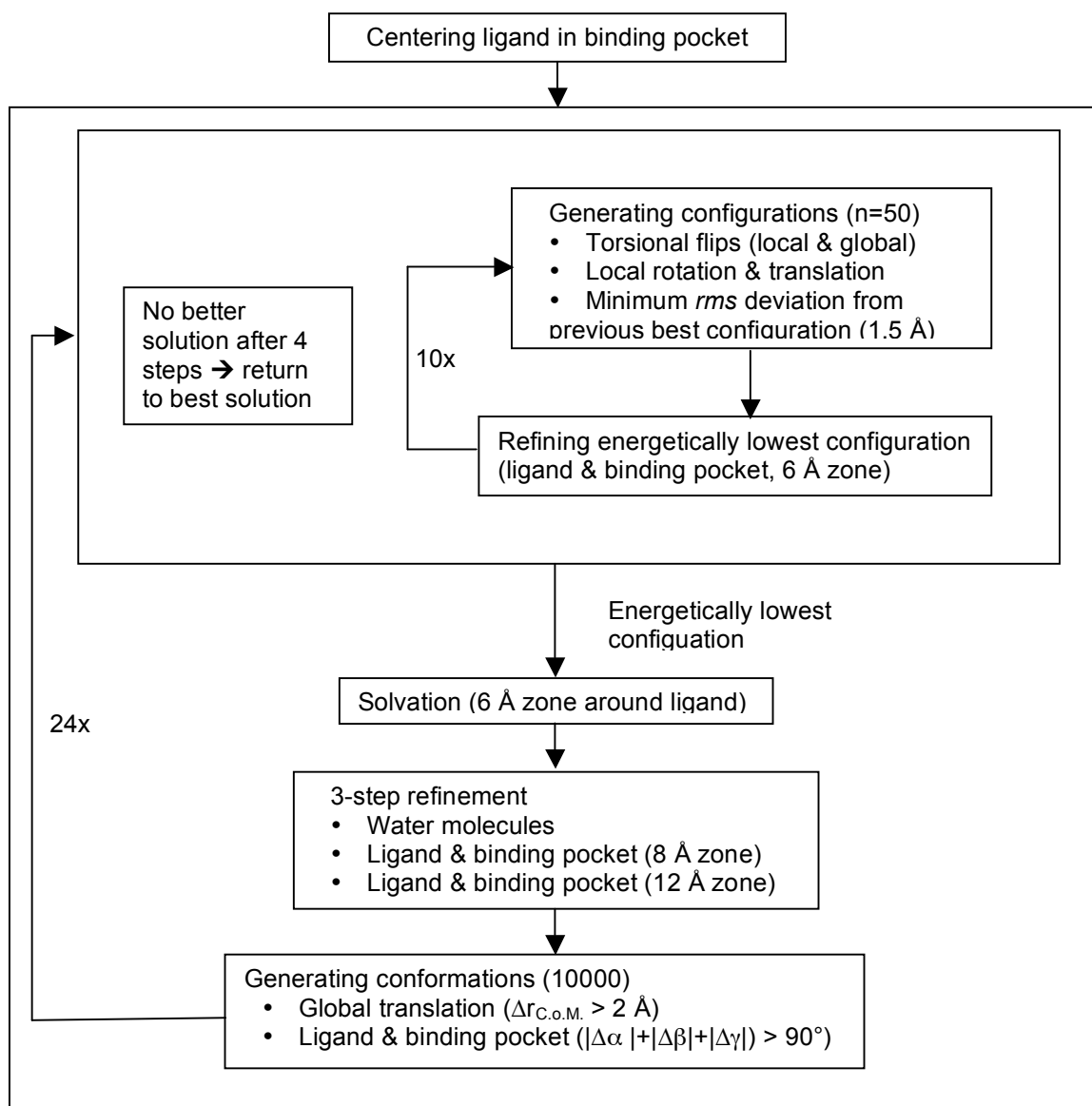
*Xcode* is Apple's development environment for MacOSX. In this thesis the GCC compiler within *Xcode* tools was used to compile the code for evaluation of hydrophobic effect in the binding site of the proteins.

### 3.2.15 Yeti

*Yeti* 7.12<sup>99-101</sup> is a molecular-mechanics program developed by Vedani and Dobler to optimize proteins complexed with small molecules, by means of flexible docking. It features also a module for automated docking, i.e. for the sampling of structurally and energetically feasible arrangements within the binding pocket of a protein using flexible docking (co-refinement of the protein) combined with a Monte-Carlo search protocol and dynamic solvation of the binding pocket. Figure 7 shows the automated docking protocol. After centering a compound in the binding pocket, 50 new ligand configurations are generated with local translation, rotation and combined local/global torsion flips. The lowest-energy ligand-protein complex is refined and serves as a template for the next Monte-Carlo step. The best previous configuration is restored if no lower-energy configuration is found after four Monte-Carlo trials. This permits a global search without spending too much time in computation for local sampling of ligand configurations in unfavorable regions of the binding pocket.

In this work, the minimization protocol was used to identify, both manually and automatically, the various conformations of the ligand molecules at the binding

site and to obtain thereby the most feasible docking mode. The energetically most feasible pose was then usually used for receptor-modeling studies.



**Figure 7.** Flexible docking protocol with Monte Carlo search

### 3.3 Methods

After data preprocessing, and after the preparation of ligands and proteins, a flexible docking was performed, both automatically and manually, in order to find minimum energy binding modes. Based on the alignment obtained from the docking, a QSAR model was subsequently built and validated. The models were finally implemented in *VirtualToxLab*<sup>TM</sup> and used to test new class of compounds for toxic potential.

#### 3.3.1 Experimental Binding Affinities for GR and LXR Ligands

The experimental affinity data for the 110 GR-binding compounds and for the 54 LXR ligands were obtained from multiple sources.<sup>217-222</sup> Some of the affinity data were available as  $IC_{50}$  values, the others as  $K_i$  values. For their use in our study, we converted the  $IC_{50}$  into  $K_i$  values using the Cheng-Prusoff relation:<sup>223</sup>

$$K_i = IC_{50}/(1 + L/K_D) \quad (5)$$

where  $K_i$  is the inhibition constant,  $IC_{50}$  the 50% of the inhibitory concentration,  $L$  the concentration of the radioligand and  $K_D$  its dissociation constant.

The  $K_i$  values of the compounds, together with their structures, are given in the appendix A and B.

#### 3.3.2 Ligand Structure Generation and Conformational Analysis

The three-dimensional structures of all the ligands studied, was generated using *MacroModel* 6.5 for the GR compounds and *MacroModel* 8.0 for the LXR compounds, and optimized in aqueous solution (GB/SA continuum water model) with the *AMBER*<sup>\*224</sup> force field. Atomic partial charges (CM1) were generated using the *AMSOL* package.

In order to identify the lowest-energy conformation for all compounds, a conformational search with *MacroModel* was performed (10,000 Monte-Carlo iterations) for all the GR ligands except for the steroids, where only the compound with the highest experimental affinity was investigated, assuming that all the steroids share a common puckering of the B and C rings.

The ligands of the LXR instead are large molecules (450+) characterized by the presence of extended branches with aromatic rings that after conformational search often pack together to give the molecule a folded conformation. To overcome this artifact, for the molecules characterized by a global minimum with a folded conformation, the minimum energy conformation among the open conformations was chosen instead of the overall global minimum.

### 3.3.3 Protein Preparation

The three-dimensional structure of the GR $\alpha$  with bound dexamethasone (PDB code: 1M2Z,<sup>68</sup> 2.5 Å resolution,  $R_{\text{free}} = 0.267$ ) was obtained from the Protein Data Bank.<sup>225</sup> For our study, we chose chain A, including 255 amino-acid residues, dexamethasone, and 118 water molecules. First, all hydrogen atoms were added to the structure with the software *Yeti*<sup>99, 100</sup> and the most probable protonation state for the histidine residues was automatically assigned: His645 and His654 were protonated at their N $\delta$  atom while His588, His726 and His775 were protonated at their N $\epsilon$  atom, respectively. Next, the hydrogen-bond network was optimized on the basis of neighboring partners (donors or acceptors), by orienting polar hydrogens in order to get optimal distance, linearity, and directionality of the resulting H-bonds. Finally, the structure was relaxed using the directional *Yeti* force field.<sup>99</sup>

The three-dimensional structures of the human LXR $\alpha$  (PDB code 1UHL,<sup>226</sup> 2.90 Å resolution,  $R_{\text{free}} = 0.326$ , chain B) and  $\beta$  (PDB code 1PQ6,<sup>83</sup> 2.40 Å resolution,  $R_{\text{free}} = 0.262$ , chain A) were obtained from the Protein Data Bank. The two structures were aligned and compared with Swiss-PdbViewer, and due to high similarity of protein structure and to the high correlation of the experimental binding affinities for the two LXR isoforms  $\alpha$  and  $\beta$ , only one, the  $\beta$  isoform, was considered for the subsequent studies. The details of the comparison will be discussed in chapter 4. The amino-acid sequence was discontinuous at two points: amino-acid residues from Phe243 to Gln246 and Leu254 to Pro258 were not resolved in the crystal structure. The structure of such amino-acid residues was adopted from other crystal structures (Phe243 to Gln246 were taken from PDB 1PQ9,<sup>83</sup> and Leu254 to Pro258 from 1UPV<sup>227</sup>), after being aligned with the template LXR $\beta$ . The complete structure was then minimized with *BioX* and with *MacroModel* (*AMBER\**, 5000 iterations, TNCG, implicit water solvent).

### 3.3.4 The Docking Protocol

Before running any minimization, the force-field parameter database had to be augmented to include all specific parameters for the studied ligands – most particularly, 1–4 (torsional) interactions. Next, the ligands were manually docked into the binding site of GR or LXR and minimized with the software *Yeti*.<sup>100</sup> The minimization of the complex was performed over two steps: first, only the protein around the ligand (within a distance of 8 Å) was minimized, then the ligand and the protein within the same zone were optimized.

In addition, an automated docking procedure was performed with *Yeti*.<sup>100</sup> The compounds were grouped by chemical classes and for each class the most active compound was manually docked into the GR structure. Such complex was used as a template for the automated docking of the remaining compounds. For each ligand 25 poses were saved and the 4 minimum energy conformations used for the QSAR.

Automated docking for the LXR ligands was also carried out with *Glide*.<sup>104</sup> Results with both automated procedures (*Yeti*,<sup>99, 100</sup> *Glide*<sup>104</sup>) were less encouraging than the manual ones and for this reason the QSAR studies here presented refer to the alignment obtained with the manual docking.

In order to improve the pose selection in the automated docking procedure, a particular consideration of the hydrophobic effect in the active site of proteins was performed through the development of a new empiric scoring function. A literature study was carried out to get an overview of the approaches in the field and to find a suitable scoring function that was used as a starting point for a subsequent development and modification, leading to a new empirical score based on hydrophobic matching between protein and ligand. A small program in C was implemented to calculate such scoring function and applied to the docking of compounds to the GR and LXR. More detailed information on the development and results is given in chapter 4.

### 3.3.5 Molecular-Dynamics Simulations

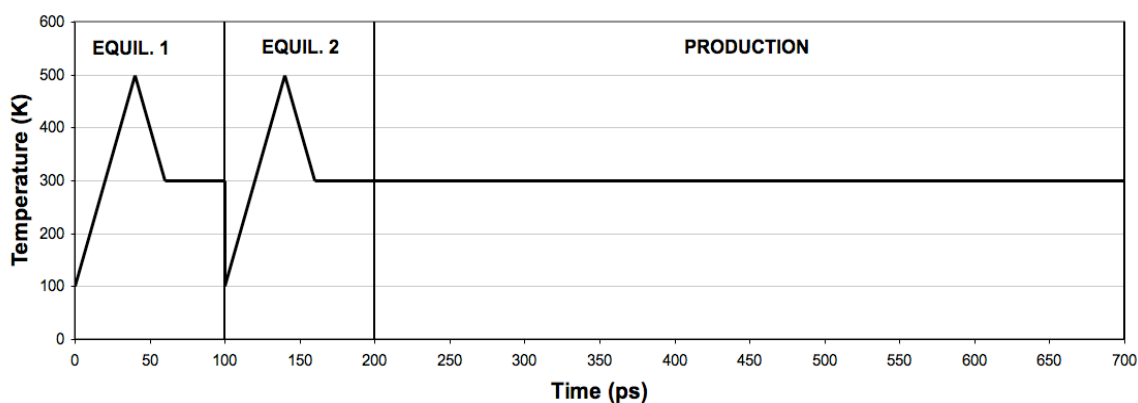
The receptors prepared as indicated above were converted to the proper format with the specific *AMBER*-conventions for atom types. The co-crystallized ligands were processed with the *Antechamber* module in *AMBER*, where atomic partial charges (AM1-BCC) were calculated with *MOPAC*.<sup>228</sup> The structures were then read into *Leap*, where it was checked for consistency and the libraries of parameters were created (using force fields *AMBER ff99* for the proteins and *GAFF* for the ligands). The complexes were solvated with a water cap of 25 Å around the respective ligands. The solvent shell was constituted of explicit TIP3P water molecules. During the MD, the bonds to hydrogens were constrained with



the *SHAKE* algorithm. The MD protocol applied (*Sander* module of *AMBER*) was the following:

- A first minimization (steepest-descent method) was carried out, using a non-bonded cutoff of 12 Å.
- Two equilibration steps were then performed. First, only water molecules were equilibrated and the protein was kept fixed. After heating from 100K to 500K for 40 ps, and cooling from 500K to 300K for 20 ps, temperature was kept constant at 300K for 40 ps.
- The second equilibration step involves the inner zone (20 Å around the ligand, where the residues were identified with the module *Carnal*). The timings and temperature management were the same as the previous step: after heating from 100K to 500K for 40 ps, and cooling from 500K to 300K for 20 ps, temperature was kept constant at 300K for 40 ps.
- The production phase of different length for the two receptors was carried out at a constant temperature of 300 K and involves the inner zone.
- A second minimization on the final structure was performed.

In Figure 8 is shown a diagram of the molecular dynamics protocol.



**Figure 8.** Diagram of the molecular dynamics protocol. The production phase is 500 ps for GR and 250 for LXR.

In order to perform meaningful and well equilibrated molecular dynamics simulation, the settings were slightly different for the two receptors: for the GR the initial and final minimization were characterized by 300 cycles, while for LXR it was necessary a longer minimization of 5000 cycles. The time-step chosen was 2 fs for the GR and 1 fs for the LXR. The production phase was performed for 500 ps for the GR and, being the time step shorter, only for 250 ps for the LXR, due to simulation length. This difference can be explained with the fact that the LXR structure was subjected to a modification, by the introduction of 9 amino-acid residues not solved in the crystal structure, that may have required a longer

minimization and a shorter time-step. In Table 3 are reported the parameters chosen for each receptor.

	Minim. steps	Time step (ps)	Equil. 1 (ps)	Equil. 2 (ps)	Production (ps)
GR	300	0.002	100	100	500
LXR	5000	0.001	100	100	250

**Table 3.** Parameter settings during the molecular dynamics.

Analysis of temperature and energy was performed with the module *Ptraj*.

The MD simulation for the GR was performed to get insight into the dynamic behaviour of the complex, but the structure used for the further processing was the one before MD, prepared as described in paragraph 3.3.3.

For the LXR, the molecular dynamics simulation was performed not only to analyze the binding mode and the dynamic behavior of the crystallized complex, but also to further minimize the energy of the protein structure after the construction of the missing amino-acid residues. For this reason, a post-processing of the protein structure was necessary. In order to have a suitable structure for the docking, the minimum energy conformation of the protein-ligand complex during the production phase was extracted from the trajectories, imported in *MacroModel* and minimized. Three other random conformations during the MD were also extracted and superimposed, together with the average structure, to compare the structural differences during the simulation. The result of the comparison is discussed in the chapter 4. Only the minimum energy structure was then further processed for the docking studies. The structure was imported in *Yeti*<sup>100</sup>, after the deletion of surface water molecules from the protein and minimized (full refinement).

### 3.3.6 QSAR Studies: Quasar

The data set was split into a training set, used to build the model, and test set, used to evaluate it, in such a way that a maximal diversity of the training set with respect to binding affinity and chemical properties was obtained. In order to achieve this goal, the compounds were grouped according to their chemical class (i.e. sharing the same scaffold) and were ranked by affinity. For each group, the most and the least active compound was assigned to the training set. From the compounds remaining in the pool, compounds with different scaffolds and functionalities were selected to be part of the training set in order to achieve maximal chemical diversity. For the QSAR simulations, we used 4:1 ratio yielding 88 training and 22 test compounds for the GR, and 40 training and 12 test compounds for the LXR.

We used the so-called mixed-modeling approach,<sup>153, 177</sup> i.e. docking the ligands to the X-ray crystal structure and quantifying their binding affinity using a quasi-atomistic binding-site model thereof. This approach yields more reliable binding energies than directly scoring the ligand-protein interactions at the experimental structure. This is due to the fact that the estimation of the binding energy on a whole protein-ligand complex is associated with errors that are much larger than the difference in binding of two small ligands, for example.

The ligand alignment obtained from the manual docking was used as input to build the QSAR model. Attempts were also made with the alignment from the automated docking procedure. In order to develop the model, together with the ligand alignment were added by the software *Biograf*<sup>X</sup> also the binding affinity values, the *AMSOL* solvation of the neutral species, the entropy values and the internal strain calculated with *MacroModel*, necessary to have a quantitative prediction of the binding affinity.

After creating the three-dimensional surface, the ligands individual envelopes were generated using all of the six different induced-fit scenarios (module *Envelope*). In order to develop a consistent QSAR-model, which could predict the binding affinity towards GR and LXR in good agreement with the experimental data, different parameters settings were explored and the model was built with *Quasar*:

- Population size (100 or 200 individuals)
- Number of crossovers (from 2000 to 100000)
- Attenuation factors (weight between 0 and 1 of solvation energies, entropies, internal strain)
- Induced-fit weight (0.25, 0.5, 1.0, 2.0)
- Number of cross-validation groups (5, 6, 7, 8, 9, 10)

### 3.3.7 Model Validation

Particularly in multidimensional QSAR (mQSAR) it is of outmost importance to challenge a model, e.g. by means of a series of scramble tests, of consensus scoring as well as an external validation set. In this work, all these tests were employed. Scrambling the training set data<sup>229</sup> (also called y-scrambling) is a widely used technique to ensure sensitivity of the model. It consists of a random shuffle of the binding data of the training set, with respect to the true biological values. If, under these circumstances, the ligands of the test set are still predicted correctly, the model is worthless, as it is not sensitive towards the biological data. To assess the robustness of the model, a second methodology (software *Raptor*<sup>153</sup>) was applied, in order to compare predictions in a consensus-scoring approach. To truly challenge the model, a new independent

set of compounds was identified from the literature, and employed only for this validation step.

### 3.3.8 Estimation of Toxic Potential through VirtualToxLab™

The models for GR and LXR were implemented in the *VirtualToxLab*<sup>TM3, 4, 230</sup>, that presently includes 12 models (androgen, aryl hydrocarbon, estrogen  $\alpha/\beta$ , glucocorticoid, liver X, mineralocorticoid, peroxisome proliferator-activated  $\gamma$  and thyroid  $\alpha/\beta$  receptors as well as for the enzymes cytochrome P450 2A13 and 3A4). The binding affinity of 1500+ compounds towards all 12 models was calculated. The results are accessible through the Internet.<sup>209</sup> To assess the validity of the approach, a series of 24 psychotropic drugs ligands were tested in the automatic procedure, and the results are discussed in the chapter 4.1.6. The protocol implemented in *VirtualToxLab*<sup>TM</sup> includes a full conformational search in aqueous solution, the identification of the most probable protonation and tautomeric state at physiological pH, followed by automated, flexible docking and calculation of the binding affinity using 6D-QSAR. The technical flowchart of *VirtualToxLab*<sup>TM</sup> is presented in Figure 9.

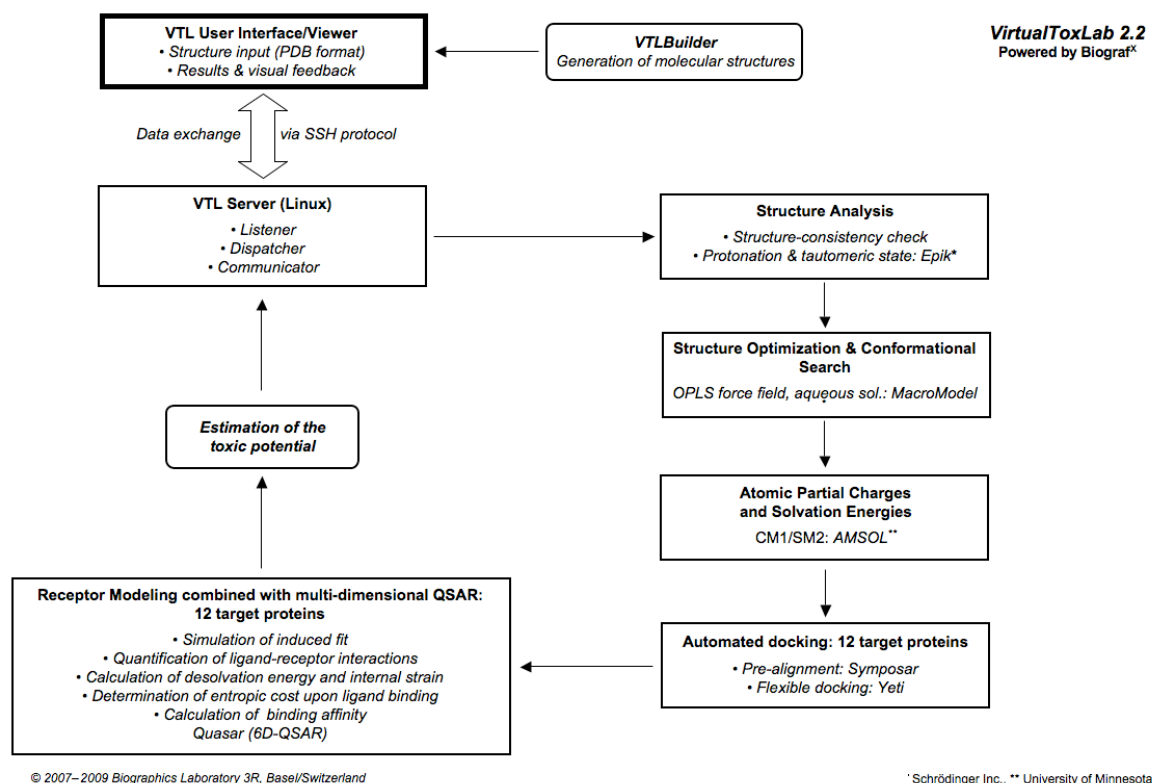


Figure 9. The technical flowchart of *VirtualToxLab*<sup>TM</sup>.

## 4 Results and Discussion

This chapter is organized in three sections: the first (4.1) comprises the glucocorticoid receptor (GR) studies, the second (4.2) the liver X receptor (LXR) studies and the third one (4.3) the development and application of a novel scoring function to the GR and LXR systems, to quantify the hydrophobic interactions between a ligand and a protein in the active site.

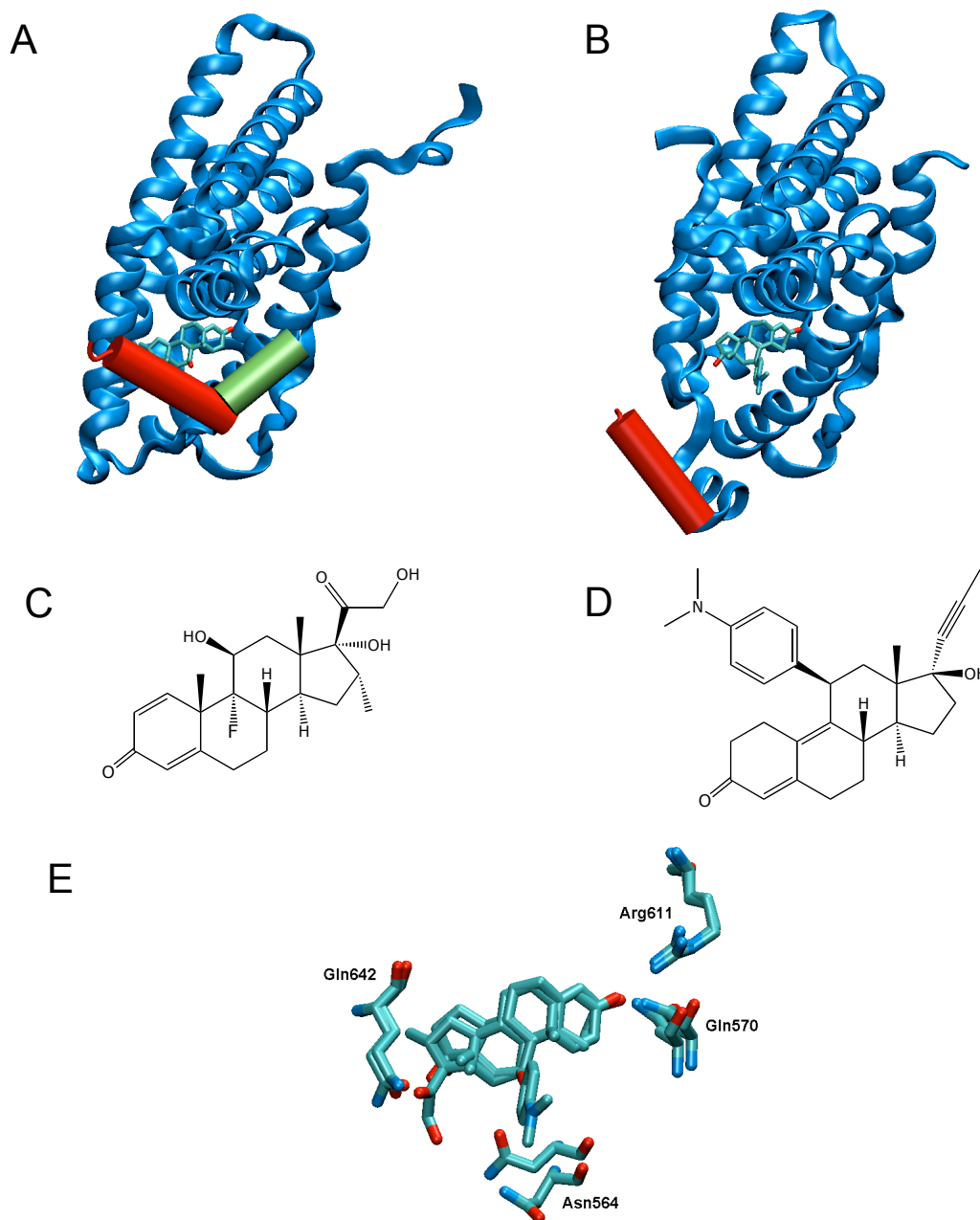
### 4.1 Glucocorticoid Receptor (GR)

In this section, the work performed on the GR is discussed: the analysis of 3D crystal structures, the retrieval of the binding affinity data and the preparation of compound structures of the GR ligands, the docking to the receptor, the MD studies, the building and validation of a QSAR model and the screening of a series of psychotropic drugs for their GR mediated toxic potential.

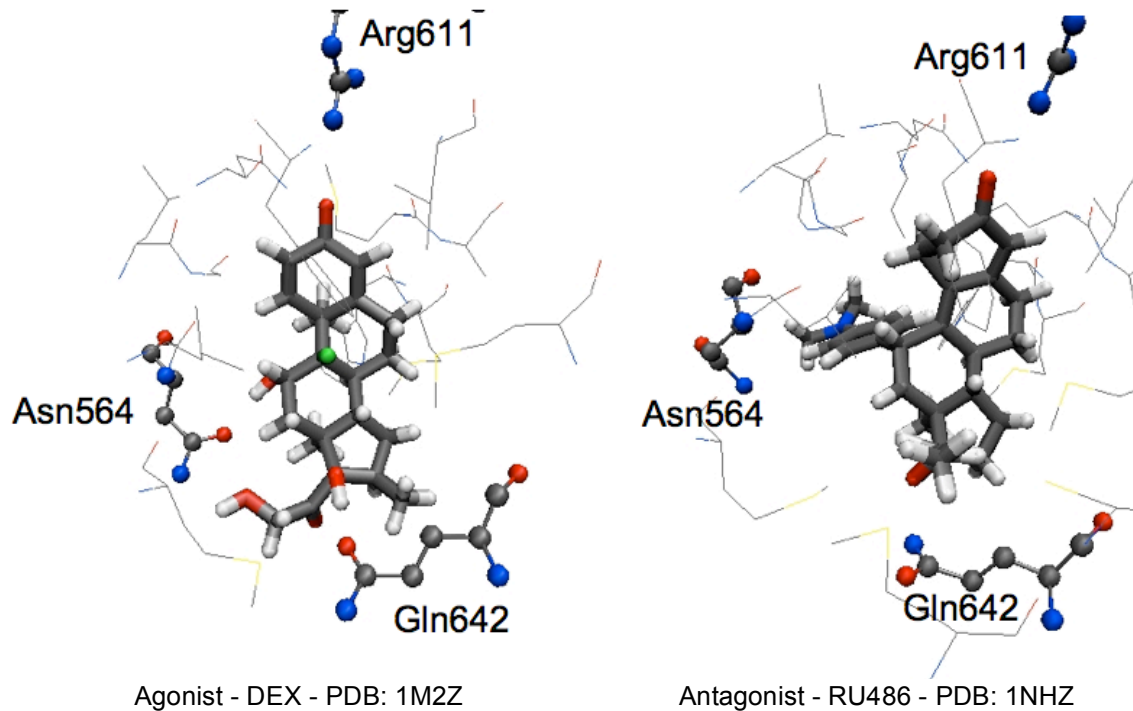
#### 4.1.1 Analysis of the GR Crystal Structures

Prior to any docking and molecular-dynamics studies on the GR, the conformation of the highest resolution crystal structure (at the time: October 2005) of the GR in agonistic form (PDB code 1M2Z<sup>68</sup>), later used for the docking of the compound library, was compared with the structure of the same receptor in the antagonistic form (PDB code 1NHZ<sup>231</sup>). The superposition of the agonistic form, bound with dexamethasone, and the antagonistic form, bound to RU-486, is shown in Figure 10. In agreement with what is seen in other nuclear receptor antagonistic complexes,<sup>232</sup> the main difference between the agonistic and the antagonistic form is found in helix 12, where, in the antagonistic form, the dimethylaniline moiety of RU-486 physically prevents helix 12 from adopting the characteristic agonist position over the ligand-binding pocket as in the agonistic structure.

The binding modes of the co-crystallized agonist and antagonist are compared in Figure 11. Both ligands (the agonist dexamethasone and the antagonist RU-486) interact with Arg611 and Gln642. Due to a different arrangement of the protein side chains of the antagonistic form, the antagonist RU-486, differently from the agonist dexamethasone, is not involved in an hydrogen bond with Asn564.



**Figure 10.** A and B: comparison between crystal structures of GR in the agonist form (A) complexed with the agonist dexamethasone and in the antagonist form (B), complexed with RU-486. In green is represented the coactivator peptide interacting with the agonist form. In red is represented helix 12. Both the receptors are represented in ribbons and the ligands as licorice. C and D: chemical structure of dexamethasone and RU-486. E: superposition of the two ligands and of the amino acid residues Arg611, Gln570, Gln642 and Asn564.

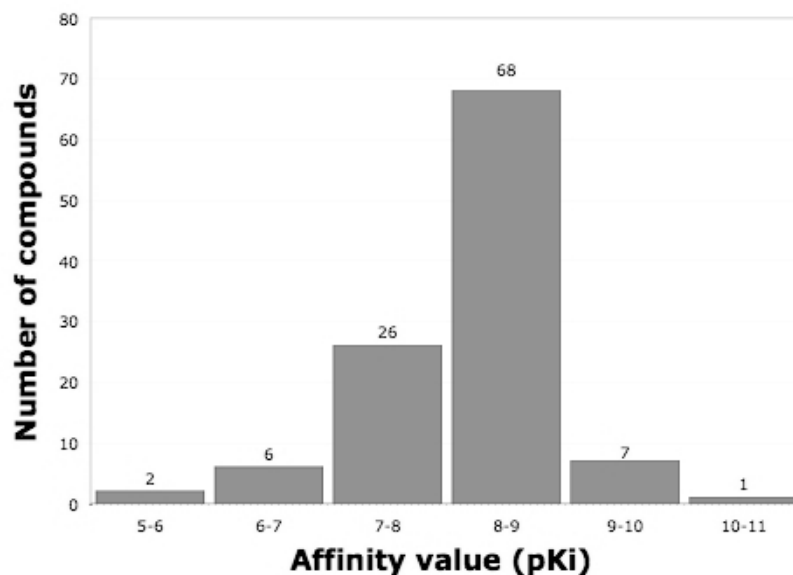


**Figure 11.** Details of the binding pocket of GR bound to the agonist dexamethasone (left) and to the antagonist RU-486 (right). The ligands are represented as licorice and completed with hydrogens, key aminoacid residues as balls and sticks and the other amino-acid residues as lines.

The crystal structure of the agonistic form is available at a slightly lower resolution than the antagonistic form (2.50 Å of 1M2Z<sup>68</sup> versus 2.30 Å of 1NHZ<sup>231</sup>), but was chosen for the docking studies, because for 1M2Z<sup>68</sup> in the sequence comprised between amino-acid residues Ala523 and Lys777 all residues were resolved, while in 1NHZ<sup>231</sup> 16 residues were unresolved.

### 4.1.2 Retrieval of the Binding Affinity Data and Preparation of the Compound Structures

The pharmacological data for the 110 GR-binding compounds were obtained from multiple sources.<sup>217-222</sup> IC<sub>50</sub> values were converted in K<sub>i</sub> values using the Cheng-Prusoff equation,<sup>223</sup> taking into account the concentration of the radioligand (<sup>3</sup>H-labelled Dexamethasone) in each individual experimental assay and its K<sub>i</sub>. Figure 12 shows the activity distribution of the binding data.

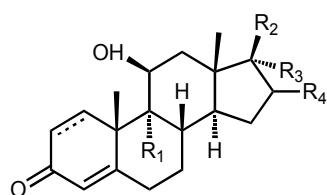
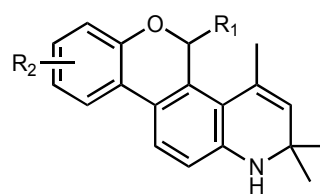
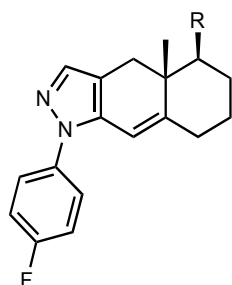
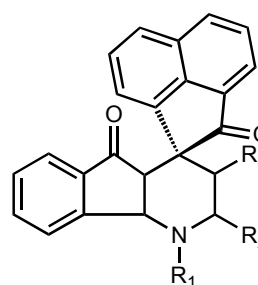


**Figure 12.** Distribution of the pK<sub>i</sub> values: 85% of the affinities cluster within two logarithmic units, while the whole data set spans six orders of magnitude (0.05 nM – 32 μM).

The experimental binding affinities range from  $3.2 \times 10^{-5}$  to  $5 \times 10^{-11}$  M. The majority of affinities, however, lies within two orders of magnitude ( $10^{-7}$  –  $10^{-9}$  M; cf. Figure 12). For prednisolone (compound code within this study: **A02**), a comparison of experimental binding affinities obtained from three different sources shows a difference of more than a factor 10 in its K<sub>i</sub> value (from 2.4 nM to 32 nM), indicating the presence of noise in our input data and making the search for a predictive model more difficult.

The ligands comprise four different chemical classes: steroids, quinoline derivatives, fluorophenyl-indazole derivatives and spirocyclic dihydropyridine derivatives (Figure 13). The compound structures and their K<sub>i</sub> values are given in the appendix A.



Steroids (**A01–A17**)Quinoline derivatives (**B01–B30**)Fluorophenyl-indazole derivatives (**C01–C55**)Spirocyclic dihydropyridine derivatives (**D01–D08**)

**Figure 13.** Compound classes A–D used in the QSAR study. All 110 structures are given in the appendix A.

All the quinoline derivatives (except compound **B02** where the substituent  $R_1$  is H) bear a stereo center. For four compounds of the data set — (R-) and (S-)2,5-dihydro-10-methoxy-2,2,4-trimethyl-5-phenyl-1H-[1]benzopyrano-[3,4-f]quinoline (**B25**, **B26**) and (R-) and (S-)2,5-dihydro-10-methoxy-2,2,4-trimethyl-5-(3,5-dichlorophenyl)-1H-[1]benzopyrano-[3,4-f]quinoline (**B29**, **B30**) — the  $K_i$  values for individual stereoisomers are available. In both cases, the S-isomer shows a higher affinity than the R counterpart (a factor 33 for **B26/B25** and 114 for **B30/B29**, respectively).<sup>218</sup> Docking studies with these stereoisomers to the GR are in agreement with this observation (data reported in section 4.1.3). For the quinoline derivatives that were only tested as racemates, only the S isomer was included in the study and the affinity values were corrected by dividing their  $K_i$  values by a factor of two, in order to account for the content of the almost inactive isomer in the sample.

The ligand structures were generated and optimized as indicated in paragraph 3.3.2.

### 4.1.3 Docking to the GR

The 110 ligands were manually docked to the receptor structure with the software *Yeti*.<sup>99, 101</sup>

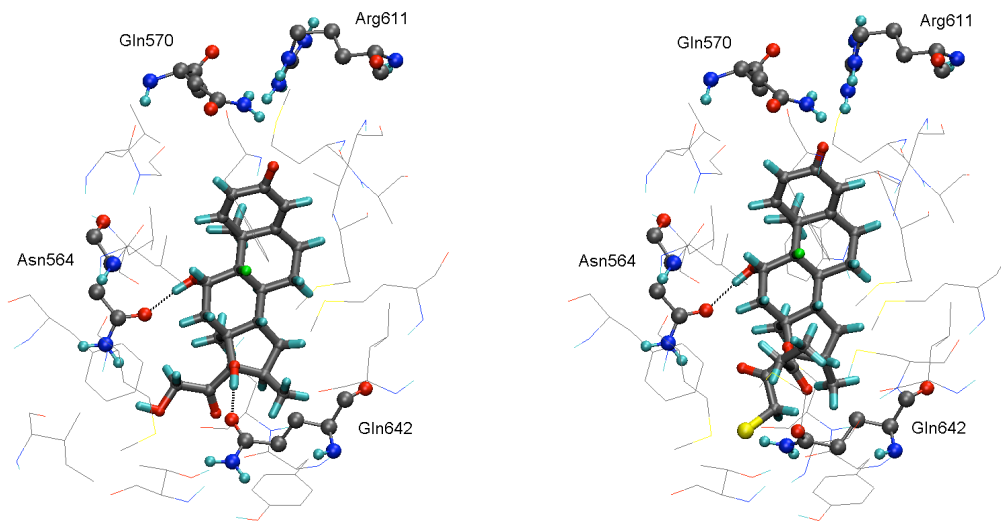
As a general rule, for each chemical class the most active compound was docked and subsequently used it as binding-mode template for the remaining compounds belonging to the same chemical class. For the steroids class dexamethasone (**A01**) was chosen as template (instead of the most active compound **A11**), because it was the ligand bound to GR in the crystal structure. For the B class, the compounds **B18** and **B02** were selected as templates. **B18** was chosen because it is the most active compound of the B class, and **B02** because it is unsubstituted: it features a hydrogen atom in position R<sub>1</sub>, (cf. Figure 13), where the remaining compounds have different functional groups. Finally, **C12** and **D02**, the most active compounds in their classes, were chosen as templates for the C and D class. The binding modes will be discussed in detail.

The general protocol used to lead the docking was the following:

- Minimization of the protein–ligand interaction energy, with special emphasis to the hydrogen bond contribution calculated by the directional *Yeti*<sup>100</sup> force field.
- Establishing the hydrogen bond pattern observed in the crystal structure (like in Figure 11), if possible.

In particular, the amino-acid residues Gln642 and Asn564, involved in hydrogen bonding in the crystal structures (see Figure 11), were considered as anchor points for hydrogen bond interactions in the search for the binding modes.

Figure 14 shows the binding modes of compounds **A01**, the ligand bound to GR in the crystal structure, and **A11**, the most active compound of the A class. In the crystal structure, dexamethasone is completely enclosed within one half of the GR ligand binding domain, and is oriented with the steroidal A ring towards the amino-acid residues Gln570 and Arg611. A similar orientation was reproduced in the docking of **A11** and of the rest of the A class. The binding mode of both ligands (**A01** and **A11**) is stabilized by hydrophobic interactions involving the steroid core (mainly with amino-acid residues Leu566, Trp600, Leu608, Phe623, Leu732 and Phe749) and by hydrophilic interactions with the amino-acid residues Asn564, Gln570, Arg611 and Gln642. Both **A01** and **A11** are hydrogen bonded to Asn564, and **A01** is bonded also to Gln642. Table 4 lists the energetic contributions calculated from the docking of the mentioned compounds. **A11** is characterized by higher (less favorable) energy, when compared with **A01**. A possible explanation of the disagreement between docking simulations and experimental values is solvation energy. **A01** contains more hydrophilic groups in the substituents located at position 17 of the steroidal scaffold (in particular two OH), and the solvation energy is almost double, when compared to **A11** (17 kcal/mol for **A01** and 9 for **A11**). By considering only protein–ligand interaction energy, penalties due to solvation energy, entropy or internal energy of the ligand are not included (but will be considered in the QSAR).



**Figure 14.** Details of the binding of the compounds dexamethasone (**A01**, left) and **A11** (right). The ligands are represented as licorice, key amino-acid residues as balls and sticks and the other amino-acid residues lining the binding pocket as lines.

**Table 4.** Quantitative aspects of ligand binding to the GR, as obtained from the manual docking.

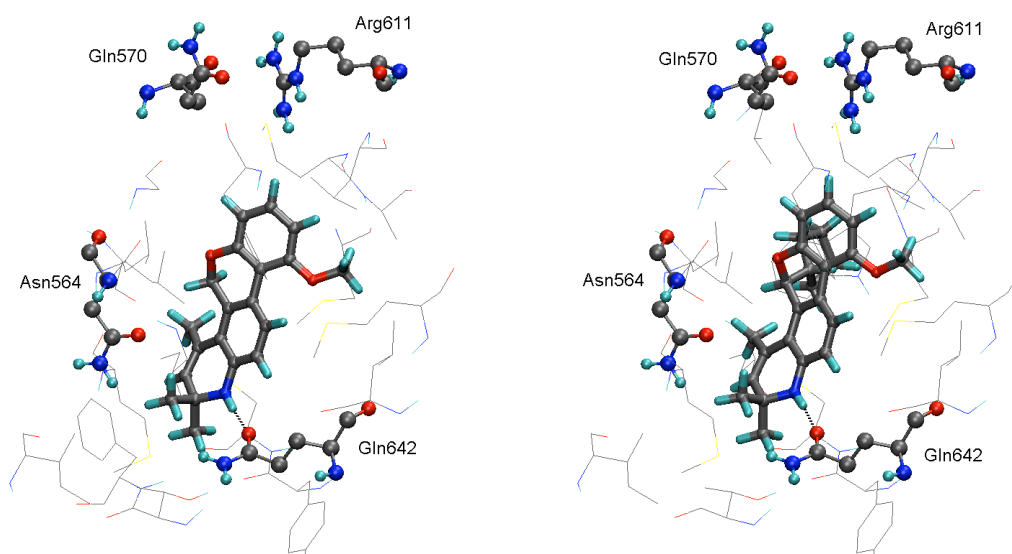
Ligand	$E_{\text{Protein-Ligand}}$	( $E_{\text{Elec}}$	$E_{\text{vdW}}$	$E_{\text{HB}}$ )	$K_i$ (nM)
<b>A01</b>	-58.9	-16.9	-33.3	-8.7	2.3
<b>A11</b>	-52.4	-15.1	-33.6	-3.7	0.047
<b>B02</b>	-33.8	-8.3	-22.5	-3.0	210
<b>B18</b>	-41.1	-7.9	-30.3	-2.8	1.1
<b>C12a</b>	-37.9	-4.6	-29.3	-4.0	0.52
<b>C12b</b>	-46.1	-7.2	-34.3	-4.6	0.52
<b>D02</b>	-37.9	-12.6	-24.3	-1.0	0.89

All energies are given in kcal/mol.  $E_{\text{Protein-Ligand}}$ : protein–ligand interaction energy;  $E_{\text{Elec}}/E_{\text{vdW}}/E_{\text{HB}}$ : electrostatic, van der Waals, and hydrogen bonding components;  $K_i$ : binding affinity. All the compound structures are given in Appendix A.

Figure 15 shows the binding modes of two compounds of the B class: **B02** and **B18**. From the docking study, both compounds form a hydrogen bond with Gln642, and locate their hydrophobic quinoline core in the GR pocket in a similar manner as **A01** and **A11** position their steroid core. Hydrophobic interactions are found between the ligands and amino-acid residues Leu608, Phe623, Mer646, Ile747 and Leu753. Compound **B18**, when compared to **B02**, features an

additional cyclopentane substituent that seems to be in contact with Trp600 and Met601. These additional hydrophobic contacts could be responsible for the more favorable van der Waals interaction energy (Table 4).

Energetic contributions and experimental values for compounds **B25**, **B26**, **B29** and **B30** are also reported (Table 5). **B25** and **B29** (*R* stereoisomers), are less active than **B26** and **B30** (the corresponding *S* stereoisomers). The docking study is in agreement with the experimental values: **B25** and **B29** are characterized by less favorable energies, when compared to **B26** and **B30**. According to this observation, for the remaining compounds belonging to the B class that were tested as racemates, the *R* stereoisomers were assumed to be the inactive isomers, and were not included in the study.



**Figure 15.** Details of the binding of the compounds **B02** (left) and **B18** (right). The ligands are represented as licorice, key amino-acid residues as balls and sticks and the other amino acid residues lining the binding pocket as lines.

