FACULTY OF HEALTH SCIENCES
DEPARTMENT OF MEDICAL BIOLOGY
MEDICAL PHARMACOLOGY AND TOXICOLOGY RESEARCH GROUP

Structure, function and ligand interactions of the ecdysone receptor from *Daphnia magna*

LINN MARI EVENSETH
Master thesis in Molecular Biotechnology (MBI-3941)
June 2014

Acknowledgements

I would like to thank Prof. Ingebrigt Sylte for giving me the opportunity to work with my main interests, modeling and toxicology. I am also very thankful that I was included on the EDRISK project.
I have met many obstacles on the road to achieving the necessary results, especially during the setup in Norway. To find solutions for these difficulties, it was necessary for me to contact other universities to search for people with more experience. I am impressed by the helpfulness all the affected parts have showed.

There are many people that have contributed to the results presented in this thesis.
I would like to express a special gratitude to my great supervisor Assoc. Prof. Kurt Kristiansen for being so patient with me even though I frequently ran down his office and often questioned his organizational skills due to a messy office (he was probably happy I spent most of my time traveling). I would also like to express a special thank to Prof. Ingebrigt Sylte for all the support, advisable talks and help in the writing process.

I also want to thank the following people:

NIVA
Knut Erik Tollefsen for sending me to Japan, all the advises and educational discussions in addition to pushing me to produce results

PhD student Maria Hultmann for helping me set up the all the necessary equipment, giving me advises, providing me articles and for making nice drawings on days were things went slow

PhD You Song for fulfilling discussions, for not asking too many questions during his training in the technique and for all the support and help in the writing process

Research assistant David Eidsvoll for interesting discussions and for always helping out

UIO
Prof. Tor Gjøen for allowing me to use the labs at the pharmacy department, providing *E.coli* and for letting me use their Victor[3]

PhD student Marianne Arnemoe for helping me setting up a new protocol on Victor[3] and guidance on how to use the machine

Dr. Ragnhild Elisabeth Paulsen for helping me setting up all necessary equipment for cloning of the plasmids and for being available during the performance of the protocol

Abstract

Background: Endocrine disruptors are of an increasing concern to the global environment due to their ability to modulate endocrine processes and cause adverse apical effects. Invertebrates are important species in the aquatic environment and are a central subject for toxicological testing, but the effects of endocrine disruptors (EDs) in invertebrates are limited due to less knowledge concerning the endocrine systems.

A molecular modeling approach can be used for high-throughput screening of potential active compounds to predict binding affinity towards a target. Homology modeling, docking and scoring studies can play an important role in risk assessment of EDs.

Two homology models of the ecdysone receptor (EcR) in *D. magna* were constructed from resolved X-ray structures of *Bemisia tabacil* and *Heliothis virescens* EcR ligand binding domain. The models were evaluated by docking studies and an *in vitro* two-hybrid reporter assay as an attempt to support the constructed models, identify a possible ED target and identify ED chemicals.

This project is as a part of the research council of Norway (RCN) funded and NIVA-led project EDRISK, which main goal is to develop and evaluate adverse outcome pathways of EDs in the crustacean *D. magna* for potential inclusion in hazard and risk assessment of EDs.

Results: Docking scores of presumed active binders were good for both models. *In vitro* data of the presumed active binders, ponasterone A and 20-hydroxyecdysone were verified to act as agonists in the reporter assay supporting the docking results. *In vitro* data of TFOA, triclosan and diethyl phthalate showed that the compounds were not able to bind to the EcR, partially opposing the predicted scores.

Conclusion: Theoretical studies predicted model II to be a more accurate representation of the EcR in *D. magna* than model I and results of experimental testing supported this prediction. The experimental testing of the selected compounds was not sufficient to fully support the predicted models since too few compounds were tested.

Abbreviations

AB	Alamar blue
ADMET	Absorption-distribution-metabolism-excretion-toxicity
AF	Activation function
ANOVA	Analysis of variance
AUC	area under the curve
B. tabaci	*Bemisia tabaci*
CFDA	5-carboxyfluorescein diacetoxymethyl ester
CHO	Chinese hamster ovary
DBD	DNA binding region
D.magna	*Daphnia magna*
D.melanogaster	*Drosophila melanogaster*
DMEM	Dulbecco´s modified Eagle´s medium
E. Coli	Escherichia coli
EcR	Ecdysone receptor
EC50	Half maximum effective concentration
EDs	Endocrine disruptors
EDTA	Ethylenediaminetetraacetic acid
FBS	Fetal bovine serum
H	Helix
HCB	Hexachlorobenzene
H. virescens	*Heliothis virescens*
HRE	Hormone response element
Hydrogen bond	HB
ICM	Internal coordinate Mechanics
IC50	Concentration of a substance required for 50% inhibition of binding
K_{AW}	Air-water partition coefficient
K_{OA}	Octanol-air partition coefficient
K_{OW}	Octanol-water partition coefficient
LBD	Ligand binding domain
MM	Molecular mechanics
nM	nanomolar