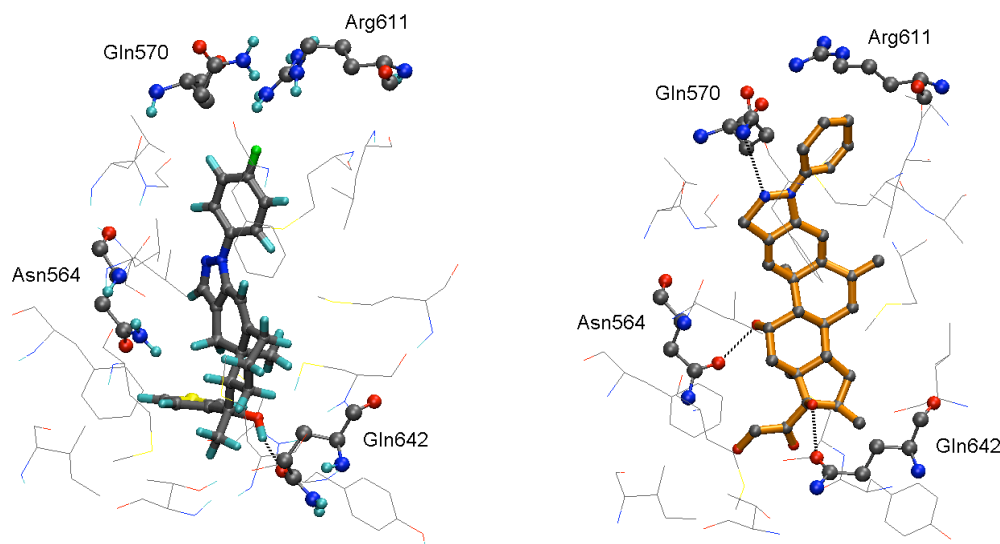
**Table 5.** Quantitative aspects of ligand binding to the GR for the stereoisomers **B25/B26** and **B29/B30**.

Ligand	$E_{\text{Protein-Ligand}}$	( $E_{\text{Elec}}$	$E_{\text{vdW}}$	$E_{\text{HB}}$ )	$K_i$ nM
<b>B25</b>	-30.0	-8.0	-21.7	-0.1	240
<b>B26</b>	-38.3	-6.2	-29.8	-2.4	2.1
<b>B29</b>	-20.7	-2.2	-18.4	-0.1	95
<b>B30</b>	-36.5	-3.9	-29.6	-3.0	3.3

All energies are given in kcal/mol.  $E_{\text{Protein-Ligand}}$ : protein–ligand interaction energy;  $E_{\text{Elec}}/E_{\text{vdW}}/E_{\text{HB}}$ : electrostatic, van der Waals, and hydrogen bonding components;  $K_i$ : binding affinity. All the compound structures are given in Appendix A.

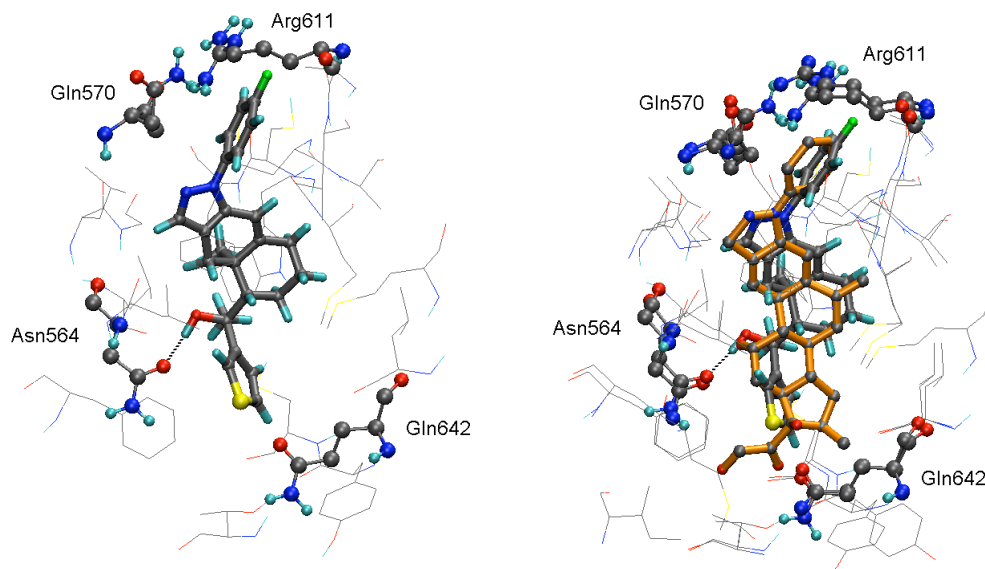
Figure 16 shows the binding mode initially considered for the ligand **C12** (C12a). The fluorophenyl moiety is directed towards amino-acid residues Gln570 and Arg611, like the A rings of the steroids. A hydrogen bond is formed between the ligand and Gln642. Hydrophobic contacts involve the amino-acid residues Met604, Leu608, Phe623 and Met646.

Following the publication in 2008 of the crystal structure of GR bound to deacylcortivalzol (PDB code 3BQD,<sup>66</sup> resolution 2.5 Å), shown in Figure 16 on the right, a second binding mode (C12b) was considered for the compound **C12** and consequently for the C class. This second binding mode was based on the observation that in the crystal structure 3BQD, the aryl-pyrazole moiety of deacylcortivalzol is located 4 Å higher (when considering the view of Figure 16) compared to the aryl-pyrazole moiety of C12a. In C12b, as it can be seen from Figure 17, a hydrogen bond is formed with Asn564 (instead of Gln642 as it is from C12a). Hydrophobic interactions still involve the amino-acid residues Met604, Leu608 and Met646, but also Trp600, Met601 and Leu753. A superposition of C12b and 3BQD is also shown in Figure 17.



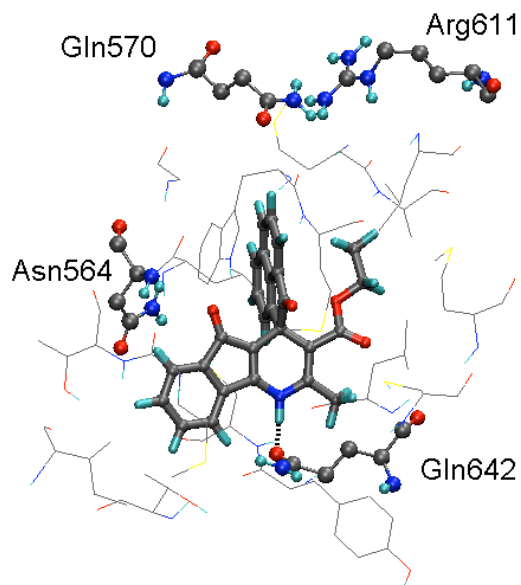
**Figure 16.** Details of the first binding mode of the compounds **C12** (C12a) on the left and the crystal structure of GR bound to deacylcortivalzol (PDB code 3BQD) on the right. The bonds of ligand deacylcortivalzol are colored in orange.

Even if the energetic contributions for C12b are more favorable than for C12a (Table 4), both binding modes were included in the QSAR study, because both binding modes are reasonable in the GR binding pocket and because the ligands from the C class (non steroidal ligands) differ significantly from deacylcortivalzol (steroid ligand), and it is feasible that they bind in a different way.



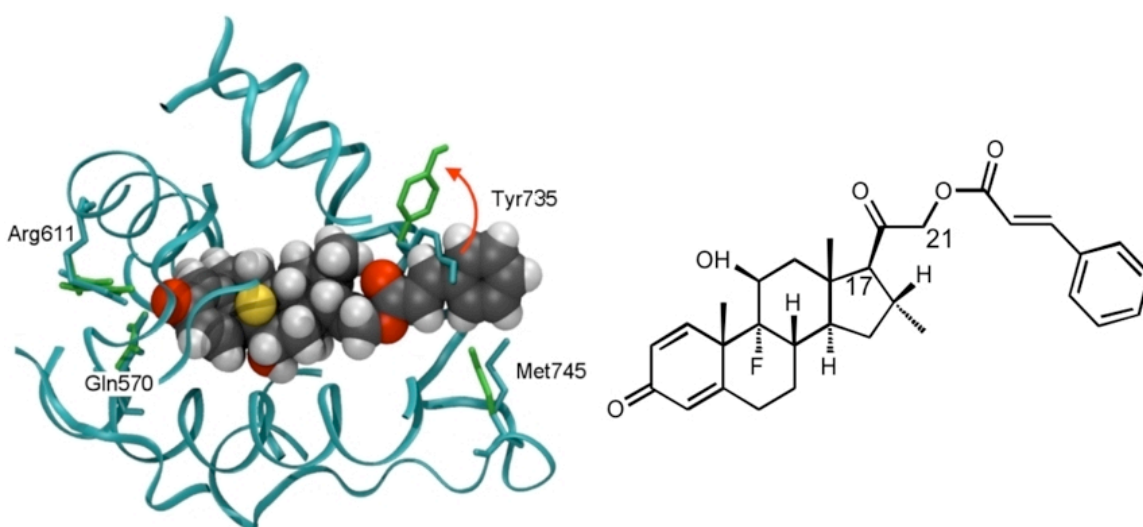
**Figure 17.** Details of the second binding mode of the compounds **C12** (C12b) on the left and the superposition of C12b on the crystal structure 3BQD on the right. The bonds of ligand deacylcortivazol are colored in orange.

Finally, the binding mode of compound **D02** is shown in Figure 18. A hydrogen bond is formed with Gln642, and hydrophobic interactions involve amino-acid residues Met560, Trp600, Met601, Met604, Ala605, Met646, Leu732 and Ile747.



**Figure 18.** Details of the binding of the compound **D02**. The ligand is represented as licorice, key amino-acid residues as balls and sticks and the other amino-acid residues lining the binding pocket as lines.

The consideration of protein flexibility by the simulation of a local induced fit was a key issue in identifying reasonable binding poses. Particularly, the docking of glucocorticoids with bulky substituents at the 17 $\beta$ -position (Figure 19) to a rigid structure of the glucocorticoid receptor may not yield the correct binding pose, because the binding site would not seem to be wide enough to accommodate large compounds such as desoxymethasone 21-cinnamate (**A08**). For other steroid receptors, such as the progesterone receptor,<sup>233</sup> there is experimental evidence of a local induced fit to accommodate bulky substituents in position 17. For the androgen receptor, a similar behavior has been postulated.<sup>234</sup> From recent studies,<sup>235</sup> induced-fit mechanisms are observed also in the GR binding pocket in correspondence of the 17 $\alpha$  position. Likewise, bulky ligands may indicate steric interference with the GR-binding pocket, if the protein structure is kept rigid during docking (cyan licorice in Figure 19). In contrast hereto, energetically favorable binding poses are found when the side chains of the protein are allowed to adjust to the ligand topology (green licorice in Figure 19).



**Figure 19.** With a rigid protein structure, docking of desoxymethasone 21-cinnamate (**A08**; shown in space-filling mode; chemical structure shown on the right) to the crystal structure of the GR, obtained from dexamethasone **A01**–GR complex (cyan ribbons and sticks), leads to unfavorable interactions. The following amino-acid residues lining the binding pocket are shown: Tyr735, Met745, Arg611 and Gln570. Cyan stick representations refer to the position of amino-acid residues as they are observed in the crystal structure, while green sticks refer to the amino-acid position after flexible docking. While the position of Arg611 and Gln570 remains almost unchanged, flexible docking allows the side-chain of Met745 and Tyr735 to move and accommodate the 21-cinnamate moiety of **A08** (the most evident movement of Tyr735 is indicated by the red arrow). On the left, **A08** is rotated vertically by 180 degrees when compared to the standard steroid depiction on the right.

A protocol of automated flexible docking with *Yeti/Autodock*<sup>100</sup> was also performed, leading to a less consistent alignment and lower energy values (the energetic contributions are reported in Table 6), when compared with the manual docking alignment. In particular, the average protein–ligand interaction energy of the investigated compounds is 4 kcal/mol lower for the poses obtained with the automated docking. The hydrogen-bond contribution, in particular, is significantly low (on the average is 3.2 kcal/mol lower for the automated docking results). The same behavior was observed for the remaining compounds of the library (data not shown).



**Table 6.** Quantitative aspects of ligand binding to the GR, as obtained with the automated docking.

Ligand	$E_{\text{Protein-Ligand}}$	$(E_{\text{Elec}}$	$E_{\text{vdW}}$	$E_{\text{HB}})$
<b>A01a</b>	-51.0	-16.0	-31.7	-4.1
<b>A01b</b>	-43.3	-15.6	-23.9	-3.7
<b>A01c</b>	-42.8	-16.2	-25.3	-1.3
<b>A01d</b>	-41.8	-11.7	-25.8	-4.2
<b>A01 average</b>	-44.7	-14.9	-26.7	-3.3
<b>A11a</b>	-50.9	-16.1	-33.8	-0.9
<b>A11b</b>	-50.7	-20.8	-29.9	-0.0
<b>A11c</b>	-42.6	-12.9	-29.7	-0.0
<b>A11d</b>	-41.4	-18.9	-22.6	-0.0
<b>A11 average</b>	-46.4	-17.2	-29.0	-0.2
<b>B02</b>	-36.6	-13.2	-23.4	-0.0
<b>B02</b>	-36.4	-12.6	-23.7	-0.0
<b>B02</b>	-36.3	-13.1	-23.2	-0.0
<b>B02</b>	-30.4	-5.9	-22.3	-2.2
<b>B02 average</b>	-34.9	-11.2	-23.2	-0.6
<b>B18</b>	-43.3	-13.2	-30.2	-0.0
<b>B18</b>	-41.6	-12.9	-28.8	-0.0
<b>B18</b>	-35.0	-10.8	-24.2	-0.0
<b>B18</b>	-34.6	-9.2	-25.3	-0.0
<b>B18 average</b>	-38.6	-11.5	-27.1	-0.0
<b>C12a</b>	-40.5	-8.6	-32.0	-0.0
<b>C12b</b>	-35.8	-5.9	-29.8	-0.1
<b>C12c</b>	-33.4	-6.0	-26.7	-0.7
<b>C12d</b>	-27.7	-3.2	-24.0	-0.5
<b>C12 average</b>	-34.4	-5.9	-28.1	-0.3
<b>D02a</b>	-41.6	-13.3	-28.4	-0.0
<b>D02b</b>	-40.8	-14.9	-25.9	-0.0
<b>D02c</b>	-40.3	-9.6	-30.6	-0.1
<b>D02d</b>	-39.9	-12.6	-27.3	-0.0
<b>D02 average</b>	-40.7	-12.6	-28.1	-0.0

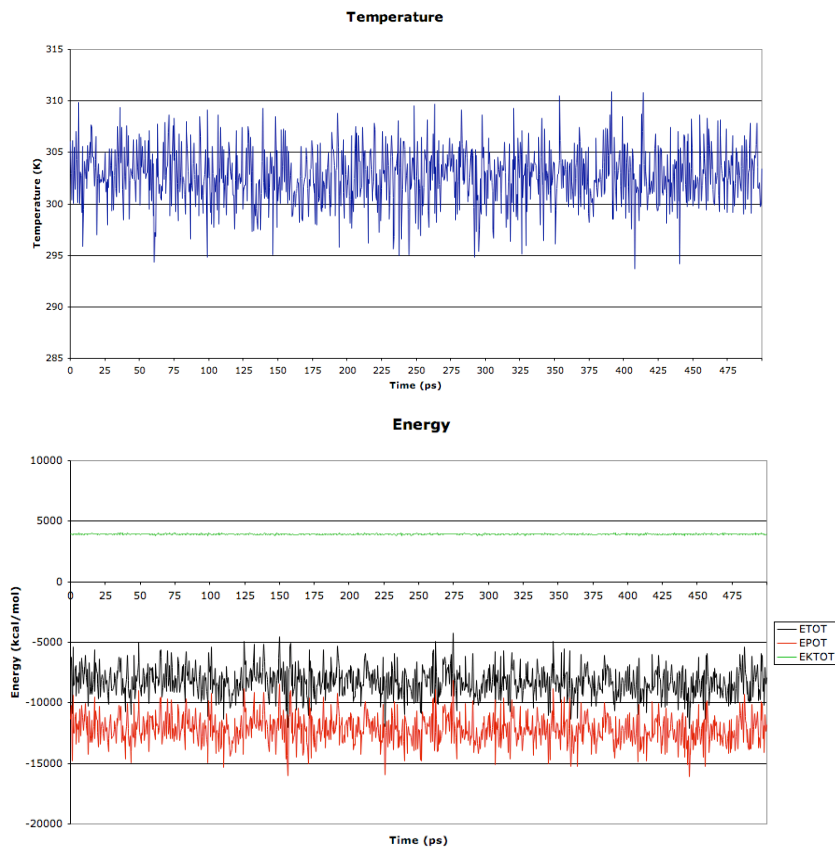
All energies are given in kcal/mol.  $E_{\text{Protein-Ligand}}$ : protein–ligand interaction energy;  $E_{\text{Elec}}/E_{\text{vdW}}/E_{\text{HB}}$ : electrostatic, van der Waals, and hydrogen bonding components. For each ligand, the energetic contributions of the 4 lowest energy poses from automated docking are reported (a, b, c, d), together with the average values.

#### 4.1.4 Molecular-Dynamics Simulations on the GR

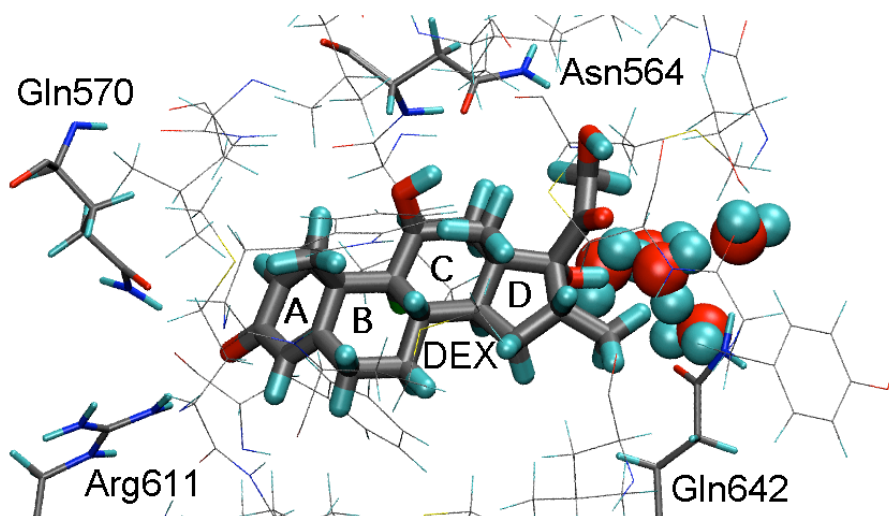
To analyze the interactions between GR and its ligands from a dynamic point of view, and to explore possible binding modes of representative ligands, molecular dynamics simulations of protein–ligand complexes were performed. Details of the simulation protocols are reported in the section 3.3.5. A total of five ligands were chosen to comprise representative subset of the studied glucocorticoids: four ligands belong to the dataset (**A01**, **B02**, **B18** **C12**, structures reported in the appendix A), and fluorocortivazol, a potent GR agonist not included in the dataset but belonging to the C class and being the rationale for the design of the C class, were investigated. As starting point for the MD studies, for each ligand the binding mode obtained from manual docking was considered. MD simulations and in particular the analysis of the trajectories allowed for a more dynamic characterization (i.e. time dependent) of the protein–ligand interactions responsible for ligand binding and to gain insight into the binding mode.

As an example of a molecular-dynamics simulation, the binding of compound **A01** (dexamethasone) to GR is reported. An analysis of temperature and energetic contributions was performed during the production phase (Figure 20). The temperature fluctuates during the simulation, but the average remains constant, being the simulation conducted at constant temperature. As there are no sudden jumps in the energy contributions, the simulation seems to have reached the equilibrium. The potential energy fluctuates more than the kinetic energy. This can be explained with the fact that the simulation was conducted at constant temperature, which determines the value of kinetic energy.

Figure 21 shows a snapshot of the trajectory and highlights the key amino-acid residues. From the analysis of the trajectory during the simulation, interactions with the amino-acid residues Asn564, Gln570, Arg611 and Gln642 are observed. In particular, a hydrogen bond between the carbonyl oxygen in the ring A of dexamethasone and the amino-acid residue Gln570 is the most present during the simulation (it occurs 96% of the time). The same oxygen atom is involved also in a hydrogen bond with Arg611, but for a shorter time (39%). The average distances (between the oxygen atom of the ligand and hydrogen atoms of the protein) are respectively 2.8 and 3.7 Å. Asn564 is also engaged as an acceptor in a very persistent hydrogen bond (it occurs 95% of the simulation time, average distance: 2.1 Å) with the hydroxyl of the C ring of the ligand, and occasionally (17%) with the hydroxyacetyl moiety. As a donor, Asn564 is hydrogen-bonded for 81% of the simulation with the hydroxyl oxygen of the hydroxyacetyl moiety. Gln642 in our simulation is involved in a hydrogen bond with the hydroxyl of the D ring only for 1% of the simulation (average distance: 3.9 Å). In our study, this can be explained with the fact that such hydroxyl group is surrounded by water molecules, able to form hydrogen bonds with the ligand. The hydrogen bond pattern observed in the simulation appears to be an intense network that stabilizes the binding mode and explains the high affinity of the ligand ( $K_i$  of 2.3 nM).



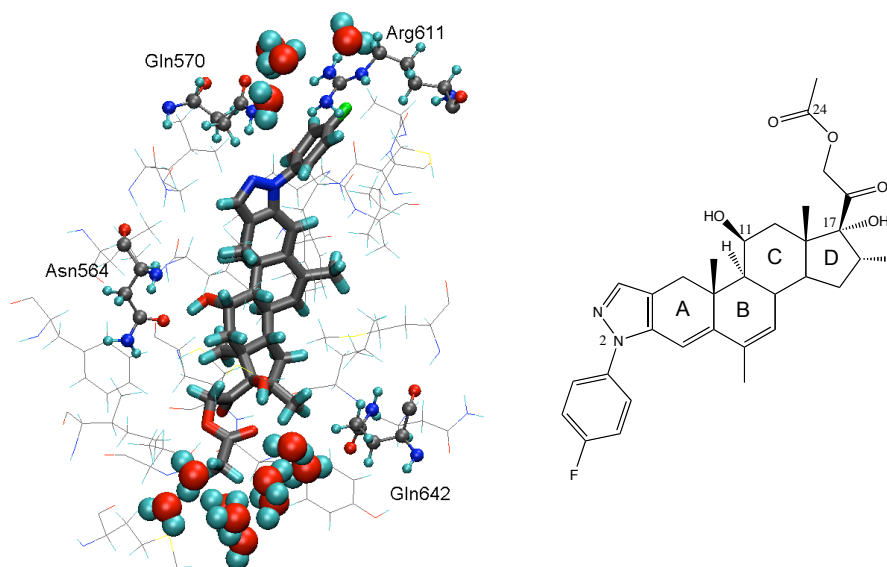
**Figure 20.** Plot of the temperature (top) and energy contributions (bottom) during the production phase (500 ps). Temperature and energies fluctuate around a constant value, indicating that the simulation is stable and has reached equilibrium. ETOT: total energy; EPOT: potential energy; EKTOT: kinetic energy.



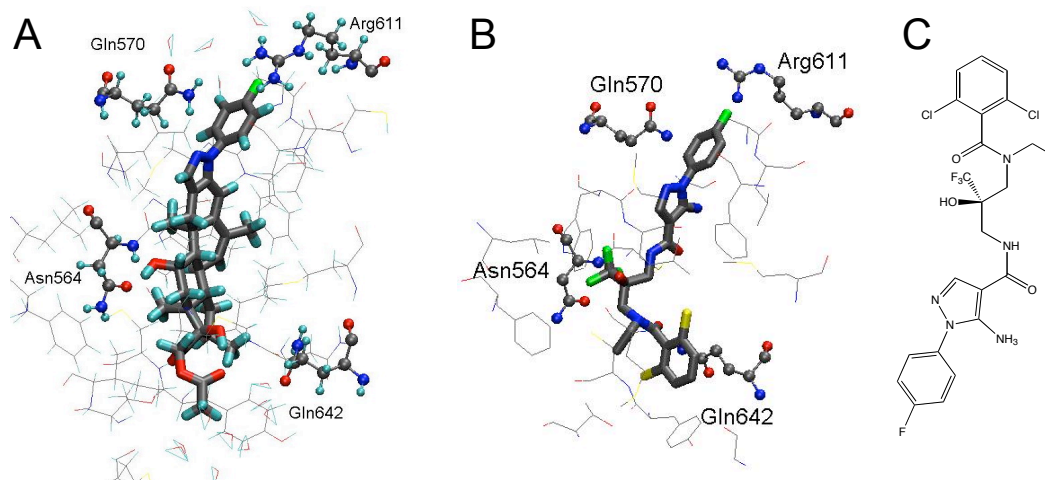
**Figure 21.** Detail from the average structure from the MD. The ligand dexamethasone is represented as thick sticks, key amino-acid residues (Gln570, Asn564, Arg611 and Gln642) are represented as thin sticks, the remaining amino-acid residues in the binding pocket as lines and the water molecules are represented as spheres.

As a second example, the simulation of fluorocortivazol is discussed. Fluorocortivazol is among the most effective agents currently available for the treatment of inflammatory and allergic diseases.<sup>236, 237</sup> The design of the library of the C class was aimed by the authors<sup>220</sup> at discovering new glucocorticoids by retaining the features of fluorocortivazol believed to be important for the binding:<sup>220</sup> the A and B ring, the 11-OH, and the pyrazole containing the 2'-aryl substituent. Therefore, the exploration of the binding mode of fluorocortivazol was of interest for this study. Figure 22 displays a frame of the MD. During the simulation, two hydrogen bonds are observed for a significant period of time: the first one between Asn564 and the 11-OH (occurring for 100% of the simulation time, average distance: 1.7 Å), and the second one between Gln570 and the nitrogen atom of the pyrazole ring (occurring 58% of the time, average distance: 2.6 Å). A parallel stacking<sup>238</sup> is observed between Arg611 and the aryl moiety of the ligand (average distance 3.0 Å). In the aryl moiety of fluorocortivazol, the acidity of the aromatic hydrogen atoms is increased by the presence of fluorine in the ring, and this could favor the interaction. An intra-molecular hydrogen bond is also observed, between the 17-OH and the 24-O (present 100% of the time, average distance 1.9 Å).

Recently a GR crystal structure was published (PDB code 3E7C,<sup>239</sup> resolution 2.15 Å), for the first time with a non-steroidal ligand bound (GSK866, structure shown in Figure 23c). A common feature can be observed between fluorocortivazol, deacylcortivazol, the C class and GSK866: the presence of 2'-aryl-pyrazole. Figure 23 shows, side by side, the binding mode of fluorocortivazol and of GSK866, as it is in the crystal structure. In both structures, the 2'-aryl-pyrazole is positioned towards Arg611, and hydrogen bonds are formed between the ligands and the amino-acid residues Asn564 and Gln570. The comparison confirms the correctness of the binding pose identified by our simulations.



**Figure 22.** Screenshot of the average structure from the MD. Fluorocortivazol is represented as licorice, key amino-acid residues (Gln570, Asn564, Arg611 and Gln642) are represented as balls and sticks, the remaining amino-acid residues in the binding pocket as lines and the water molecules are represented as spheres. The chemical structure of fluorocortivazol is shown on the right. The structure is rotated vertically by 180 degrees when compared to the depiction on the left.



**Figure 23.** A: binding mode of fluorocortivazol. The structure has been completed with hydrogens. B: binding mode of GSK866, as it from the crystal structure. C: chemical structure of GSK866. The structure is rotated vertically by 180 degrees when compared to the depiction on the left. Ligands are represented as licorice, key amino-acid residues (Gln570, Asn564, Arg611 and Gln642) are represented as balls and sticks, and the remaining amino-acid residues in the binding pocket as lines.

#### 4.1.5 Building and Validation of QSAR Models for GR

The ensemble of poses of the 110 glucocorticoids (88 training and 22 test compounds), identified using flexible docking as described in the section 3.2.15 and 3.3.4, was used as input for the quantitative structure-activity relationship technology *Quasar*.<sup>214</sup> The mixed-modeling approach allows for both the correct identification of the binding mode and the reliable estimation of binding affinities.

Two different alignments (from manual and automated docking) and different parameter settings were explored. Table 7 lists some of the performed simulations (sorted according to the predictive  $r^2$  value), the parameter settings (number of crossovers, population size, attenuation factors, induced-fit weight, number of cross-validation groups) and the corresponding results cross-validation  $r^2$ , predictive  $r^2$ , *rms* on the training set, *rms* on the test set. All parameters not specified in this table were taken as default.<sup>240</sup>

The simulations obtained from the manual docking alignment show higher  $r^2$  values and lower *rms* (upper part of Table 7). The impact of the different parameters on the predictivity of the models is briefly discussed here:

##### *Number of crossovers*

This number indicates the length of a simulation. To avoid overfitting, each simulation was stopped at the point where the predictive  $r^2$  starts to drop (while the cross-validated  $r^2$  continues to raise). For comparative reasons, some simulations are reported also with higher number of crossovers (MD01–MD02, MD07–MD10, AD05–07 AD13–AD18). The simulations from the automated docking alignment reach a maximum value of predictive  $r^2$  earlier than the ones from manual docking, but the corresponding value of predictive  $r^2$  is never greater than 0.5. From Table 7 we observe that, to obtain a good predictivity, a number of crossovers between 30,000 and 60,000 (corresponding to 150–300 generations) is required.

##### *Population size*

Population size of 100 and 200 individuals led to comparable results. Bigger populations lead in general to longer simulations, which were however characterized by comparable values of predictive  $r^2$  and *rms*.

##### *Attenuation factors*

Attenuation factors are relative weights applied to the various correction terms (in this thesis they were used to scale solvation, entropy and internal energy). They allow to reduce or enhance a particular effect or even to switch it off ( $w=0.0$ ). By default, all values are set to 1.0. Two sets were used: the default values, and a different one (0.5 for solvation, 1.0 for entropy and 0.25 for internal energy). From the simulations we observe that the best models underweight solvation and even

more internal energy. This is due to the fact that the range of solvation (18 kcal/mol for the studied ligands) and internal (27 kcal/mol) energies is higher than the range in binding affinity (6 kcal/mol), and without an appropriate scaling such energies might influence too strongly the model evolution.

#### *Induced-fit weight*

This value allows for a scaling of the induced-fit energy.<sup>240</sup> The results suggest that the value is appropriate when used as it is (weight 1.0), since the underweighted values don't produce models with improved performance.

#### *Cross-validation groups*

This parameter refers to the number of groups the training set is split into, in order to train the models for prediction during the evolution: a value of  $n=10$  results in defining ten subgroups of the ligand molecules comprising the training set. 9/10 of these groups are used to derive the classical correlation while the last 1/10 is predicted based on the corresponding correlation. A low number of cross-validation groups result in more restrictive conditions. As it can be expected, for models where a good predictivity is found (upper part of Table 7), higher number of cross validation groups leads to more predictive models.

**Table 7.** Summary of the *Quasar* simulations.

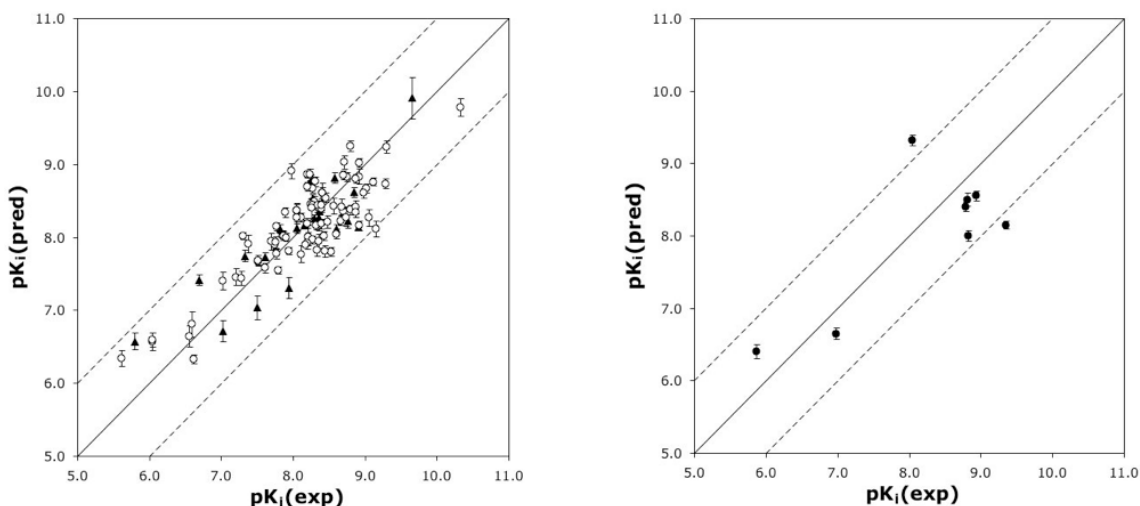
Simulation	#CO	POP	AF	IFW	CVG	$q^2$	$p^2$	Training rms avg/max	Test rms avg/max
MD01	60,000	200	0.5, 1.0, 0.25	1.0	10	0.702	0.719	1.5/9.9	1.6/4.9
MD02	70,000	200	0.5, 1.0, 0.25	1.0	10	0.708	0.717	1.4/9.8	1.6/5.0
MD03	40,000	100	0.5, 1.0, 0.25	1.0	6	0.722	0.701	1.4/7.9	1.8/6.9
MD04	50,000	200	0.5, 1.0, 0.25	1.0	6	0.700	0.693	1.5/8.3	1.8/5.1
MD05	35,000	100	0.5, 1.0, 0.25	1.0	7	0.688	0.687	1.5/9.4	1.8/6.7
MD06	35,000	200	0.5, 1.0, 0.25	1.0	8	0.639	0.686	1.7/9.9	1.8/11.5
MD07	30,000	100	0.5, 1.0, 0.25	1.0	10	0.700	0.681	1.5/9.7	1.8/8.9
MD08	40,000	200	0.5, 1.0, 0.25	1.0	7	0.655	0.678	1.6/8.6	1.8/6.9
MD09	30,000	200	0.5, 1.0, 0.25	0.5	5	0.670	0.670	1.6/10.1	1.9/7.8
MD10	35,000	100	0.5, 1.0, 0.25	1.0	10	0.712	0.664	1.4/8.8	1.9/8.1
MD11	30,000	100	1.0, 1.0, 1.0	1.0	5	0.610	0.630	1.8/10.5	2.1/10.8
MD12	30,000	100	1.0, 1.0, 1.0	0.5	10	0.602	0.622	1.9/10.5	2.1/8.1
MD13	25,000	200	1.0, 1.0, 1.0	0.5	5	0.498	0.614	2.3/98.6	2.1/16.3
MD14	25,000	100	1.0, 1.0, 1.0	0.5	5	0.495	0.591	2.3/96.7	2.3/18.3
MD15	35,000	200	1.0, 1.0, 1.0	0.5	5	0.482	0.581	2.3/98.5	2.3/17.5
MD16	25,000	200	1.0, 1.0, 1.0	1.0	10	0.492	0.575	2.3/117	2.3/20.3
MD17	25,000	200	1.0, 1.0, 1.0	1.0	5	0.476	0.569	2.3/95.5	2.4/19.0
MD18	60,000	100	1.0, 1.0, 1.0	1.0	10	0.655	0.555	1.7/10.1	2.5/10.1
AD01	10,000	200	0.5, 1.0, 0.25	0.5	5	0.641	0.446	1.3/5.9	3.9/56.1
AD02	4,000	100	1.0, 1.0, 1.0	1.0	10	0.741	0.445	1.1/3.4	4.1/22.1
AD03	5,000	100	0.5, 1.0, 0.25	1.0	6	0.688	0.413	1.2/5.6	4.2/101
AD04	5,000	100	1.0, 1.0, 1.0	1.0	5	0.743	0.398	1.1/3.8	4.4/21.9
AD05	10,000	200	0.5, 1.0, 0.25	1.0	10	0.628	0.390	1.3/5.2	4.4/100
AD06	5,000	200	1.0, 1.0, 1.0	1.0	5	0.447	0.366	1.9/44.2	4.7/72.2
AD07	15,000	200	0.5, 1.0, 0.25	1.0	10	0.785	0.349	0.9/2.3	4.7/70.6
AD08	5,000	200	0.5, 1.0, 0.25	1.0	6	0.436	0.324	2.0/30.9	5.0/149
AD09	10,000	200	0.5, 1.0, 0.25	1.0	7	0.613	0.285	1.3/5.2	5.2/127
AD10	20,000	200	1.0, 1.0, 1.0	0.5	5	0.945	0.284	0.3/0.8	5.4/51.5
AD11	2,000	100	1.0, 1.0, 1.0	0.5	5	0.431	0.274	1.9/20.6	5.4/93.7
AD12	4,000	100	0.5, 1.0, 0.25	1.0	7	0.550	0.258	1.6/11.4	5.5/207
AD13	4,000	100	0.5, 1.0, 0.25	1.0	10	0.456	0.235	1.8/18.9	5.8/207
AD14	5,000	200	1.0, 1.0, 1.0	0.5	5	0.667	0.224	1.2/4.0	5.7/120
AD15	2,000	100	1.0, 1.0, 1.0	0.5	10	0.516	0.216	1.7/13.6	5.9/116
AD16	5,000	200	0.5, 1.0, 0.25	1.0	8	0.462	0.214	1.9/18.5	5.9/268
AD17	1,000	200	1.0, 1.0, 1.0	1.0	10	0.169	0.205	2.7/95.7	6.0/147
AD18	5,000	100	0.5, 1.0, 0.25	1.0	10	0.599	0.202	1.4/9.3	6.0/200

Simulation: MD: manual docking; AD: automated docking; #CO, number of crossovers; POP, population size; AF, attenuation factors (weight of solvation, entropy, internal energy); IFW, induced-fit weights; CVG, cross validation groups;  $q^2$ : cross-validated  $r^2$ ,  $p^2$ : predictive  $r^2$ ; *Training rms avg/max*, *Test rms avg/max*, the *rms* and maximal deviation from the experimental binding affinity respectively for training set and test set (given as a factor (off) in  $K_i$ ).



The model MD01, characterized by the highest predictive  $r^2$  and lowest  $rms$  on the test set, will be discussed in detail. The family of receptor models was evolved for 60,000 crossover cycles, corresponding to 300 generations. For cross-validation we selected a *leave-n-out* ( $n = 10$ ) protocol. Protein flexibility was mimicked using a total of six induced-fit scenarios. Throughout the entire simulation, a static mutation rate of 0.02 was applied during transcription of the quasi-atomistic properties.

The model family converged at a cross-validated  $r^2$  of 0.702 for the 88 training compounds and a predictive  $r^2$  of 0.719 for the 22 test ligands (cf. Figure 24 and Table 8).



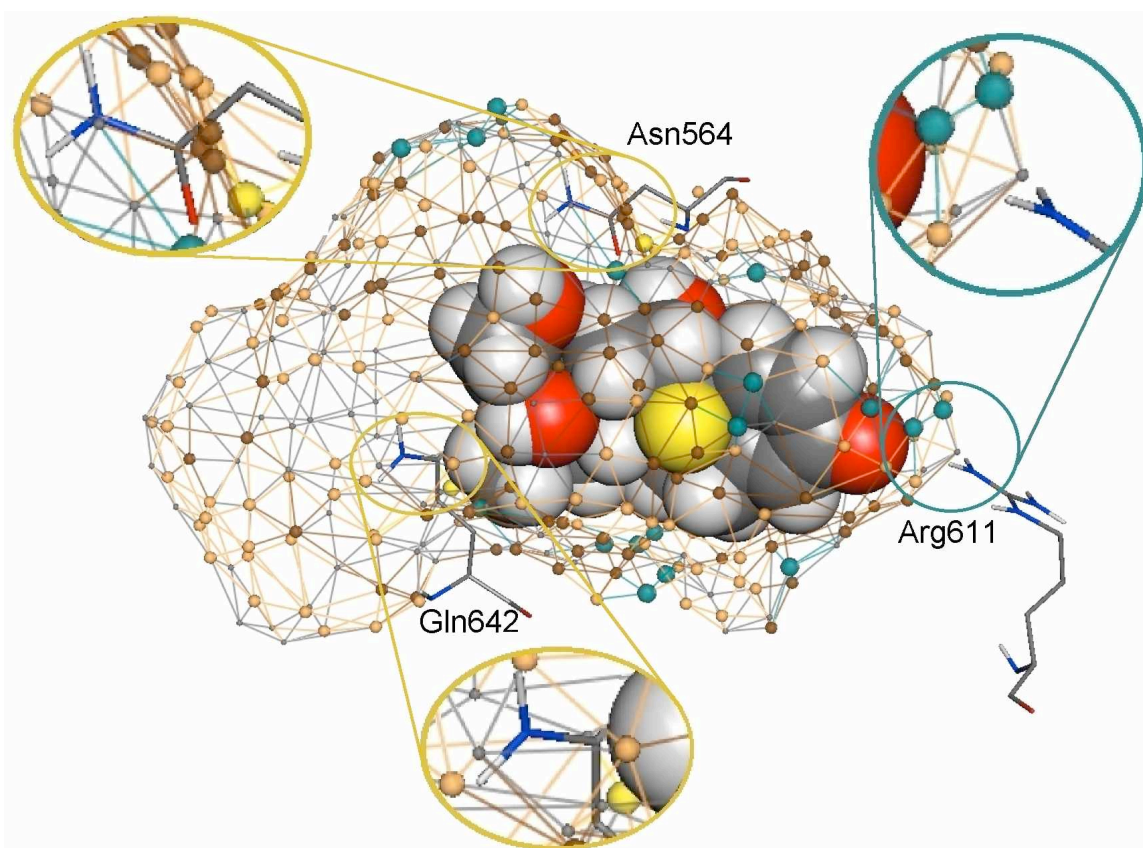
**Figure 24.** Comparison of experimental and predicted binding affinities of the training and test set left and the external test set right towards the GR, as obtained with *Quasar*. Ligands of the training set are depicted as open circles, those of the test set as filled triangles and those of the prediction set as filled circles. Dashed lines are drawn at a factor 10 from the experimental value.

On the average ( $rms$ ), the calculated binding affinity of the training and test ligands deviate by a factor 1.5 and 1.6, respectively, from the experimental  $K_i$  value. The maximal observed deviation of an individual compound corresponds to a factor of 9.9 in  $K_i$  for the training set and 4.9 for the test set, respectively. A representation of the receptor surrogate with bound dexamethasone (**A01**), along with key amino-acid residues is depicted in Figure 25. When compared to the crystal structure of the GR complexed with dexamethasone, the receptor surrogate generated by *Quasar* properly reproduces the corresponding properties at the position of the key amino-acid residues Asn564, Arg611 and Gln642 as identified by X-ray at the quasi-atomistic level: a H-bond donating domain (green) corresponding to the guanidinium moiety of Arg611, and H-bond accepting domains (yellow) mimicking the carbonyl groups of Asn564 and Gln642 at the true biological receptor.

**Table 8.** Summary of the *Quasar* and *Raptor* simulations including 88 training and 22 test compounds

Simulation	$r^2$	$q^2$	$rms$ training	max. training	$p^2$	$rms$ test	max. test
<i>Quasar</i>	0.710	0.702	1.5	9.9	0.719	1.6	4.9
<i>Raptor</i>	0.680	n/a	2.1	5.3	0.519	6.1	20.4

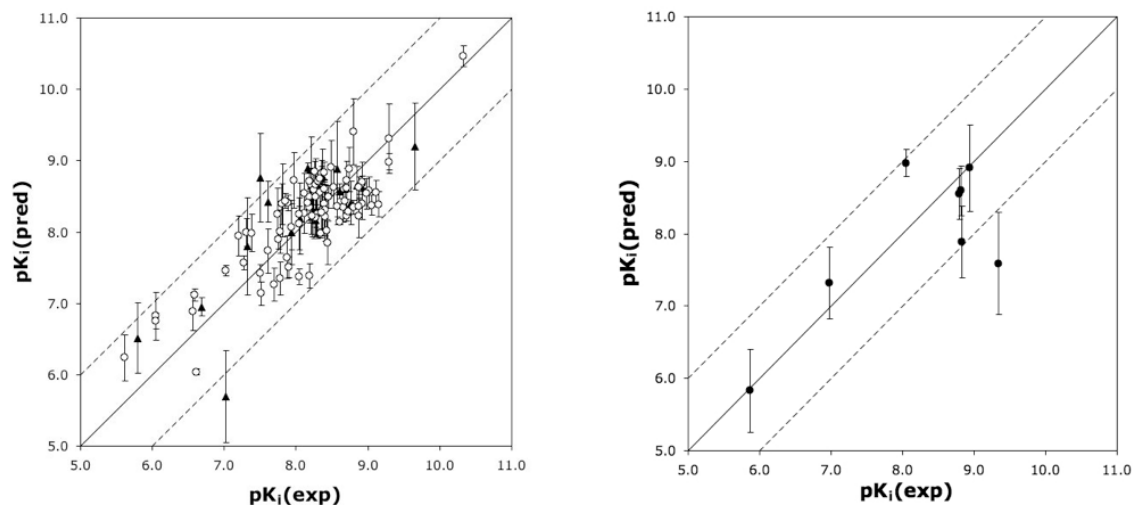
$r^2$ : correlation coefficient,  $q^2$ : cross-validated  $r^2$ ,  $p^2$ : predictive  $r^2$ ; the  $rms$  and maximal deviation from the experimental binding affinity is given as a factor (off) in  $K_i$ .



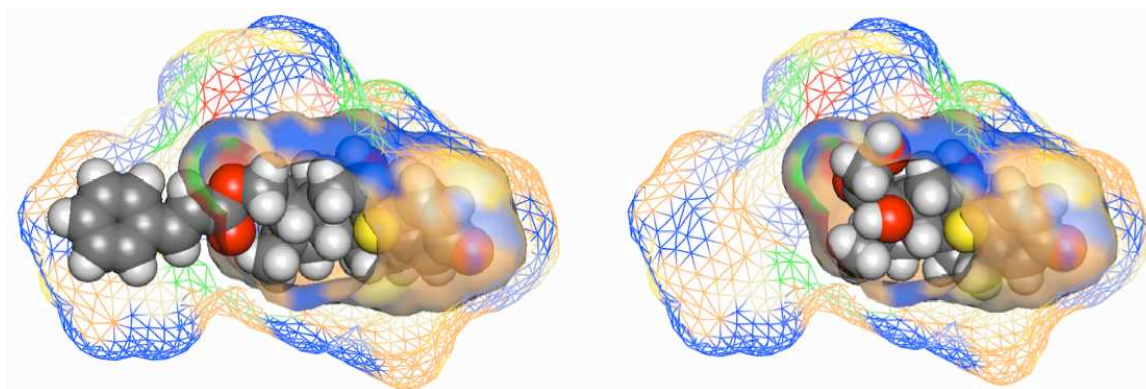
**Figure 25.** Representation of the GR surrogate (*Quasar*) with bound dexamethasone (space-filling), compound code in our study: **A01**. The mapped quasi-atomistic properties are colored as follows: blue (salt bridge, positively charged), red (salt bridge, negatively charged), green (H-bond donor), yellow (H-bond acceptor), saddle brown (hydrophobic, positively charged), chocolate brown (hydrophobic, negatively charged), grey (hydrophobic, neutral). Key amino-acid residues from the X-ray structure (Asn564, Arg611 and Gln642) are shown as sticks. The analogies between model and receptor are encircled and shown in bigger size: green (H-bond donor) corresponding to Arg611, yellow (H-bond acceptor) corresponding to the carbonyl of Gln642, hidden in the back, and to the carbonyl of Asn564.

Particularly, in order to validate a model in multidimensional QSAR (mQSAR), it is of utmost importance<sup>187, 188, 191</sup> to challenge the model, e.g. by means of a scramble test, an external prediction set or by consensus scoring. The scramble test, which is frequently used to assess the sensitivity of a model,<sup>229</sup> consists of a random shuffling of the binding affinities of the training-set ligands with respect to the true affinity values. If, under these circumstances, the ligands of the test set are still predicted correctly, the model is worthless, as it is not sensitive towards the biological data. In our study, 20 scramble tests were performed using a different shuffling of the biological data for each scramble test. They yielded an average predictive  $r^2$  of  $-0.241$  compared to  $+0.719$  for the simulation using unscrambled values. Only a single simulation shows a positive predictive  $r^2$  value ( $0.375$ ), all others found no correlation (predictive  $r^2 < 0.0$ ), demonstrating that the model for the GR is indeed sensitive towards the biological data. Another sensitive issue is the possible overfitting of the model. For this purpose, we stopped the simulation at the maximum value of predictive  $r^2$ . As our *Quasar* model family consists of 200 individual models, the scattering of the individual predictions around the mean value (cf. Figure 24, left) is another indicator for possible overfitting. Although the ligands of the test set show a broader distribution, those of the training set scatter, on the average, a factor 2–3 about their mean value. An overfitted model would result in rather small standard deviation for the ligands of the training set.

To challenge the model, we applied a second methodology (software *Raptor*<sup>153</sup>). The *Raptor* simulation — using the same ligand alignment and selection — yielded an  $r^2$  of  $0.680$  and a predictive  $r^2$  of  $0.519$ . The comparison of predicted and experimental binding affinities is shown in Figure 26, and the performance coefficients are given in Table 8. When compared with *Quasar*, the *Raptor* simulation would seem to yield only a modest predictive power, but considering the limited range of experimental activity (85% of compounds cluster within two orders of magnitude), the compounds' chemical diversity and the different literature source for the affinities, the *Raptor* model can be considered acceptable in terms of quality. Moreover, its dual-shell representation can simulate induced fit more realistically, particularly with compounds substantially different in bulkiness, or in the presence of two induced-fit mechanisms. For example, the cinnamate substituent at position 21 of **A08** (cf. Figure 19) is snugly accommodated by the outer shell (Figure 27, left), indicating the necessity of side-chain rearrangement in the binding pocket of the protein, in order to allocate additional space for the large substituent. On the other hand, the inner shell, hosts compounds such as dexamethasone (**A01**), characterized by a smaller volume (Figure 27, right).



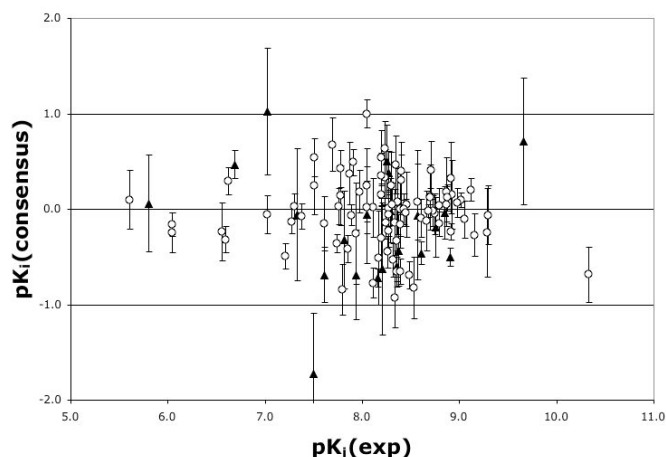
**Figure 26.** Comparison of experimental and predicted binding affinities of the training and test set (left) and the external test set (right) towards the GR as obtained with *Raptor*. Ligands of the training set are depicted as open circles, those of the test set as filled triangles and those of the prediction set as filled circles. Dashed lines are drawn at a factor 10 from the experimental value.



**Figure 27.** *Raptor* model of the GR binding-site model with compound **A08** (left) and dexamethasone (**A01**) (right) bound (hydrophobic fields = beige, hydrogen-bond donating propensity = blue, hydrogen-bond accepting propensity = red, hydrogen-bond flip-flop = green). The inner shell (transparent surface) and outer shell (wire frame) are shown in different style to highlight the two shells of the surrogate (for clarity, the front part of the receptor model has been clipped).

The comparison of predicted binding affinities obtained by the two approaches is shown in Figure 28. The average deviation is 0.32 logarithmic units (a factor 2.1 in  $K_i$ ). For only two compounds out of 110 compounds (1.8%), both of them belonging to the test set, the ratio is greater than a factor of 10 (in  $K_i$ ). Thus, consensus between *Quasar* and *Raptor* predictions has been achieved. While for prednisolone (**A02**), the threshold for acceptance is only slightly exceeded: 1.02 logarithmic units a factor 10.5 in  $K_i$ , the disagreement for 2,5'-dioxo-2'-phenyl-3'-ethoxycarbonyl-spiro[1,4'-acenaphthene-1',4'-dihydroindeno[3,2b]pyridine (**D06**) is 1.73 logarithmic units (a factor 53.7 in  $K_i$ ). Compound **D06** features a phenyl ring

where the other compounds belonging to the D class feature a methyl group. This different feature may have been interpreted in a different way by the two approaches, and this could have led to a different result.



**Figure 28.** Consensus scoring using *Quasar* and *Raptor*. The quantity is expressed as  $pK_i$ ,  $\text{consensus} = -\log(K_{i\text{-}Quasar}/K_{i\text{-}Raptor})$ . Error bars indicate the cumulative standard deviation,  $\text{esd}_{\text{cumulative}} = \sqrt{(\text{esd}_{\text{Quasar}}^2 + \text{esd}_{\text{Raptor}}^2)}$ .

While the test set was not employed for generating and optimizing the model, its performance was, of course, considered as a criterion for selecting the final model among all those generated. To further challenge the model, a new independent set of compounds was identified in the literature,<sup>241</sup> and employed only for this validation step (the compound structures and their  $K_i$  values are given in the Appendix A).

The chemical property domain of the prediction set (**E01–E08**) is comprised in the model's property space, because the scaffold of the new compounds is already represented in the training set, and because the activity range is within the broader range of the activities of the training set, allowing therefore a reliable prediction instead of an extrapolation of binding affinities. The  $K_i$  values of the external compounds (as predicted by *Quasar*) are shown in Figure 24 (right). The predictive  $r^2$  is 0.538, the *rms* deviation for this external set corresponds to a factor of 4.7 off in  $K_i$ , and the maximal deviation is 17.9 — an appreciable result considering that these ligands were not used for model construction and selection. The  $K_i$  values predicted by *Raptor* are shown in Figure 26 (right). The predictive  $r^2$  is 0.488, the *rms* deviation for this external set corresponds to a factor of 20.3 in  $K_i$ , and the maximal deviation is 56.4.

#### 4.1.6 Binding of Psychotropic Drugs to the GR

To assess the validity of the approach, we have simulated the binding of 24 psychotropic drugs to the GR. For this task we employed the protocol as implemented in the *VirtualToxLab*<sup>TM3</sup> the protocol of which (flowchart in Figure 9) includes a full conformational search in aqueous solution, the identification of the most probable protonation and tautomeric state at physiological pH, followed by automated, flexible docking and calculation of the binding affinity using 6D-QSAR.<sup>177</sup> The resulting binding affinities are given in Table 9. Because our GR model was trained using almost exclusively neutral species (87:1), we calculated the binding affinities of the psychotropic drugs for both the neutral and charged state (where applicable) and observed that the charged species — corresponding to the protonation state in aqueous solution at pH 7.4 — typically yield higher affinities. This might, however, represent an artifact caused by the electrostatic contribution to the protein–ligand interaction. It is well known, of course, that the dielectric properties in the interior of a receptor may differ significantly from those in aqueous solution. Six compounds — bupropion, fluoxetine, lorazepam, methylphenidate, trimipramine and venlafaxine — are marketed as racemic mixtures. Here, we simulated both stereoisomers (cf. Table 9). As an example, lorazepam is discussed in detail (Figure 29). Although the R-isomer engages in hydrogen bonds with the GR, it features a nine-fold weaker activity (cf. Table 10). This is a consequence of the hydrophobic ligand–protein interactions, which are more pronounced for the S-isomer. Table 10 lists the scaled interaction energies as obtained from the mQSAR simulation (*Quasar*): the S-isomer is characterized by more favorable electrostatic and van der Waals energies.

Figure 30 (top) reveals details of the binding of clomipramine to the GR. In the most contributing pose (a total of eight were considered in the mQSAR), the proton of the dimethylammonium group points towards its own phenyl moiety, thus forming a charge- $\pi$  interaction. Another pose found in the conformational search (not shown here) sees the very H atom engaging in a weak hydrogen bond with the nearby Asn564 residue (highlighted in Figure 30). Likewise, the ammonium H atom of mirtazapine engages in a hydrogen bond with Asn564 (Figure 30: center). For other tested compounds, however, e.g. buspirone, fluoxetin, trimipramin and trazodone (Figure 30: bottom), no such possibility exists and binding in the neutral state would make more sense. As our GR model was implemented in the *VirtualToxLab*<sup>TM</sup>, we calculated the binding affinity towards all 12 models (androgen, aryl hydrocarbon, estrogen  $\alpha/\beta$ , glucocorticoid, liver X, mineralocorticoid, peroxisome proliferator-activated  $\gamma$  and thyroid  $\alpha/\beta$  receptors as well as for the enzymes cytochrome P450 2A13 and 3A4). The results, along with those of 1500+ other compounds, are accessible through the Internet.<sup>3</sup>

A wealth of information on the side effects of psychotropic drugs is published; summaries can be found, for example, in Wikipedia.<sup>242</sup> Some of the compounds analyzed in this study trigger adverse effects via the GR. Those include

alprazolam (increased glucocorticoid levels),<sup>243</sup> amitriptyline (induction of GR-mRNA),<sup>244</sup> clomipramine (regulates GR expression),<sup>245</sup> desipramine (GR translocation, increased GR-mRNA),<sup>246</sup> escitalopram (reversion of GR immunoreactivity),<sup>246</sup> fluoxetine (enhances GR function),<sup>245, 247</sup> imipramine (partial agonist-like effects),<sup>248</sup> moclobemide (increased GR-mRNA),<sup>247</sup> reboxetine (increased cortisol levels),<sup>249</sup> sertraline (high GR responsiveness).<sup>250</sup> For other compounds, no or less clear effects were observed.<sup>251, 252</sup> All this suggests, that our computational protocol as implemented in the *VirtualToxLab*<sup>TM</sup> might well be suited for identifying such effects *in silico*.

**Table 9.** Predicted binding affinities for 24 psychotropic drugs towards the GR

<i>Compound</i>	<i>Formal charge</i>	<i>Calculated binding affinity</i>	<i>Relative binding affinity</i> <sup>1</sup>
Alprazolam	neutral	$1.8 \times 10^{-8}$ M	0.16
Amitriptyline	neutral +1	$2.9 \times 10^{-7}$ M $6.5 \times 10^{-9}$ M	0.010 0.46
Bupropion (R/S)	neutral +1	$8.6 \times 10^{-7}$ M / $1.7 \times 10^{-6}$ M $3.6 \times 10^{-8}$ M / $2.0 \times 10^{-8}$ M	0.0034 / 0.0018 0.082 / 0.15
Buspirone	neutral +1	$1.3 \times 10^{-6}$ M $1.3 \times 10^{-10}$ M	0.0023 23
Clomipramine	neutral +1	$5.0 \times 10^{-9}$ M $2.3 \times 10^{-9}$ M	0.59 1.3
Desipramine	neutral +1	$2.7 \times 10^{-8}$ M $8.2 \times 10^{-9}$ M	0.11 0.36
Doxepin	neutral +1	$1.1 \times 10^{-8}$ M $2.3 \times 10^{-8}$ M	0.27 0.13
Duloxetine	neutral +1	$1.7 \times 10^{-6}$ M $6.1 \times 10^{-9}$ M	0.0017 0.49
Escitalopram	neutral +1	$5.1 \times 10^{-8}$ M $4.8 \times 10^{-9}$ M	0.058 0.62
Flunitrazepam	neutral	$6.9 \times 10^{-8}$ M	0.043
Fluoxetine (R/S)	neutral +1	$1.8 \times 10^{-6}$ M / $4.7 \times 10^{-7}$ M $2.7 \times 10^{-10}$ M / $1.7 \times 10^{-8}$ M	0.0016/ 0.0063 11 / 0.17
Imipramine	neutral +1	$6.6 \times 10^{-7}$ M $1.5 \times 10^{-9}$ M	0.0045 2.0
Lorazepam (R/S)	neutral	$8.3 \times 10^{-8}$ M / $9.1 \times 10^{-9}$ M	0.036 / 0.33
Methylphenidate (R/S)	neutral +1	$3.9 \times 10^{-6}$ M / $1.0 \times 10^{-3}$ M $1.1 \times 10^{-7}$ M / $1.2 \times 10^{-8}$ M	0.00076/ 0.0030 0.027 / 2.5
Mirtazapine	neutral +1	$5.5 \times 10^{-7}$ M $2.0 \times 10^{-10}$ M	0.0054 15
Moclobemide	neutral	$8.4 \times 10^{-7}$ M	0.0035
Modafinil	neutral	$1.2 \times 10^{-8}$ M	0.25

**Table 9.** Continued from previous page.

Compound	Formal charge	Calculated binding affinity	Relative binding affinity <sup>1</sup>
Nortriptyline	neutral	$3.5 \times 10^{-7}$ M	0.0035
	+1	$8.3 \times 10^{-9}$ M	0.36
Paroxetine	neutral	$1.4 \times 10^{-8}$ M	0.21
	+1	$6.8 \times 10^{-10}$ M	4.4
Reboxetine	neutral	$5.1 \times 10^{-8}$ M	0.058
	+1	$4.8 \times 10^{-9}$ M	0.62
Sertraline	neutral	$9.0 \times 10^{-8}$ M	0.033
	+1	$1.6 \times 10^{-9}$ M	1.9
Trazodone	neutral	$2.8 \times 10^{-7}$ M	0.011
	+1	$6.4 \times 10^{-10}$ M	4.6
Trimipramine (R/S)	neutral	$4.6 \times 10^{-7}$ M / $1.1 \times 10^{-8}$ M	0.0064 / 0.27
	+1	$4.4 \times 10^{-10}$ M / $1.1 \times 10^{-8}$ M	6.7 / 0.27
Venlafaxine (R/S)	neutral	$6.6 \times 10^{-7}$ M / $1.1 \times 10^{-6}$ M	0.0045 / 0.0027
	+1	$7.3 \times 10^{-9}$ M / $1.3 \times 10^{-8}$ M	0.41 / 0.23

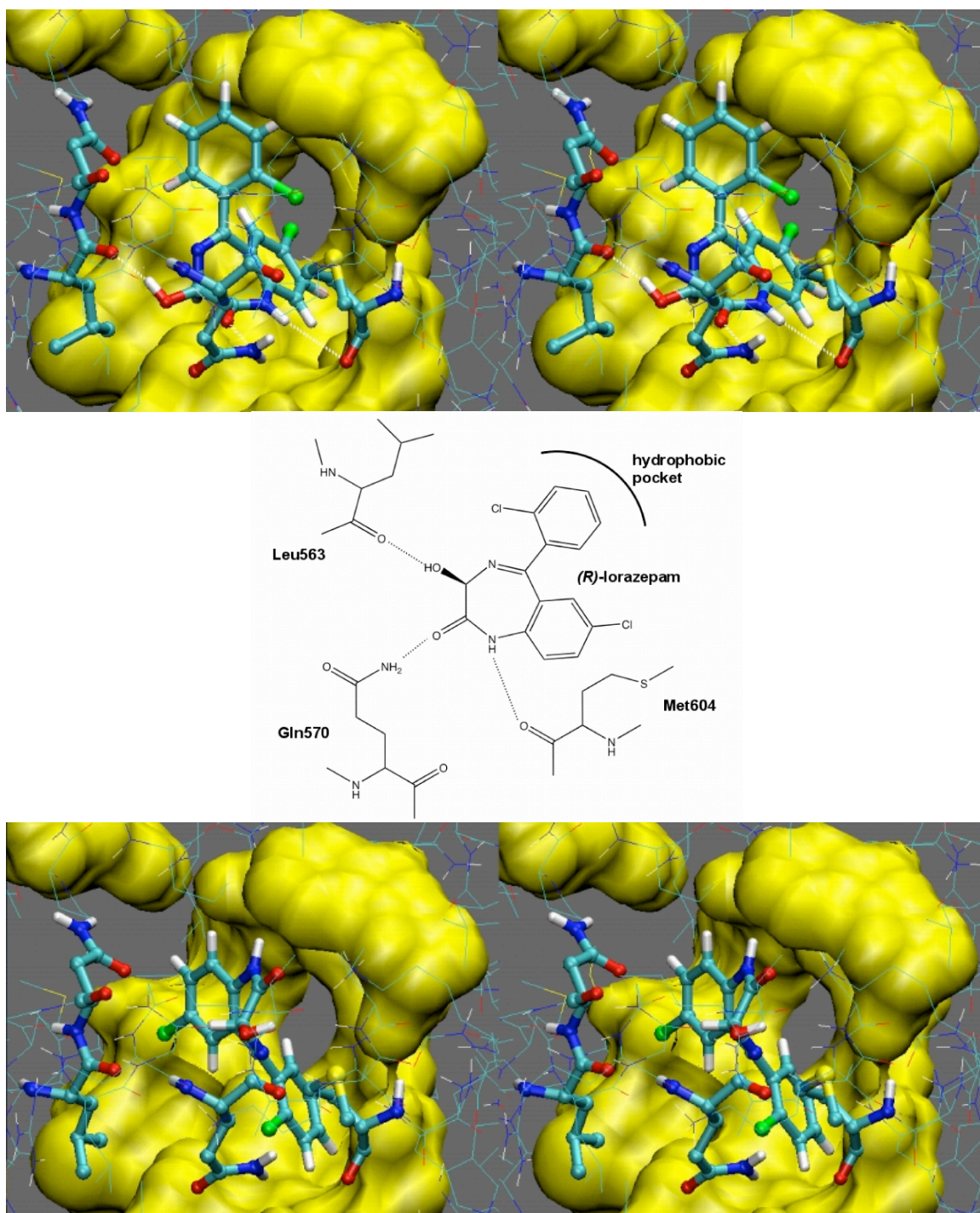
<sup>1</sup> Relative binding affinity with dexamethasone (**A01**) serving as reference compound:  $K_i$  (exp.) =  $2.963 \times 10^{-9}$  M  $\rightarrow$  RBA = 1.0. RBA > 1.0 indicates a more active compound than dexamethasone; RBA < 1.0 marks entities less active than dexamethasone.

**Table 10.** Quantitative aspects of (R/S)-lorazepam binding to the GR.

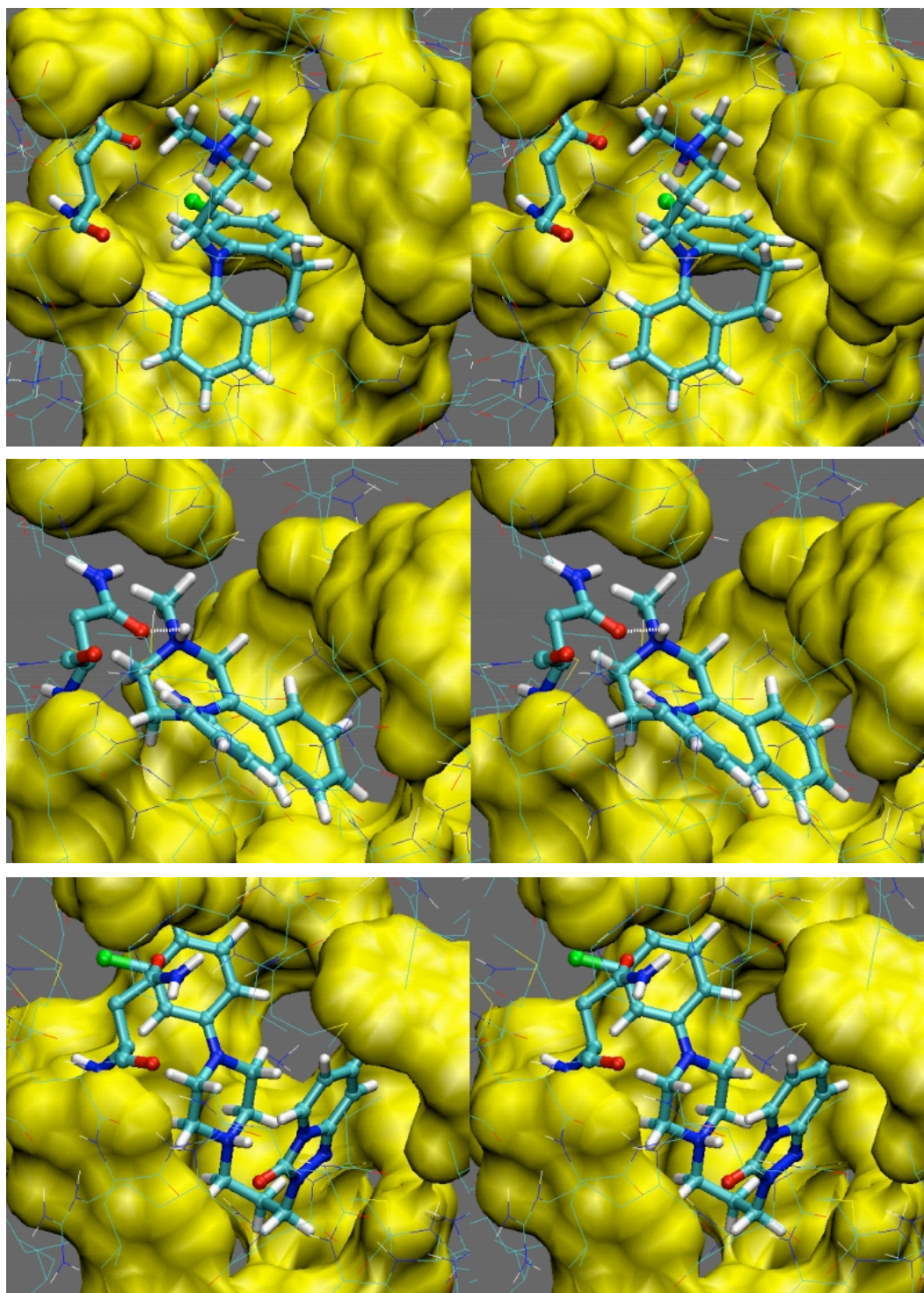
Isomer	$E_{PL}$	$E_{Elec}$	$E_{vdW}$	$E_{HB}$	$E_{Pol}$	$E_{Desolv}$	$T\Delta S$	$E_{int}$	$E_{IndFit}$	$E_{Bdg}$	$K_i$
R	-14.3	-6.3	-2.1	-2.5	-3.4	+3.6	+0.5	+0.4	+0.3	-9.5	83 nM
S	-15.2	-6.7	-4.4	-0.0	-4.1	+3.6	+0.5	+0.1	+0.2	-10.8	9.1 nM

All energies are given in kcal/mol.  $E_{PL}$ : ligand–protein interaction energy;  $E_{Elec}/E_{vdW}/E_{HB}/E_{Pol}$ : electrostatic, van der Waals, hydrogen bonding and polarization components;  $E_{Desolv}$ : ligand desolvation;  $T\Delta S$ : entropy;  $E_{int}$ : internal strain;  $E_{IndFit}$ : induced fit;  $E_{Bdg}$ : calculated binding energy;  $K_i$ : calculated binding affinity (cf. equation 2).





**Figure 29.** Details of the binding mode of lorazepam (top and bottom: stereo view). Lorazepam, Leu 563 and Asn564 are colored by atom type, the residues lining the hydrophobic pocket are depicted in yellow. Top: R-lorazepam engages in a strong hydrogen bond with Leu563 and two weak interactions with Gln570 and Met604, respectively (dashed lines). Center: Schematic representation of the hydrogen-bonding pattern associated with the binding of R-lorazepam to the GR. Bottom: S-lorazepam does not engage in hydrogenbonds but features stonger hydrophobic interactions.



**Figure 30.** Details of the binding mode of three psychotropic drugs to the GR (stereo view). The ligands and Asn564 are colored by atom type, the residues lining the hydrophobic pocket are depicted in yellow. Clomipramine (top): the ammonium H atom does not engage in a hydrogen bond with the GR. Mirtazapine (center): the hydrogen bond the ammonium H atom with Asn564 (white dashed line) is too long (2.37 Å) and not linear (127°) and, therefore, very weak. Trazodone (bottom): again, the ammonium H atom cannot engage in a hydrogen bond with the receptor.

#### 4.1.7 Conclusions and Applicability of the GR Model

A QSAR model for a series of 118 ligands of the GR (88 training set and 22 test set, belonging to 4 different chemical classes, and 8 prediction set compounds) was built with *Quasar* and successfully validated. The best *Quasar* simulation was obtained with an alignment from the manual docking and yielded a cross-validated  $r^2$  of 0.702 for the training set and a predictive  $r^2$  of 0.719 for the internal test set, respectively.

Robustness was proved by applying a second methodology (*Raptor*) and verifying the agreement on the predictions. The *Raptor* simulation yielded a cross-validated  $r^2$  of 0.680 for the training set and a predictive  $r^2$  of 0.519 for the internal test set, respectively. For 98.2% of the compounds the predictions by the two methodologies are in agreement (the ratio being below a factor of 10, in  $K_i$ ).

The predictivity was evaluated by 8 new compounds: the predictive  $r^2$  is 0.538 with *Quasar*, and 0.488 with *Raptor* (the *rms* deviation is a factor in  $K_i$  of 4.7 and 20 with the two methodologies, respectively).

Model sensitivity was assessed by 20 scramble tests: they gave an average predictive  $r^2$  value of  $-0.241$ , demonstrating that the model for the GR is indeed sensitive towards the biological data.

The model was then used for simulating and quantifying the binding of 24 psychotropic drugs to the GR.

The results suggest that the model, can be applied to predict the binding affinity of new drug candidates, in order to design new glucocorticoids, or of existing compounds, in order to check out possible interactions. The model is, at least in part, interpretable and correlated with structural properties. In contrast to other modeling studies on the GR<sup>217, 253</sup> induced fit, a key mechanism for ligand binding, was explicitly simulated both in the docking phase and during the QSAR simulations.

Limitations of the current model for the GR are the sensitivity to the formal charge and the molecular weight of the compounds to be tested. In the former instance, charged species would currently seem to be overestimated as the model was trained using predominantly neutral species. The size of a compound matters as for very small ligands the automated docking protocol might not sample enough poses (default = 25) while for large molecules leading to a significant induced-fit ( $rms > 5 \text{ \AA}$ ), the underlying protocol is unable to simulate and quantify such large conformational changes at the protein.

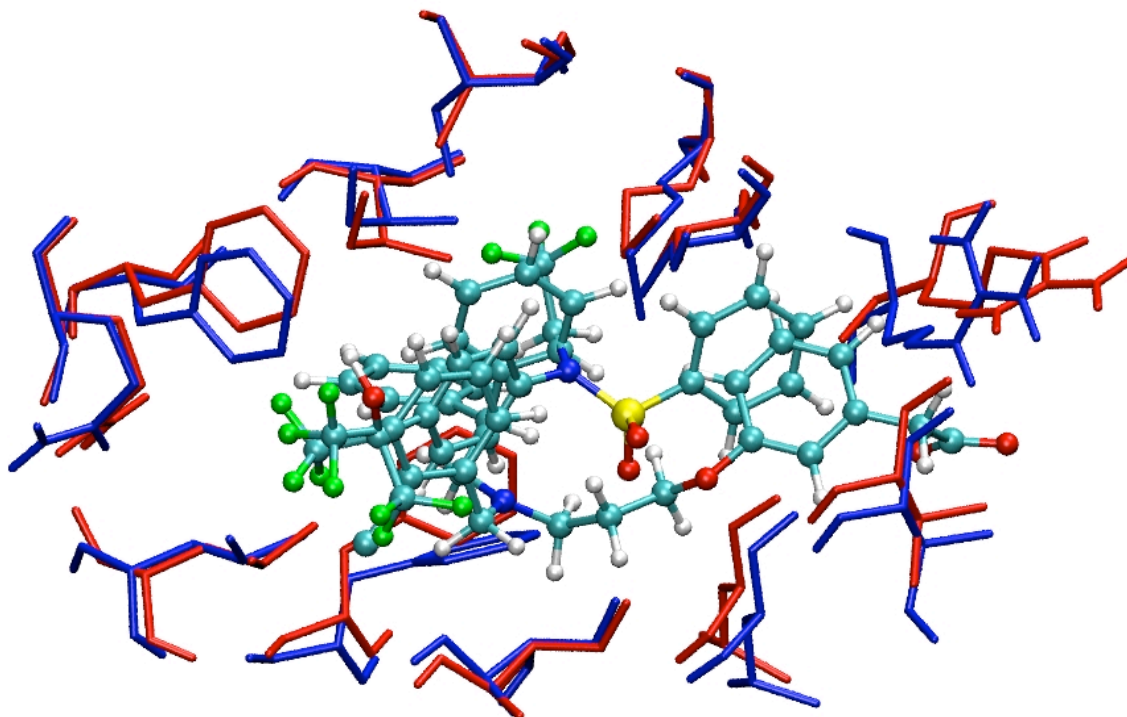
The GR model has been added to the *VirtualToxLab*<sup>TM</sup> developed by the Biophysics Laboratory 3R.<sup>4, 254</sup> Presently, it includes 12 validated models for the androgen, aryl hydrocarbon, estrogen  $\alpha/\beta$ , glucocorticoid, liver X, mineralocorticoid, peroxisome proliferator-activated  $\gamma$  and thyroid  $\alpha/\beta$ , receptors as well as for the enzymes cytochrome P450 2A13 and 3A4.

## 4.2 Liver X Receptor (LXR)

As for the glucocorticoid receptor, the following sections describe the work on the liver X receptor (LXR): the analysis on the crystal structures, the MD studies, the preparation of the binding affinity data and compound structures, the docking to the receptor, and the building and validation of a QSAR model.

### 4.2.1 Analysis of the LXR Crystal Structures and Receptor Preparation

Two isoforms of LXR have been identified and are referred to as LXR $\alpha$  and LXR $\beta$ . The sequence identity in the ligand binding domains between the two isoforms is high (77%). The amino-acid residues lining the binding domain that differ in the two isoforms are either located at least 5 Å away from the bound ligands or conservative in nature (e.g. Val263 in LXR $\alpha$  Ile277) in LXR $\beta$  and in any case all amino-acid residues that interact with the ligands are conserved (Figure 31).



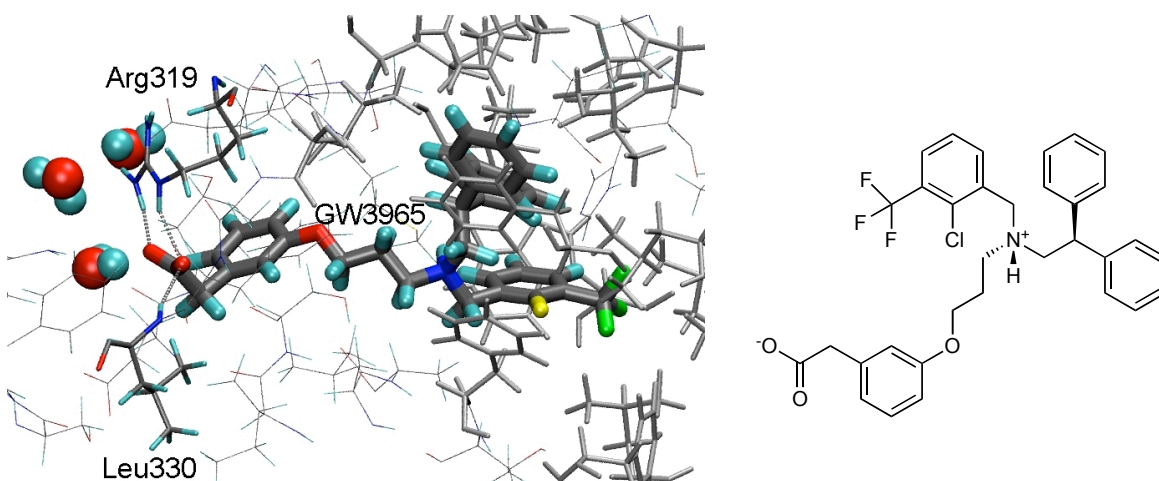
**Figure 31.** Superposition of the amino-acid residues lining the binding pocket of LXR $\alpha$  (PDB code: 1UHL) colored in red, and LXR $\beta$  (PDB code: 1PQ6) colored in blue. The co-crystallized ligands are also shown, colored by atom type. For clarity, some of the amino-acid residues are not shown. The figure demonstrates the high similarity of structure of the binding pocket between the  $\alpha$  and  $\beta$  isoforms.

Due to the high similarity of the two isoforms, and to the non-selectivity of the studied compounds (discussed in the section 4.2.3), the crystal structure of LXR $\beta$  (PDB code: 1PQ6, 2.40 Å resolution,  $R_{\text{free}} = 0.262$ ) was selected as a “representative receptor structure” of the LXRs for the docking of the compound library. The  $\beta$  isoform was chosen because of the higher resolution of the crystal structure for the human receptor (2.40 Å for the  $\beta$  isoform compared to 2.90 Å for the highest resolution for a human structure of LXR $\alpha$ ). Among the available human LXR $\beta$  crystal structures, 1PQ6 was chosen because of the similarity of the co-crystallized ligand with compounds of the studied library. From now on, the term LXR refers to the LXR $\beta$  isoform.

As described in paragraph 3.3.3, discontinuous parts of the crystal structure were modeled on the basis of other crystal structures available (Phe243–Gln246 from 1PQ9,<sup>83</sup> and Leu254–Pro258 from 1UPV<sup>227</sup>) and the whole structure was minimized with *Macromodel*. The complete minimized structure was then used as input for the MD studies.

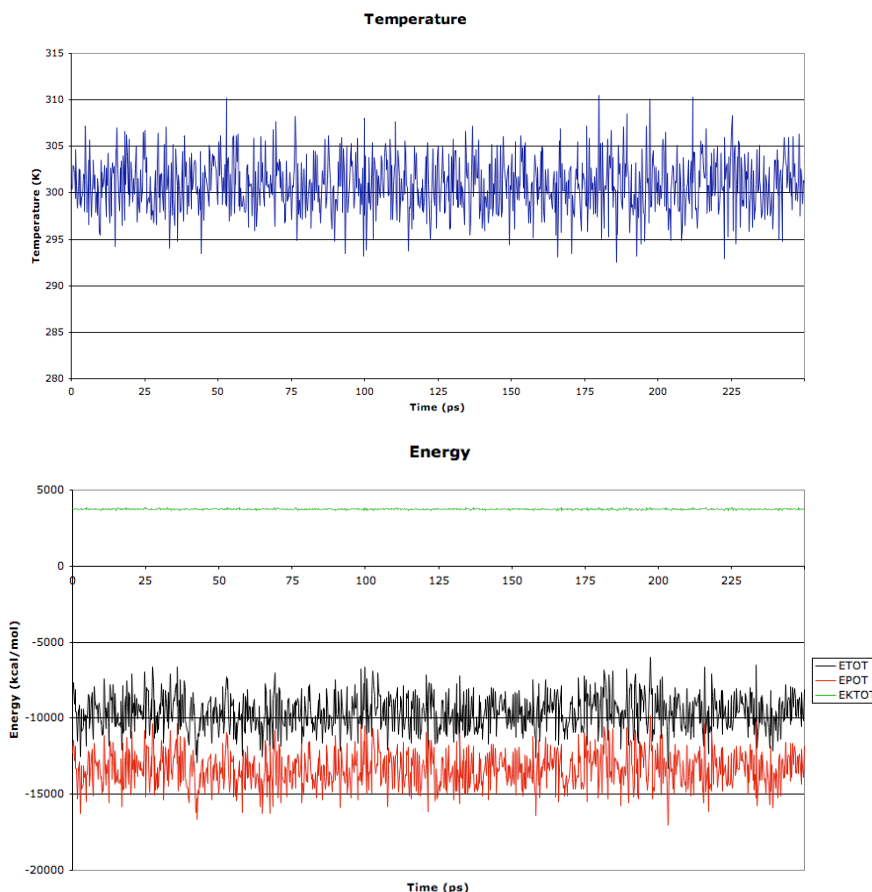
#### 4.2.2 Molecular-Dynamics Simulations on the LXR

Due to the addition of nine amino-acid residues not resolved in the LXR crystal structure (Phe243–Gln246 and Leu254–Pro258), MD studies were performed for the LXR not only to gain insight on the binding mode and on the dynamic behavior of the complex, but also to obtain a receptor structure further relaxed by MD simulations. GW3965, the co-crystallized ligand (depicted in Figure 32), was studied in the MD simulations.



**Figure 32.** Depiction of the frame from the MD (left), corresponding to the minimum energy structure. Ligand GW3965 is represented as thick sticks (chemical structure shown on the right), key amino-acid residues involved in hydrogen bonding or hydrophobic interactions are represented as thin sticks, the remaining amino-acid residues in the binding pocket as lines and the water molecules are represented as spheres. Hydrophobic amino-acid residues are colored in gray.

The plot of temperature and energetic contributions during the production phase is depicted in Figure 33: the fluctuations don't show any drift and the average value of temperature and energies remains constant. It can be assumed that the simulation reached equilibrium, being this constant behavior a prerequisite for a well equilibrated dynamic.



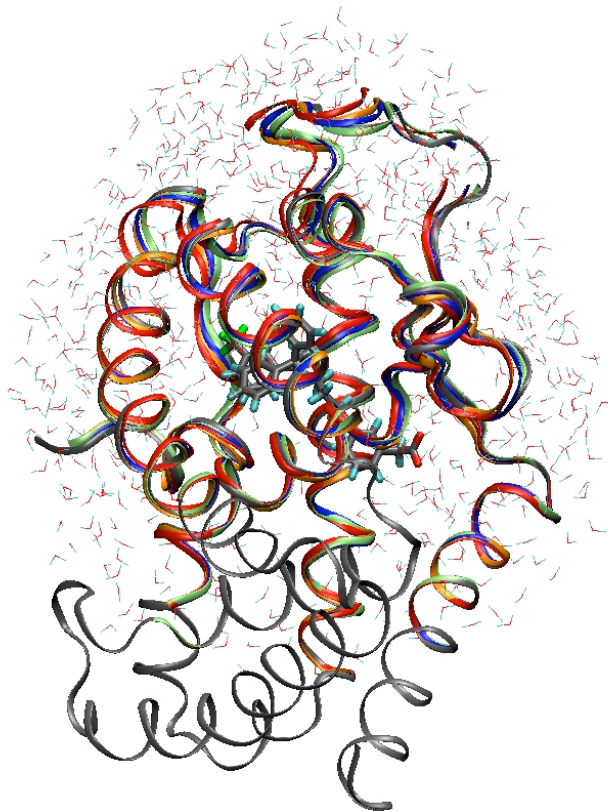
**Figure 33.** Plot of the temperature (top) and energy contributions (bottom) during the production phase (250 ps). Temperature and energies fluctuate around a constant value, indicating that the simulation is stable and has reached equilibrium. ETOT: total energy; EPOT: potential energy; EKTOT: kinetic energy.

The average structure and the whole trajectory during the production phase were observed. If a distance cutoff ( $< 3.5 \text{ \AA}$  between heavy donor–acceptor atoms and angle cutoff donor-hydrogen-acceptor angle  $> 120^\circ$ ) are set as requirement for hydrogen bond identification, then two hydrogen bonds are observed between the carboxylic group of the ligand molecule and the amino-acid residue Arg319 side-chain. A third hydrogen bond is observed between the ligand and Leu330 backbone. In particular, one of the two carboxylic oxygen atoms of GW3965 is interacting with Arg319 (100% of the simulation time, average oxygen–hydrogen distance  $1.81 \text{ \AA}$ ). The other carboxylic-oxygen atom is also interacting with Arg319 (93% of the time, average distance  $2.18 \text{ \AA}$ ) but also with the backbone of Leu330 (73% of the time, average distance  $2.35 \text{ \AA}$ ). The distance and angle

cutoff anyway don't take into account of the position of the oxygen lone pairs. Furthermore, we observe that the plane identified by the guanidinium group is often nearly perpendicular to the carboxylate plane (the average angle between the planes is  $65^\circ$ ). This angle is at the upper limit of the observed range for hydrogen bonds in nature.<sup>255-258</sup> The interaction between GW3965 and Arg319, lacking the directional component defining a strong hydrogen bond,<sup>256</sup> can then be considered most of all electrostatic in nature.

The major part of the ligand molecule is however hydrophobic in nature and is engaged in hydrophobic interactions. The binding pocket features two major hydrophobic cavities, one occupied by the chloro-trifluoromethyl benzyl group and the second one around the diphenylethyl moiety of GW3965. The number of hydrophobic amino-acid residues (Figure 32) in the area is very high, and many of them are involved in hydrophobic interactions: Phe268, Phe271, Leu274, Ala275, Ile309, Met312, Leu313, Phe329, Phe340, Leu345, Phe349, Ile350, Ile353, Phe354, Leu442, Val439, Leu449 and Trp457 are within 4 Å of the ligand. The frame corresponding to the minimum energy structure of the simulation is shown in Figure 32.

The minimum energy structure during the MD trajectory, occurring after 203 ps from the beginning of the production phase, was extracted from the trajectory, minimized, and used as input structure for the following docking studies. Figure 34 shows a superposition of the minimum energy structure (gray), the average structure (blue), and three conformations randomly chosen during the dynamics (orange, red, gray). All the structures are quite similar and don't change much during the simulation. This behavior can be deduced also from the low average of the *RMSD* of the backbone (0.51 Å).



**Figure 34.** Superposition of the minimum energy structure (gray), average structure (blue), and three conformations randomly chosen during the dynamics (orange, red, green). Proteins are represented as ribbons, the ligand as sticks and water molecules as lines. Only the minimum energy structure is shown entirely, while for the other conformations is shown only the substructure that during the MD was not constrained (like described in paragraph 3.3.4, the residue not constrained are the ones within 20 Å from the ligand).

### 4.2.3 Preparation of the Binding Affinity Data and Compound Structures

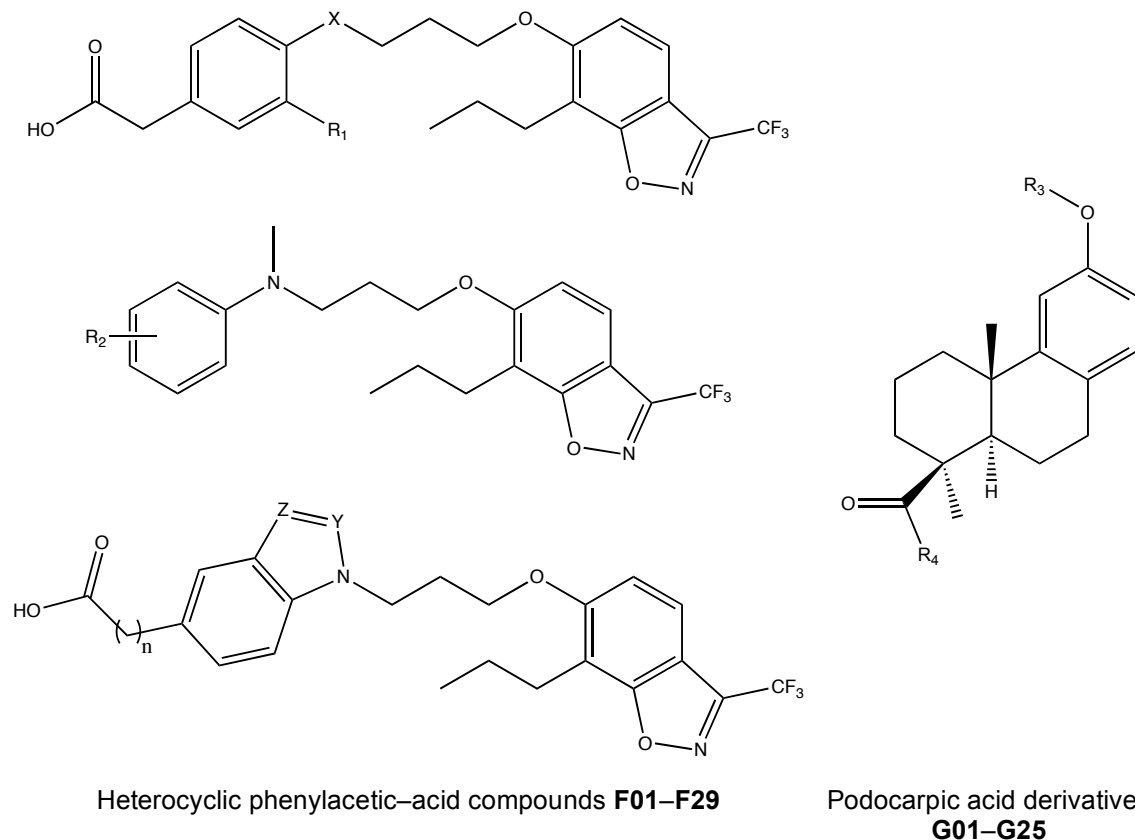
The structural information and pharmacological data for 54 compounds binding to the LXRs were obtained from literature.<sup>259</sup>

The compounds can be classified in two chemical classes: heterocyclic phenylacetic-acid compounds (compound code **F01–F29**) and compounds derived from podocarpic acid (compound code **G01–G25**). Their chemical structure is shown in Figure 35.