NMR	Nuclear magnetic resonance
NR	Nuclear receptor
PAH	Polycyclic aromatic hydrocarbons
PBDE	Polybrominated diphenyl ethers
PBS	Phosphate buffered saline
PCBs	Polychlorinated biphenyls
PCDFs	Polychlorinated dibenzofurans
PDB	Protein data bank
PLB	Passive lysis buffer
POPs	Persistent organic pollutants
QM	Quantum mechanics
RMSD	Root mean square deviation
ROC	Receiver characteristics operator
Rpm	Revolutions per minute
SAVES	Structural Analysis and Verification Server
TFOA	1H, 1H, 2H, 2H-perfluorooctyl acrylate
T. castaneum	*Tribolium castaneum*
USP	Ultraspiracle Protein
VDW	Van der Waals
VLS	Virtual ligand screening
Å	Ångstrøm
3D	Three-dimensional
20E	20-Hydroxyecdysone

Table of contents

1 Introduction

1.1 Persistent organic pollutions and endocrine disruptors in the environment

The ability of natural and synthetic compounds to interfere with the endogenous hormone receptors was known already in the early 20th century. Pollutants such as polychlorinated biphenyls (PCB) and polychlorinated dibenzofurans (PCDFs) were found to affect the birth weight and neurological development in infants of mothers exposed to these compounds. Despite the correlation between exposure of chemicals and endocrine-mediated toxicity, the chemicals received no attention beyond the mechanism known at that time (Marty *et al.,* 2011). In the early 90`s, researchers found a correlation between exposure of man-made chemicals and developmental and reproductive toxicology threatening humans and wildlife (Marty *et al.,* 2011).

Persistent organic pollutants (POPs) are defined as a group of chemical compounds with the similar characteristics like (1) bioaccumulation, (2) ability to travel long distances through the atmosphere, (3) toxic and persistent in the environment (Hansen *et al.,* 2004).

POPs and other endocrine disruptors are of a great concern to the global environment because of their potential to target and disrupt endocrine processes (Wang *et al.,* 2010 and Tyler *et al.,* 1998). Many POPs are highly lipophilic and tend to accumulate in adipose tissue (bioaccumulation). This characteristic makes POPs able to concentrate in the food chain and can often be detected in high levels in top predators (Verreault *et al.,* 2007*)*. Adverse outcomes of POPs and other endocrine disruptors (ED) should not be underestimated in smaller animals despite the lower degree of bioaccumulation (due to less adipose tissue). Small crustaceans such as the water flea *D.magna* play an important role in the ecosystem by being a significant component of fish diets. *D.magna* also contributes to clearing water by grazing algae and bacteria (Kato *et al.,* 2007).

The structure and chemical properties of many POPs and other ED are similar to that of endogenous hormones, which enables the chemicals to interfere with normal hormonal signalling. They can function as agonists, partial agonists or antagonists by

occupying the same receptor binding sites as the endogenous hormones. The functional outcome of exposure can be disruption of processes such as growth, reproduction and fertility. Exposure has also been associated with immune/autoimmune diseases and a variety of different of cancer forms. Studies have also shown that ED and endogenous hormones can interfere with the function of each other, leading to a joint toxicity that give rise to additive and/or synergistic effects (Rajapakse *et al.,* 2002).

POPs and other EDs are mainly man-made chemicals produced for a variety of industrial purposes and can be found in routine products such as hygiene and cosmetics, food articles and pesticides. POPs can be divided in two groups based on their source: (1) Intentionally produced or (2) Accidentally produced (Breivik *et al.,* 2002).

Most POPs are organohalogenated aromatic compounds. This group includes brominated, chlorinated and fluorinated chemicals among others (Safe, 1990). Polybrominated diphenyl ethers (PBDE) are an example of intentionally produced chemicals, which are very common. These compounds consist of phenyl rings connected by an ether bridge and have different degrees of bromination (Figure 1).