Two compounds of the class G (**G10** and **G16**) bear stereocenters, and in the biological assay were tested as racemates. From docking studies emerged the fact that for both compounds, both stereoisomers can be equally well accommodated in the binding pocket, with no clear preference for one stereo-



mer. Therefore both isomers for both compounds were modeled in the docking studies and in the QSAR.



**Figure 35.** The three different possibilities for the scaffold of compounds in the class F (left), and the scaffold of the compounds in the class G (right) are shown. All the compound structures and  $K_i$  values are given in the appendix B.

$R_1 = \text{H, Cl};$

$X = \text{S, O, CH}_2, \text{NMe, NH, NEt, N-Formyl, NAc};$

$R_2 = \text{H, } m\text{-CH}_2\text{CO}_2\text{H, } p\text{-CO}_2\text{H};$

$Y = \text{C, CO, N};$

$Z = \text{C, CH}_2, \text{N};$

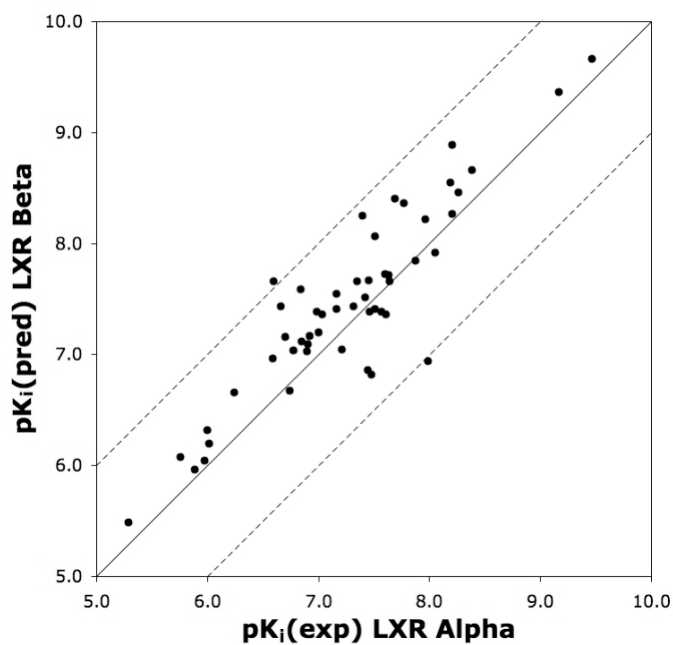
$R_3 = \text{H, COCH}_3;$

$R_4 = \text{adamantane derivatives, bicycloheptane derivatives, different substituents derived by 5-, 6-, or 7-membered rings, aromatic or aliphatic.}$

All measured  $\text{IC}_{50}$  values were converted to  $K_i$  values by means of the Cheng-Prusoff relation.<sup>223</sup> Two compounds belonging to the F class (**F09** and **F10**) are weak actives for  $\text{LXR}\beta$  (<50% activation at 10  $\mu\text{M}$ ) and could therefore not be used in the model development with *Quasar*, for which a finite value of binding affinity is required, instead they were used as a predictive set during the model validation (discussed in the paragraph 4.2.5). Figure 36 compares the experimental binding affinities for the 52 compounds used in this study towards the  $\text{LXR}\alpha$  and  $\beta$ . Only for two compounds (3.8%) the difference in binding affinity exceeds 1 logarithmic unit. The selectivity of the compounds for the two subtypes of LXR is therefore not sufficient to identify specific models for each subtype: the

error in prediction would be comparable to the difference in experimental values between the two subtypes. Therefore a representative model has been built and validated on a single LXR isoform: such a LXR $\beta$  model may be used to predict the binding affinity value for both subtypes.

The ligand structures were generated and optimized as indicated in paragraph 3.3.2.

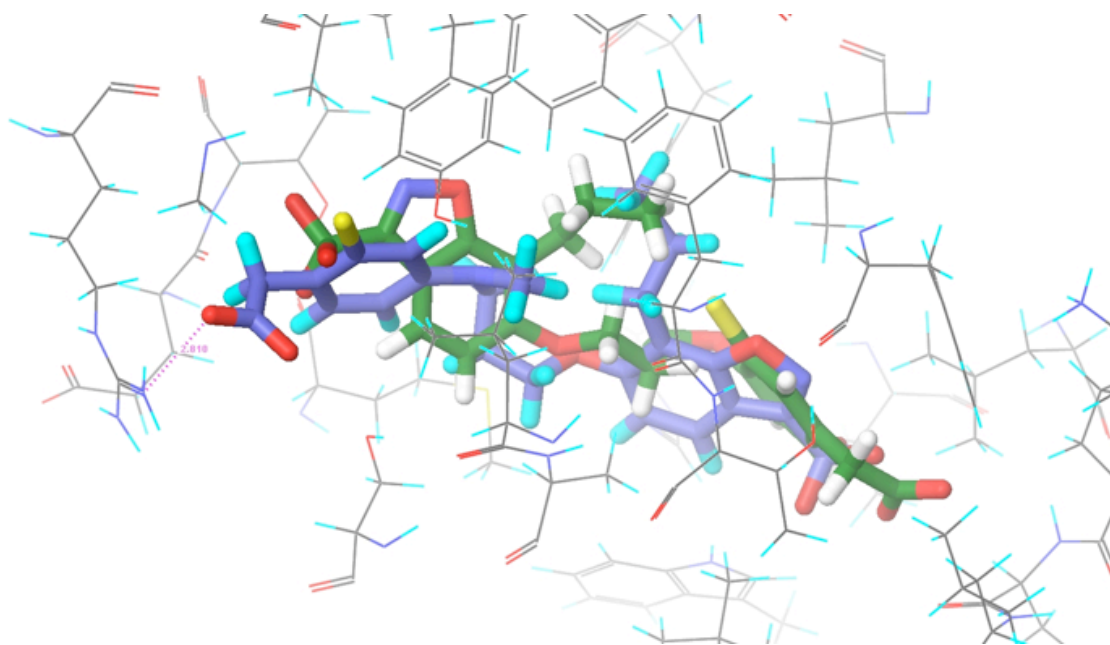


**Figure 36.** Comparison of binding affinities towards LXR $\alpha$  and  $\beta$  for the 52 investigated compounds

#### 4.2.4 Docking to the LXR

In order to have a suitable input structure for the docking of the 52 ligands, the position of the hydrogen atoms was re-calculated with *Yeti*,<sup>99, 101</sup> and added to the structure obtained from the molecular-dynamics simulation. The correct protonation state for the histidine residues was then assigned: His270, His322, His341, His435 and His460 are all protonated at their N<sub>δ</sub> atom. His435 is part of the binding pocket and the protonation state depend on the ligand bound that may present in the vicinity of the histidine residue hydrogen-bond donors, acceptors or groups that are not able to form hydrogen bonds. Different functional groups of the bound ligand may induce different protonation states for His435. For the LXR bound to the ligand GW3965, as well as to the other LXR ligands studied in this thesis, the most probable protonation state would be the N<sub>δ</sub> atom. After the assignment of the protonation state, the hydrogen-bond network was optimized and the structure was relaxed using the directional force field implemented in *Yeti*.<sup>99</sup>

Automated docking using different software was initially performed for each of the 52 ligands, with the automated docking module of *Yeti*,<sup>99</sup> with *Glide*,<sup>104</sup> and *eHiTS*<sup>109</sup> but all of them led to a less satisfactory result than the manual docking because of the higher protein–ligand interaction energies (when compared to the results of the manual docking, see Table 11), and the non-homogeneous alignment in the output: similar molecules resulted in different binding poses, as it is shown in Figure 37.



**Figure 37.** Comparison of results obtained with the automated docking procedure performed with *Glide* for two similar compounds: **F24** (carbon atoms colored in blue) and **F02** (carbon atoms colored in green). The two similar compounds result in very different binding poses, almost flipped 180 degrees on the horizontal axis.

**Table 11.** Quantitative aspects of ligand binding to the LXR, as obtained from the manual and automatic docking (*Yeti* software).

Ligand	Docking	$E_{\text{Protein-Ligand}}$	$(E_{\text{Elec}}$	$E_{\text{vdW}}$	$E_{\text{HB}})$
F01a	M	-57.9	-27.5	-28.6	-1.7
F01b	M	-60.1	-31.5	-25.7	-3.1
F01c	M	-59.5	-32.1	-25.8	-1.6
F01d	M	-58.6	-29.6	-27.4	-1.6
<b>F01 average</b>	<b>M</b>	<b>-59.0</b>	<b>-30.2</b>	<b>-26.9</b>	<b>-2.0</b>
F01a	A	-41.9	-10.6	-31.4	-0.0
F01b	A	-46.5	-13.0	-31.6	-1.9
F01c	A	-40.9	-15.8	-25.1	-0.0
F01d	A	-39.7	-16.8	-21.5	-1.4
<b>F01 average</b>	<b>A</b>	<b>-42.3</b>	<b>-14.1</b>	<b>-27.4</b>	<b>-0.8</b>
F13a	M	-53.4	-26.6	-24.7	-2.1
F13b	M	-53.7	-25.2	-28.2	-0.2
<b>F13 average</b>	<b>M</b>	<b>-53.6</b>	<b>-25.9</b>	<b>-26.5</b>	<b>-1.2</b>
F13a	A	-50.8	-26.5	-22.3	-2.1
F13b	A	-51.4	-26.2	-22.2	-2.9
F13c	A	-44.6	-18.5	-26.1	-0.0
F13d	A	-59.2	-28.1	-29.0	-2.0
<b>F13 average</b>	<b>A</b>	<b>-51.5</b>	<b>-24.8</b>	<b>-24.9</b>	<b>-1.8</b>
F24a	M	-55.2	-30.4	-23.6	-1.2
F24b	M	-52.9	-29.3	-22.3	-1.2
F24c	M	-63.6	-36.9	-25.5	-1.2
F24d	M	-64.9	-37.7	-26.0	-1.2
<b>F24 average</b>	<b>M</b>	<b>-59.2</b>	<b>-33.6</b>	<b>-24.4</b>	<b>-1.2</b>
F24a	A	-53.0	-20.1	-29.2	-3.7
F24b	A	-51.2	-26.2	-25.0	-0.0
F24c	A	-51.4	-17.2	-34.3	-0.0
F24d	A	-54.0	-22.0	-30.5	-1.5
<b>F24 average</b>	<b>A</b>	<b>-52.4</b>	<b>-21.4</b>	<b>-29.8</b>	<b>-1.3</b>
G19a	M	-34.2	-13.0	-20.6	-0.5
G19b	M	-48.3	-17.4	-29.8	-1.2
<b>G19 average</b>	<b>M</b>	<b>-41.3</b>	<b>-15.2</b>	<b>-25.2</b>	<b>-0.9</b>
G19a	A	-42.8	-13.3	-26.1	-3.4
G19b	A	-40.1	-10.3	-29.8	-0.0
G19c	A	-42.2	-7.5	-34.7	-0.0
G19d	A	-40.3	-9.5	-30.1	-0.0
<b>G19 average</b>	<b>A</b>	<b>-41.4</b>	<b>-10.2</b>	<b>-30.2</b>	<b>-0.9</b>

All energies are given in kcal/mol. Docking: M: manual; A: automated.  $E_{\text{Protein-Ligand}}$ : protein-ligand interaction energy;  $E_{\text{Elec}}/E_{\text{vdW}}/E_{\text{HB}}$ : electrostatic, van der Waals, and hydrogen bonding components. a, b, c, d refer to different binding modes: 4 for the automated docking and from 1 to 4 for the manual docking. All the compound structures are given in Appendix B.

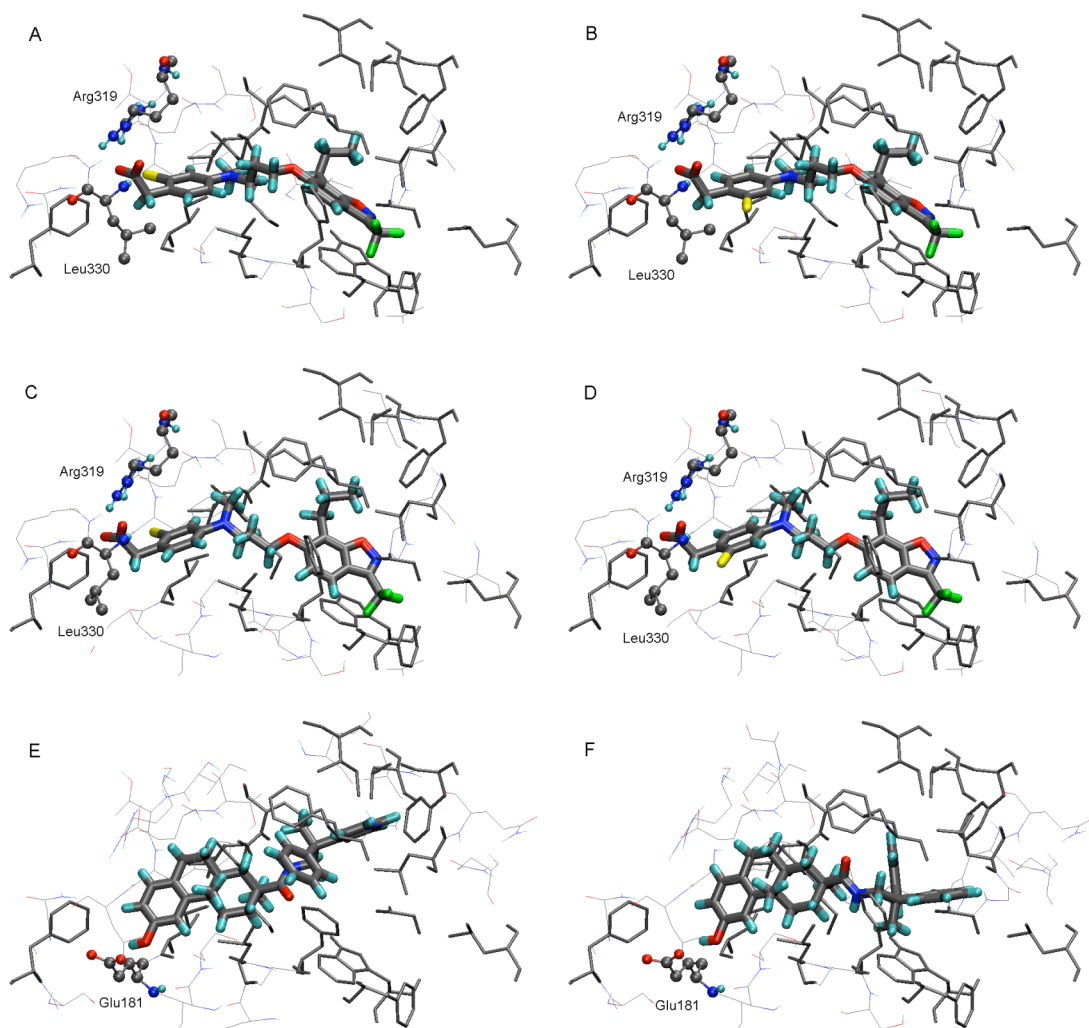
In the binding poses of ligands **F02** and **F24**, as obtained from the docking with *Glide*, the ligands' carboxylate moiety is located in two opposite sides of the binding pocket (Figure 37). Each of the 52 ligands was then manually docked to the receptor structure with the software *Yeti*.<sup>99, 101</sup> Table 11 lists energetic contributions calculated on the poses from manual and automated docking (with *Yeti*<sup>99, 101</sup>) for four compounds of the dataset (one for each chemical class of Figure 35): **F01**, **F13**, **F24** and **G19**. The poses obtained with manual docking are characterized on average by a more favorable (or in the worst case comparable) protein–ligand interaction energy.

The manual docking was aimed at finding binding poses that could reproduce, when possible, the interactions observed in the crystal structure and in the molecular-dynamics simulations between GW3965 and LXR: a hydrogen bond with Leu330, electrostatic interactions with the guanidinium group of Arg319, and hydrophobic interactions in the two major hydrophobic cavities (one occupied by the chloro-trifluoromethyl benzyl group and the second one around the diphenylethyl moiety of GW3965).

For all the 52 compounds, from a minimum of one to a maximum of four binding poses were selected.

More in detail, the carboxylate moiety of the ligands in class F was located in the vicinity of amino-acid residues Arg319 and Leu330: two different binding modes were considered for this class, depending on the position of the carboxylate group with respect to amino-acid residues Arg319 and Leu330, and on different orientations of the benzo-isoxazole moiety. For each of these binding modes, symmetrical features of the ligands allowed for two possible orientations, leading to a maximum of four different binding poses. As an example for class F, compound **F24** is discussed. The four selected binding modes (F24a, F24b, F24c and F24d) are depicted in Figure 38 A, B, C and D, respectively. In F24a and F24b the carboxylate moiety is located between Arg319 and Leu330 and creates a hydrogen bond with both amino-acid residues, while in F24c and F24d is located 2 Å closer to Arg319 and form hydrogen bonds only with Arg319. The difference between F24a and F24b, or between F24c and F24d is the orientation of the chlorine atom in the aromatic ring, that is rotated by 180°. F24c and F24d are characterized by more favorable protein–ligand interaction energies (Table 11) than F24a and F24b. All four binding modes were included in the final alignment.

The compounds belonging to class G locate the hydroxyl group of the podocarpate moiety towards the amino-acid residue Glu281, while the hydrophobic podocarpate substituents can occupy the two hydrophobic cavities mentioned above, leading to two different binding modes. Details of the binding of the compound **G19** are shown in Figure 38 E and F: G19a and G19b are two binding modes. In both cases, a hydrogen bond is observed between the ligand and Glu281.



**Figure 38.** Details of the binding modes for representative compound of classes F and G used for this QSAR study. The ligands (**F24**: A, B, C, D; **G19**: E, F) are represented as licorice, the residues Leu330, Arg319 and Asp181 as balls and sticks and the residues lining the hydrophobic pocket as lines. Hydrophobic amino-acid residues lining the binding pocket are represented as grey sticks.

#### 4.2.5 Building and Validation of a QSAR Models for the LXR

The alignment of the 52 compounds, divided in 40 training and 12 test ligands, identified using flexible docking, was used as input for the quantitative structure-activity relationship software *Quasar*.<sup>177</sup>

Two different alignments (from manual and automated docking) and different parameter settings were explored. Table 12 lists some of the performed simulations (sorted according to the predictive  $r^2$  value), the parameter settings and the corresponding results. For the parameters not specified in this table, the default values<sup>240</sup> were used.

The simulations from the manual docking alignment provide the highest predictive  $r^2$  values, and the lowest *rms* they are located in the upper part of the table.

##### *Number of crossovers*

The number of crossovers leading to simulations characterized by satisfactory  $r^2$  values ( $> 0.5$ ) are significantly lower (1,000–3,000) for the LXR than for the GR (30,000–60,000). This can be explained with the fact that the number of compounds is lower (52 ligands for the LXR and 110 for the GR), the library is more homogeneous (two chemical classes instead of four) and the source of the data as well (the  $K_i$  values were measured for all the compounds in the same laboratory).

##### *Population size*

The simulations with highest predictivity suggest that the optimal population is constituted by 200 individuals, even if some simulations (MD06, MD12) with 100 individuals can reach equally high predictivity.

##### *Attenuation factors*

The most predictive simulation (MD01) uses weights that downscale the contribution of solvation and internal energy (0.5,0.25). However, also some models with default attenuation factors (1.0,1.0,1.0) lead to comparable results.

##### *Induced-fit weight*

The two weights used (1.0 and 0.5) resulted in comparable results, indicating that even with different treating of the induced fit energy, is possible to obtain predictive models.

### *Cross-validation groups*

A high number of cross validation groups, meaning less restrictive conditions for the internal predictivity, is correlated with high predictive  $r^2$  values. Low number of cross validation groups lead occasionally to good predictivity (MD05, MD07, MD09).

The model characterized by the highest predictive  $r^2$  (0.701) is MD01. However, MD01 reaches the maximum value of predictive  $r^2$  only at a late stage of the evolution, after 20,000 crossovers, corresponding to 100 generations. At this point, the value of cross-validated  $r^2$  (0.924) is much higher (+0.223) than the predictive  $r^2$ . Being such behavior symptomatic of overfitting,<sup>188, 191</sup> the simulation MD02 was chosen instead for further analysis and validation.



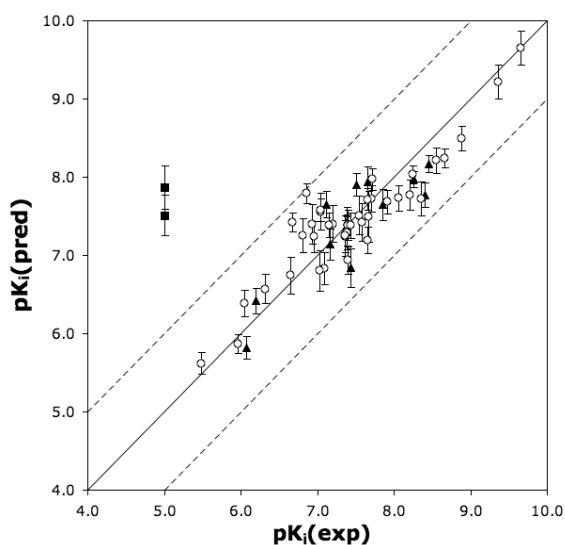
**Table 12.** Summary of the *Quasar* simulations.

Simulation	#CO	POP	AF	IFW	CVG	$q^2$	$p^2$	Training rms avg/max	Test rms avg/max
MD01	20,000	200	0.5, 1.0, 0.25	1.0	10	0.924	0.701	0.6/5.3	1.5/6.1
MD02	3,000	200	0.5, 1.0, 0.25	0.5	8	0.763	0.697	1.2/7.5	1.3/3.3
MD03	4,000	200	1.0, 1.0, 1.0	0.5	7	0.742	0.660	1.4/7.9	1.5/5.0
MD04	3,000	200	1.0, 1.0, 1.0	0.5	8	0.746	0.657	1.3/6.8	1.5/3.8
MD05	5,000	200	1.0, 1.0, 1.0	1.0	5	0.747	0.627	1.4/6.7	1.6/6.3
MD06	2,000	100	0.5, 1.0, 0.25	1.0	8	0.784	0.621	1.1/6.4	1.6/5.1
MD07	2,000	200	0.5, 1.0, 0.25	0.5	5	0.706	0.614	1.6/8.0	1.6/5.1
MD08	4,000	200	0.5, 1.0, 0.25	1.0	7	0.770	0.609	1.2/6.3	1.7/6.5
MD09	3,000	200	1.0, 1.0, 1.0	0.5	5	0.705	0.592	1.5/6.9	1.8/7.5
MD10	3,000	200	1.0, 1.0, 1.0	1.0	7	0.708	0.591	1.6/8.6	1.8/4.7
MD11	3,000	200	1.0, 1.0, 1.0	1.0	8	0.750	0.587	1.3/6.8	1.7/7.5
MD12	1,000	100	1.0, 1.0, 1.0	1.0	8	0.605	0.503	1.9/9.0	2.1/6.5
MD13	2,000	100	1.0, 1.0, 1.0	1.0	5	0.753	0.496	1.4/8.4	2.2/6.8
MD14	1,000	100	1.0, 1.0, 1.0	0.5	8	0.662	0.466	1.7/8.4	2.2/10.2
MD15	4,000	200	0.5, 1.0, 0.25	1.0	6	0.815	0.425	1.1/4.9	2.4/9.5
MD16	1,000	100	1.0, 1.0, 1.0	0.5	5	0.632	0.368	1.9/11.5	2.5/11.6
AD01	5,000	200	0.5, 1.0, 0.25	0.5	5	0.744	0.252	1.2/4.7	2.7/7.4
AD02	1,000	100	1.0, 1.0, 1.0	0.5	8	0.641	0.226	1.6/11.2	2.9/10.0
AD03	2,000	200	0.5, 1.0, 0.25	1.0	6	0.632	0.216	1.6/9.8	2.4/5.8
AD04	10,000	200	0.5, 1.0, 0.25	0.5	8	0.813	0.190	1.0/3.8	2.9/17.3
MD17	1,000	100	0.5, 1.0, 0.25	1.0	6	0.704	0.186	1.6/6.9	3.2/12.3
AD05	3,000	200	1.0, 1.0, 1.0	0.5	5	0.641	0.177	1.6/6.4	2.9/11.6
AD06	1,000	200	0.5, 1.0, 0.25	1.0	10	0.523	0.172	1.9/11.7	2.4/8.1
AD07	1,000	200	0.5, 1.0, 0.25	1.0	7	0.538	0.144	1.8/8.9	2.7/15.6
AD08	1,000	100	0.5, 1.0, 0.25	1.0	8	0.684	0.141	1.6/8.8	3.3/10.6
AD09	1,000	100	0.5, 1.0, 0.25	1.0	6	0.618	0.138	1.7/10.8	2.9/13.2
AD10	1,000	200	1.0, 1.0, 1.0	1.0	7	0.471	0.130	2.2/13.6	2.9/14.7
AD11	2,000	200	1.0, 1.0, 1.0	0.5	8	0.593	0.127	1.9/16.1	3.0/15.4
AD12	1,000	100	1.0, 1.0, 1.0	0.5	5	0.564	0.116	2.0/13.0	3.2/8.6
AD13	1,000	200	1.0, 1.0, 1.0	1.0	8	0.458	0.088	2.1/12.8	2.8/12.7
AD14	1,000	100	1.0, 1.0, 1.0	1.0	8	0.535	0.071	2.0/12.0	3.1/21.5
MD15	35,000	100	0.5, 1.0, 0.25	1.0	7	0.669	0.051	1.6/9.9	3.8/15.1
AD16	4,000	200	1.0, 1.0, 1.0	0.5	7	0.699	0.028	1.5/7.9	3.7/15.0
AD17	1,000	200	1.0, 1.0, 1.0	1.0	5	0.470	0.007	2.1/12.4	3.2/17.1
AD18	4,000	100	0.5, 1.0, 0.25	1.0	7	0.827	-0.024	1.1/7.4	4.4/10.2
AD19	1,000	100	1.0, 1.0, 1.0	1.0	5	0.587	-0.162	1.9/14.1	4.5/26.8

Simulation: MD: manual docking; AD: automated docking; #CO, number of crossovers; POP, population size; AF, attenuation factors (weight of solvation, entropy, internal energy); IFW, induced-fit weights; CVG, cross validation groups;  $q^2$ : cross-validated  $r^2$ ,  $p^2$ : predictive  $r^2$ ; *Training rms avg/max*, *Test rms avg/max*, the rms and maximal deviation from the experimental binding affinity respectively for training set and test set (given as a factor (off) in  $K_i$ ).

The *Quasar* simulation MD02 was based on a family of 200 receptor models that differ in respect to the properties mapped on their surface. The family of receptor models was evolved for 3,000 crossover cycles, corresponding to 15 generations. For the cross-validation, we selected a *leave-n-out* ( $n = 8$ ) protocol. Protein flexibility was taken into account by six induced-fit scenarios. During the simulation, a mutation rate of 0.02 was applied in the evolution of the quasi-atomistic properties.

The model family converged at a cross-validated  $r^2$  of 0.763 for the 40 training compounds and yielded a predictive  $r^2$  of 0.697 for the 12 test ligands (cf. Figure 39 and Table 13).



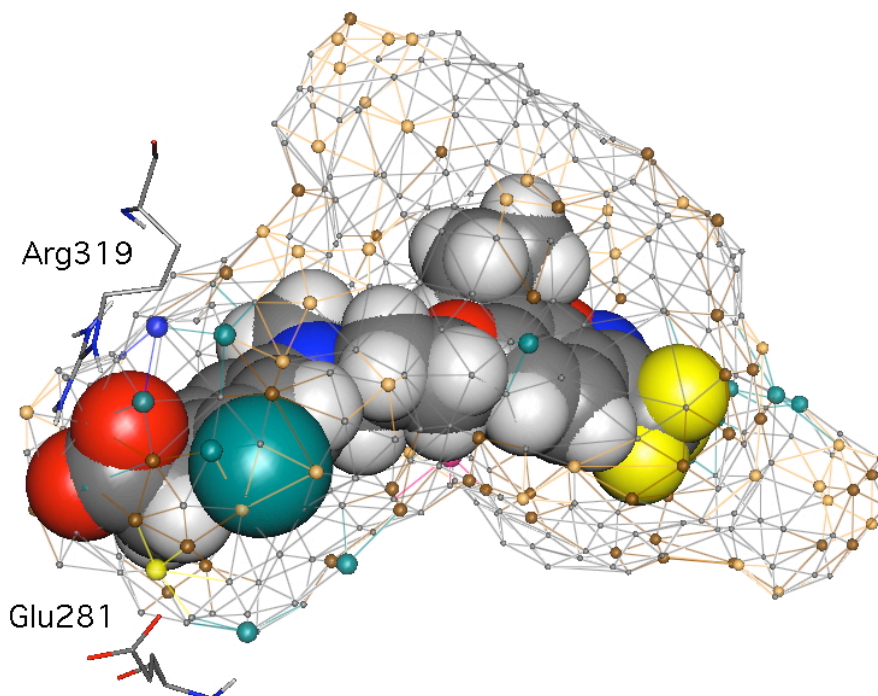
**Figure 39.** Comparison of experimental and predicted binding affinities of the training and test set (open circles and filled triangles, respectively) and of the compounds **F09** and **F10** used as external test set (filled squares) towards the LXR, as obtained with *Quasar*. Dashed lines are drawn at a factor 10 from the experimental value. The experimental values of compounds **F09** and **F10** are only indicatively set at a value of 10  $\mu\text{M}$ , the measured  $K_i$  being higher than 10  $\mu\text{M}$ .

On the average (*rms*), the calculated binding affinity of the training and test ligands deviates by a factor 1.2 and 1.3, respectively, from the experimental  $K_i$  value. The maximal observed deviation of an individual compound corresponds to a factor of 7.5 in  $K_i$  for the training set and 3.3 for the test set, respectively. The predicted binding affinities for the compounds **F09** and **F10**, belonging to the prediction set and not employed in the model building, are 7.3 and 4.9 nM, respectively: their activity is overestimated (more than a factor of 30 in  $K_i$ ) by the model. A possible reason for this poor prediction could be the fact that none of the compounds of the training and test set has a very low activity (the least active compound of class F is 210 nM).

**Table 13.** Summary of the *Quasar* and *Raptor* simulations for the 40 training and 12 test compounds

Simulation	$r^2$	$q^2$	$rms$ training	max. training	$p^2$	$rms$ test	max. test
<i>Quasar</i>	0.783	0.763	1.2	7.5	0.697	1.3	3.3
<i>Raptor</i>	0.876	n/a	1.2	3.1	0.484	3.6	8.7

$r^2$ : correlation coefficient,  $q^2$ : cross-validated  $r^2$ ,  $p^2$ : predictive  $r^2$ ; the  $rms$  and maximal deviation from the experimental binding affinity is given as a factor (off) in  $K_i$ .

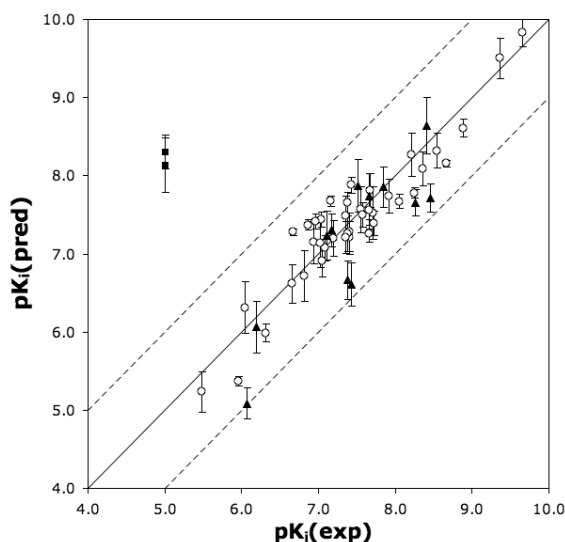


**Figure 40.** Representation of the LXR surrogate (*Quasar*) with bound compound **F24** (space-filling). The mapped quasi-atomistic properties are colored as follows: blue (salt bridge, negatively charged), green (H-bond donor), yellow (H-bond acceptor), saddle brown (hydrophobic, positively charged), chocolate brown (hydrophobic, negatively charged), grey (hydrophobic, neutral). Residues Glu281 and Arg319 are shown as sticks.

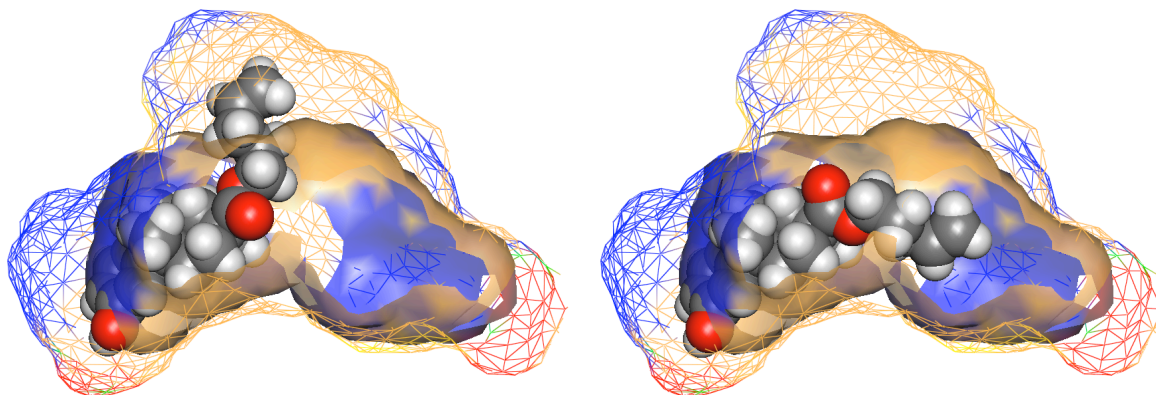
A representation of the receptor surrogate with bound the compound **F24** is depicted in Figure 40. When compared to the crystal structure of the LXR complexed with GW3965, the receptor surrogate generated by *Quasar* properly reproduces properties observed for the amino-acid residues in the binding pocket: a salt bridge, positively charged (blue) is located in proximity of the guanidinium moiety of Arg319, and a hydrogen bond acceptor (yellow) is located close to the position occupied by the Glu281. Moreover, hydrophobic properties (brown) populate great part of the surface, correctly reflecting the hydrophobic character of the binding pocket.

In order to assess the sensitivity of the model, 10 scramble tests were performed using a different shuffle of the biological data (of the training set) for each simulation. They yielded an average predictive  $r^2$  of 0.018 compared to +0.697 for the simulation using un-scrambled values, demonstrating that the model for the LXR is indeed sensitive towards the biological data.

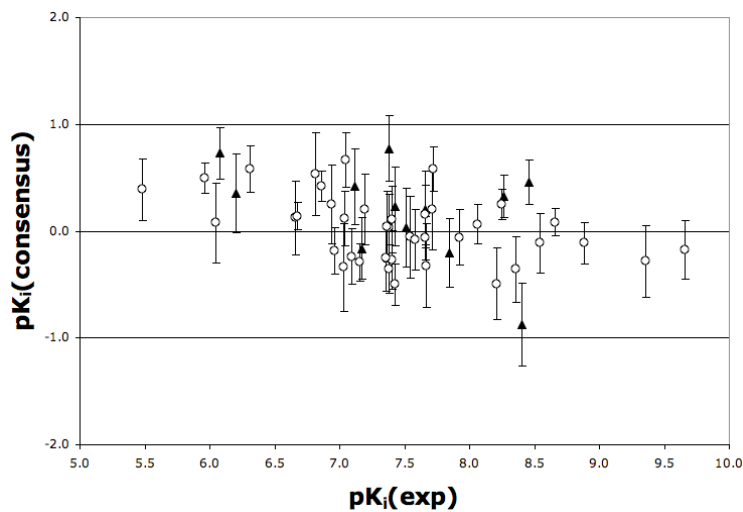
To test for consensus, we applied a second methodology (software *Raptor*). The *Raptor* simulation –using the same ligand alignment and selection– yielded an  $r^2$  of 0.876 and a predictive  $r^2$  of 0.484. The comparison of predicted and experimental binding affinities is shown in Figure 41, and the performance coefficients are given in Table 13. The *Raptor* simulation yielded a lower predictive power, but the performance can be considered satisfactory. The activity of compounds **F09** and **F10**, like for the *Quasar*'s model, is overestimated by the *Raptor*'s model. *Raptor*'s dual-shell model can simulate ligand-dependent induced-fit mechanisms. For example, the alkylic substituent of **G06** (cf. Figure 42) can be accommodated in two different hydrophobic cavities present in the binding pocket. The outer shell and the inner shell can represent the two possibilities (Figure 42, left and right). The comparison of predicted binding affinities obtained by the two approaches is shown in Figure 43. The average deviation is 0.27 logarithmic units (a factor 1.9 in  $K_i$ ). For none of the compounds the ratio is greater than a factor of 10 (in  $K_i$ ), the maximal ratio being 0.88 logarithmic units (a factor 7.6 in  $K_i$ ). Thus, consensus between *Quasar* and *Raptor* predictions has been achieved.



**Figure 41.** Comparison of experimental and predicted binding affinities of the training (open circles), test set (filled triangles) and the external test set (filled squares) towards the LXR as obtained with *Raptor*.



**Figure 42.** *Raptor* model of the LXR binding-site with compound **G06** in two different binding modes (left and right; hydrophobic fields = beige, hydrogen-bond donating propensity = blue, hydrogen-bond accepting propensity = red, hydrogen-bond flip-flop = green). The inner shell (transparent surface) and outer shell (wire frame) are shown in different style to highlight the two shells of the surrogate (for clarity, the front part of the receptor model has been clipped).



**Figure 43.** Consensus scoring using *Quasar* and *Raptor*. The quantity is expressed as  $pK_i$ ,  $\text{consensus} = -\log(K_{i-\text{Quasar}}/K_{i-\text{Raptor}})$ . Error bars indicate the cumulative standard deviation,  $\text{esd}_{\text{cumulative}} = \sqrt{(\text{esd}_{\text{Quasar}}^2 + \text{esd}_{\text{Raptor}}^2)}$ .

#### 4.2.6 Conclusions and Applicability of the LXR Model

A QSAR model for a series of 54 ligands of the LXR (40 training set, 12 test set and 2 prediction set compounds), belonging to 2 different chemical classes was built and validated. The *Quasar* simulation yielded a cross-validated  $r^2$  of 0.763 for the training set and a predictive  $r^2$  of 0.697 for the internal test set, respectively. A second methodology *Raptor* was employed to test for consensus: the simulation yielded a cross-validated  $r^2$  of 0.876 for the training set and a predictive  $r^2$  of 0.484 for the internal test set, respectively. For all the compounds the predictions by the two methodologies are in agreement (the ratio being below a factor of 10, in  $K_i$ ).

Two weakly active compounds ( $K_i > 10 \mu\text{M}$ ) were also evaluated and predicted more active than the experimental value assesses: probably because low-activity compounds are not sufficiently represented in the dataset.

The model sensitivity was assessed by 10 scramble tests: the average predictive  $r^2$  is 0.018, demonstrating that the model for the LXR is sensitive towards the biological data.

The results suggest that the model can be used in binding affinity prediction of new drug candidates, in the design of new LXR agonists, or of existing compounds, in the investigation of possible toxicological mechanisms. The model shows to be correlated with structural properties, and therefore is to some extent interpretable.

Limitations of the approach include the smaller number of compounds (when compared to the GR) and of the chemical classes: the diversity of the compounds is limited to the two chemical classes of the presented study. The molecular weight of the compounds is also a limitation in the automated procedure used to test new compounds: the automated docking for bulky compounds that involve significant induced-fit might not sample sufficient poses.

The LXR model, like for the GR one, has been added to the *VirtualToxLab*<sup>TM3, 4</sup> developed by the Biographics Laboratory 3R.<sup>3</sup>

## 4.3 Hydrophobic Effect Estimation and Evaluation

### 4.3.1 Development of an Empirical Hydrophobicity Function

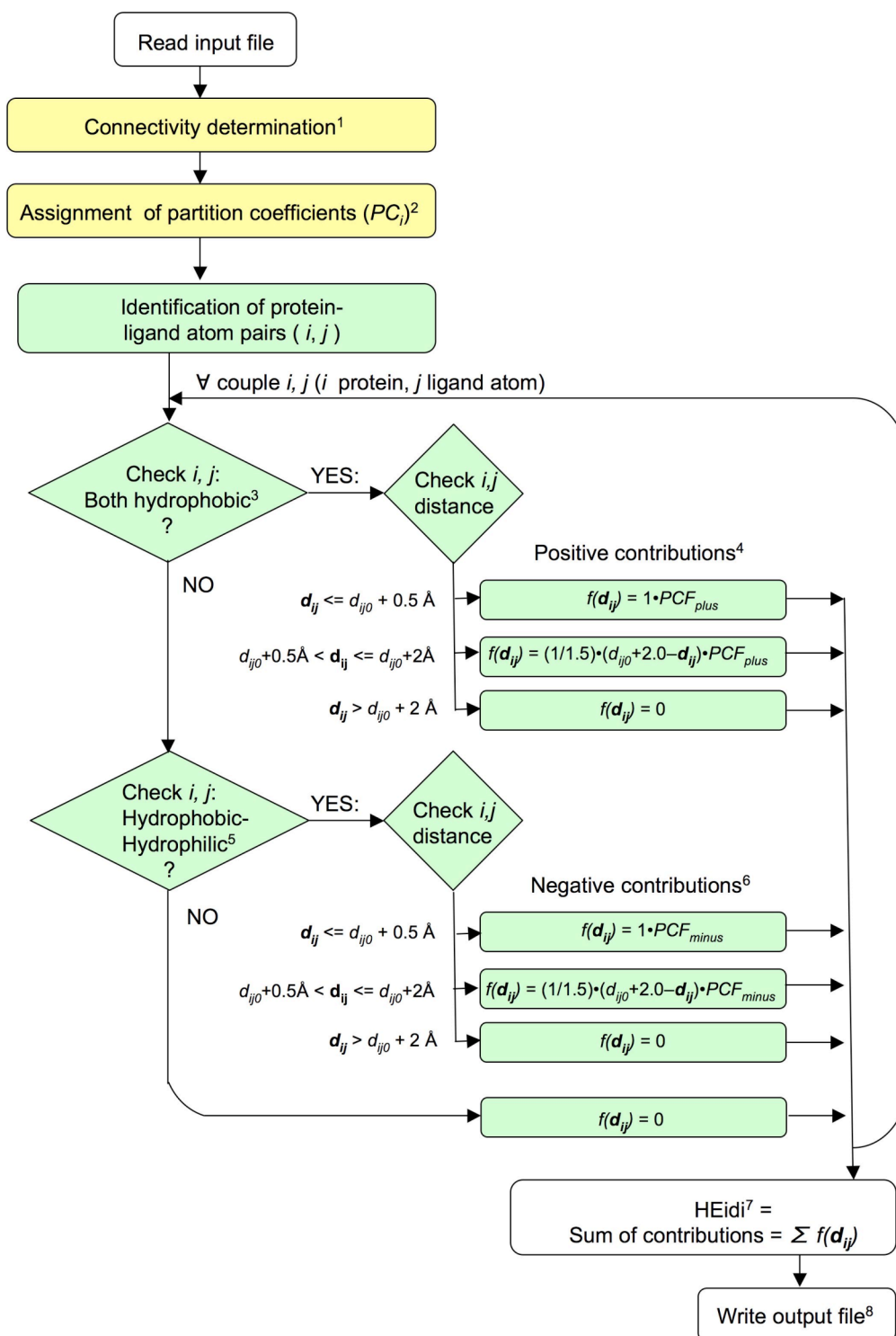
A novel empirical scoring function has been devised and implemented in a small program (*HEidi*: Hydrophobic Effect in Drug Interactions), written in C, aiming at the quantification of the hydrophobic effect for scoring protein–ligand binding energies.

The **input** of *HEidi* are the coordinates of a protein–ligand complex including all atoms (the hydrogen atoms are explicitly represented).

The **output** of *HEidi* is a scalar value of a hydrophobicity function. This value can be used to score how good the ligand, in a specific conformation and orientation, matches hydrophobic cavities of a protein binding pocket.

The basic idea is to give a bonus (a positive contribution to the score) for each “good hydrophobic match” and a penalty a negative contribution to the score for each “bad hydrophobic match”. In *HEidi*, a “good hydrophobic match” is defined as a pair of atoms, one belonging to the protein and one to the ligand, that are both hydrophobic and within a certain distance. A “bad hydrophobic match” is defined as a pair of atoms, one belonging to the protein and one to the ligand, that are one hydrophobic and the other hydrophilic, within a certain distance. The classification of the atoms as hydrophobic or hydrophilic is determined on the basis of partition coefficients determined by Ghose and Crippen.<sup>260</sup> These partition coefficients are, in turn, determined by the atom type and connectivity for each atom in a molecule<sup>260</sup>. The amount of the bonus or penalty depends on the distance between the pair of atoms (the closer, the higher prize or penalty) and on their hydrophobicity (high partition coefficients for both atoms give high prize, and very different partition coefficients for the two atoms gives high penalty).

Bonuses and penalties for each pair of protein–ligand atoms are then summed to give the final score. The algorithm is explained more in detail in a flowchart (Figure 44).



**Figure 44.** Flowchart of the program for the calculation of the hydrophobic effect score. The white boxes are performed only once during the algorithm, the yellow ones require operations repeated for each atom and the green for each pair of protein–ligand atoms of the input file.



Comments to the flowchart:

1. **Connectivity determination:** for each atom, the atom type, and the number and type of the bonded atoms are stored. This information is used to assign the partition coefficients in the next step. Atom types are classifications of the atoms based on element and bonding environment.
2. **Assignment of the partition coefficient:** for each atom  $i$  a partition coefficient ( $PC_i$ ) is assigned, according to Ghose and Crippen,<sup>260</sup> on the basis of the atom types and bonded atoms identified. Each atom is classified into 110 atom types, and a specific partition coefficient is assigned to each of them.
3. **Check if  $i, j$  are both hydrophobic:** for each pair of atoms  $i$  and  $j$ ,  $i$  belonging to the protein and  $j$  to the ligand, the atoms  $i$  and  $j$  are considered hydrophobic if they have positive partition coefficients ( $PC_i > 0$ ,  $PC_j > 0$ ).
4. **Positive contributions:** if both atoms are hydrophobic, then a good hydrophobic match between protein and ligand is found, and a positive contribution to the score is calculated ( $f(d_{ij})$ ). The contribution depends on the distance between the two atoms ( $d_{ij}$ ), and on a scaling factor (Partition Coefficient Factor,  $PCF_{i,j-plus}$ ):

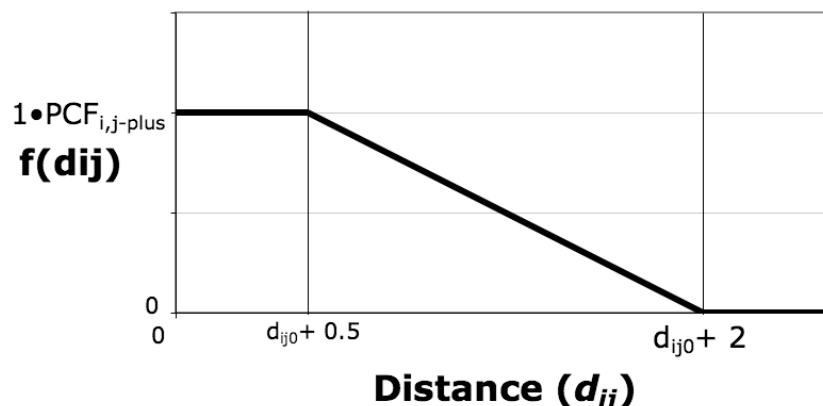
$$\begin{array}{lll}
 \text{if } d_{ij} > d_{ij0} + 2 \text{ \AA} & \text{then} & f(d_{ij}) = 0 \\
 \text{if } d_{ij} \leq d_{ij0} + 0.5 \text{ \AA} & \text{then} & f(d_{ij}) = 1 \cdot PCF_{i,j-plus} \\
 \text{if } d_{ij0} + 0.5 \text{ \AA} < d_{ij} \leq d_{ij0} + 2 \text{ \AA} & \text{then} & f(d_{ij}) = (1/1.5) \cdot (d_{ij0} + 2.0 - d_{ij}) \cdot PCF_{i,j-plus}
 \end{array}$$

where  $d_{ij0}$  is the sum of their van der Waals radii (the list of the van der Waals radii is reported in Appendix C)

$PCF_{i,j-plus}$  is the average of the partition coefficients, scaled by interval:

$$PCF_{i,j-plus} = (((PC_i + PC_j)/2) + PC_{min}) / (PC_{max} - PC_{min})$$

The thresholds for the distance, as well as the type of dependence from the distance, are chosen according to examples from Wang *et Al.*<sup>147</sup> The scaling factor ( $PCF_{i,j-plus}$ ) has been chosen to give the maximal prize positive contribution to the atoms that are most hydrophobic, or, in other words, the ones having highest partition coefficients. The relationship between the positive contributions and the distance is shown in Figure 45.



**Figure 45.** Dependence of each positive contribution from the distance between the pair of atoms  $i, j$  ( $i$  belonging to the protein and  $j$  to the ligand).

5. **Check  $i, j$ , hydrophobic–hydrophilic pair:** for each pair of atoms  $i$  and  $j$ ,  $i$  belonging to the protein and  $j$  to the ligand, the atoms  $i$  and  $j$  are one hydrophobic and the other hydrophilic if they have very different partition coefficients ( $PC_i > 0.4$ , and  $PC_j < -0.4$ ) or vice versa. The choice of the threshold was subsequent to a statistical analysis of the distribution of contributions (discussed in the paragraph 4.3.2) and was intended to consider only pairs where the hydrophobicity was very different between the two atoms (a hydrophobic atom in contact with a hydrophilic atom, resulting in a bad match of properties).
6. **Negative contributions:** if the atoms are one hydrophobic and the other hydrophilic, then a bad hydrophobic match between protein and ligand is found, and a negative contribution to the score is calculated  $f(d_{ij})$ . When compared to the positive contribution, a negative contribution has in module the same dependence on the distance between the two atoms ( $d_{ij}$ ), and on a scaling factor ( $PCF_{i,j-minus}$ ), but opposite in the sign:

$$\begin{array}{lll} \text{if } d_{ij} > d_{ij0} + 2 \text{ \AA} & \text{then} & f(d_{ij}) = 0 \\ \text{if } d_{ij} \leq d_{ij0} + 0.5 \text{ \AA} & \text{then} & f(d_{ij}) = -1 \cdot PCF_{i,j-minus} \\ \text{if } d_{ij0} + 0.5 \text{ \AA} < d_{ij} \leq d_{ij0} + 2 \text{ \AA} & \text{then} & f(d_{ij}) = -(1/1.5) \cdot (d_{ij0} + 2.0 - d_{ij}) \cdot PCF_{i,j-minus} \end{array}$$

where

$d_{ij0}$  is the sum of their van der Waals radii (the list of the van der Waals radii is reported in Appendix C)

$PCF_{i,j-minus}$  is the half of the average of the partition coefficients, scaled by interval:

$$PCF_{i,j-minus} = (((PC_i + PC_j)/4) + PC_{min}) / (PC_{max} - PC_{min})$$

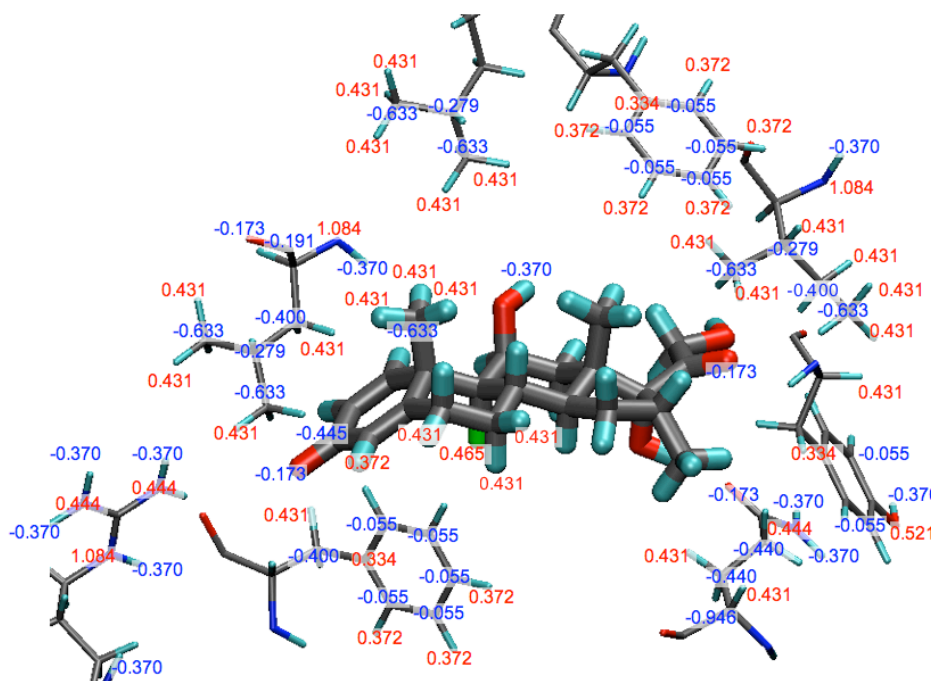
In the case of the negative contributions, when compared to the positive contributions, the sum of the partition coefficients is divided by 4 instead of 2 (in order to take into account the fact that on a random distribution of the pairs, there's double probability to have a hydrophobic–hydrophilic pair than a hydrophobic–hydrophobic pair).

7. ***HEidi* calculation:** all the positive and negative contributions are summed to give the final hydrophobic effect score (*HEidi*).
8. **Write output file:** The value of *HEidi* and of the individual contributions are written to a file.

### 4.3.2 Preliminary Analysis on a Protein–Ligand Complex

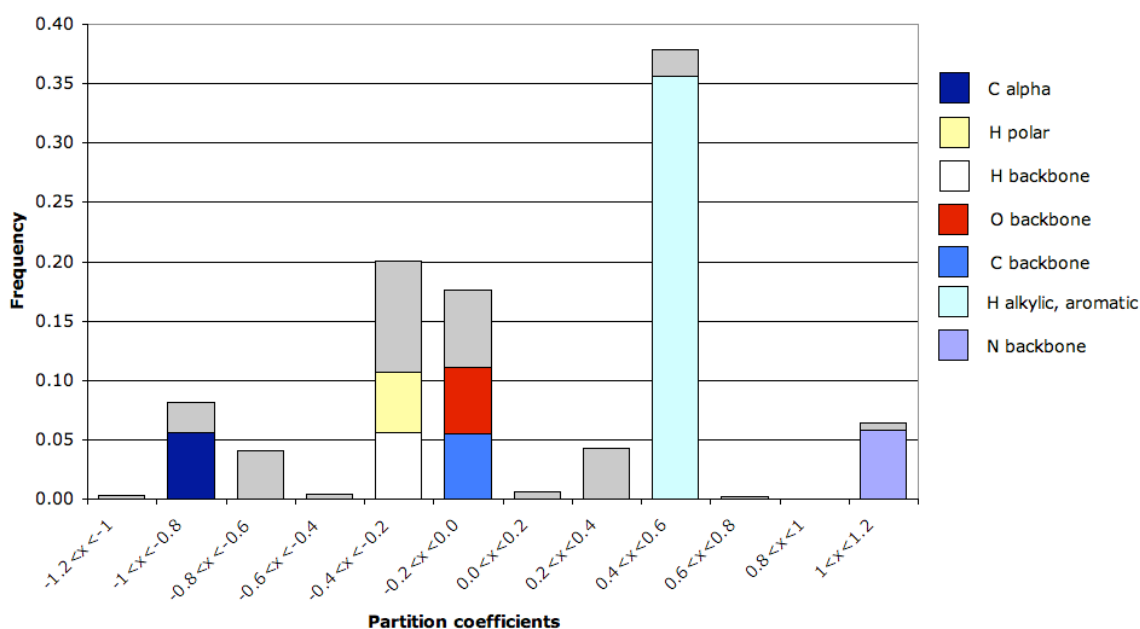
The program, after a test phase on very simple systems (for example a ligand and very few amino-acid residues), was then applied to a real system: dexamethasone bound to the GR, in order to check partition coefficients and contributions assignment.

In Figure 46, dexamethasone and few amino-acid residues lining the binding pocket are shown. Labels indicate the partition coefficients of the atoms. A great part of the positive coefficients (indicating hydrophobic atoms according to the program) is assigned to hydrogens of alkylic or aromatic units. Negative coefficients are assigned to hydrogen, oxygen, and carbon atoms of hydroxyl and carbonyl groups, and to carbon atoms of alkylic or aromatic units (for example in amino-acid residues like leucine, isoleucine, valine, or phenylalanine) but only if connected to at least one hydrogen atom. When an  $sp^3$  carbon atom is tetrasubstituted (e.g. position 10 and 13 in the steroids), or when is connecting two condensed aromatic rings, then the partition coefficient is positive. Nitrogen atoms in protein backbone or in arginine residues have a positive coefficient, but all the hydrogen atoms connected to such nitrogen atoms, have negative partition coefficient.



**Figure 46.** Partition coefficients in a real system: GR bound to dexamethasone. In sticks dexamethasone (thicker) and some amino-acid residues in the binding pocket (thinner). The numbers indicate the partition coefficient of atoms: colored in red the positive values (classified in the program as hydrophobic atoms) and in blue negative values (classified in the program as hydrophilic atoms).

Figure 47 shows a statistic of the partition coefficients for all the atoms of the GR-dexamethasone structure. The atoms of the proteins are almost the totality of the atoms (99%) of the complex. Hydrogens belonging to alkylic chains or to aromatic rings, as anticipated previously, constitute great part of the positive coefficients and 36% of all the atoms. They have partition coefficients between 0.4 and 0.6. For this reason, in the calculation of the negative contributions, a threshold of  $\pm 0.4$  was chosen. The idea was to consider for the negative contributions only atoms with very different partition coefficients (and therefore with different hydrophobicities), but being such hydrogen atoms a relevant part of the overall atoms, they couldn't be excluded and the threshold was set at a value of 0.4. The polar hydrogens (in arginine, serine, glutamine, asparagine, lysine, tryptophan, histidine, cysteine, threonine), and the backbone hydrogens, receive a negative partition coefficient. Oxygen and carbon atoms of the backbone, forming the carbonyl group, have slightly negative coefficients. The highest and lowest coefficients are assigned respectively to the nitrogen atoms of the protein backbone and to the carbon alpha.

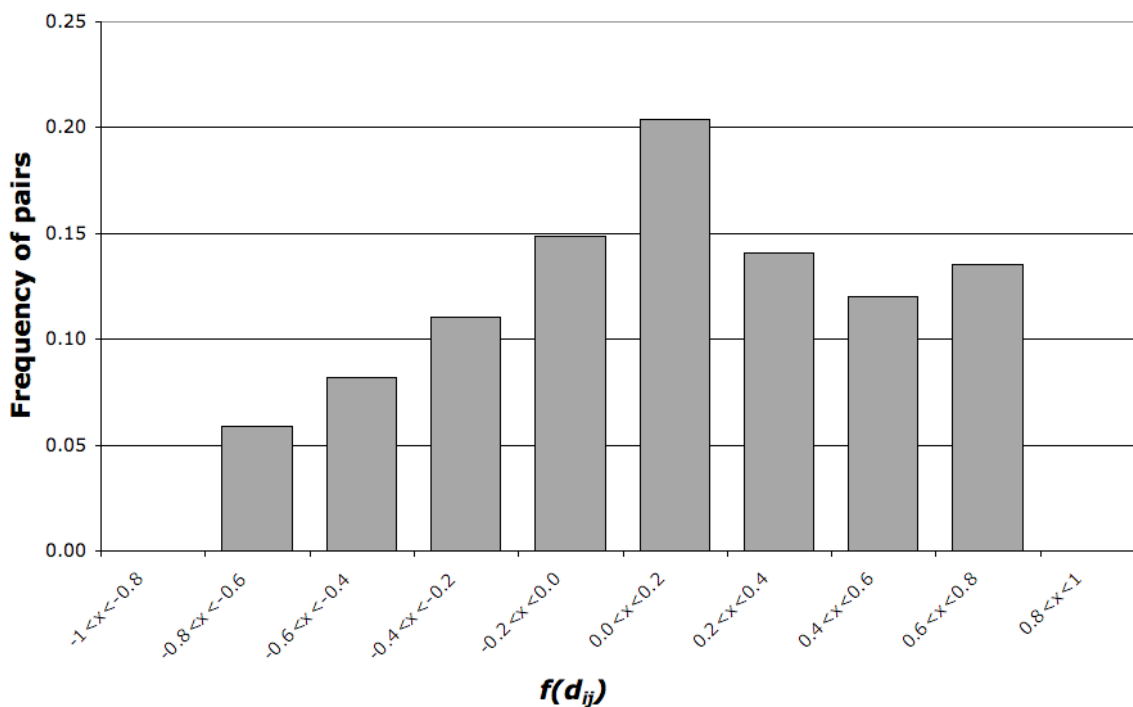


**Figure 47.** Frequency of the partition coefficients. In different colors are shown the most abundant atom types.

Partition coefficients are used in the program to determine whether an atom is classified hydrophobic (partition coefficient  $> 0$ ) or not, and to determine the total amount of the (positive or negative) contribution. Figure 48 shows the distribution of the contributions received by the protein–ligand atom pairs. The distribution is not symmetrical because of the choice to weight the positive matches (the good hydrophobic matches) higher when compared to the negative matches (the bad hydrophobic matches). Many pairs of protein–ligand atoms receive a low

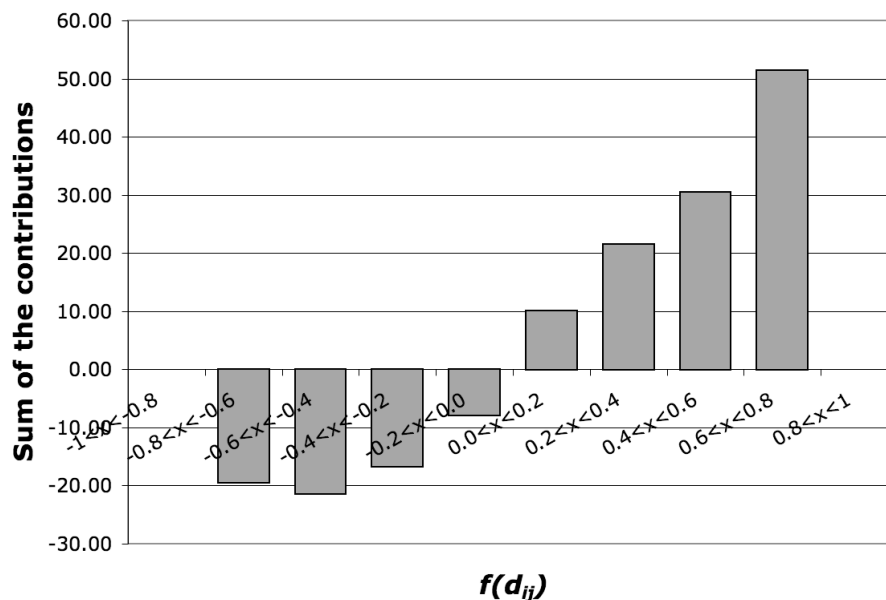
contribution (positive or negative) and few pairs receive a high contribution (positive or negative).

Ideally, the maximal positive contribution of +1 is assigned to a pair of protein–ligand atoms that both have the maximal possible value of partition coefficient and a distance that is lower than  $d_{ij0}+0.5\text{\AA}$ . Similar behavior, but opposite in sign, can be described for the negative contributions.



**Figure 48.** Frequency of the contributions. The histograms indicate the frequency of pairs (protein–ligand atoms) receiving the corresponding contribution indicated in the x axis. Most pairs of protein–ligand atoms receive a small contribution (50% are within -0.2 and 0.2), few pairs receive a contribution > 0.2.

Since the individual contributions are summed to give the final score, it is more important to evaluate the sum of the contributions for each interval, more than the individual contributions. The histograms of Figure 49 indicate that the higher contributions (the very good matches) are more relevant in the total score than the lower contributions. In other words, few good matches are more influential than many average matches. The same trend for the negative contributions is not so evident, and reflects the choice to give more relevance to the positive contributions due to good matches than to the negative contributions due to bad matches.



**Figure 49.** Sum of the contributions for each interval of individual contributions.

### 4.3.3 Application to the GR and LXR

The hydrophobic effect function was then applied to score the poses obtained from automated docking (performed with *Yeti/Autodock*<sup>100</sup>) on the GR-dexamethasone and LXR-GW3965 complexes.

With *Yeti/Autodock*, up to 25 different possible binding poses for each ligand are stored in the output file, together with the corresponding values of protein–ligand interaction energies and the individual partitioning into van der Waals, electrostatic and hydrogen bond components (according to *Yeti* force field<sup>99</sup>). The hydrophobic effect function was also applied to calculate a score (*HEidi*) for each of the binding poses and the results were analyzed.

A good scoring function in a docking algorithm should be able to identify the correct binding mode within the top rank of the output poses. Assuming that the pose from the crystal structure identifies a real binding mode, then it should have also maximal absolute value of the docking scoring function, or, referring to energy values, should have the minimum negative value. In principle, we can assume that low *RMSD* (root mean square distance) values of each pose (calculated from the ligand of the crystal structure) should correspond to favorable energies and high *RMSD* values correspond to unfavorable energies: at least in principle, the closer to the crystal structure, the better the docking score.

Table 14 and 15 report the results for the GR and LXR. The crystal structure and the 25 output poses are reported, together with the total protein–ligand energies, the individual energetic contributions (van der Waals, electrostatic and hydrogen bond), the *HEidi* and the *RMSD* values. For the GR (Table 14) low values of *RMSD* are clearly associated also to lower (more favorable) total protein–ligand

interaction energies and lower values for the *HEidi*. The same behavior, but less pronounced, can be observed for the LXR (Table 15). When observing the *RMSD* values for LXR (Table 15), it can be noticed that most of the poses are associated with high *RMSD*. That reflects a difficulty in identifying the correct binding mode, but also the fact that the ligand GW3965 (depicted in Figure 32) is a bulky ligand and different docking poses can lead to very high *RMSD* values.

The hydrophobic effect is only one aspect in the complex binding process where other important contributions play a relevant role, and therefore can't be used alone in the poses evaluation. For this reason, a linear combination of the van der Waals, electrostatic, hydrogen bond energies, together with the *HEidi* would be desirable as a docking scoring function (*DSF*):

$$DSF = a \cdot E_{Ele} + b \cdot E_{vdW} + c \cdot E_{Hbond} + d \cdot HEidi \quad (6)$$

where  $E_{Ele}$ : electrostatic energy,  $E_{vdW}$ : van der Waals energy,  $E_{Hbond}$ : hydrogen bond energy, *HEidi*: score calculated from the hydrophobic effect function, *a*, *b*, *c*, *d*: coefficients.



**Table 14.** Docking results for GR–dexamethasone complex

Pose	<i>Protein-Ligand total</i> (kcal/mol)	<i>Ele</i> (kcal/mol)	<i>vdW</i> (kcal/mol)	<i>Hbond</i> (kcal/mol)	<i>-HEidi</i>	<i>RMSD</i> (Å)
X-Ray	-58.05	-15.99	-33.49	-8.57	-48.46	0.00
Pose 20	-56.71	-17.29	-29.88	-9.53	-50.92	0.81
Pose 14	-50.17	-14.87	-31.30	-4.00	-49.17	0.85
Pose 12	-59.93	-17.62	-30.41	-11.90	-47.78	0.86
Pose 09	-50.31	-13.63	-32.38	-4.30	-51.56	0.90
Pose 10	-51.39	-16.45	-30.84	-4.10	-49.69	0.94
Pose 13	-52.27	-15.42	-30.98	-5.87	-49.22	0.97
Pose 22	-51.44	-15.73	-30.34	-5.37	-49.31	0.97
Pose 24	-51.52	-15.12	-31.04	-5.36	-49.25	0.97
Pose 08	-51.11	-16.57	-30.18	-4.36	-49.85	1.01
Pose 11	-52.59	-15.86	-30.81	-5.92	-46.88	1.01
Pose 23	-49.08	-12.71	-31.89	-4.48	-48.60	1.07
Pose 19	-55.25	-13.42	-33.44	-8.39	-48.11	1.11
Pose 06	-52.09	-16.06	-32.16	-3.88	-53.93	1.12
Pose 25	-34.08	-8.37	-25.21	-0.50	-41.88	3.15
Pose 17	-36.27	-15.53	-17.85	-2.89	-43.30	3.88
Pose 05	-40.27	-12.42	-23.77	-4.08	-41.40	6.98
Pose 21	-41.96	-9.36	-25.69	-6.91	-42.51	6.99
Pose 18	-36.04	-6.76	-26.82	-2.47	-43.60	7.02
Pose 16	-35.30	-9.78	-25.40	-0.13	-43.18	7.05
Pose 01	-40.66	-8.59	-27.85	-4.22	-42.45	7.08
Pose 15	-32.13	-8.57	-22.08	-1.47	-41.48	7.12
Pose 02	-39.17	-11.19	-27.96	-0.03	-40.24	7.34
Pose 07	-30.73	-3.48	-27.15	-0.10	-44.33	7.35
Pose 04	-39.66	-10.27	-26.01	-3.39	-41.85	7.44
Pose 03	-41.20	-7.73	-26.69	-6.77	-38.29	7.45

*Protein-Ligand total*: total protein–ligand interaction energy ( $Protein-Ligand\ total = Ele + vdW + Hbond$ ), *Ele*: electrostatic energy *vdW*: van der Waals energy, *Hbond*: hydrogen bond energy, *-HEidi*: score calculated from the hydrophobic effect function (taken with negative sign in order to be coherent with energies, where negative numbers mean more favorable energies), *RMSD*: root mean square distance.

The table is sorted by increasing *RMSD*.

**Table 15.** Docking results for LXR-GW3965 complex

Pose	<i>Protein-Ligand total (kcal/mol)</i>	<i>Ele (kcal/mol)</i>	<i>vdW (kcal/mol)</i>	<i>Hbond (kcal/mol)</i>	<i>-HEidi</i>	<i>RMSD (Å)</i>
X-Ray	-86.05	-30.04	-55.25	-0.75	-173.79	0.00
Pose 16	-70.83	-31.25	-38.22	-1.36	-104.14	3.77
Pose 15	-70.79	-33.38	-36.73	-0.68	-101.38	3.78
Pose 12	-69.84	-29.72	-38.76	-1.36	-116.38	3.83
Pose 19	-30.68	-7.15	-23.53	0.00	-129.50	4.48
Pose 14	-49.61	-21.51	-28.09	-0.02	-112.02	4.95
Pose 13	-41.18	-18.26	-22.80	-0.12	-109.58	5.02
Pose 17	-66.59	-25.15	-41.17	-0.27	-121.67	5.28
Pose 25	-48.35	-20.99	-27.24	-0.12	-127.69	5.74
Pose 05	-59.60	-21.29	-36.61	-1.70	-155.11	6.05
Pose 04	-52.53	-19.52	-32.97	-0.04	-156.53	6.05
Pose 18	-62.22	-29.43	-29.99	-2.80	-157.66	6.17
Pose 07	-41.02	-9.69	-31.33	0.00	-158.85	7.11
Pose 06	-42.60	-7.43	-35.17	0.00	-155.22	7.31
Pose 02	-35.75	-7.67	-28.07	0.00	-167.75	7.49
Pose 01	-47.58	-8.36	-39.22	0.00	-164.69	7.59
Pose 08	-37.84	-16.87	-20.97	0.00	-141.73	7.59
Pose 11	-26.83	-6.11	-20.72	0.00	-140.19	7.64
Pose 24	-48.38	-9.96	-38.39	-0.02	-135.34	7.74
Pose 09	-31.97	-3.44	-28.52	0.00	-130.53	7.76
Pose 03	-39.34	-3.94	-35.40	0.00	-136.20	8.17
Pose 22	-37.71	-6.69	-31.02	0.00	-139.39	8.26
Pose 23	-36.75	-6.70	-30.06	0.00	-139.72	8.28
Pose 21	-37.35	-2.53	-34.82	0.00	-140.22	8.29
Pose 10	-36.27	-11.95	-24.32	0.00	-142.10	8.40
Pose 20	-47.58	-9.04	-38.37	-0.17	-156.29	8.50

*Protein-Ligand total*: total protein–ligand interaction energy ( $Protein-Ligand\ total = Ele + vdW + Hbond$ ), *Ele*: electrostatic energy *vdW*: van der Waals energy, *Hbond*: hydrogen bond energy, *-HEidi*: score calculated from the hydrophobic effect function (taken with negative sign in order to be coherent with energies, where negative numbers mean more favorable energies), *RMSD*: root mean square distance.

The table is sorted by increasing *RMSD*.

If we further assume that docking poses that are close to the pose identified with the crystal structure (low *RMSD*) should be high in the rank given by the docking function, and docking poses that are away from the crystal (high *RMSD*) should get a lower rank, then to evaluate a docking function we could calculate a correlation between the two ranks (obtained one from the docking function and one from *RMSD*).

**Spearman's rank correlation coefficient**<sup>261</sup> is a non-parametric measure of correlation – that is, it assesses how well an arbitrary monotonic function could describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables. In practice, the raw scores are converted to ranks, and the difference  $r_i$  between the ranks of each observation on the two variables are calculated. If there are no tied ranks, then the Spearman's rank correlation coefficient is calculated as:

$$\rho = 1 - \frac{6 \sum r_i^2}{n(n^2 - 1)} \quad (7)$$

where  $r_i$  is the difference between the ranks, and  $n$  the number of objects (the 26 poses = 25 from the docking plus the crystal structure).

A series of possible linear combinations of the equation (6) have been calculated on the docking poses and compared with the original *Yeti* function.<sup>99</sup> Results for GR are reported in Table 16.

Different combinations of coefficients lead to scoring functions that gave often similar, and sometimes higher, correlation coefficients, compared to the original *Yeti* scoring function ( $\rho = 0.776$ ). In the automated docking, it is possible with *Yeti* to give different weights to the different energetic contributions. In this study however, all the weights were set to the default value of 1.0, and a different weighting system was considered only in this final step. Analyzing the combinations that lead to higher correlation coefficients, it would seem a good idea, at least in the studied complex, to underweight van der Waals energy, that is always the highest contribution to the total energy (for the GR, on the average 62% of the overall energy), while the electrostatic contribution should be only slightly underweighted (on the average it represents the 28% of the total energy), and it is often advantageous to overweight the hydrogen bond term, that in this case is responsible only for 10% of the total energy but is often an important contribution to identify correct binding poses. The *HEidi*, as it is derived by an empirical scoring function, has to be largely downscaled to reach the values comparable to the other terms (with a scaling of 0.3 it reaches values comparable, on the average, to the electrostatic energy).

**Table 16.** Different possible linear combinations for the function SDF reported in (5)

<b>GR</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>	<b><math>\rho</math></b>
Yeti	1	1	1	0	0.776
SDF01	1	0.8	2	-0.4	0.780
SDF02	1	0.8	2	-0.2	0.792
SDF03	1	0.8	2	-0.1	0.787
SDF04	1	0.8	1.5	-0.4	0.772
SDF05	1	0.8	1.5	-0.2	0.782
SDF06	1	0.8	1.5	-0.1	0.782
SDF07	1	0.8	1	-0.4	0.790
SDF08	1	0.8	1	-0.2	0.783
SDF09	1	0.8	1	-0.1	0.773
SDF10	1	0.4	2	-0.4	0.793
SDF11	1	0.4	2	-0.2	0.797
SDF12	1	0.4	2	-0.1	0.797
SDF13	1	0.4	1.5	-0.4	0.796
SDF14	1	0.4	1.5	-0.2	0.796
SDF15	1	0.4	1.5	-0.1	0.796
SDF16	1	0.4	1	-0.4	0.804
SDF17	1	0.4	1	-0.2	0.798
SDF18	1	0.4	1	-0.1	0.800
SDF19	1	0.3	2	-0.4	0.793
SDF20	1	0.3	2	-0.2	0.796
SDF21	1	0.3	2	-0.1	0.802
SDF22	1	0.3	1.5	-0.4	0.801
SDF23	1	0.3	1.5	-0.2	0.809
SDF24	1	0.3	1.5	-0.1	0.802
SDF25	1	0.3	1	-0.4	0.796
SDF26	1	0.3	1	-0.2	0.798
SDF27	1	0.3	1	-0.1	0.799
SDF28	0.8	0.8	2	-0.4	0.783
SDF29	0.8	0.8	2	-0.2	0.790
SDF30	0.8	0.8	2	-0.1	0.787
SDF31	0.8	0.8	1.5	-0.4	0.776
SDF32	0.8	0.8	1.5	-0.2	0.783
SDF33	0.8	0.8	1.5	-0.1	0.783
SDF34	0.8	0.8	1	-0.4	0.778
SDF35	0.8	0.8	1	-0.2	0.772
SDF36	0.8	0.8	1	-0.1	0.776
SDF37	0.8	0.4	2	-0.4	0.789
SDF38	0.8	0.4	2	-0.2	0.792
SDF39	0.8	0.4	2	-0.1	0.792
SDF40	0.8	0.4	1.5	-0.4	0.784

Table 16. Continued from previous page

<i>GR</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	$\rho$
SDF41	0.8	0.4	1.5	-0.2	0.794
SDF42	0.8	0.4	1.5	-0.1	0.788
SDF43	0.8	0.4	1	-0.4	0.801
SDF44	0.8	0.4	1	-0.2	0.796
SDF45	0.8	0.4	1	-0.1	0.800
SDF46	0.8	0.3	2	-0.4	0.796
SDF47	0.8	0.3	2	-0.2	0.796
SDF48	0.8	0.3	2	-0.1	0.792
SDF49	0.8	0.3	1.5	-0.4	0.794
SDF50	0.8	0.3	1.5	-0.2	0.797
SDF51	0.8	0.3	1.5	-0.1	0.797
SDF52	0.8	0.3	1	-0.4	0.798
SDF53	0.8	0.3	1	-0.2	0.798
SDF54	0.8	0.3	1	-0.1	0.802
SDF55	0.5	0.8	2	-0.4	0.778
SDF56	0.5	0.8	2	-0.2	0.786
SDF57	0.5	0.8	2	-0.1	0.787
SDF58	0.5	0.8	1.5	-0.4	0.774
SDF59	0.5	0.8	1.5	-0.2	0.782
SDF60	0.5	0.8	1.5	-0.1	0.785
SDF61	0.5	0.8	1	-0.4	0.781
SDF62	0.5	0.8	1	-0.2	0.764
SDF63	0.5	0.8	1	-0.1	0.771
SDF64	0.5	0.4	2	-0.4	0.784
SDF65	0.5	0.4	2	-0.2	0.783
SDF66	0.5	0.4	2	-0.1	0.763
SDF67	0.5	0.4	1.5	-0.4	0.776
SDF68	0.5	0.4	1.5	-0.2	0.785
SDF69	0.5	0.4	1.5	-0.1	0.785
SDF70	0.5	0.4	1	-0.4	0.786
SDF71	0.5	0.4	1	-0.2	0.780
SDF72	0.5	0.4	1	-0.1	0.792
SDF73	0.5	0.3	2	-0.4	0.785
SDF74	0.5	0.3	2	-0.2	0.772
SDF75	0.5	0.3	2	-0.1	0.763
SDF76	0.5	0.3	1.5	-0.4	0.785
SDF77	0.5	0.3	1.5	-0.2	0.786
SDF78	0.5	0.3	1.5	-0.1	0.783
SDF79	0.5	0.3	1	-0.4	0.794
SDF80	0.5	0.3	1	-0.2	0.780
SDF81	0.5	0.3	1	-0.1	0.787

The poses obtained by the automated docking of dexamethasone to the GR and of GW3965 to the LXR were specifically evaluated by the following equation (coefficients of *DSF21*), chosen as a valuable function to score docking poses in the studied systems:

$$DSF21 = 1.0 \cdot E_{Eie} + 0.3 \cdot E_{vdW} + 2 \cdot E_{Hbond} - 0.1 \cdot HEidi \quad (8)$$

where  $E_{Eie}$ : electrostatic energy,  $E_{vdW}$ : van der Waals energy,  $E_{Hbond}$ : hydrogen bond energy,  $HEidi$ : score calculated from the hydrophobic effect function.

The new scoring function *DSF21* yields a ranking of the poses that has a higher correlation with the *RMSD*, when compared to the original *Yeti* scoring function (the correlation coefficients are reported in Table 17). However, the combination of coefficients that leads to *DSF21* might be not appropriate for other systems, with different distribution of energetic contributions (e.g. when the electrostatic contributions are more relevant than the van der Waals contributions). To assess the applicability of *DSF21* further studies are required.

**Table 17.** Correlation coefficient of the *Yeti* function and of the new function *DSF21* with *RMSD*, for GR and LXR complexes

	<i>Yeti</i> $\rho$	<i>DSF21</i> $\rho$
<b>GR</b>	0.776	0.802
<b>LXR</b>	0.646	0.681

The total energies and the contributions obtained from the new scoring function *DSF21* are listed for the GR in Table 18 and for the LXR in Table 19, for all the poses from the automated docking and for the crystal structure. To low values of *RMSD* correspond in general low (favorable) values of *DSF21* and of *HEidi*. This is desirable, because poses that are very close to the crystal structure are expected to have, at least in principle, good protein–ligand energies and good hydrophobic interactions.

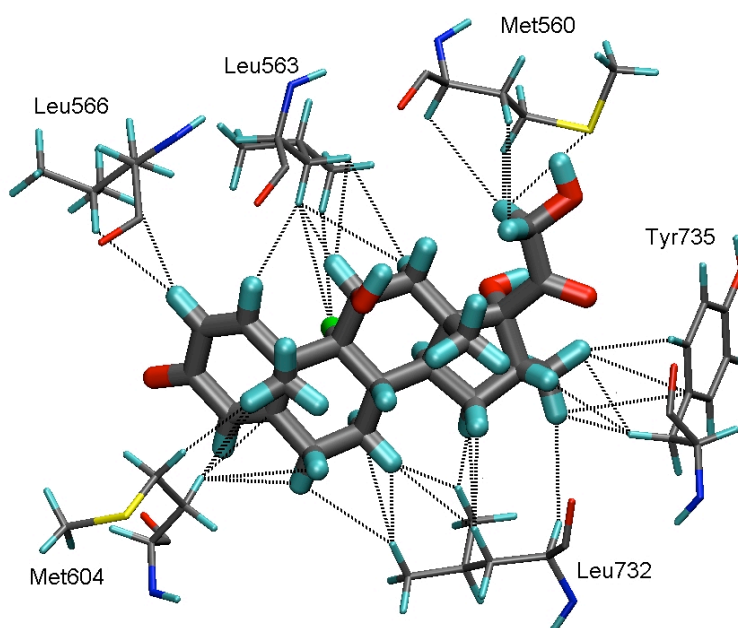
Figures 50 and 52 show some of the good hydrophobic matches that contribute to *HEidi*, as it is calculated for the crystal structures of GR bound to dexamethasone and LXR bound to GW3965. Both ligands can engage in numerous hydrophobic interactions, where, according to *HEidi*, hydrogen atoms play a major role. A visual representation of hydrophobic interactions is also possible through *Maestro* (from the Schrodinger Suite), and it is shown in Figures 51 and 53. Differently from *HEidi*, with *Maestro* the main actors are carbon atoms (and therefore carbon–carbon interactions), and don't involve some atom types (like nitrogen of the protein backbone) that are instead taken into account with *HEidi*. Moreover the calculated interactions are slightly longer ranged. With such exceptions, a similar pattern of hydrophobic interactions is found in both programs, for both receptors.

**Table 18.** The new scores and energetic contributions obtained by *DSF21* on the GR-dexamethasone complex

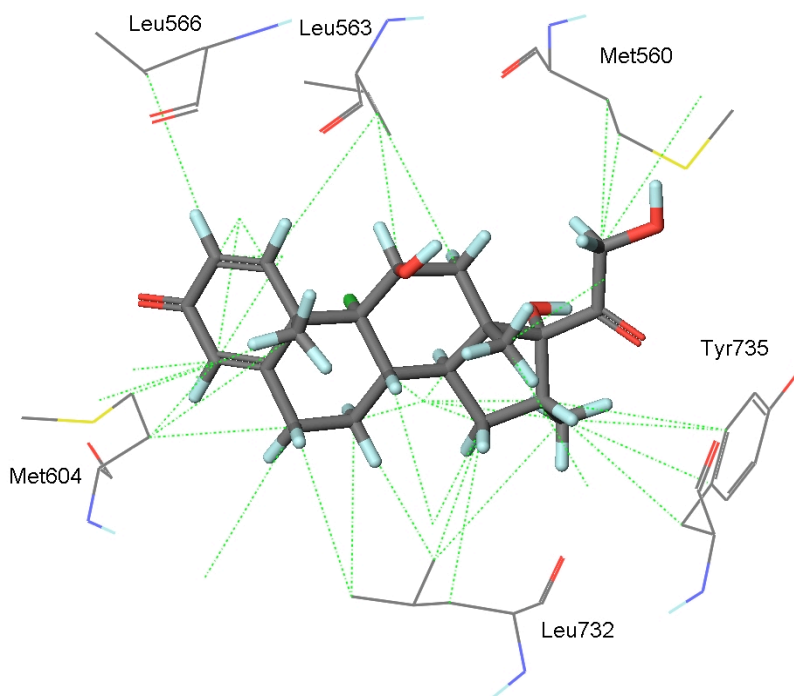
Pose	<i>a•Ele</i>	<i>b•vdW</i>	<i>c•Hbond</i>	<i>-d•HEidi</i>	<i>DSF21</i>	<i>RMSD</i>	<i>R<sub>RMSD</sub></i>	<i>R<sub>DSF21</sub></i>
X-Ray	-15.99	-10.05	-17.14	-4.85	-48.02	0	1	3
Pose 20	-17.29	-8.96	-19.06	-5.09	-50.41	0.81	2	2
Pose 14	-14.87	-9.39	-8.00	-4.92	-37.18	0.85	3	12
Pose 12	-17.62	-9.12	-23.80	-4.78	-55.32	0.86	4	1
Pose 09	-13.63	-9.71	-8.59	-5.16	-37.09	0.9	5	13
Pose 10	-16.45	-9.25	-8.20	-4.97	-38.87	0.94	6	10
Pose 13	-15.42	-9.29	-11.75	-4.92	-41.38	0.97	7	6
Pose 22	-15.73	-9.10	-10.73	-4.93	-40.50	0.97	9	7
Pose 24	-15.12	-9.31	-10.72	-4.93	-40.08	0.97	8	8
Pose 08	-16.57	-9.05	-8.72	-4.99	-39.33	1.01	10	9
Pose 11	-15.86	-9.24	-11.83	-4.69	-41.62	1.01	11	5
Pose 23	-12.71	-9.57	-8.96	-4.86	-36.10	1.07	12	14
Pose 19	-13.42	-10.03	-16.78	-4.81	-45.05	1.11	13	4
Pose 06	-16.06	-9.65	-7.76	-5.39	-38.86	1.12	14	11
Pose 25	-8.37	-7.56	-0.99	-4.19	-21.11	3.15	15	25
Pose 17	-15.53	-5.35	-5.79	-4.33	-31.00	3.88	16	18
Pose 05	-12.42	-7.13	-8.15	-4.14	-31.84	6.98	17	17
Pose 21	-9.36	-7.71	-13.82	-4.25	-35.14	6.99	18	15
Pose 18	-6.76	-8.05	-4.93	-4.36	-24.10	7.02	19	21
Pose 16	-9.78	-7.62	-0.25	-4.32	-21.96	7.05	20	24
Pose 01	-8.59	-8.35	-8.45	-4.25	-29.63	7.08	21	19
Pose 15	-8.57	-6.62	-2.95	-4.15	-22.29	7.12	22	23
Pose 02	-11.19	-8.39	-0.05	-4.02	-23.65	7.34	23	22
Pose 07	-3.48	-8.15	-0.19	-4.43	-16.25	7.35	24	26
Pose 04	-10.27	-7.80	-6.77	-4.19	-29.03	7.44	25	20
Pose 03	-7.73	-8.01	-13.55	-3.83	-33.12	7.45	26	16

*DSF21*: new docking scoring function ( $DSF21 = a \cdot Ele + b \cdot vdW + c \cdot Hbond + d \cdot HEidi$  where  $a=1$ ;  $b=0.3$ ;  $c=2$ ;  $d=-0.1$ ), *Ele*: electrostatic energy *vdW*: van der Waals energy, *Hbond*: hydrogen bond energy, *HEidi*: score calculated from the hydrophobic effect function, *RMSD*: root mean square distance, *R<sub>RMSD</sub>*: rank according *RMSD*, *R<sub>DSF21</sub>*: rank according to *DSF21*.

The table is sorted by increasing *RMSD*.



**Figure 50.** GR bound to dexamethasone. In sticks are shown dexamethasone (thick) and some amino-acid residues (thin) in the binding pocket that contribute to *HEidi*. The black dotted lines connect the pair of protein–ligand atoms that give a good hydrophobic match. For clarity, only some of the protein–ligand matches are shown.



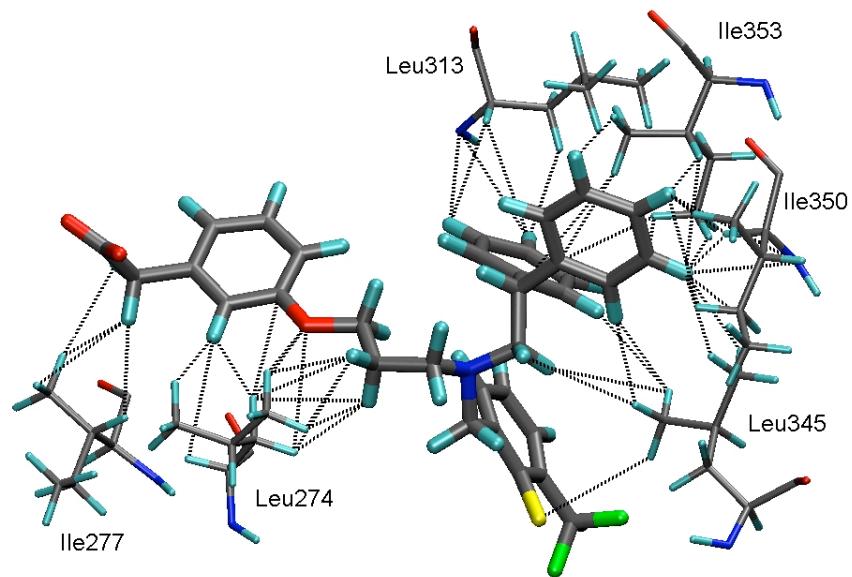
**Figure 51.** GR bound to dexamethasone: with green dotted lines are represented the hydrophobic interactions as identified by *Maestro* (in the protein, only polar hydrogen atoms are displayed).