The compounds are widely used as flame-retardants and have very low water solubility (Darnerud *et al.,* 2001). Polychlorinated biphenyls (PBC) are also intentionally produced POPs that consists of paired aromatic phenyl rings with different degree of chlorination (Figure 1). These compounds are often used as flame-retardant due to their non-flammable ability and high boiling point, but are in addition commonly used in the

Polybrominated diphenyl ethers (PBDE)

Polychlorinated biphenyls (PCB)
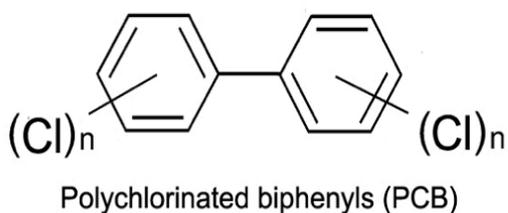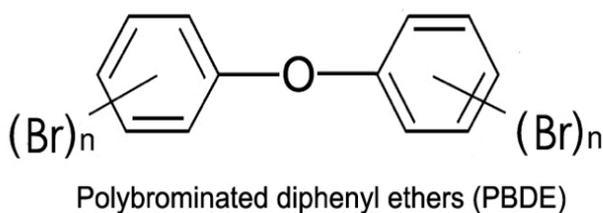
Figure 1 –The general structure of commonly found intentionally produced Persistent organic pollutants (POPs). Top: The general structure of PBDE. Bottom: General structure of PCB. Source picture: http://endocrinevet. blogspot.no/2012/07/flame-retardant-chemicals-in-house-dust.html

industry as an ingredient in e.g. plastic production (Safe, 1990). Fluorinated chemicals include both organic and inorganic aliphatic compounds with different degree of fluorination. These compounds are characterized by lipophilicity and are often used for this quality as e.g. oil or water repellents (Buck *et al.,* 2011).

Other chemicals with potential ED effect that are used as e.g. ingredients in personal care products include siloxanes and less persistent and bioaccumulative compounds such as triclosan and bisphenol A among many other compounds (Clayton *et al.,* 2011). Accidentally formed chemicals are by-products from different sources such as pesticide production and incomplete combustion of oil and coal like dioxins and furans (Breivik *et al.,* 2002).

1.1.1 Distribution of POPs to the environment

Persistent organic pollutants are widely distributed to the environment by natural processes involving soil, water and air. Emission of endocrine disruptors to the atmosphere, atmospheric deposition to the ocean and transport via ocean currents can contribute to spreading of these hazardous chemicals to the environment where they are accumulated and magnified in the food chain (Hansen *et al.,* 2004).

Partitions coefficients for air, water and octanol are used to describe physiochemical properties of chemicals. The octanol-water coefficient (K_{OW}) is used to describe the ratio of the solubility of a compound in octanol to its solubility in water. It is inversely related to solubility and proportional to molecular weight (Hawker *et al.,* 1998). The air-water coefficient (K_{AW}) describes the ratio of the solubility of a compound in air to its solubility in water. A high K_{AW} specifies the compounds ability to evaporate (Tancréde *et al.,* 1990). The octanol-air (K_{OA}) partition coefficient is used to predict the behaviour of a compound in the air and environment. It can be described as the solubility of a chemical in octanol to its ration of soluble concentration in air at equilibrium (Meyland *et al.,* 2005).

1.2 Nuclear receptors

Nuclear receptors (NR) are a superfamily of proteins consisting of approximately 150 members divided in 6 subfamilies (Wurtz *et al.,* 2000). Sequencing of the human genome has led to identification of 48 possible NR, but ligands have only been identified for 24 of them (Hashimoto *et al.,* 2005). The receptors are thought to

originate from a common ancient ancestor due to arthropod homologues (Aranda *et al.,* 2001).

The cellular action of many hormones is mediated through binding to nuclear receptors. NRs function as ligand-inducible transcription factors by interacting as monomers, homodimers or heterodimers mediating hormonal functions such as signalling for growth, development and reproduction. They are mainly localized in the nucleus, but some are also located in the cytoplasm and are translocated to the nucleus upon ligand binding (Figure 2) (Robinson *el al.* 2003).
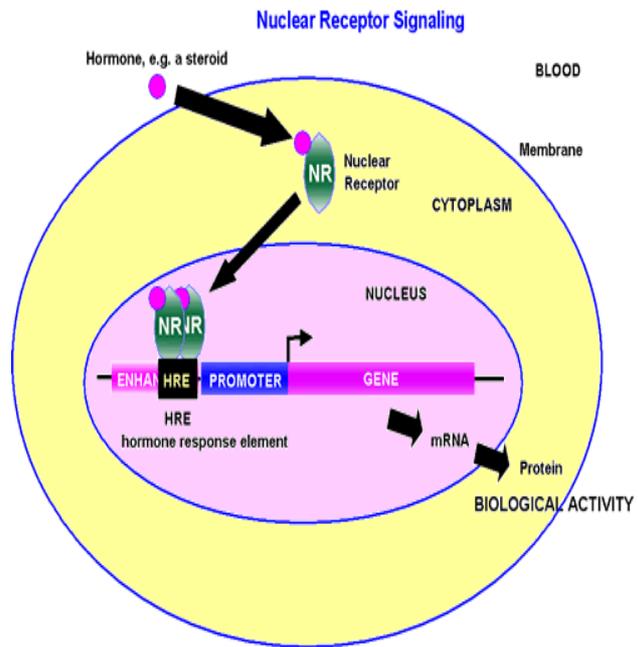
Figure 2 –General mechanism of nuclear receptors. The hormone fuses through the plasma membrane and bind to the receptor in cytoplasm (or nucleus). The receptor binds to the hormone responsive element on DNA and regulate gene transcription. Source picture: http://www.angelfire.com/sc3/ toxchick/endocrinology/endocrinology04.html