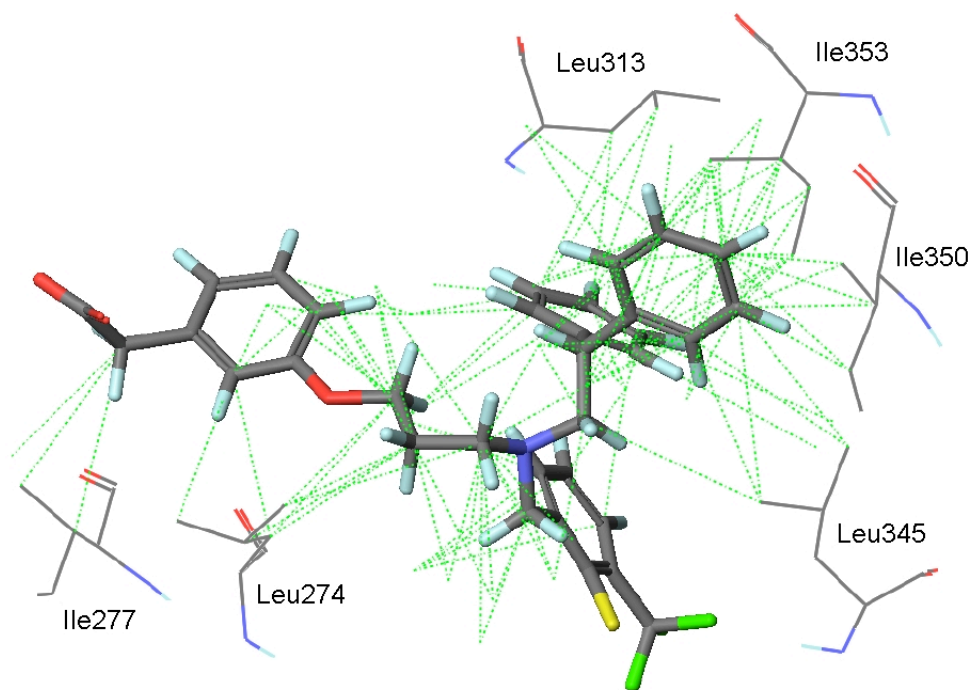
**Table 19.** The new scores and energetic contributions obtained by *DSF21* on the LXR–GW3965 complex

<i>Pose</i>	<i>a•Ele</i>	<i>b•vdW</i>	<i>c•Hbond</i>	<i>d•HEidi</i>	<i>DSF21</i>	<i>RMSD</i>	<i>R<sub>RMSD</sub></i>	<i>R<sub>DSF21</sub></i>
X-Ray	-30.04	-16.58	-1.51	-17.38	-65.51	0.00	1	1
Pose 16	-31.25	-11.47	-2.71	-10.41	-55.85	3.77	2	4
Pose 15	-33.38	-11.02	-1.36	-10.14	-55.89	3.78	3	3
Pose 12	-29.72	-11.63	-2.73	-11.64	-55.71	3.83	4	5
Pose 19	-7.15	-7.06	0.00	-12.95	-27.16	4.48	5	23
Pose 14	-21.51	-8.43	-0.05	-11.20	-41.18	4.95	6	10
Pose 13	-18.26	-6.84	-0.24	-10.96	-36.30	5.02	7	14
Pose 17	-25.15	-12.35	-0.53	-12.17	-50.20	5.28	8	7
Pose 25	-20.99	-8.17	-0.24	-12.77	-42.17	5.74	9	9
Pose 05	-21.29	-10.98	-3.39	-15.51	-51.18	6.05	10	6
Pose 04	-19.52	-9.89	-0.08	-15.65	-45.14	6.05	11	8
Pose 18	-29.43	-9.00	-5.59	-15.77	-59.79	6.17	12	2
Pose 07	-9.69	-9.40	0.00	-15.89	-34.97	7.11	13	16
Pose 06	-7.43	-10.55	0.00	-15.52	-33.51	7.31	14	17
Pose 02	-7.67	-8.42	0.00	-16.77	-32.87	7.49	15	19
Pose 01	-8.36	-11.76	0.00	-16.47	-36.60	7.59	16	12
Pose 08	-16.87	-6.29	0.00	-14.17	-37.33	7.59	17	11
Pose 11	-6.11	-6.22	0.00	-14.02	-26.34	7.64	18	25
Pose 24	-9.96	-11.52	-0.04	-13.53	-35.06	7.74	19	15
Pose 09	-3.44	-8.56	0.00	-13.05	-25.05	7.76	20	26
Pose 03	-3.94	-10.62	0.00	-13.62	-28.18	8.17	21	22
Pose 22	-6.69	-9.31	0.00	-13.94	-29.93	8.26	22	20
Pose 23	-6.70	-9.02	0.00	-13.97	-29.69	8.28	23	21
Pose 21	-2.53	-10.45	0.00	-14.02	-26.99	8.29	24	24
Pose 10	-11.95	-7.30	0.00	-14.21	-33.46	8.40	25	18
Pose 20	-9.05	-11.51	-0.34	-15.63	-36.53	8.50	26	13

*DSF21*: new docking scoring function ( $DSF21 = a \cdot Ele + b \cdot vdW + c \cdot Hbond + d \cdot HEidi$  where  $a=1$ ;  $b=0.3$ ;  $c=2$ ;  $d=-0.1$ ), *Ele*: electrostatic energy *vdW*: van der Waals energy, *Hbond*: hydrogen bond energy, *HEidi*: score calculated from the hydrophobic effect function, *RMSD*: root mean square distance,  $R_{RMSD}$ : rank according *RMSD*,  $R_{DSF21}$ : rank according to *DSF21*. The table is sorted by increasing *RMSD*.



**Figure 52.** LXR bound to GW4965. In sticks are shown GW4965 (thick) and some amino-acid residues (thin) in the binding pocket that contribute to *HEidi*. The black dotted lines connect the pair of protein–ligand atoms that give a good hydrophobic match. For clarity, only some of the protein–ligand matches are shown.



**Figure 53.** LXR bound to GW4965: with green dotted lines are represented the hydrophobic interactions as identified by Maestro (in the protein, only polar hydrogen atoms are displayed).

#### 4.3.4 Applicability of the Hydrophobic Effect Function

A new empirical function has been proposed to evaluate the hydrophobic effect in protein–ligand complexes and implemented in a small computer program. Such function is based on the assignment of a contribution for the favorable or unfavorable hydrophobic matches between atoms of the ligand and atoms of the protein, in a specific conformation. All the contributions are then summed up to give a hydrophobicity score (*HEidi*).

*HEidi* has been added to the electrostatic, van der Waals and hydrogen bond energies (as calculated with *Yeti*) to give a scoring function used to evaluate the poses obtained with automated docking.

The advantage of *HEidi*, when combined to a scoring function including electrostatic, van der Waals and hydrogen bond contributions, consists in the possibility of taking into account, during the docking of compounds, the hydrophobic effect that, especially for lipophilic binding pockets, may be relevant in the identification of feasible binding modes. With appropriate weights for the energetic contributions, it was possible to derive a scoring function (*DSF21*) that led to a better correlation with *RMSD*, when compared to the default *Yeti* scoring function.

However, *HEidi* was applied in the analysis of the docking results to only two real systems (GR bound to dexamethasone and LXR bound to GW3965). To assess the applicability of the method, many different complexes should be considered and their poses evaluated by *HEidi*. A statistical analysis could be then performed to assess if the use of *HEidi* ameliorates the performance in the identification of the correct binding mode or if it improves the enrichment factor. For example, having at disposal many complexes it would be possible to check how often the pose from the crystal structure is associated to the best score, or how often it is within the first three top ranked poses. Moreover, a rank correlation of the docking scoring function with *RMSD* (as done in this thesis) is only an indicative value of performance, being sometimes high values of *RMSD* associated with feasible binding modes (for example for compounds characterized by structural symmetries).

Possible improvements could be achieved with the incorporation of *HEidi* in the docking routine of *Yeti*, to drive the docking algorithm, instead of being used only in the final evaluation of the poses. In this case, when integrated in *Yeti* automated docking, the program should be able to deal with structures with implicit apolar hydrogen atoms (as it is required by *Yeti*), rather than, as it is now required in *HEidi*, with explicit hydrogens. This might be a problem of non-trivial solution, because on one hand it would be computationally quite demanding the hydrogen–addition at each step of the docking routine, and on the other hand, hydrogen atoms and short distances are relevant in *HEidi*.

## 5 Conclusions and Outlook

A mixed-modeling approach was employed to establish a mQSAR model and to quantify the binding affinity of a series of 110 glucocorticoids and 52 liver X receptor ligands, respectively.

The binding mode of the investigated compounds was identified through flexible docking, using both automated and manual protocols. The simulation of local induced fit was achieved by allowing amino-acid side chain flexibility, and proved to be a prerequisite to dock large ligands. The docking to a rigid structure of the glucocorticoid receptor of steroidal glucocorticoids with bulky substituents at the 17 $\beta$ -position did not yield the correct binding mode (i.e. desoxymethasone 21-cinnamate), without a rearrangement of amino-acid side chains in the binding pocket. Molecular-dynamics simulations, performed for a representative of each compounds class, allowed a dynamical characterization of the binding mode. The interactions between the receptors (both the glucocorticoid and the liver X receptors) and their ligands are characterized by numerous hydrophobic interactions. Hydrophilic interactions and hydrogen bonds are responsible for specific recognition of the ligands.

Quantitative structure-activity relationships (QSARs), aimed at the quantification of the binding affinity towards the glucocorticoid and the liver X receptors, were derived on the basis of the alignments obtained from the docking. For the glucocorticoid and liver X receptors, a good correlation was obtained between the experimental and the calculated binding affinities, for both the training set and the test set compounds. Special attention was paid to the model validation: the predictivity was further evaluated by external sets of compounds. For the glucocorticoid receptor, the prediction for a set of eight additional glucocorticoids yielded a good agreement with the experimental binding affinities. The QSAR model was used for simulating and quantifying the binding of 24 psychotropic drugs to the GR. Although their experimental binding affinity is not available in literature, there is evidence that some of the compounds analyzed in this study trigger adverse effects via the GR (Chapter 4.1.6). For the liver X receptor, the model was used to predict the binding affinity of two compounds that were not included in the model development, because of their weak activity. In this case there was only a moderate agreement between predicted and experimental value. Robustness of the models (for the glucocorticoid and liver X receptors) was verified by the consensus with a second QSAR methodology (*Raptor*), which features a fundamentally different scoring function, and sensitivity was tested with a series of scramble tests.

The results suggest that the models can be applied to predict the binding affinity of new drug candidates, in order to design new ligands active towards the glucocorticoid or liver X receptors, or of existing compounds, in order to check out possible interactions. Induced fit, a key mechanism for ligand binding, was explicitly simulated in the docking phase and accounted for in the QSAR models.

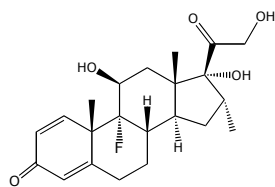
Limitations of the models involve the applicability domain: the predictions for compounds that exceed the physico-chemical, structural or biological space, knowledge or information on which the training set of the model has been developed, are extrapolated. In the QSAR towards the glucocorticoid receptor, the activity of charged compounds seem to be overestimated as the model was trained using predominantly neutral species. The compounds' size is another limitation, because for smaller ligands the automated docking protocol that is used in the prediction of new compounds might not sample enough poses while for large molecules leading to a significant induced-fit, the underlying protocol is unable to simulate and quantify such large conformational changes at the protein.

Limitations concerning the applicability domain can be reduced by the selection of a wider training set, if activity data were available from literature, including a broader diversity of the compounds. The availability of high quality experimental data is in this case the limiting factor for the choice of a wider dataset. In my opinion, an evaluation of the similarity between the compounds of the training set and the new compounds to be tested would be useful to discriminate between interpolated and extrapolated data. The *in silico* prediction of the toxic potential through the *VirtualToxLab*<sup>TM</sup> will become more reliable with the inclusion of a higher number of receptor models, allowing a wider spectrum of adverse effect mechanisms. The application of manual docking is not adequate for the screening of a large number of compounds, where automated procedures are necessary. Improvements of the automated docking process could be obtained by the consideration of neglected aspects, such as the role of entropy, solvent-stripping or the hydrophobic effect in protein-ligand binding. The development and implementation in a small program of an empirical function (*HEidi*: Hydrophobic Effect in Drug Interactions) aiming at the quantification of the hydrophobic effect for scoring protein-ligand binding energies, was an attempt in this direction. Encouraging results were obtained in the evaluation of the docking poses of ligands binding to the glucocorticoid and liver X receptors, when a linear combination of electrostatic, van der Waals, hydrogen bond and *HEidi* contributions was used as a docking scoring function. A more extensive evaluation of *HEidi* is still necessary. The application to many different protein-ligand systems is mandatory to evaluate whether *HEidi* improves the ability to find the bioactive conformation. Possible improvements could be achieved by incorporating *HEidi* in the docking routine, to drive the docking algorithm, instead of being used only in the final evaluation of the poses. In this case, the function should be modified in order to deal with structures characterized by implicit apolar hydrogen atoms.

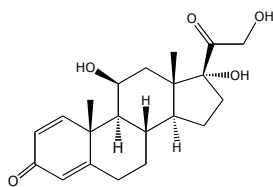
## 6 Appendix A: Chemical Structures, $pK_i$ Values and Scramble Tests of Glucocorticoid Ligands

### A. Chemical Structures of Glucocorticoid Ligands

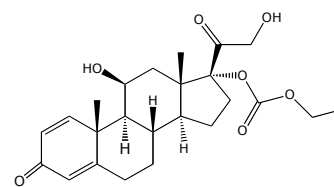
#### 1. Steroids (A01–A17)



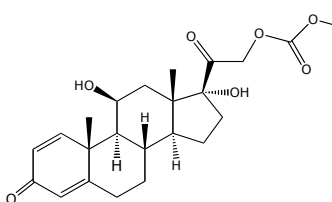
A01



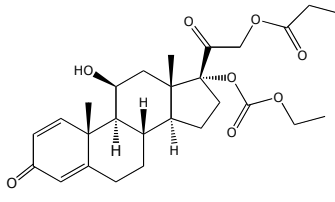
A02



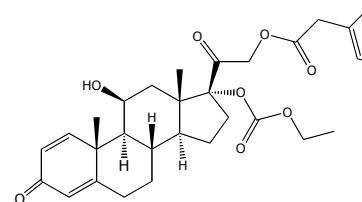
A03



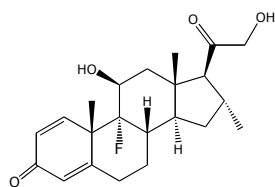
A04



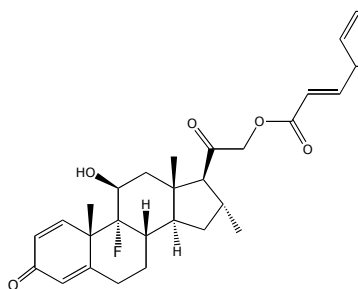
A05



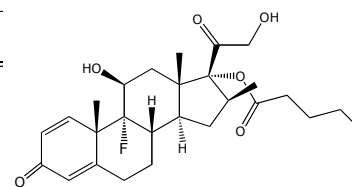
A06



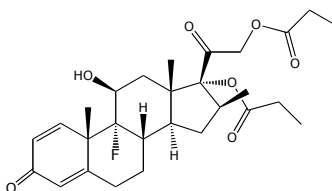
A07



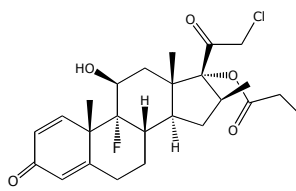
A08



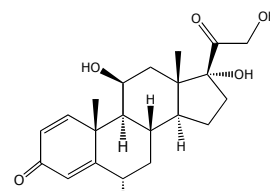
A09



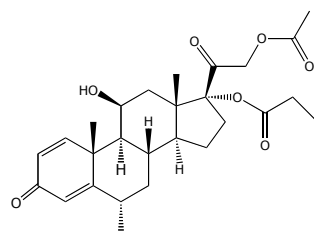
A10



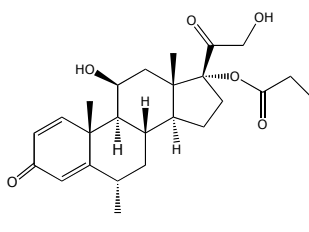
A11



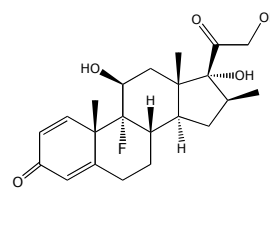
A12



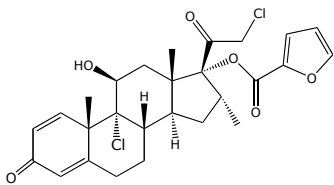
A13



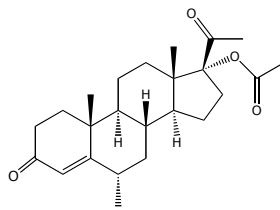
A14



A15

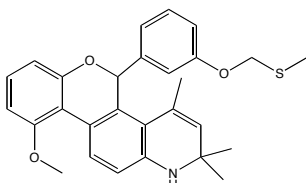


A16

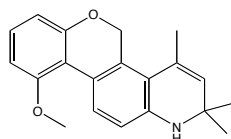


A17

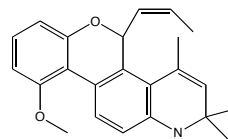
## 2. Quinoline derivatives (B01–B30)



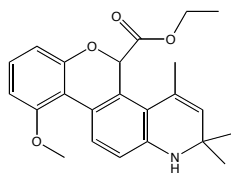
B01



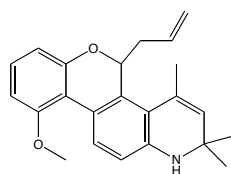
B02



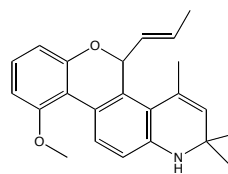
B03



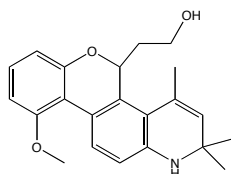
B04



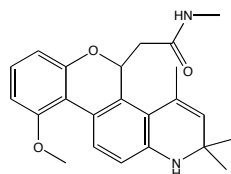
B05



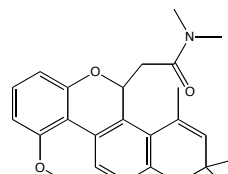
B06



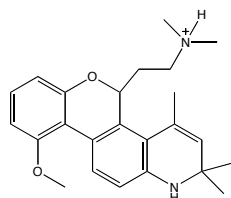
B07



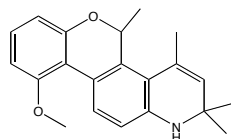
B08



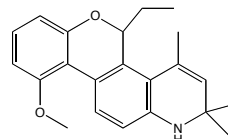
B09



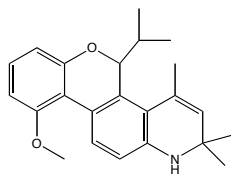
B10



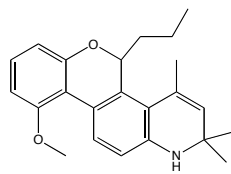
B11



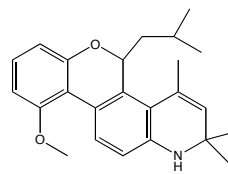
B12



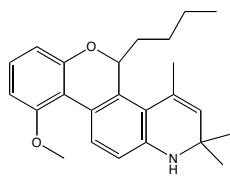
B13



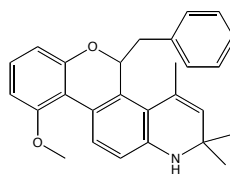
B14



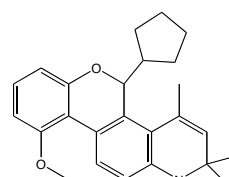
B15



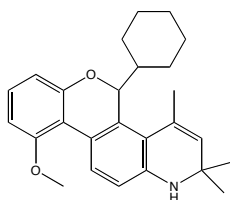
B16



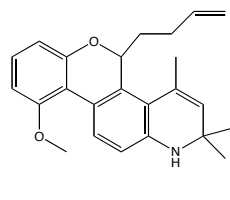
B17



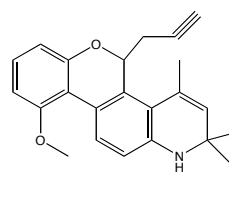
B18



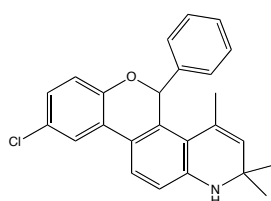
B19



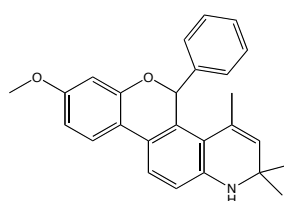
B20



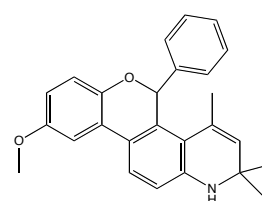
B21



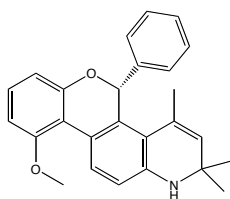
B22



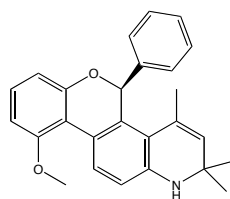
B23



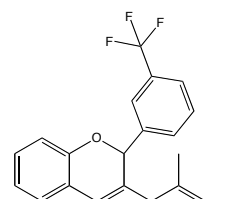
B24



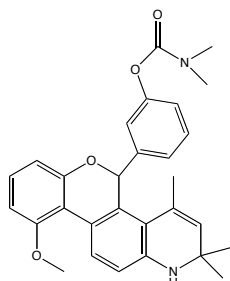
B25



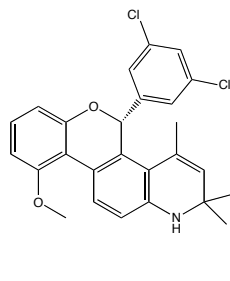
B26



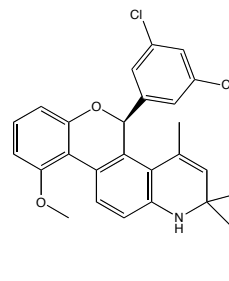
B27



B28



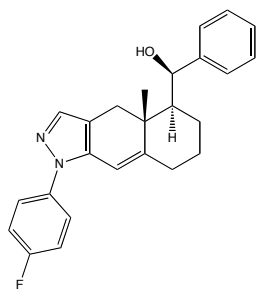
B29



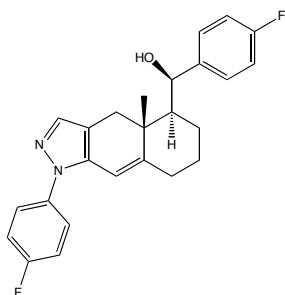
B30



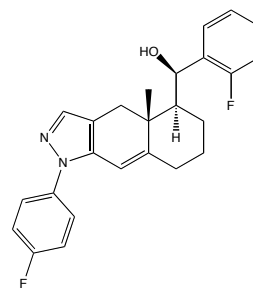
## 3. Fluorophenyl indazole derivatives (C01–C55)



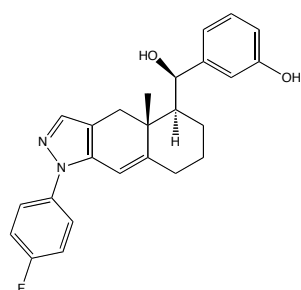
C01



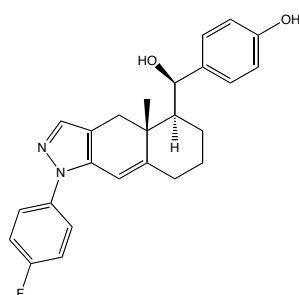
C02



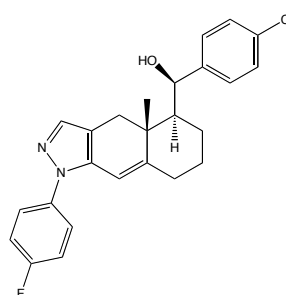
C03



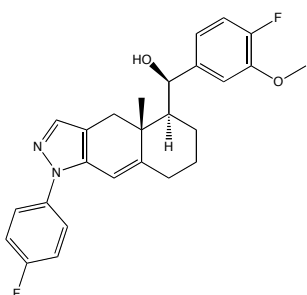
C04



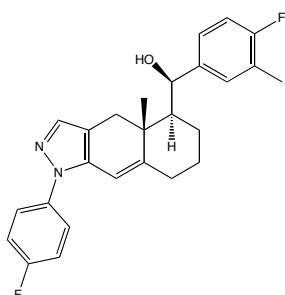
C05



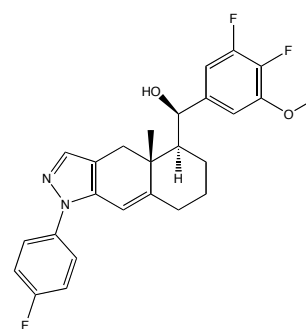
C06



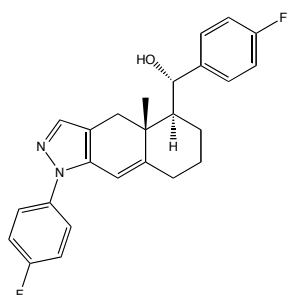
C07



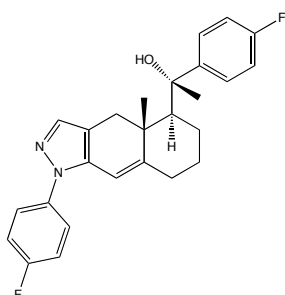
C08



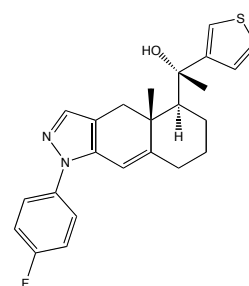
C09



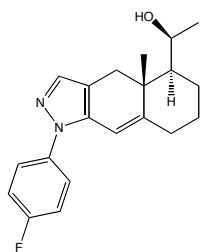
C10



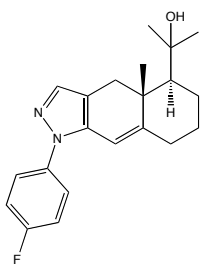
C11



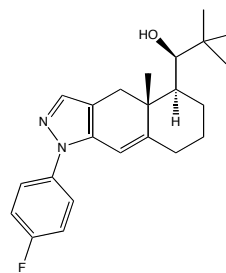
C12



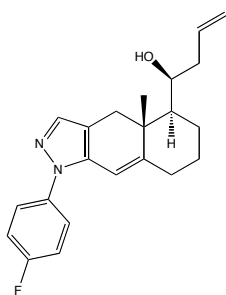
C13



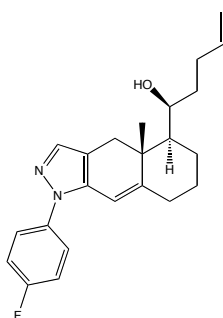
C14



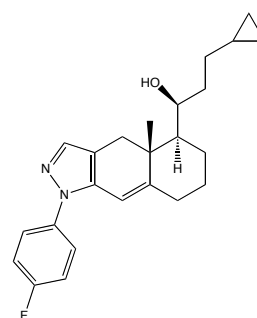
C15



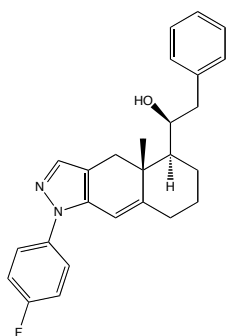
C16



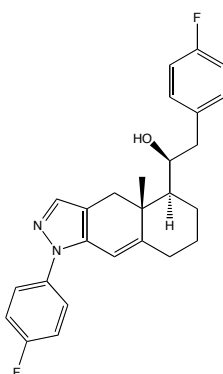
C17



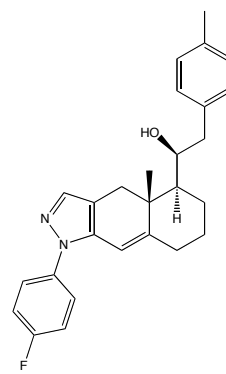
C18



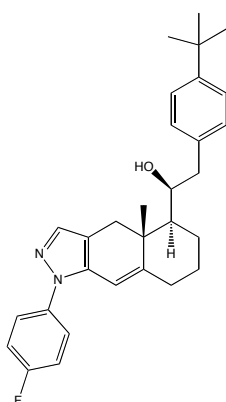
C19



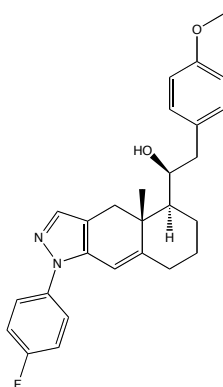
C20



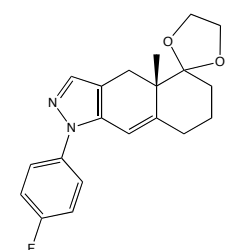
C21



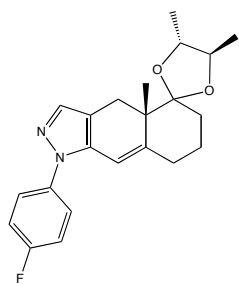
C22



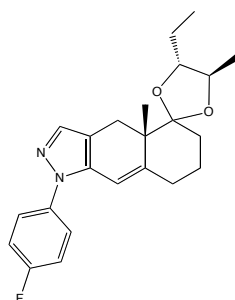
C23



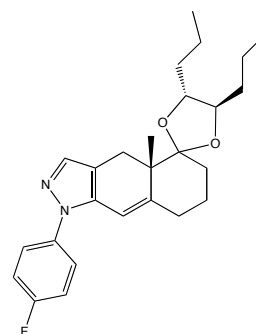
C24



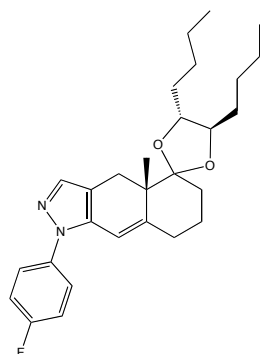
C25



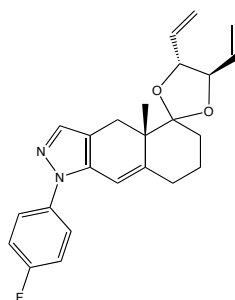
C26



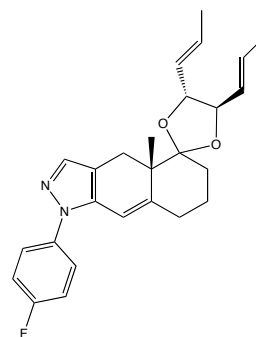
C27



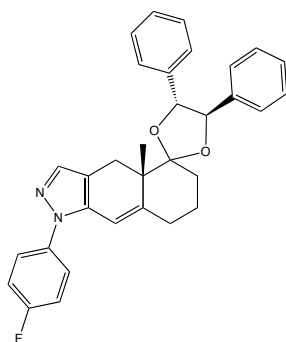
C28



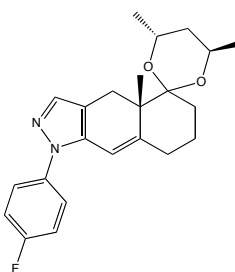
C29



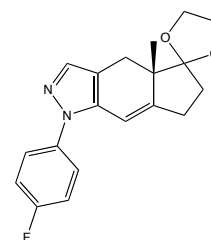
C30



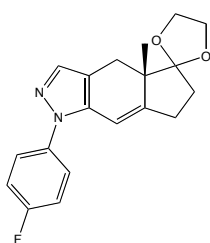
C31



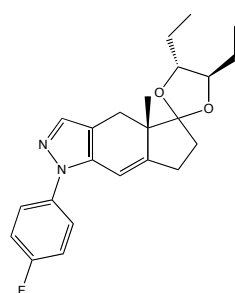
C32



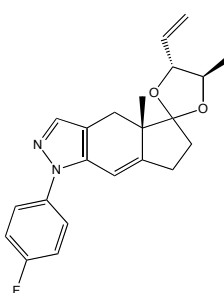
C33



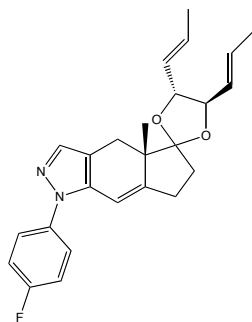
C34



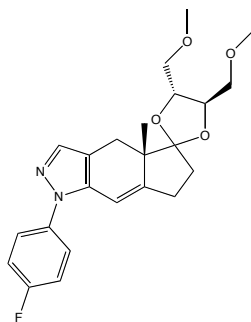
C35



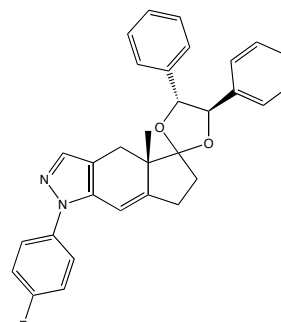
C36



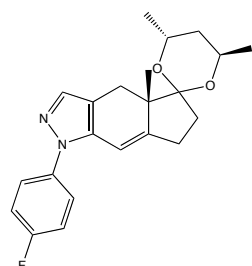
C37



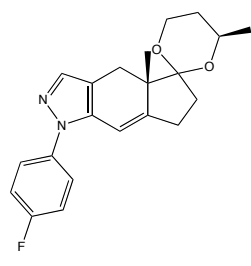
C38



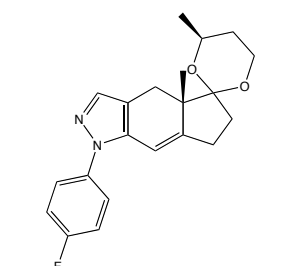
C39



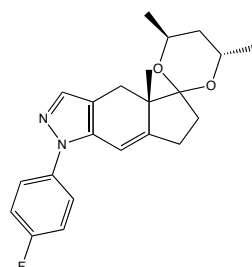
C40



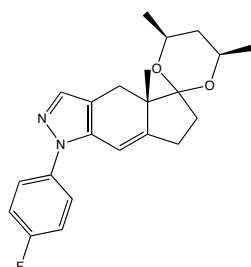
C41



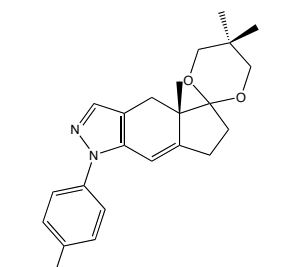
C42



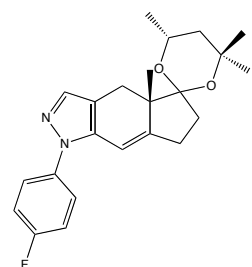
C43



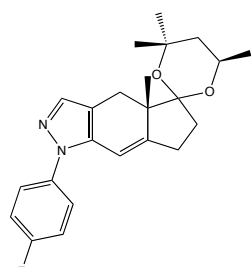
C44



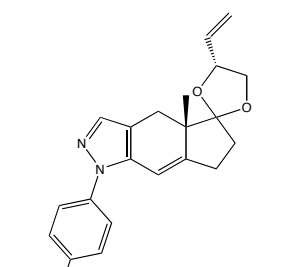
C45



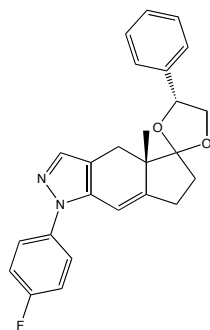
C46



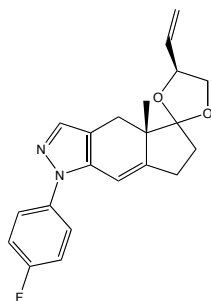
C47



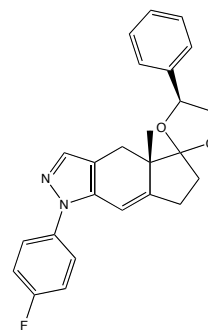
C48



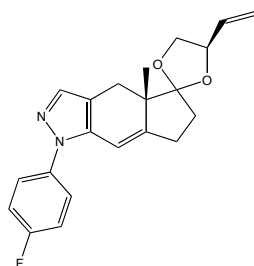
C49



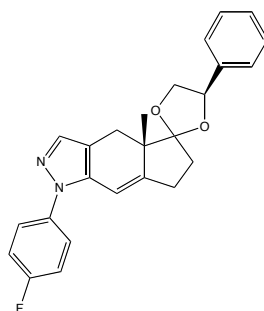
C50



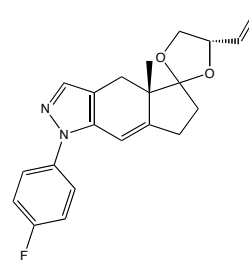
C51



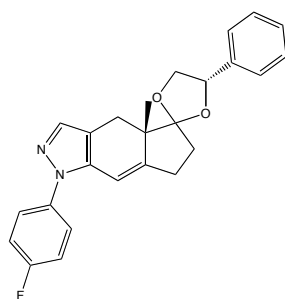
C52



C53

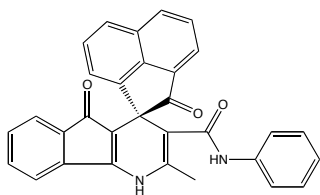


C54

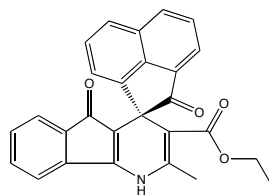


C55

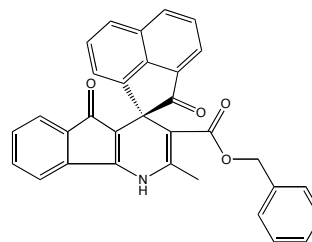
#### 4. Spirocyclic dihydropyridine derivatives (D01–D08)



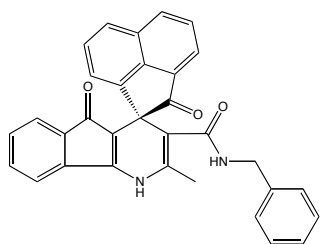
D01



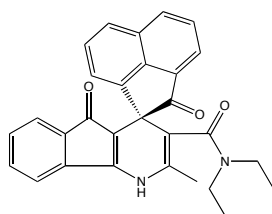
D02



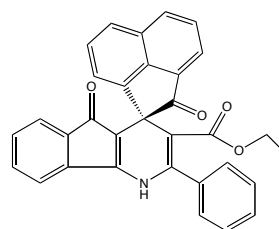
D03



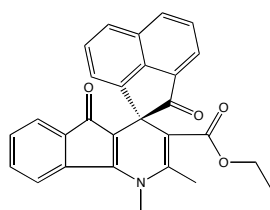
D04



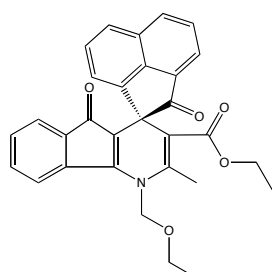
D05



D06

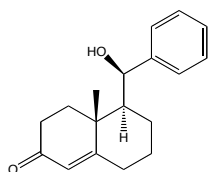


D07

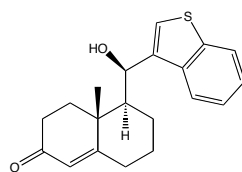


D08

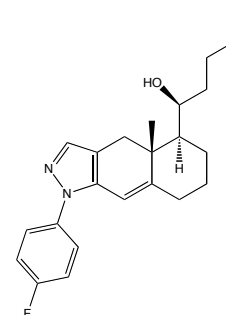
## 5. Prediction set (E01–E08)



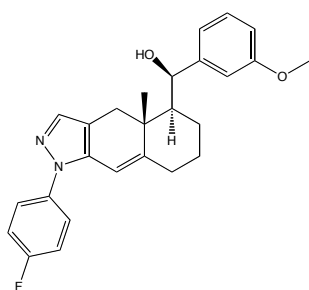
E01



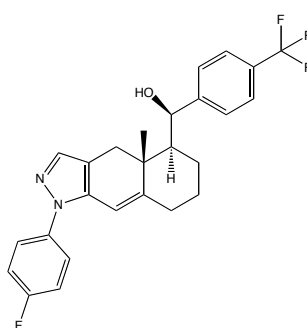
E02



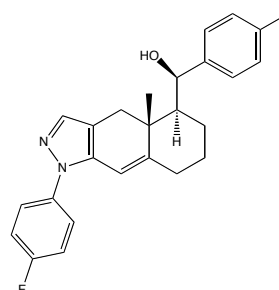
E03



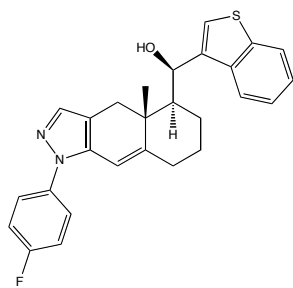
E04



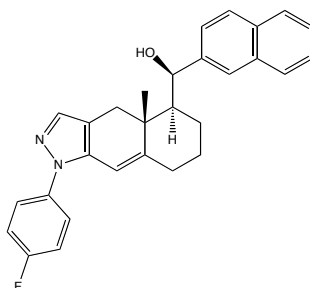
E05



E06

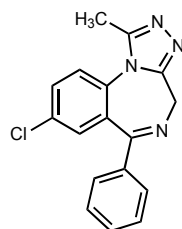


E07

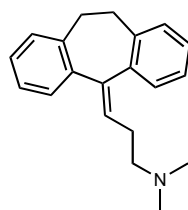


E08

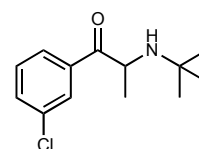
## 6. Psychotropic drugs



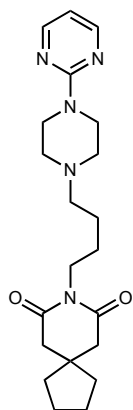
Alprazolam



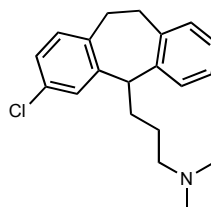
Amitriptyline



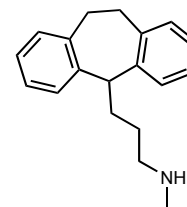
Bupropion (R/S)



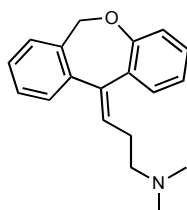
Buspirone



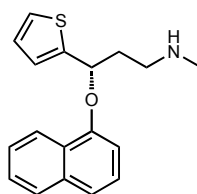
Clomipramine



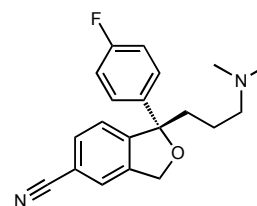
Desipramine



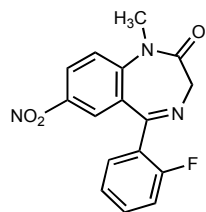
Doxepin



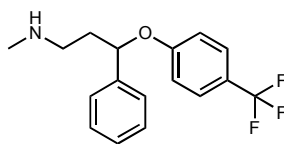
Duloxetine



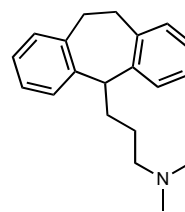
Escitalopram



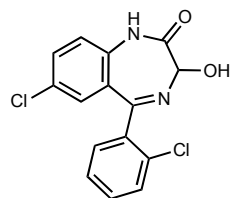
Flunitrazepam



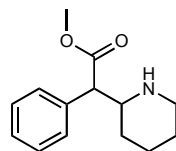
Fluoxetine (R/S)



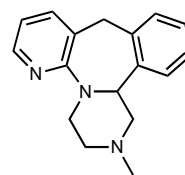
Imipramine



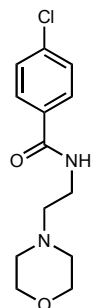
Lorazepam (R/S)



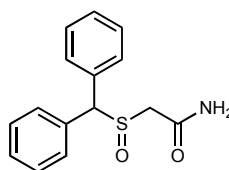
Methylphenidate (R/S)



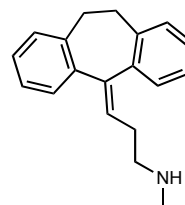
Mirtazapine



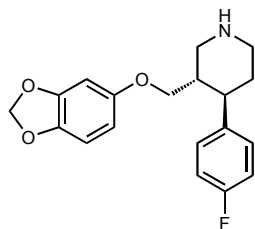
Moclobemide



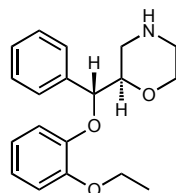
Modafinil



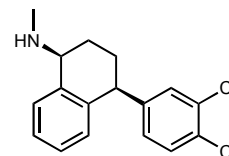
Nortriptyline



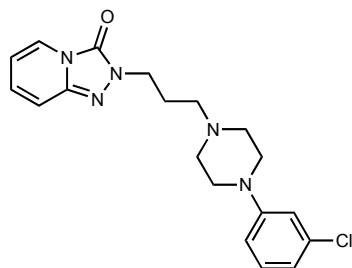
Paroxetine



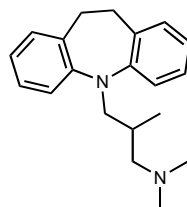
Reboxetine



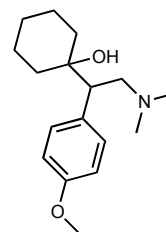
Sertraline



Trazodone



Trimipramine (R/S)



Venlafaxine (R/S)



## B. Comparison of Experimental and Calculated pK<sub>i</sub> Values

Training ligand	Experiment	Quasar: calc	Quasar: res.	Raptor: calc.	Raptor: res.
A01	8.53	7.80	0.72	8.63	-0.10
A03	8.57	8.43	0.14	8.36	0.21
A04	7.69	7.94	-0.25	7.26	0.42
A05	8.40	8.61	-0.21	8.31	0.09
A06	8.05	8.36	-0.31	8.11	-0.06
A07	8.36	7.94	0.42	8.28	0.08
A08	8.23	8.86	-0.63	8.22	0.00
A10	8.71	9.03	-0.32	8.62	0.09
A11	10.33	9.78	0.55	10.46	-0.14
A13	7.76	7.93	-0.17	7.90	-0.14
A15	8.11	7.77	0.34	8.54	-0.43
A16	9.30	9.24	0.06	9.31	-0.01
A17	8.17	7.89	0.27	8.40	-0.24
B01	8.70	8.85	-0.15	8.72	-0.02
B03	8.04	8.27	-0.22	8.24	-0.20
B04	7.90	8.00	-0.09	7.51	0.39
B06	8.35	8.17	0.18	8.83	-0.47
B07	7.28	7.44	-0.16	7.57	-0.29
B09	6.59	6.81	-0.21	7.12	-0.53
B10	7.51	7.68	-0.17	7.14	0.37
B15	8.32	8.17	0.15	8.70	-0.38
B16	8.73	8.27	0.46	8.29	0.44
B17	8.40	8.18	0.21	8.84	-0.44
B18	8.98	8.61	0.37	8.54	0.44
B19	8.31	8.77	-0.46	8.71	-0.41
B20	8.40	8.45	-0.05	8.60	-0.21
B21	8.87	8.34	0.53	8.22	0.65
B22	8.34	7.82	0.51	8.75	-0.41
B23	6.05	6.59	-0.55	6.75	-0.71
B24	8.04	8.37	-0.33	7.37	0.67
B25	6.62	6.33	0.29	6.04	0.58
B26	8.68	8.41	0.26	8.43	0.24
B27	8.26	8.40	-0.14	8.85	-0.59
B28	8.92	8.83	0.09	8.68	0.24
B29	7.02	7.40	-0.38	7.46	-0.44
B30	8.48	8.21	0.27	8.90	-0.42
C02	8.40	8.61	-0.21	8.22	0.19
C03	8.33	8.41	-0.08	8.55	-0.22
C04	7.80	7.55	0.25	8.39	-0.60
C05	8.21	8.00	0.21	8.66	-0.45
C06	8.24	8.85	-0.61	8.52	-0.28

C07	8.30	8.49	-0.20	8.25	0.05
C08	8.36	8.39	-0.03	8.31	0.05
C09	8.91	9.02	-0.11	8.70	0.22
C10	7.89	8.34	-0.45	8.40	-0.52
C11	8.80	9.25	-0.46	9.40	-0.61
C12	9.29	8.73	0.55	8.98	0.30
C14	8.91	8.16	0.76	8.40	0.52
C15	7.30	8.02	-0.72	7.99	-0.69
C16	8.60	8.05	0.56	8.14	0.46
C17	9.15	8.12	1.04	8.39	0.77
C18	9.02	8.67	0.35	8.57	0.44
C20	8.74	8.83	-0.09	8.88	-0.14
C21	8.66	8.23	0.43	8.35	0.31
C22	7.97	8.91	-0.93	8.72	-0.75
C23	8.39	8.40	0.00	8.40	0.00
C25	8.19	8.19	0.00	8.50	-0.31
C26	8.45	8.54	-0.09	8.49	-0.04
C27	8.24	8.45	-0.21	8.59	-0.35
C28	7.85	8.02	-0.17	8.44	-0.59
C29	9.12	8.76	0.36	8.56	0.56
C30	8.80	8.38	0.42	8.34	0.45
C31	8.19	8.86	-0.67	8.71	-0.52
C32	8.27	8.67	-0.40	8.66	-0.39
C38	7.77	7.77	0.00	7.35	0.42
C39	8.19	8.69	-0.50	8.35	-0.16
C40	8.27	8.44	-0.18	8.50	-0.23
C41	7.94	7.81	0.12	8.07	-0.13
C42	7.87	8.02	-0.15	7.65	0.22
C43	8.34	8.44	-0.10	7.98	0.36
C44	8.70	8.35	0.35	8.22	0.48
C45	7.74	7.90	-0.15	8.25	-0.51
C47	8.87	8.81	0.06	8.63	0.24
C48	8.43	8.01	0.42	8.02	0.41
C49	8.27	7.97	0.30	8.20	0.07
C50	8.44	7.81	0.64	7.85	0.60
C51	8.11	8.27	-0.16	8.26	-0.15
C52	7.61	7.58	0.03	7.73	-0.12
C53	7.77	8.15	-0.37	8.00	-0.23
C54	7.50	7.67	-0.16	7.42	0.09
C55	7.38	7.91	-0.53	7.99	-0.61
D01	8.19	7.93	0.26	7.39	0.80
D02	9.05	8.27	0.78	8.38	0.67
D03	8.87	8.41	0.46	8.36	0.51
D04	7.21	7.45	-0.25	7.95	-0.74

D05	6.56	6.64	-0.08	6.88	-0.32
D07	6.04	6.57	-0.53	6.82	-0.78
D08	5.61	6.34	-0.73	6.24	-0.63

Test ligand	Experiment	Quasar: calc	Quasar: res.	Raptor: calc.	Raptor: res.
A02	7.50	7.03	0.46	8.76	-1.26
A09	9.66	9.91	-0.25	9.20	0.46
A12	7.94	7.31	0.63	7.99	-0.06
A14	8.21	8.03	0.18	8.65	-0.44
B02	6.68	7.42	-0.73	6.95	-0.27
B05	8.90	8.14	0.76	8.64	0.26
B08	5.80	6.57	-0.77	6.51	-0.71
B11	7.60	7.73	-0.13	8.42	-0.82
B12	8.16	8.17	-0.02	8.89	-0.74
B13	8.35	8.27	0.09	8.86	-0.51
B14	8.60	8.11	0.49	8.57	0.03
C01	8.27	8.54	-0.27	8.16	0.11
C13	8.25	8.78	-0.53	8.28	-0.03
C19	8.57	8.82	-0.24	8.88	-0.31
C24	8.05	8.13	-0.08	8.19	-0.14
C33	7.33	7.74	-0.42	7.80	-0.47
C34	7.81	8.12	-0.31	8.44	-0.63
C35	8.37	8.36	0.01	8.79	-0.42
C36	8.29	8.26	0.03	8.28	0.01
C37	8.76	8.22	0.54	8.41	0.35
C46	8.85	8.62	0.23	8.66	0.19
D06	7.02	6.71	0.31	5.69	1.33

Prediction ligand	Experiment	Quasar: calc	Quasar: res.	Raptor: calc.	Raptor: res.
E01	5.87	6.40	-0.53	5.83	0.04
E02	6.97	6.65	0.33	7.32	-0.34
E03	9.35	8.15	1.20	7.59	1.76
E04	8.83	8.00	0.83	7.89	0.95
E05	8.05	9.32	-1.28	8.98	-0.93
E06	8.79	8.40	0.40	8.54	0.25
E07	8.94	8.55	0.38	8.91	0.03
E08	8.81	8.50	0.31	8.59	0.22

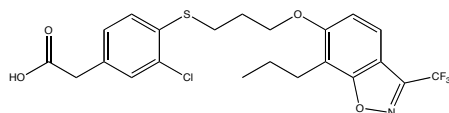
### C. Details of the 20 Scramble Tests

#	Cross-validated $r^2$ (not relevant)	Predictive $r^2$	Mean deviation test set (factor in $K_i$ )	Maximal deviation test set (factor in $K_i$ )
1	0.390	-0.060	6.8	270
2	0.581	-0.411	9.1	530
3	0.664	-0.289	8.2	90
4	0.020	-0.029	6.7	240
5	0.554	-0.699	11	750
6	0.044	-0.029	6.7	240
7	0.780	-0.402	8.9	1,100
8	0.693	-0.475	9.5	180
9	0.676	-0.287	8.2	120
10	0.020	-0.028	6.7	240
11	0.621	-0.162	7.3	230
12	0.398	-0.175	7.5	650
13	0.660	-0.638	11	1,400
14	0.675	-0.425	9.1	120
15	0.266	-0.167	7.5	120
16	0.046	-0.028	6.7	240
17	0.071	-0.029	6.7	240
18	0.417	-0.838	12	1,300
19	0.622	0.375	4.3	36
20	0.044	-0.029	6.7	240

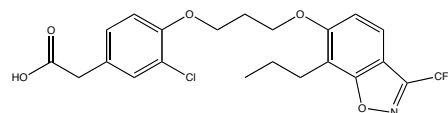
## 7 Appendix B: Chemical Structures, $pK_i$ Values and Scramble Tests of Liver X Receptor Ligands

### A. Chemical Structures of Liver X Receptor Ligands

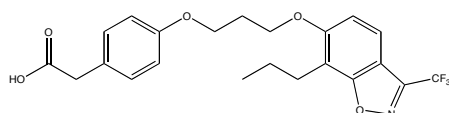
#### 1. Heterocyclic phenylacetic-acid compounds (F01–F29)



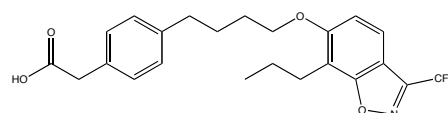
F01



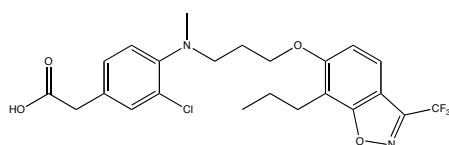
F02



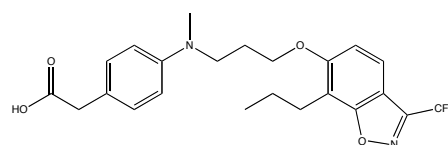
F03



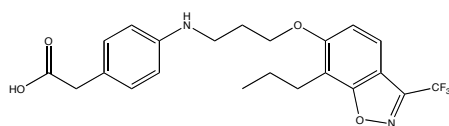
F04



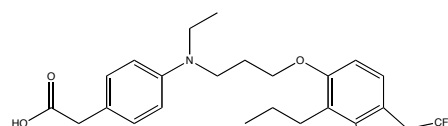
F05



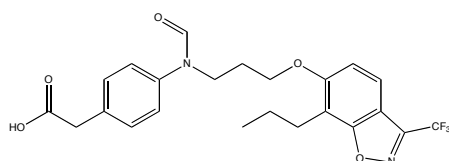
F06



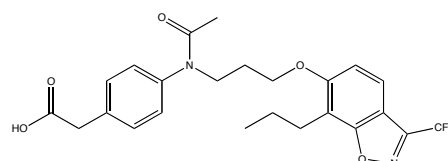
F07



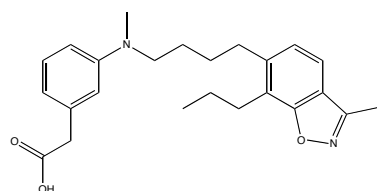
F08



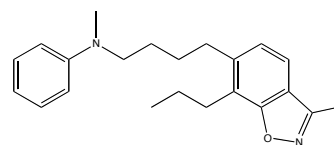
F09



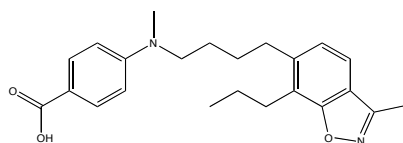
F10



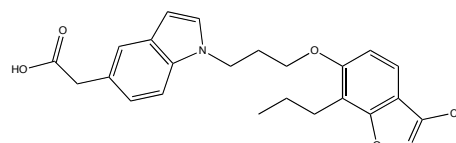
F11



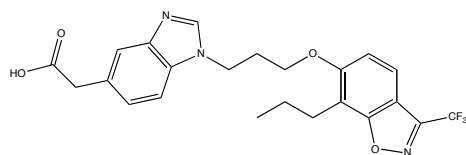
F12



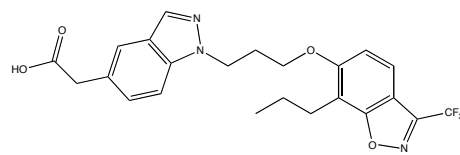
F13



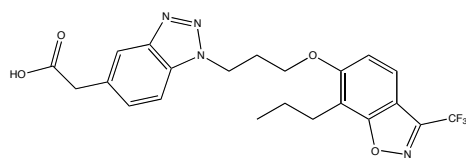
F14



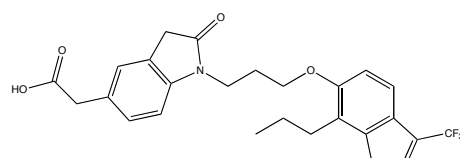
F15



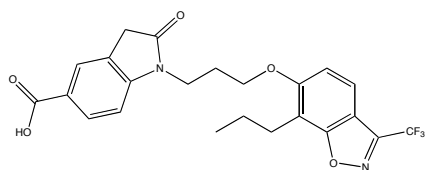
F16



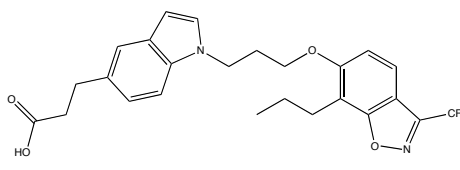
F17



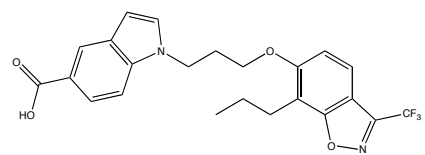
F18



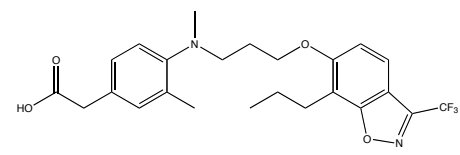
F19



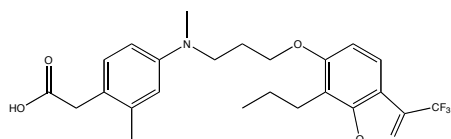
F20



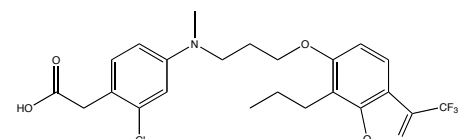
F21



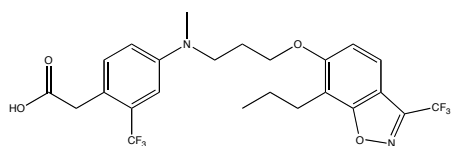
F22



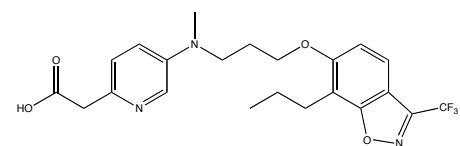
F23



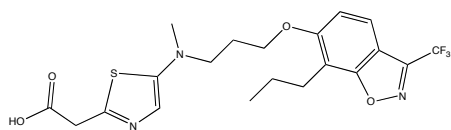
F24



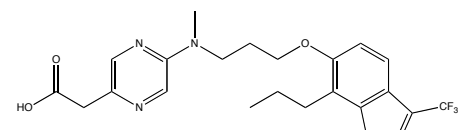
F25



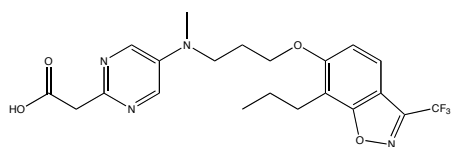
F26



F27

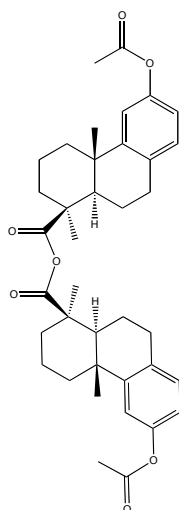


F28

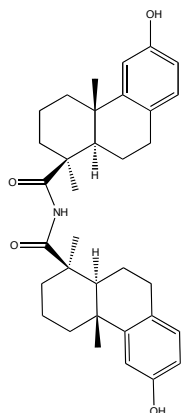


F29

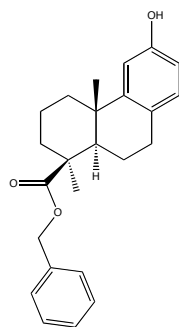
## 2. Heterocyclic phenylacetic-acid compounds (G01–G25)



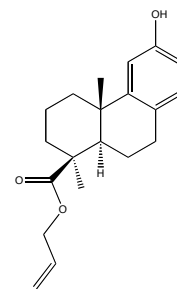
G01



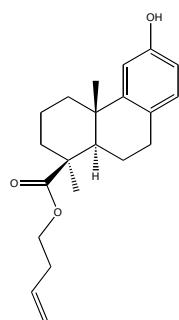
G02



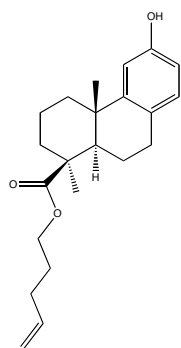
G03



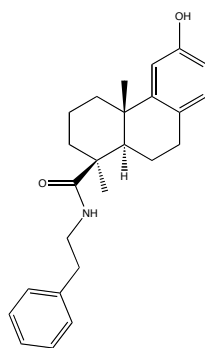
G04



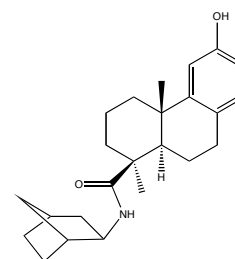
G05



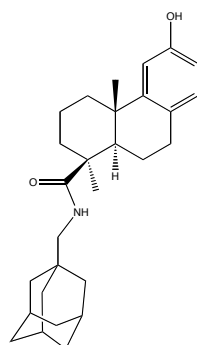
G06



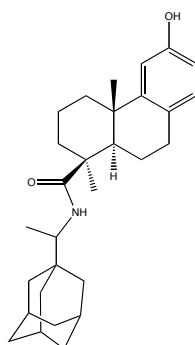
G07



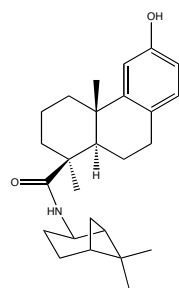
G08



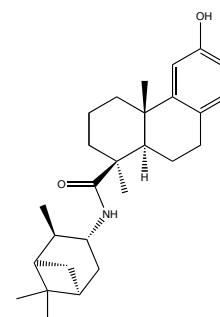
G09



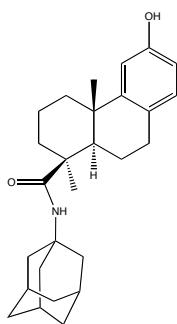
G10



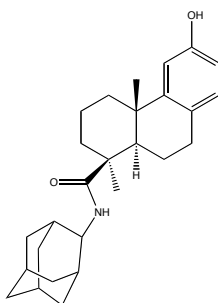
G11



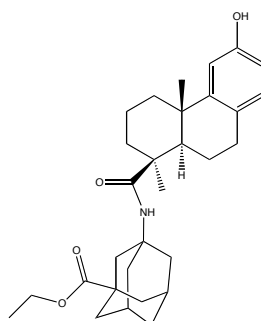
G12



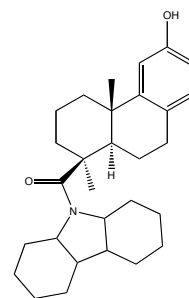
G13



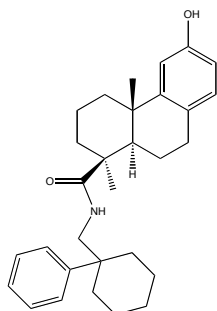
G14



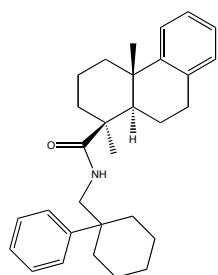
G15



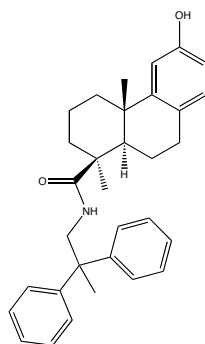
G16



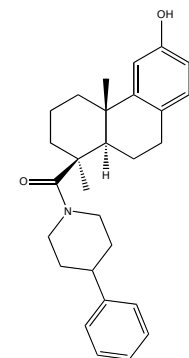
G17



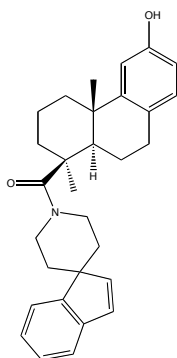
G18



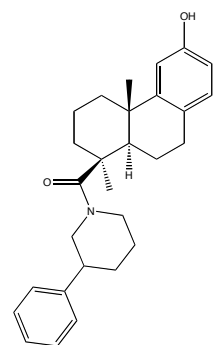
G19



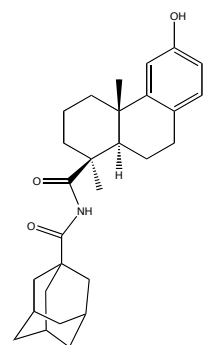
G20



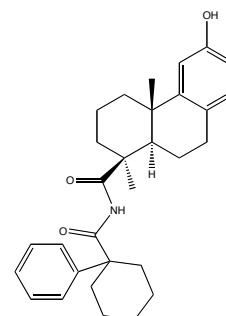
G21



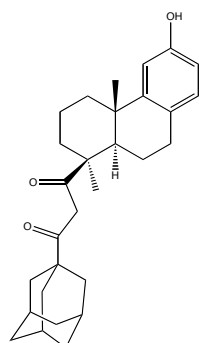
G22



G23



G24



G25



## B. Comparison of Experimental and Calculated $pK_i$ Values

Training ligand	Experiment	Quasar: calc	Quasar: res.	Raptor: calc.	Raptor: res.
F01	8.21	7.77	0.44	8.26	-0.05
F04	6.67	7.42	-0.75	7.28	-0.61
F05	7.43	7.38	0.05	7.88	-0.45
F07	7.71	7.73	-0.02	7.52	0.19
F11	7.67	7.49	0.18	7.81	-0.15
F12	7.03	6.80	0.22	7.14	-0.11
F13	7.38	7.29	0.09	7.26	0.12
F15	7.05	7.57	-0.53	6.91	0.14
F16	7.72	7.97	-0.25	7.39	0.33
F18	8.55	8.21	0.33	8.32	0.22
F19	6.81	7.25	-0.44	6.72	0.10
F20	8.88	8.50	0.38	8.61	0.27
F21	7.36	7.25	0.11	7.21	0.16
F22	7.15	7.38	-0.23	7.67	-0.52
F23	8.24	8.03	0.21	7.78	0.47
F24	8.66	8.24	0.42	8.16	0.51
F26	6.86	7.79	-0.93	7.37	-0.51
F27	7.66	7.71	-0.05	7.56	0.10
F28	8.06	7.73	0.33	7.67	0.39
F29	7.92	7.68	0.24	7.74	0.18
G01	9.36	9.22	0.14	9.50	-0.14
G02	9.66	9.65	0.01	9.83	-0.17
G04	5.48	5.62	-0.14	5.23	0.25
G06	5.96	5.87	0.09	5.37	0.59
G07	6.31	6.57	-0.26	5.99	0.32
G08	6.04	6.39	-0.34	6.31	-0.27
G10	7.40	6.94	0.47	7.21	0.20
G12	7.36	7.24	0.12	7.49	-0.13
G14	7.66	7.19	0.47	7.26	0.40
G15	6.94	7.39	-0.46	7.14	-0.21
G16	7.38	7.31	0.07	7.66	-0.28
G17	8.36	7.73	0.63	8.09	0.27
G18	7.55	7.51	0.04	7.57	-0.02
G19	7.04	7.56	-0.52	7.43	-0.40
G20	7.40	7.39	0.02	7.28	0.13
G21	6.96	7.24	-0.27	7.42	-0.46
G22	7.09	6.83	0.26	7.07	0.02
G23	7.20	7.40	-0.21	7.20	0.00
G24	7.58	7.42	0.16	7.50	0.08
G25	6.66	6.74	-0.09	6.62	0.04

Test ligand	Experiment	Quasar: calc	Quasar: res.	Raptor: calc.	Raptor: res.
F02	7.85	7.65	0.20	7.85	-0.01
F03	7.66	7.94	-0.28	7.73	-0.07
F06	8.26	7.97	0.29	7.65	0.61
F08	7.51	7.91	-0.39	7.87	-0.36
F14	8.46	8.17	0.28	7.72	0.74
F17	7.12	7.65	-0.54	7.24	-0.12
F25	8.40	7.77	0.64	8.64	-0.24
G03	6.20	6.42	-0.22	6.06	0.14
G05	6.07	5.82	0.26	5.09	0.99
G09	7.38	7.44	-0.06	6.67	0.71
G11	7.17	7.14	0.03	7.30	-0.14
G13	7.43	6.84	0.59	6.61	0.82

Prediction ligand	Experiment	Quasar: calc	Quasar: res.	Raptor: calc.	Raptor: res.
F09	< 5.00	7.51	-	8.14	-
F10	< 5.00	7.87	-	8.31	-

### C. Details of the 10 Scramble Tests

#	Cross-validated $r^2$ (not relevant)	Predictive $r^2$	Mean deviation test set (factor in $K_i$ )	Maximal deviation test set (factor in $K_i$ )
1	0.082	0.001	4.4	23.3
2	0.128	0.021	4.1	26.4
3	0.127	0.095	3.6	18.2
4	0.092	-0.349	6.0	24.9
5	0.150	0.328	2.9	16.2
6	0.247	0.437	2.3	11.5
7	0.354	-0.154	4.9	13.4
8	0.141	-0.395	6.2	24.0
9	0.271	0.150	3.5	14.7
10	0.082	0.045	4.2	21.0

## 8 Appendix C: List of Van der Waals Radii

Van der Waals radius [Å]	Equivalent atom types
1.375	Hydrogen bound to carbon
1.000	Any other hydrogen atom
1.800	Carbon sp <sup>3</sup>
1.850	Carbon sp <sup>2</sup>
1.875	Carbon sp
1.850	Nitrogen sp <sup>3</sup>
1.750	Nitrogen sp <sup>2</sup>
1.725	Nitrogen sp
1.650	Oxygen sp <sup>3</sup>
1.600	Oxygen sp <sup>2</sup>
1.550	Fluorine atom
2.100	Phosphorus
2.000	Sulfur
1.950	Chlorine
2.150	Bromine
2.350	Iodine

## 9 References

1. Damstra, T.; Barlow, S.; Bergman, A.; Kavlock, R.; Van der Kraak, G. Global Assessment of the State-of-the-Science of Endocrine Disruptors, IPCS-WHO. In **2002**; p 1.
2. Weatherman, R. V.; Fletterick, R. J.; Scanlan, T. S. Nuclear-receptor ligands and ligand-binding domains. *Annual Review of Biochemistry* **1999**, *68*, 559-581.
3. Vedani, A.; Spreafico, M.; Peristera, O.; Dobler, M.; Smiesko, M. VirtualToxLab - in silico prediction of the endocrine-disrupting potential of drugs and chemicals. *Chimia* **2008**, *62*, 322-328.
4. Vedani, A.; Lill, M. A.; Dobler, M. Predicting the toxic potential of drugs and chemicals in silico. *Altex-Alternativen Zu Tierexperimenten* **2007**, *24*, 63-66.
5. Carson, R. Silent spring. Houghton Mifflin Publisher: Washington D.C., 1962.
6. Foster, W. G.; Agzarian, J. Toward less confusing terminology in endocrine disruptor research. *Journal of Toxicology and Environmental Health-Part B-Critical Reviews* **2008**, *11*, 152-161.
7. Kavlock, R. J.; Daston, G. P.; DeRosa, C.; FennerCrisp, P.; Gray, L. E.; Kaattari, S.; Lucier, G.; Luster, M.; Mac, M. J.; Maczka, C.; Miller, R.; Moore, J.; Rolland, R.; Scott, G.; Sheehan, D. M.; Sinks, T.; Tilson, H. A. Research needs for the risk assessment of health and environmental effects of endocrine disruptors: A report of the US EPA-sponsored workshop. *Environmental Health Perspectives* **1996**, *104*, 715-740.
8. Dosis, I.; Kamarianos, A. Environmental toxic Endocrine Disrupting Compounds (EDCs): Effects on environment, animal production and human. *Journal of the Hellenic Veterinary Medical Society* **2007**, *58*, 321-334.
9. DeRosa, C.; Richter, P.; Pohl, H.; Jones, D. E. Environmental exposures that affect the endocrine system: Public health implications. *Journal of Toxicology and Environmental Health-Part B-Critical Reviews* **1998**, *1*, 3-26.
10. Damstra, T. Potential effects of certain persistent organic pollutants and endocrine disrupting chemicals on the health of children. *Journal of Toxicology-Clinical Toxicology* **2002**, *40*, 457-465.
11. Damstra, T.; Page, S. W.; Herrman, J. L.; Meredith, T. Persistent organic pollutants: potential health effects? *Journal of Epidemiology and Community Health* **2002**, *56*, 824-825.
12. Choi, S. M.; Yoo, S. D.; Lee, B. M. Toxicological characteristics of endocrine-disrupting chemicals: Developmental toxicity, carcinogenicity, and mutagenicity. *Journal of Toxicology and Environmental Health-Part B-Critical Reviews* **2004**, *7*, 1-32.
13. Solomon, G. M.; Schettler, T. Environment and health: 6. Endocrine disruption and potential human health implications. *Canadian Medical Association Journal* **2000**, *163*, 1471-1476.
14. Kaltreider, R. C.; Davis, A. M.; Lariviere, J. P.; Hamilton, J. W. Arsenic alters the function of the glucocorticoid receptor as a transcription factor. *Environmental Health Perspectives* **2001**, *109*, 245-251.

15. Giusti, R. M.; Iwamoto, K.; Hatch, E. E. Diethylstilbestrol revisited - A review of the long-term health-effects. *Annals of Internal Medicine* **1995**, 122, 778-788.
16. Swartz, W. J.; Corkern, M. Effects of methoxychlor treatment of pregnant mice on female offspring of the treated and subsequent pregnancies. *Reproductive Toxicology* **1992**, 6, 431-437.
17. Vomsaal, F. S.; Nagel, S. C.; Palanza, P.; Boechler, M.; Parmigiani, S.; Welshons, W. V. Estrogenic pesticides - binding relative to estradiol in MCF-7 cells and effects of exposure during fetal life on subsequent territorial behavior in male-mice. *Toxicology Letters* **1995**, 77, 343-350.
18. Gray, L. E., Jr. Xenoendocrine disrupters: Laboratory studies on male reproductive effects. *Toxicology Letters (Shannon)* **1998**, 102-103, 331-335.
19. Gray, L. E.; Ostby, J.; Monosson, E.; Kelce, W. R. Environmental antiandrogens: low doses of the fungicide vinclozolin alter sexual differentiation of the male rat. *Toxicology and Industrial Health* **1999**, 15, 48-64.
20. Johansson, M.; Nilsson, S.; Lund, B. O. Interactions between methylsulfonyl PCBs and the glucocorticoid receptor. *Environmental Health Perspectives* **1998**, 106, 769-772.
21. Jacobson, J. L.; Jacobson, S. W. Intellectual impairment in children exposed to polychlorinated biphenyls in utero. *New England Journal of Medicine* **1996**, 335, 783-789.
22. Zoeller, R. T.; Dowling, A. L. S.; Vas, A. A. Developmental exposure to polychlorinated biphenyls exerts thyroid hormone-like effects on the expression of RC3/neurogranin and myelin basic protein messenger ribonucleic acids in the developing rat brain. *Endocrinology* **2000**, 141, 181-189.
23. Bradlow, H. L.; Davis, D. L.; Lin, G.; Sepkovic, D.; Tiwari, R. Effects of pesticides on the ratio of 16-alpha/2-hydroxyestrone - a biologic marker of breast-cancer risk. *Environmental Health Perspectives* **1995**, 103, 147-150.
24. Tran, D. Q.; Kow, K. Y.; McLachlan, J. A.; Arnold, S. F. The inhibition of estrogen receptor-mediated responses by chloro-S-triazine-derived compounds is dependent on estradiol concentration in yeast. *Biochemical and Biophysical Research Communications* **1996**, 227, 140-146.
25. Cooper, R. L.; Goldman, J. M.; Stoker, T. E. Neuroendocrine and reproductive effects of contemporary-use pesticides. *Toxicology and Industrial Health* **1999**, 15, 26-36.
26. Donna, A.; Crosignani, P.; Robutti, F.; Betta, P. G.; Bocca, R.; Mariani, N.; Ferrario, F.; Fissi, R.; Berrino, F. Triazine herbicides and ovarian epithelial neoplasms. *Scandinavian Journal of Work Environment & Health* **1989**, 15, 47-53.
27. Kettles, M. A.; Browning, S. R.; Prince, T. S.; Horstman, S. W. Triazine herbicide exposure and breast cancer incidence: An ecologic study of Kentucky counties. *Environmental Health Perspectives* **1997**, 105, 1222-1227.
28. Brown, N. M.; Manzollilo, P. A.; Zhang, J. X.; Wang, J.; Lamartiniere, C. A. Prenatal TCDD and predisposition to mammary cancer in the rat. *Carcinogenesis* **1998**, 19, 1623-1629.

29. Mably, T. A.; Moore, R. W.; Peterson, R. E. In utero and lactational exposure of male-rats to 2,3,7,8-tetrachlorodibenzo-para-dioxin. 1.Effects on androgenic status. *Toxicology and Applied Pharmacology* **1992**, 114, 97-107.
30. Egeland, G. M.; Sweeney, M. H.; Fingerhut, M. A.; Wille, K. K.; Schnorr, T. M.; Halperin, W. E. Total serum testosterone and gonadotropins in workers exposed to dioxin. *American Journal of Epidemiology* **1994**, 139, 272-281.
31. Nagayama, J.; Okamura, K.; Iida, T.; Hirakawa, H.; Matsueda, T.; Tsuji, H.; Hasegawa, M.; Sato, K.; Ma, H. Y.; Yanagawa, T.; Igarashi, H.; Fukushige, J.; Watanabe, T. Postnatal exposure to chlorinated dioxins and related chemicals on thyroid hormone status in Japanese breast-fed infants. *Chemosphere* **1998**, 37, 1789-1793.
32. Steenland, K.; Piacitelli, L.; Deddens, J.; Fingerhut, M.; Chang, L. I. Cancer, heart disease, and diabetes in workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Journal of the National Cancer Institute* **1999**, 91, 779-786.
33. Endocrine Disruptor Screening Committee, E. D. S. a. T. A. Endocrine disruptor screening and testing advisory committee final report. <http://www.epa.gov/scipoly/oscpendo/>
34. OECD. Endocrine disruptor testing and assessment. [http://www.oecd.org/document/62/0,2340,en\\_2649\\_34377\\_2348606\\_1\\_1\\_1\\_1,%2000.html](http://www.oecd.org/document/62/0,2340,en_2649_34377_2348606_1_1_1_1,%2000.html)
35. European Commission. REACH. [http://ec.europa.eu/environment/chemicals/reach/reach\\_intro.htm](http://ec.europa.eu/environment/chemicals/reach/reach_intro.htm)
36. Swiss National Science Foundation. National Research Programme 50. <http://www.nrp50.ch/>
37. Laudet, V.; Gronemeyer, H. The nuclear receptors factsbook. Academic Press Publisher: London, **2002**.
38. Evans, R. M. The steroid and thyroid-hormone receptor superfamily. *Science* **1988**, 240, 889-895.
39. Chawla, A.; Repa, J. J.; Evans, R. M.; Mangelsdorf, D. J. Nuclear receptors and lipid physiology: Opening the X-files. *Science* **2001**, 294, 1866-1870.
40. Kastner, P.; Mark, M.; Chambon, P. Nonsteroid nuclear receptors - what are genetic-studies telling us about their role in real-life. *Cell* **1995**, 83, 859-869.
41. Mangelsdorf, D. J.; Thummel, C.; Beato, M.; Herrlich, P.; Schutz, G.; Umesono, K.; Blumberg, B.; Kastner, P.; Mark, M.; Chambon, P.; Evans, R. M. The nuclear receptor superfamily - the 2nd decade. *Cell* **1995**, 83, 835-839.
42. Glass, C. K.; Ogawa, S. Combinatorial roles of nuclear receptors in inflammation and immunity. *Nature Reviews Immunology* **2006**, 6, 44-55.
43. Galon, J.; Franchimont, D.; Hiroi, N.; Frey, G.; Boettner, A.; Ehrhart-Bornstein, M.; O'Shea, J. J.; Chrousos, G. P.; Bornstein, S. R. Gene profiling reveals unknown enhancing and suppressive actions of glucocorticoids on immune cells. *Faseb Journal* **2002**, 16, 61-71.
44. Joseph, S. B.; Castrillo, A.; Laffitte, B. A.; Mangelsdorf, D. J.; Tontonoz, P. Reciprocal regulation of inflammation and lipid metabolism by liver X receptors. *Nature Medicine* **2003**, 9, 213-219.
45. Ogawa, S.; Lozach, J.; Benner, C.; Pascual, G.; Tangirala, R. K.; Westin, S.; Hoffmann, A.; Subramaniam, S.; David, M.; Rosenfeld, M. G.; Glass, C. K.

- Molecular determinants of crosstalk between nuclear receptors and toll-like receptors. *Cell* **2005**, 122, 707-721.
46. Lu, N. Z.; Wardell, S. E.; Burnstein, K. L.; Defranco, D.; Fuller, P. J.; Giguere, V.; Hochberg, R. B.; McKay, L.; Renoir, J. M.; Weigel, N. L.; Wilson, E. M.; McDonnell, D. P.; Cidlowski, J. A. International Union of Pharmacology. LXV. The pharmacology and classification of the nuclear receptor superfamily: Glucocorticoid, mineralocorticoid, progesterone, and androgen receptors. *Pharmacological Reviews* **2006**, 58, 782-797.
  47. Barnes, P. J.; Pedersen, S.; Busse, W. W. Efficacy and safety of inhaled corticosteroids - New developments. *American Journal of Respiratory and Critical Care Medicine* **1998**, 157, S1-S53.
  48. Rhen, T.; Cidlowski, J. A. Antiinflammatory action of glucocorticoids - New mechanisms for old drugs. *New England Journal of Medicine* **2005**, 353, 1711-1723.
  49. Oakley, R. H.; Sar, M.; Cidlowski, J. A. The human glucocorticoid receptor beta isoform - Expression, biochemical properties, and putative function. *Journal of Biological Chemistry* **1996**, 271, 9550-9559.
  50. Lewis-Tuffin, L. J.; Jewell, C. M.; Bienstock, R. J.; Collins, J. B.; Cidlowski, J. A. Human glucocorticoid receptor beta binds RU-486 and is transcriptionally active. *Molecular and Cellular Biology* **2007**, 27, 2266-2282.
  51. Bamberger, C. M.; Bamberger, A. M.; Decastro, M.; Chrousos, G. P. Glucocorticoid receptor-beta, a potential endogenous inhibitor of glucocorticoid action in humans. *Journal of Clinical Investigation* **1995**, 95, 2435-2441.
  52. Oakley, R. H.; Jewell, C. M.; Yudt, M. R.; Bofetiado, D. M.; Cidlowski, J. A. The dominant negative activity of the human glucocorticoid receptor beta isoform - Specificity and mechanisms of action. *Journal of Biological Chemistry* **1999**, 274, 27857-27866.
  53. Lewis-Tuffin, L. J.; Cidlowski, J. A. The physiology of human glucocorticoid receptor beta (hGR beta) and glucocorticoid resistance. *Basic and Clinical Aspects of Neuroendocrine Immunology in Rheumatic Diseases* **2006**, 1069, 1-9.
  54. Beato, M.; Herrlich, P.; Schutz, G. Steroid-hormone receptors - many actors in search of a plot. *Cell* **1995**, 83, 851-857.
  55. Karin, M. New twists in gene regulation by glucocorticoid receptor: Is DNA binding dispensable? *Cell* **1998**, 93, 487-490.
  56. Yang-Yen, H. F.; Chambard, J. C.; Sun, Y. L.; Smeal, T.; Schmidt, T. J.; Drouin, J.; Karin, M. Transcriptional interference between C-JUN and the glucocorticoid receptor - mutual inhibition of DNA-binding due to direct protein-protein interaction. *Cell* **1990**, 62, 1205-1215.
  57. Jonat, C.; Rahmsdorf, H. J.; Park, K. K.; Cato, A. C. B.; Gebel, S.; Ponta, H.; Herrlich, P. Antitumor promotion and antiinflammation: down-modulation of AP-1 (Fos/Jun) activity by glucocorticoid hormone. *Cell* **1990**, 62, 1189-1204.
  58. vanderBurg, B.; Liden, J.; Okret, S.; Delaunay, F.; Wissink, S.; vanderSaag, P. T.; Gustafsson, J. A. Nuclear factor-kappa B repression in antiinflammation and immunosuppression by glucocorticoids. *Trends in Endocrinology and Metabolism* **1997**, 8, 152-157.



59. Brattsand, R.; Linden, M. Cytokine modulation by glucocorticoids: Mechanisms and actions in cellular studies. *Alimentary Pharmacology & Therapeutics* **1996**, *10*, 81-90.
60. Scott, D. K.; Stromstedt, P. E.; Wang, J. C.; Granner, D. K. Further characterization of the glucocorticoid response unit in the phosphoenolpyruvate carboxykinase gene. The role of the glucocorticoid receptor-binding sites. *Molecular Endocrinology* **1998**, *12*, 482-491.
61. Meyer, T.; Gustafsson, J. A.; CarlstedtDuke, J. Glucocorticoid-dependent transcriptional repression of the osteocalcin gene by competitive binding at the TATA box. *DNA and Cell Biology* **1997**, *16*, 919-927.
62. Hofbauer, L. C.; Gori, F.; Riggs, B. L.; Lacey, D. L.; Dunstan, C. R.; Spelsberg, T. C.; Khosla, S. Stimulation of osteoprotegerin ligand and inhibition of osteoprotegerin production by glucocorticoids in human osteoblastic lineage cells: Potential paracrine mechanisms of glucocorticoid-induced osteoporosis. *Endocrinology* **1999**, *140*, 4382-4389.
63. Sasaki, N.; Kusano, E.; Ando, Y.; Yano, K.; Tsuda, E.; Asano, Y. Glucocorticoid decreases circulating osteoprotegerin (OPG): possible mechanism for glucocorticoid induced osteoporosis. *Nephrology Dialysis Transplantation* **2001**, *16*, 479-482.
64. Max, S. R. Glucocorticoid-mediated induction of glutamine-synthetase in skeletal-muscle. *Medicine and Science in Sports and Exercise* **1990**, *22*, 325-330.
65. Dhaese, J.; Rutschmann, M.; Dahlmann, B.; Hinssen, H. Activity of a gelsolin-like actin modulator in rat skeletal-muscle under protein catabolic conditions. *Biochemical Journal* **1987**, *248*, 397-402.
66. Suino-Powell, K.; Xu, Y.; Zhang, C.; Tao, Y. G.; Tolbert, W. D.; Simons, S. S.; Xu, H. E. Doubling the size of the glucocorticoid receptor ligand binding pocket by deacylcortivazol. *Molecular and Cellular Biology* **2008**, *28*, 1915-1923.
67. Litwack, G. *Nuclear receptor coregulators*, 68. Elsevier: Philadelphia, **2004**.
68. Bledsoe, R. K.; Montana, V. G.; Stanley, T. B.; Delves, C. J.; Apolito, C. J.; McKee, D. D.; Consler, T. G.; Parks, D. J.; Stewart, E. L.; Willson, T. M.; Lambert, M. H.; Moore, J. T.; Pearce, K. H.; Xu, H. E. Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **2002**, *110*, 93-105.
69. Lin, C. W.; Nakane, M.; Stashko, M.; Falls, D.; Kuk, J.; Miller, L.; Huang, R.; Tyree, C.; Miner, J. N.; Rosen, J.; Kym, P. R.; Coghlan, M. J.; Carter, G.; Lane, B. C. Trans-activation and repression properties of the novel nonsteroid glucocorticoid receptor ligand 2,5-dihydro-9-hydroxy-10-methoxy-2,2,4-trimethyl-5-(1-methylcyclohexen-3 -y1)-1H-[1]benzopyrano[3,4-f]quinoline (A276575) and its four stereoisomers. *Molecular Pharmacology* **2002**, *62*, UNSP 1606/997371.
70. Pleyer, U.; Yang, J.; Knapp, S.; Schacke, H.; Schmees, N.; Orlic, N.; Otasevic, L.; De Ruijter, M.; Ritter, T.; Keipert, S. Effects of a selective glucocorticoid receptor agonist on experimental keratoplasty. *Graefes Archive for Clinical and Experimental Ophthalmology* **2005**, *243*, 450-455.
71. Schacke, H.; Schottelius, A.; Docke, W. D.; Strehlke, P.; Jaroch, S.; Schmees, N.; Rehwinkel, H.; Hennekes, H.; Asadullah, K. Dissociation of transactivation from transrepression by a selective glucocorticoid receptor agonist leads to

- separation of therapeutic effects from side effects. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, 101, 227-232.
72. Janowski, B. A.; Willy, P. J.; Devi, T. R.; Falck, J. R.; Mangelsdorf, D. J. An oxysterol signalling pathway mediated by the nuclear receptor LXR alpha. *Nature* **1996**, 383, 728-731.
  73. Zhang, Y.; Mangelsdorf, D. J. LuXuRies of lipid homeostasis: the unity of nuclear hormone receptors, transcription regulation, and cholesterol sensing. *Mol Interv* **2002**, 2, 78-87.
  74. Tontonoz, P.; Mangelsdorf, D. J. Liver X receptor signaling pathways in cardiovascular disease. *Molecular Endocrinology* **2003**, 17, 985-993.
  75. Wojcicka, G.; Jamroz-Wisniewska, A.; Horoszewicz, K.; Beltowski, J. Liver X receptors (LXRs). Part I: structure, function, regulation of activity, and role in lipid metabolism. *Postepy Hig Med Dosw (Online)* **2007**, 61, 736-59.
  76. Fowler, A. J.; Sheu, M. Y.; Schmuth, M.; Kao, J.; Fluhr, J. W.; Rhein, L.; Collins, J. L.; Willson, T. M.; Mangelsdorf, D. J.; Elias, P. M.; Feingold, K. R. Liver X receptor activators display anti-inflammatory activity in irritant and allergic contact dermatitis models: Liver-X-receptor-specific inhibition of inflammation and primary cytokine production. *Journal of Investigative Dermatology* **2003**, 120, 246-255.
  77. Hindinger, C.; Hinton, D. R.; Kirwin, S. J.; Atkinson, R. D.; Burnett, M. E.; Bergmann, C. C.; Stohman, S. A. Liver X receptor activation decreases the severity of experimental autoimmune encephalomyelitis. *Journal of Neuroscience Research* **2006**, 84, 1225-1234.
  78. Valledor, A. F.; Hsu, L. C.; Ogawa, S.; Sawka-Verhelle, D.; Karin, M.; Glass, C. K. Activation of liver X receptors and retinoid X receptors prevents bacterial-induced macrophage apoptosis. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, 101, 17813-17818.
  79. Wassif, C. A.; Yu, J. H.; Cui, J. S.; Porter, F. D.; Javitt, N. B. 27-hydroxylation of 7-and 8-dehydrocholesterol in Smith-Lemli-Opitz syndrome: a novel metabolic pathway. *Steroids* **2003**, 68, 497-502.
  80. Frolov, A.; Zielinski, S. E.; Crowley, J. R.; Dudley-Rucker, N.; Schaffer, J. E.; Ory, D. S. NPC1 and NPC2 regulate cellular cholesterol homeostasis through generation of low density lipoprotein cholesterol-derived oxysterols. *Journal of Biological Chemistry* **2003**, 278, 25517-25525.
  81. Jamroz-Wisniewska, A.; Wojcicka, G.; Horoszewicz, K.; Beltowski, J. Liver X receptors (LXRs). Part II: non-lipid effects, role in pathology, and therapeutic implications. *Postepy Hig Med Dosw (Online)* **2007**, 61, 760-85.
  82. Chuu, C. P.; Kokontis, J. M.; Hiipakka, R. A.; Liao, S. S. Modulation of liver X receptor signaling as novel therapy for prostate cancer. *Journal of Biomedical Science* **2007**, 14, 543-553.
  83. Farnegardh, M.; Bonn, T.; Sun, S.; Ljunggren, J.; Ahola, H.; Wilhelmsson, A.; Gustafsson, J. A.; Carlquist, M. The three-dimensional structure of the liver X receptor beta reveals a flexible ligand-binding pocket that can accommodate fundamentally different ligands. *Journal of Biological Chemistry* **2003**, 278, 38821-38828.

84. Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *Journal of Medicinal Chemistry* **2006**, 49, 5851-5855.
85. Steindl, T.; Langer, T. Docking versus pharmacophore model generation: A comparison of high-throughput the search of human rhinovirus virtual screening strategies for coat protein inhibitors. *Qsar & Combinatorial Science* **2005**, 24, 470-479.
86. Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound Ligands and their use as virtual screening filters. *Journal of Chemical Information and Modeling* **2005**, 45, 160-169.
87. Vedani, A.; Smiesko, M. Structure-based computational pharmacology and toxicology. In *Computational structural biology. Methods and applications*, Schwede, T.; Peitsch, M., Eds. World Scientific Publishing Co. Pte. Ltd.: Singapore, **2008**; pp 549–572.
88. R., L. A. *Molecular modelling. Principles and applications*. Pearson Education Ltd.: Harlow, **2001**.
89. Kubinyi, H. *3D-QSAR in drug design. Theory, methods, and applications*. ESCOM Science Publisher B. V.: Leiden, **1993**.
90. Kubinyi, H. From narcosis to hyperspace: The history of QSAR. *Quantitative Structure-Activity Relationships* **2002**, 21, 348-356.
91. Kubinyi, H.; Folkers, G.; Martin, Y. C. 3D QSAR in drug design: Ligand-protein interactions and molecular similarity - Preface. *Perspectives in Drug Discovery and Design* **1998**, 9-11, V-VII.
92. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews Drug Discovery* **2004**, 3, 935-949.
93. Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. Docking: Successes and challenges. *Current Pharmaceutical Design* **2005**, 11, 323-333.
94. Krovat, E. M.; Steindl, T.; Langer, T. Recent advances in docking and scoring. *Current Computer-Aided Drug Design* **2005**, 1, 93-102.
95. Muegge, I.; Rarey, M. Small molecule docking and scoring. In *Reviews in computational chemistry*, Lipkowitz, K. B.; Boyd, D. B., Eds. VCH: New York, **2001**.
96. Jones, G.; Willett, P.; Glen, R. C. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *Journal of Computer-Aided Molecular Design* **1995**, 9, 532-549.
97. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **1997**, 267, 727-748.
98. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *Journal of Computer-Aided Molecular Design* **1996**, 10, 293-304.
99. Vedani, A.; Huhta, D. W. A new force-field for modeling metalloproteins. *Journal of the American Chemical Society* **1990**, 112, 4759-4767.
100. Vedani, A. Yeti manual. <http://www.biograf.ch/downloads/yeti.pdf>

101. Vedani, A.; Huhta, D. W. An algorithm for the systematic solvation of proteins based on the directionality of hydrogen-bonds. *Journal of the American Chemical Society* **1991**, 113, 5860-5862.
102. Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L. Assessing energy functions for flexible docking. *Journal of Computational Chemistry* **1998**, 19, 1612-1622.
103. Lill, M. A.; Dobler, M.; Vedani, A. Prediction of small-molecule binding to cytochrome P450 3A4: Flexible docking combined with multidimensional QSAR. *Chemmedchem* **2006**, 1, 73-81.
104. Bembenek, S. D.; Keith, J. M.; Letavic, M. A.; Apodaca, R.; Barbier, A. J.; Dvorak, L.; Aluisio, L.; Miller, K. L.; Lovenberg, T. W.; Carruthers, N. I. Lead identification of acetylcholinesterase inhibitors-histamine H-3 receptor antagonists from molecular modeling. *Bioorganic & Medicinal Chemistry* **2008**, 16, 2968-2973.
105. McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A.; Brown, F. K. Gaussian docking functions. *Biopolymers* **2003**, 68, 76-90.
106. McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kretsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J. F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *Journal of Chemical Information and Modeling* **2007**, 47, 1504-1519.
107. Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *Journal of Molecular Graphics & Modelling* **2003**, 21, PII S1093-3263(02)00164-X.
108. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *Journal of Medicinal Chemistry* **2006**, 49, 5912-5931.
109. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, B. S.; Johnson, A. P. eHITS: An innovative approach to the docking and scoring function problems. *Current Protein & Peptide Science* **2006**, 7, 421-435.
110. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology* **1982**, 161, 269-288.
111. Bohm, H. J.; Stahl, M. Rapid empirical scoring functions in virtual screening applications. *Medicinal Chemistry Research* **1999**, 9, 445-462.
112. Yin, S.; Biedermannova, L.; Vondrasek, J.; Dokholyan, N. V. MedusaScore: An accurate force field-based scoring function for virtual drug screening. *Journal of Chemical Information and Modeling* **2008**, 48, 1656-1662.
113. Gohlke, H.; Klebe, G. Statistical potentials and scoring functions applied to protein-ligand binding. *Current Opinion in Structural Biology* **2001**, 11, 231-235.
114. Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. *Journal of Medicinal Chemistry* **1999**, 42, 2498-2503.
115. Muegge, I. PMF scoring revisited. *Journal of Medicinal Chemistry* **2006**, 49, 5895-5902.

116. Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease - Conformationally flexible docking by evolutionary programming. *Chemistry & Biology* **1995**, *2*, 317-324.
117. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions .1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design* **1997**, *11*, 425-445.
118. Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins-Structure Function and Genetics* **1998**, *33*, 367-382.
119. Rarey, M.; Wefing, S.; Lengauer, T. Placement of medium-sized molecular fragments into active sites of proteins. *Journal of Computer-Aided Molecular Design* **1996**, *10*, 41-54.
120. Schulz-Gasch, T.; Stahl, M. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *Journal of Molecular Modeling* **2003**, *9*, 47-57.
121. Apostolakis, J.; Pluckthun, A.; Caflisch, A. Docking small ligands in flexible binding sites. *Journal of Computational Chemistry* **1998**, *19*, 21-37.
122. Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Current Opinion in Chemical Biology* **2001**, *5*, 375-382.
123. Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *Journal of Molecular Biology* **2004**, *337*, 209-225.
124. Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *Journal of Molecular Biology* **1994**, *235*, 345-356.
125. Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient molecular docking considering protein structure variations. *Journal of Molecular Biology* **2001**, *308*, 377-395.
126. Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins-Structure Function and Bioinformatics* **2002**, *46*, 34-40.
127. Knegt, R. M. A.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *Journal of Molecular Biology* **1997**, *266*, 424-440.
128. Carlson, H. A.; Masukawa, K. M.; McCammon, J. A. Method for including the dynamic fluctuations of a protein in computer-aided drug design. *Journal of Physical Chemistry A* **1999**, *103*, 10213-10219.
129. Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *Journal of the American Chemical Society* **2002**, *124*, 5632-5633.
130. Pei, J. F.; Wang, Q.; Zhou, J. J.; Lai, L. H. Estimating protein-ligand binding free energy: Atomic solvation parameters for partition coefficient and solvation free energy calculation. *Proteins-Structure Function and Bioinformatics* **2004**, *57*, 651-664.
131. Gilson, M. K.; Zhou, H. X. Calculation of protein-ligand binding affinities. *Annual Review of Biophysics and Biomolecular Structure* **2007**, *36*, 21-42.

132. Jackson, J. D. *Classical Electrodynamics*. Wiley Publisher: New York, **1975**.
133. Landau, L. D.; Lifshitz, E. M. *Electrodynamics of Continuous Media*. Pergamon Publisher: New York, **1960**.
134. Cramer, C. J.; Truhlar, D. G. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chemical Reviews* **1999**, 99, 2161-2200.
135. Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free-energies using macroscopic solvent models. *Journal of Physical Chemistry* **1994**, 98, 1978-1988.
136. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *Journal of Physical Chemistry A* **1997**, 101, 3005-3014.
137. Bordner, A. J.; Cavasotto, C. N.; Abagyan, R. A. Accurate transferable model for water, n-octanol, and n-hexadecane solvation free energies. *Journal of Physical Chemistry B* **2002**, 106, 11009-11015.
138. Hamelberg, D.; McCammon, J. A. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method. *Journal of the American Chemical Society* **2004**, 126, 7683-7689.
139. Li, Z.; Lazaridis, T. The effect of water displacement on binding thermodynamics: Concanavalin A. *Journal of Physical Chemistry B* **2005**, 109, 662-670.
140. Olano, L. R.; Rick, S. W. Hydration free energies and entropies for water in protein interiors. *Journal of the American Chemical Society* **2004**, 126, 7991-8000.
141. Tashiro, M.; Stuchebrukhov, A. A. Thermodynamic properties of internal water molecules in the hydrophobic cavity around the catalytic center of cytochrome c oxidase. *Journal of Physical Chemistry B* **2005**, 109, 1015-1022.
142. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *Journal of Physical Chemistry* **1987**, 91, 6269-6271.
143. Jorgensen, W. L.; Jenson, C. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *Journal of Computational Chemistry* **1998**, 19, 1179-1186.
144. Ren, P. Y.; Ponder, J. W. Temperature and pressure dependence of the AMOEBA water model. *Journal of Physical Chemistry B* **2004**, 108, 13427-13437.
145. Cortis, C. M.; Friesner, R. A. An automatic three-dimensional finite element mesh generation system for the Poisson-Boltzmann equation. *Journal of Computational Chemistry* **1997**, 18, 1570-1590.
146. Eisenberg, D.; McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* **1986**, 319, 199-203.
147. Wang, R. X.; Lai, L. H.; Wang, S. M. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design* **2002**, 16, 11-26.
148. Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Fundamentals and applications in chemistry and biology*. American Chemical Society: Washington, DC, **1995**.

149. Tang, N. H.; Blum, D. J. W.; Nirmalakhandan, N.; Speece, R. E. QSAR parameters for toxicity of organic-chemicals to nitrobacter. *Journal of Environmental Engineering-Asce* **1992**, 118, 17-37.
150. Cronin, M. T. D.; Dearden, J. C.; Duffy, J. C.; Edwards, R.; Manga, N.; Worth, A. P.; Worgan, A. D. P. The importance of hydrophobicity and electrophilicity descriptors in mechanistically-based QSAEs for toxicological endpoints. *Sar and Qsar in Environmental Research* **2002**, 13, 167-176.
151. Akahane, K.; Nagano, Y.; Umeyama, H. Hydrophobic effect on the protein-ligand interaction - Hydrophobic field-effect index and hydrophobic correlation index. *Chemical & Pharmaceutical Bulletin* **1989**, 37, 86-92.
152. Kellogg, G. E.; Burnett, J. C.; Abraham, D. J. Very empirical treatment of solvation and entropy: a force field derived from Log P-o/w. *Journal of Computer-Aided Molecular Design* **2001**, 15, 381-393.
153. Lill, M. A.; Vedani, A.; Dobler, M. Raptor: Combining dual-shell representation, induced-fit simulation, and hydrophobicity scoring in receptor modeling: Application toward the simulation of structurally diverse ligand sets. *Journal of Medicinal Chemistry* **2004**, 47, 6174-6186.
154. Leach, A. R. A survey of methods for searching the conformational space of small and medium-sized molecules. In *Reviews in Computational Chemistry*, Lipkowitz, K. B.; Boyd, D. B., Eds. VCH Publisher: New York, **1991**.
155. Heiden, W.; Moeckel, G.; Brickmann, J. A new approach to analysis and display of local lipophilicity hydrophilicity mapped on molecular-surfaces. *Journal of Computer-Aided Molecular Design* **1993**, 7, 503-514.
156. Gunnarsson, G. T.; Desai, U. R. Designing small, nonsugar activators of antithrombin using hydrophobic interaction analyses. *Journal of Medicinal Chemistry* **2002**, 45, 1233-1243.
157. Galatin, P. S.; Abraham, D. J. QSAR: Hydrophobic analysis of inhibitors of the p53-mdm2 interaction. *Proteins-Structure Function and Genetics* **2001**, 45, 169-175.
158. Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. Molecular lipophilicity potential, a tool in 3D QSAR- Method and applications. *Journal of Computer-Aided Molecular Design* **1994**, 8, 83-96.
159. Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT - A new method of empirical hydrophobic field calculation for CoMFA. *Journal of Computer-Aided Molecular Design* **1991**, 5, 545-552.
160. Gaillard, P.; Carrupt, P. A.; Testa, B.; Schambel, P. Binding of arylpiperazines, (aryloxy)propanolamines, and tetrahydropyridylindoles to the 5-HT<sub>1A</sub> receptor: Contribution of the molecular lipophilicity potential to three-dimensional quantitative structure-affinity relationship models. *Journal of Medicinal Chemistry* **1996**, 39, 126-134.
161. Kneubuhler, S.; Thull, U.; Altomare, C.; Carta, V.; Gaillard, P.; Carrupt, P. A.; Carotti, A.; Testa, B. Inhibition of monoamine oxidase-B by 5H-indeno[1,2-c]pyridazines - biological-activities, quantitative structure-activity-relationships (QSARs) and 3D-QSARs. *Journal of Medicinal Chemistry* **1995**, 38, 3874-3883.

162. Folkers, G.; Merz, A. Hydrophobic fields in quantitative structure–activity relationships. In *Lipophilicity in drug action and toxicology*, Pliska, V.; Testa, B.; van de Waterbeemd, H., Eds. VCH Publisher: Weinheim, 1996.
163. Hammett, L. P. *Physical Organic Chemistry*. McGraw-Hill: New York, 1970.
164. Hammett, L. P. Some relations between reaction rates and equilibrium constants. *Chemical Reviews* **1935**, 17, 125-136.
165. Taft, R. W. Polar and steric substituent constants for aliphatic and o-benzoate groups from rates of esterification and hydrolysis of esters. *Journal of the American Chemical Society* **1952**, 74, 3120-3128.
166. Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*. American Chemical Society Publisher: Washington, DC, **1995**.
167. Ramsden, C. A. Quantitative drug design. In *Comprehensive Medicinal Chemistry*, Hansch, C., Pergamon Publisher: Oxford, UK, **1990**; Vol. Vol. 4.
168. Free, S. M.; Wilson, J. W. Mathematical contribution to structure-activity studies. *Journal of Medicinal Chemistry* **1964**, 7, 395-&.
169. Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular-field analysis (CoMFA) .1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society* **1988**, 110, 5959-5967.
170. Lill, M. A. Multi-dimensional QSAR in drug discovery. *Drug Discovery Today* **2007**, 12, 1013-1017.
171. Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRIND-INdependent descriptors (GRIND): A novel class of alignment-independent three-dimensional molecular descriptors. *Journal of Medicinal Chemistry* **2000**, 43, 3233-3243.
172. Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; Wikel, J. H.; Wrighton, S. A. Three- and four-dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors. *Drug Metabolism and Disposition* **2000**, 28, 994-1002.
173. Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B. Q.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *Journal of the American Chemical Society* **1997**, 119, 10509-10524.
174. Vedani, A.; Briem, K.; Dobler, M.; Dollinger, H.; McMasters, D. R. Multiple-conformation and protonation-state representation in 4D-QSAR: The neurokinin-1 receptor system. *Journal of Medicinal Chemistry* **2000**, 43, 4416-4427.
175. Lukacova, V.; Balaz, S. Multimode ligand binding in receptor site modeling: Implementation in CoMFA. *Journal of Chemical Information and Computer Sciences* **2003**, 43, 2093-2105.
176. Vedani, A.; Dobler, M. 5D-QSAR: The key for simulating induced fit? *Journal of Medicinal Chemistry* **2002**, 45, 2139-2149.
177. Vedani, A.; Dobler, M.; Lill, M. A. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *Journal of Medicinal Chemistry* **2005**, 48, 3700-3703.
178. Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*. Wiley-WCH Publisher: Weinheim, DE, **2000**.
179. Freedman, D. A. *Statistical Models: Theory and Practice*. Cambridge University Press Publisher: Cambridge, **2005**.



180. Luco, J. M.; Ferretti, F. H. QSAR based on multiple linear regression and PLS methods for the anti-HIV activity of a large group of HEPT derivatives. *Journal of Chemical Information and Computer Sciences* **1997**, 37, 392-401.
181. Devillers, J. *Genetic Algorithms in Molecular Modeling. Principles of QSAR and Drug Design*. Academic Press Inc. Publisher: San Diego, CA, **1996**.
182. Hasegawa, K.; Funatsu, K. GA strategy for variable selection in QSAR studies: GAPLS and D-optimal designs for predictive QSAR model. *Theochem-Journal of Molecular Structure* **1998**, 425, 255-262.
183. Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design: An Introduction*. Wiley-VCH Publisher: Weinheim, DE, **1999**.
184. Kahn, I.; Sild, S.; Maran, U. Modeling the toxicity of chemicals to *Tetrahymena pyriformis* using heuristic multilinear regression and heuristic back-propagation neural networks. *Journal of Chemical Information and Modeling* **2007**, 47, 2271-2279.
185. Topliss, J. G.; Costello, R. J. Chance correlations in structure-activity studies using multiple regression-analysis. *Journal of Medicinal Chemistry* **1972**, 15, 1066-1068.
186. Topliss, J. G.; Edwards, R. P. Chance factors in studies of quantitative structure-activity-relationships. *Journal of Medicinal Chemistry* **1979**, 22, 1238-1244.
187. Golbraikh, A.; Tropsha, A. Beware of  $q(2)!$  *Journal of Molecular Graphics & Modelling* **2002**, 20, 269-276.
188. Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *Qsar & Combinatorial Science* **2003**, 22, 69-77.
189. Gramatica, P. Principles of QSAR models validation: internal and external. *Qsar & Combinatorial Science* **2007**, 26, 694-701.
190. Miller, A. J. *Subset Selection in Regression*. Chapman & Hall Publisher: London, UK, **1990**.
191. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. Detecting "bad" regression models: multicriteria fitness functions in regression analysis. *Analytica Chimica Acta* **2004**, 515, 199-208.
192. Kubinyi, H.; Hamprecht, F. A.; Mietzner, T. Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. *Journal of Medicinal Chemistry* **1998**, 41, 2553-2564.
193. Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indexes in a comparative-analysis (CoMSIA) of drug molecules to correlate and predict their biological-activity. *Journal of Medicinal Chemistry* **1994**, 37, 4130-4146.
194. van Drie, J. H. Pharmacophore discovery - Lessons learned. *Current Pharmaceutical Design* **2003**, 9, 1649-1664.
195. van Drie, J. H. Pharmacophore discovery: a critical review. In *Computational Medicinal Chemistry for Drug Discovery*, Bultinck, P.; de Winter, H.; Langenaeker, W.; Tollenaere, J. P., Eds. Dekker Inc. Publisher: New York, **2004**.
196. Faber, N. M.; Rajko, R. How to avoid over-fitting in multivariate calibration - The conventional validation approach and an alternative. *Analytica Chimica Acta* **2007**, 595, 98-106.

197. Walker, J. D. *Handbook on quantitative structure activity relationships (QSARs) for pollution prevention, toxicity screening, risk assessment and world wide web applications*. SETAC Press Publisher: Pensacola, FL, **2002**.
198. Walker, J. D. *Handbook on quantitative structure activity relationships (QSARs) for predicting ecological effects of chemicals*. SETAC Press Publisher: Pensacola, FL, **2002**.
199. Walker, J. D. *Handbook on quantitative structure activity relationships (QSARs) for predicting effects of chemicals on environmental-human health interactions*. SETAC Press Publisher: Pensacola, FL, **2002**.
200. Walker, J. D. *Handbook on quantitative structure activity relationships (QSARs) for predicting endocrine disruption potential of chemicals*. SETAC Press Publisher: Pensacola, FL, **2002**.
201. Walker, J. D. *Handbook on quantitative structure activity relationships (QSARs) for predicting physical properties, bioconcentration, potential and environmental fate of chemicals*. SETAC Press Publisher: Pensacola, FL, **2002**.
202. OECD. OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models. <http://www.oecd.org/data-oecd/33/37/37849783.pdf>
203. Asikainen, A. H.; Ruuskanen, J.; Tuppurainen, K. A. Consensus kNN QSAR: A versatile method for predicting the estrogenic activity of organic compounds in silico. A comparative study with five estrogen receptors and a large, diverse set of ligands. *Environmental Science & Technology* **2004**, 38, 6724-6729.
204. Votano, J. R.; Parham, M.; Hall, L. H.; Kier, L. B.; Oloff, S.; Tropsha, A.; Xie, Q. A.; Tong, W. Three new consensus QSAR models for the prediction of Ames genotoxicity. *Mutagenesis* **2004**, 19, 365-377.
205. Vedani, A.; Zumstein, M.; Lill, M. A.; Ernst, B. Simulating alpha/beta selectivity at the human thyroid hormone receptor: Consensus scoring using multidimensional QSAR. *Chemmedchem* **2007**, 2, 78-87.
206. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995**, 91, 1-41.
207. Storer, J.; Giesen, D.; Cramer, C.; Truhlar, D. Class IV charge models: a new semiempirical approach in quantum chemistry. *J Comput Aided Mol Des* **1995**, 9, 87-110.
208. Giesen, D. J.; Storer, J. W.; Cramer, C. J.; Truhlar, D. G. General Semiempirical Quantum-Mechanical Solvation Model For Nonpolar Solvation Free-Energies - N-Hexadecane. *Journal of the American Chemical Society* **1995**, 117, 1057-1068.
209. Vedani, A. Biograf3R. <http://www.biograf.ch/>
210. Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. In *Epik: a software program for pK (a) prediction and protonation state generation for drug-like molecules*, 232nd National Meeting of the American-Chemical-Society, San Francisco, CA, Sep 10-14, 2006; San Francisco, CA, **2006**; pp 681-691.

211. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment Of Solvation For Molecular Mechanics And Dynamics. *Journal of the American Chemical Society* **1990**, 112, 6127-6129.
212. Guimaraes, C. R. W.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *Journal of Chemical Information and Modeling* **2008**, 48, 958-970.
213. Zbinden, P.; Dobler, M.; Folkers, G.; Vedani, A. PrGen: Pseudoreceptor modeling using receptor-mediated ligand alignment and pharmacophore equilibration. *Quantitative Structure-Activity Relationships* **1998**, 17, 122-130.
214. Vedani, A.; Dobler, M. Multidimensional QSAR: Moving from three- to five-dimensional concepts. *Quantitative Structure-Activity Relationships* **2002**, 21, 382-390.
215. Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, 18, 2714-2723.
216. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, 14, 33-&.
217. Hammer, S.; Spika, I.; Sippl, W.; Jessen, G.; Kleuser, B.; Holtje, H. D.; Schafer-Korting, M. Glucocorticoid receptor interactions with glucocorticoids: evaluation by molecular modeling and functional analysis of glucocorticoid receptor mutants. *Steroids* **2003**, 68, 329-339.
218. Coghlan, M. J.; Kym, P. R.; Elmore, S. W.; Wang, A. X.; Luly, J. R.; Wilcox, D.; Stashko, M.; Lin, C. W.; Miner, J.; Tyree, C.; Nakane, M.; Jacobson, P.; Lane, B. C. Synthesis and characterization of non-steroidal ligands for the glucocorticoid receptor: Selective quinoline derivatives with prednisolone-equivalent functional activity. *Journal of Medicinal Chemistry* **2001**, 44, 2879-2885.
219. Elmore, S. W.; Coghlan, M. J.; Anderson, D. D.; Pratt, J. K.; Green, B. E.; Wang, A. X.; Stashko, M. A.; Lin, C. W.; Tyree, C. M.; Miner, J. N.; Jacobson, P. B.; Wilcox, D. M.; Lane, B. C. Nonsteroidal selective glucocorticoid modulators: The effect of C-5 alkyl substitution on the transcriptional activation/repression profile of 2,5-dihydro-10-methoxy-2,2,4-trimethyl-1H-[1]benzopyrano[3,4-f]quinoline s. *Journal of Medicinal Chemistry* **2001**, 44, 4481-4491.
220. Ali, A.; Thompson, C. F.; Balkovec, J. M.; Graham, D. W.; Hammond, M. L.; Quraishi, N.; Tata, J. R.; Einstein, M.; Ge, L.; Harris, G.; Kelly, T. M.; Mazur, P.; Pandit, S.; Santoro, J.; Sitlani, A.; Wang, C. L.; Williamson, J.; Miller, D. K.; Thompson, C. M.; Zaller, D. M.; Forrest, M. J.; Carballo-Jane, E.; Luell, S. Novel N-arylpyrazolo[3,2-c]-based ligands for the glucocorticoid receptor: Receptor binding and in vivo activity. *Journal of Medicinal Chemistry* **2004**, 47, 2441-2452.
221. Einstein, M.; Greenlee, M.; Rouen, G.; Sitlani, A.; Santoro, J.; Wang, C. L.; Pandit, S.; Mazur, P.; Smalera, I.; Weaver, A. P. M.; Zeng, Y. Y.; Ge, L.; Kelly, T.; Paiva, T.; Geissler, W.; Mosley, R. T.; Williamson, J.; Ali, A.; Balkovec, J.; HARRIS, G. Selective glucocorticoid receptor nonsteroidal ligands completely antagonize the dexamethasone mediated induction of enzymes involved in gluconeogenesis and glutamine metabolism. *Journal of Steroid Biochemistry and Molecular Biology* **2004**, 92, 345-356.

222. Smith, C. J.; Ali, A.; Balkovec, J. M.; Graham, D. W.; Hammond, M. L.; Patel, G. F.; Rouen, G. P.; Smith, S. K.; Tata, J. R.; Einstein, M.; Ge, L.; Harris, G. S.; Kelly, T. M.; Mazur, P.; Thompson, C. M.; Wang, C. L. F.; Williamson, J. M.; Miller, D. K.; Pandit, S.; Santoro, J. C.; Sitlani, A.; Yamin, T. D.; O'Neill, E. A.; Zaller, D. M.; Carballo-Jane, E.; Forrest, M. J.; Luell, S. Novel ketal ligands for the glucocorticoid receptor: in vitro and in vivo activity. *Bioorganic & Medicinal Chemistry Letters* **2005**, *15*, 2926-2931.
223. Cheng, H. C. The power issue: determination of K-B or K-i from IC50 - A closer look at the Cheng-Prusoff equation, the Schild plot and related power equations. *Journal of Pharmacological and Toxicological Methods* **2001**, *46*, 61-71.
224. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A New Force-Field For Molecular Mechanical Simulation Of Nucleic-Acids And Proteins. *Journal of the American Chemical Society* **1984**, *106*, 765-784.
225. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. <http://www.rcsb.org/>
226. Svensson, S.; Ostberg, T.; Jacobsson, M.; Norstrom, C.; Stefansson, K.; Hallen, D.; Johansson, I. C.; Zachrisson, K.; Ogg, D.; Jendeborg, L. Crystal structure of the heterodimeric complex of LXR alpha and RXR beta ligand-binding domains in a fully agonistic conformation. *Embo Journal* **2003**, *22*, 4625-4633.
227. Hoerer, S.; Schmid, A.; Heckel, A.; Budzinski, R. M.; Nar, H. Crystal structure of the human liver X receptor beta ligand-binding domain in complex with a synthetic agonist. *Journal of Molecular Biology* **2003**, *334*, 853-861.
228. Stewart, J. J. P. Special issue - MOPAC - A semiempirical molecular-orbital program. *Journal of Computer-Aided Molecular Design* **1990**, *4*, 1-45.
229. Wold, S.; Eriksson, L. *Chemometrics Methods in Molecular Design*. 1995.
230. Vedani, A. VirtualToxLab. [http://www.biograf.ch/data/projects/virtualtoxlab\\_results.php](http://www.biograf.ch/data/projects/virtualtoxlab_results.php)
231. Kauppi, B.; Jakob, C.; Farnegardh, M.; Yang, J.; Ahola, H.; Alarcon, M.; Calles, K.; Engstrom, O.; Harlan, J.; Muchmore, S.; Ramqvist, A. K.; Thorell, S.; Ohman, L.; Greer, J.; Gustafsson, J. A.; Carlstedt-Duke, J.; Carlquist, M. The three-dimensional structures of antagonistic and agonistic forms of the glucocorticoid receptor ligand-binding domain - RU-486 induces a transconformation that leads to active antagonism. *Journal of Biological Chemistry* **2003**, *278*, 22748-22754.
232. Brzozowski, A. M.; Pike, A. C. W.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engstrom, O.; Ohman, L.; Greene, G. L.; Gustafsson, J. A.; Carlquist, M. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **1997**, *389*, 753-758.
233. Madauss, K. P.; Deng, S. J.; Austin, R. J. H.; Lambert, M. H.; McLay, I.; Pritchard, J.; Short, S. A.; Stewart, E. L.; Uings, I. J.; Williams, S. P. Progesterone receptor ligand binding pocket flexibility: Crystal structures of the norethindrone and mometasone furoate complexes. *Journal of Medicinal Chemistry* **2004**, *47*, 3381-3387.
234. Lill, M. A.; Winiger, F.; Vedani, A.; Ernst, B. Impact of induced fit on ligand binding to the androgen receptor: A multidimensional QSAR study to predict


- endocrine-disrupting effects of environmental chemicals. *Journal of Medicinal Chemistry* **2005**, 48, 5666-5674.
235. Wang, H.; Aslanian, R.; Madison, V. Induced-fit docking of mometasone furoate and further evidence for glucocorticoid receptor 17alpha pocket flexibility. *J Mol Graph Model* **2008**, 27, 512-21.
236. Fried, J. H.; Arth, G. E.; Mrozik, H.; Steinberg, N. G.; Hirschmann, R. F.; Steelman, S. L.; Bry, T. S.; Tishler, M. 16-Methylated steroids. IV. 6,16alpha-dimethyl-delta6- hydroxycortisone and related compounds. *Journal of the American Chemical Society* **1963**, 85, 236-&.
237. Hirschmann, R.; Buchschacher, P.; Fried, J. H.; Steinberg, N. G.; Tishler, M.; Kent, G. J. Synthesis and structure of steroidal 4-pregneno [3,2-c] pyrazoles - a novel class of potent anti-inflammatory steroids. *Journal of the American Chemical Society* **1963**, 85, 120-&.
238. Flocco, M. M.; Mowbray, S. L. Planar stacking interactions of arginine and aromatic side-chains in proteins. *Journal of Molecular Biology* **1994**, 235, 709-717.
239. Madauss, K. P.; Bledsoe, R. K.; McLay, I.; Stewart, E. L.; Uings, I. J.; Weingarten, G.; Williams, S. P. The first X-ray crystal structure of the glucocorticoid receptor bound to a non-steroidal agonist. *Bioorganic & Medicinal Chemistry Letters* **2008**, 18, 6097-6099.
240. Vedani, A. Quasar manual. <http://www.biograf.ch/downloads/quasar.pdf>
241. Shah, N.; Scanlan, T. S. Design and evaluation of novel nonsteroidal dissociating glucocorticoid receptor ligands. *Bioorganic & Medicinal Chemistry Letters* **2004**, 14, 5199-5203.
242. Wikipedia. Wikipedia. <http://www.en.wikipedia.org/wiki>
243. Ottenweller, J. E.; Tapp, W. N.; Natelson, B. H. Effects of chronic alprazolam treatment on plasma-concentrations of glucocorticoids, thyroid-hormones, and testosterone in cardiomyopathic hamsters. *Psychopharmacology* **1989**, 98, 369-371.
244. Vedder, H.; Bening-Abu-Shach, U.; Lanquillon, S.; Krieg, J. C. Regulation of glucocorticoidreceptor-mRNA in human blood cells by amitriptyline and dexamethasone. *Journal of Psychiatric Research* **1999**, 33, 303-308.
245. Pariante, C. M.; Hye, A.; Williamson, R.; Makoff, A.; Lovestone, S.; Kerwin, R. W. The antidepressant clomipramine regulates cortisol intracellular concentrations and glucocorticoid receptor expression in fibroblasts and rat primary neurones. *Neuropsychopharmacology* **2003**, 28, 1553-1561.
246. Uys, J. D. K.; Muller, C. J. F.; Marais, L.; Harvey, B. H.; Stein, D. J.; Daniels, W. M. U. Early life trauma decreases glucocorticoid receptors in rat dentate gyrus upon adult re-stress: Reversal by escitalopram. *Neuroscience* **2006**, 137, 619-625.
247. Yau, J. L. W.; Noble, J.; Chapman, K. E.; Seckl, J. R. Differential regulation of variant glucocorticoid receptor mRNAs in the rat hippocampus by the antidepressant fluoxetine. *Brain Res Mol Brain Res* **2004**, 129, 189-92.
248. Mukherjee, K.; Knisely, A.; Jacobson, L. Partial glucocorticoid agonist-like effects of imipramine on hypothalamic-pituitary-adrenocortical activity, thymus weight,

- and hippocampal glucocorticoid receptors in male C57BL/6 mice. *Endocrinology* **2004**, 145, 4185-4191.
249. Hill, S. A.; Taylor, M. J.; Harmer, C. J.; Cowen, P. J. Acute reboxetine administration increases plasma and salivary cortisol. *Journal of Psychopharmacology* **2003**, 17, 273-275.
250. Yehuda, R.; Yang, R. K.; Golier, J. A.; Grossman, R. A.; Bierer, L. M.; Tischler, L. Effect of sertraline on glucocorticoid sensitivity of mononuclear leukocytes in post-traumatic stress disorder. *Neuropsychopharmacology* **2006**, 31, 189-196.
251. Lee, M. S.; Yang, J. W.; Ko, Y. H.; Han, C.; Kim, S. H.; Joe, S. H.; Jung, I. K. Effects of methylphenidate and bupropion on DHEA-S and cortisol plasma levels in attention-deficit hyperactivity disorder. *Child Psychiatry & Human Development* **2008**, 39, 201-209.
252. Sharpe, C. R.; Collet, J. P.; Belzile, E.; Hanley, J. A.; Boivin, J. F. The effects of tricyclic antidepressants on breast cancer risk. *British Journal of Cancer* **2002**, 86, 92-97.
253. Buchwald, P. General linearized biexponential model for QSAR data showing bilinear-type distribution. *Journal of Pharmaceutical Sciences* **2005**, 94, 2355-2379.
254. Vedani, A.; Descloux, A. V.; Spreatico, M.; Ernst, B. Predicting the toxic potential of drugs and chemicals in silico: A model for the peroxisome proliferator-activated receptor-gamma (PPAR gamma). *Toxicology Letters* **2007**, 173, 17-23.
255. Murray-Rust, P.; Glusker, J. P. Directional hydrogen-bonding to sp<sup>2</sup>-hybridized and sp<sup>3</sup>-hybridized oxygen-atoms and its relevance to ligand macromolecule interactions. *Journal of the American Chemical Society* **1984**, 106, 1018-1025.
256. Taylor, R.; Kennard, O. Hydrogen-bond geometry in organic-crystals. *Accounts of Chemical Research* **1984**, 17, 320-326.
257. Vedani, A.; Dunitz, J. D. Lone-pair directionality in hydrogen-bond potential functions for molecular mechanics calculations - the inhibition of human carbonic anhydrase-II by sulfonamides. *Journal of the American Chemical Society* **1985**, 107, 7653-7658.
258. Xu, D.; Tsai, C. J.; Nussinov, R. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Engineering* **1997**, 10, 999-1012.
259. Szewczyk, J. W.; Huang, S.; Chin, J.; Tian, J.; Mitnal, L.; Rosa, R. L.; Peterson, L.; Sparrow, C. P.; Adams, A. D. SAR studies: Designing potent and selective LXR agonists. *Bioorganic & Medicinal Chemistry Letters* **2006**, 16, 3055-3060.
260. Ghose, A. K.; Crippen, G. M. Atomic physicochemical parameters for 3-dimensional structure-directed quantitative structure-activity-relationships .1. Partition-coefficients as a measure of hydrophobicity. *Journal of Computational Chemistry* **1986**, 7, 565-577.
261. Lehmann, E. L.; D'Abrera, H. J. M. Nonparametrics: statistical methods based on ranks. Englewood Cliffs Publisher: New Jersey, **1998**; p 292.

## 10 Curriculum Vitae

# Morena Spreafico

---

<b>PERSONAL DETAILS</b>	<p>Born in Lecco (Italy), 30/05/1977</p> <p>Current address: Stöberstr. 5. CH–4055 Basel</p> <p>Phone +41 (0) 61 267 15 03 Mobile +41 78 823 86 48</p> <p>Mail: <a href="mailto:morena.spreafico@unibas.ch">morena.spreafico@unibas.ch</a></p> <p>Nationality: Italian</p>	
<b>EDUCATION</b> 2005–09  2003  2002	<ul style="list-style-type: none"> <li>• PhD in <b>Molecular Modelling</b> at the Institute of Molecular Pharmacy, University of Basel (Prof. A. Vedani). Thesis title: “Mixed model QSAR at the glucocorticoid and liver X receptors”.</li> <li>• Master in <b>Bioinformatics</b>, University of Milano–Bicocca. Thesis title: “Preliminary steps of virtual screening: analysis of two different software” (<i>Nikem Research</i>).</li> <li>• Graduated <i>cum laude</i> in <b>Environmental Sciences</b>, University of Milano–Bicocca. Thesis title “Application of QSAR methodologies to estimate ecotoxicological parameters of pesticides” (Prof. R. Todeschini).</li> </ul>	
<b>PROFESSIONAL EXPERIENCE</b> 2005–09  2004–2005  2003–2004  2003 2001–2002 1995–2001	<ul style="list-style-type: none"> <li>• <b>University of Basel</b>. Molecular modelling on Nuclear Receptors, and endocrine disruptors (Prof. A. Vedani, Prof. B. Ernst).</li> <li>• <b>National Institute of Chemistry</b>, Ljubljana. Artificial Neural Networks and Genetic Algorithms on endocrine disruptors (Prof. Jure Zupan, Dr. Marjana Novic).</li> <li>• <b>Pharmacological Research Institute Mario Negri</b>, Milano (Dr. Emilio Benfenati): QSAR on pesticides, docking on endocrine disruptors.</li> <li>• Co-op with <b>Nikem Research</b>, Milano (Dr. Massimo Dondio).</li> <li>• Tutoring in multimedia room in University of Milano - Bicocca.</li> <li>• Private lessons in scientific subjects to high school students.</li> </ul>	

**SKILLS***Professional skills:*

- Expertise with molecular modelling software for docking, QSARs, molecular mechanics and dynamics simulations (MacroModel, Glide, AMBER, Sybyl, ChemOffice, Quasar, Raptor, Yeti), and visualisation (VMD, Swiss PDB Viewer).
- Main interests and experience concern molecular modelling, chemometrics and multivariate analysis (Statistica, MATLAB).

*Computational skills:*

- Very good knowledge of Windows and Macintosh O.S. and good knowledge of Unix O.S.
- Basic skills in C programming. Few notions of HTML.

*Languages:*

- Italian: native speaker.
- English: very good.
- French: good.
- German: intermediate level.

**INTERESTS**

Travelling to get a broader view on the world and to know different cultures, hiking to feel contact with nature and friends, photography, arts, movies and reading.

**REFEREES**

- Prof. Angelo Vedani, *Biographics Laboratory 3R*. Friedensgasse 35, 4056 Basel (Switzerland). Tel: +41 (0)61 261 42 56  
e-mail: [angelo@biograf.ch](mailto:angelo@biograf.ch)
- Dr Emilio Benfenati, Head, Laboratory of Environmental Chemistry and Toxicology. *Pharmacological Research Institute Mario Negri* - Via La Masa 19, 20156 Milano (Italy) Tel: +39 02 39014420  
e-mail: [benfenati@marionegri.it](mailto:benfenati@marionegri.it)
- Prof. Roberto Todeschini, *Milano Chemometrics & QSAR Research Group* Dept. of Environmental Sciences - P.za della Scienza 1, 20126 Milano (Italy) Tel: +39 02 64482820  
email: [roberto.todeschini@unimib.it](mailto:roberto.todeschini@unimib.it)



**SCIENTIFIC MEETINGS AND PRESENTATIONS**

- **European Workshop in Drug design.** *Siena* (Italy). 3–9 Jun 2007 (poster).
- **7th Swiss Course on Medicinal Chemistry.** *Leysin* (Switzerland), 1–6 Oct 2006.
- **MGMS International Meeting 2005 on Biomolecular Simulation.** *Dublin* (Ireland). 11–14 Sept 2005 (poster).
- **Workshop of Chemometrics.** *Varenna* (Italy), 26–27 May 2005.
- **Summer School On Nuclear Receptors.** *Lyon* (France), 13–17 Sept 2004 (poster).
- **Summer School On Endocrinology: Nuclear Receptors, Endocrine Disruptors and Metabolic Effects.** *Bregenz* (Austria), 25–29 July 2004 (poster).
- Collaborating with *Milano Chemometrics & QSAR Research Group* for the **II Italian School on molecular Descriptors.** *Milano* (Italy), 3–5 March 2004.
- *Rational Drug Design* class, faculty of Biology, **University of Insubria**, *Varese* (Italy), 3 March 2004 (oral presentation).

**PUBLICATIONS**

- O. Peristera, M. Spreafico, M. Smiesko, B. Ernst and A. Vedani. Mixed-model QSAR at the mineralocorticoid receptor: Predicting the binding mode and affinity of anabolic steroids. Submitted to **Tox. Lett.**
- O. Schwardt, H. Gäthje, A. Vedani, S. Mesch, G. P. Gao, M. Spreafico, J. von Orelli, S. Kelm and B. Ernst. Examination of the biological role of the  $\alpha(2\rightarrow6)$ -linked sialic acid in gangliosides binding to the myelin-associated glycoprotein. **J. Med. Chem.** (2009) 52, 989–1004.
- M. Spreafico, B. Ernst, M. Lill, M. Smiesko, and A. Vedani. Mixed-model QSAR at the glucocorticoid receptor: Predicting the binding mode and affinity of psychotropic drugs. **ChemMedChem** (2009) 4, 100–109.
- A. Vedani, M. Spreafico, O. Peristera, M. Dobler, M. Smiesko. VirtualToxLab - in silico Prediction of the Endocrine-Disrupting Potential

---

of Drugs and Chemicals. **CHIMIA** (2008) 5, 322–328.

- A. Vedani, M. Dobler, M. Spreafico, O. Peristera and M. Smiesko. VirtualToxLab — In silico prediction of the toxic potential of drugs and environmental chemicals: Validation status and Internet access protocol. **ALTEX** (2007) 24, 153–161.
- A. Vedani, A. V. Descloux, M. Spreafico and B. Ernst. Predicting the toxic potential of drugs and chemicals in silico: A model for the peroxisome proliferator-activated receptor  $\gamma$  (PPAR  $\gamma$ ). **Tox. Lett.** (2007) 173, 17–23.
- M. Spreafico, E. Boriani, E. Benfenati, M. Novic. Structural features of diverse ligands influencing Binding Affinities to Estrogen  $\alpha$  and Estrogen  $\beta$  receptors. Part II. Molecular Descriptors Calculated from Conformation of the Ligands in the Complex Resulting from Previous Docking Study. **Mol. Divers.** (2007) 11, 171–181.
- E. Benfenati, M. Casalegno, J. Cotterill, N. Price, M. Spreafico, A. Toropov. Characterization of chemical structures. Chapter in the book: Quantitative Structure-Activity Relationship (QSAR) for Pesticide Regulatory Purposes. (2007) Amsterdam: **Elsevier**.
- E. Boriani, M. Spreafico, E. Benfenati, M. Novic. Structural features of diverse ligands influencing Binding Affinities to Estrogen  $\alpha$  and Estrogen  $\beta$  receptors. Part I: Molecular Descriptors Calculated from Minimal Energy Conformation of Isolated Ligands. **Mol. Divers.** (2007) 11, 153–169