The modular structure of NRs exhibit functional domains that are conserved among related receptors. The N-terminal region of a receptor contains a non-conserved A/B region that codes for a transcriptional activation domain called AF-1 (Aranda *et al.,* 2001). The A/B domain is connected to the central C region, which contains a DNA binding domain (DBD). The DBD is the most conserved domain and experimental studies have shown that the receptor uses this domain to recognize a hormone response element (HRE) on the DNA (Germain *et al.,* 2006). The DBD is connected to the ligand-binding domain (LBD) through a poorly conserved hinge region (the D region). The hinge region allows rotation of the DBD relative to the LBD thereby avoiding steric hindrance. The hinge is thought to encode a nuclear localization signal, despite that it is not conserved between receptors (Germain *et al.,* 2006). The E region of NRs contains the LBD that features a dimerization surface, which mediates interactions with DBD, a co-regulator domain responsible for modulation of transcriptional activity and an activation function helix (AF-2, helix nr. 12)

responsible for ligand dependent transactivation. The C-terminal of the protein is less conserved and is referred to as the F domain (Figure 3) (Aranda *et al.,* 2001).

The crystal structures of multiple LBD have been resolved and show that the overall structure of proteins that belong to different subfamilies is similar within the NR superfamily. The LBD consists of 12 conserved α-helices and a conserved β-turn situated between Helix 5 (H5) and Helix 6 (H6). The overall structure is folded as a three layered antiparallel helical sandwich. A layer of three helices packed between two additional layers forms a cavity buried in the bottom of the structure, the ligand-binding pocket. The volume of this pocket varies among different receptors (Aranda *et al.,* 2001). Analysis of LBD in the X-ray crystal structure of several receptors have revealed an important structural feature concerning folding of H12 upon agonist binding. When the receptor is unbound to a ligand, it obtains an open conformation (apo-form) and upon ligand binding of an agonists a conformational change ensures H12 to close the pocket (holo-form) (Figure 4) (Hashimoto *et al.,* 2005)
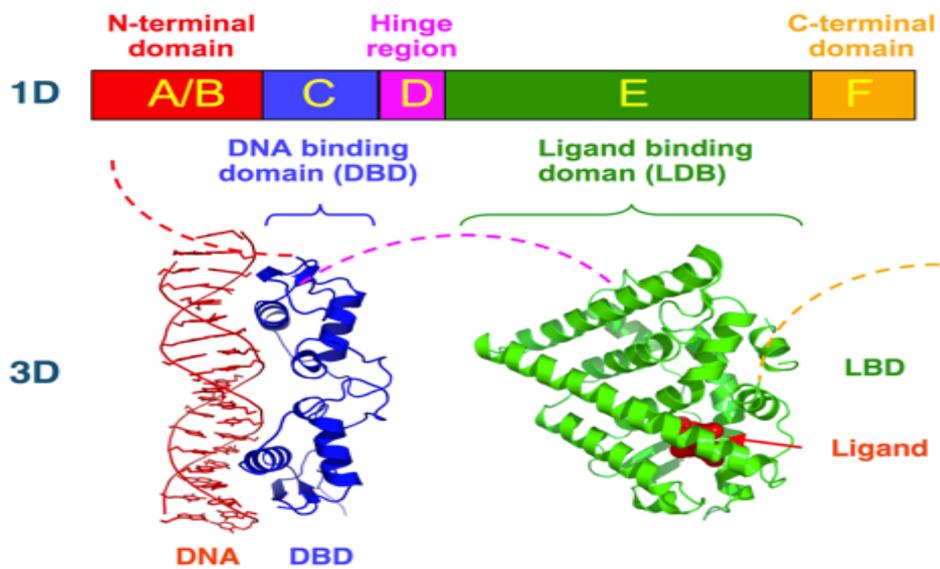
Figure 3 – Structural and functional organization of nuclear receptors. The top section shows the structural domains of a nuclear receptor with annotations. The lower section shows the 3D structure of the corresponding domains. Source picture: http://en.wikipedia.org/wiki/File: Nuclear_Receptor _Structure.png
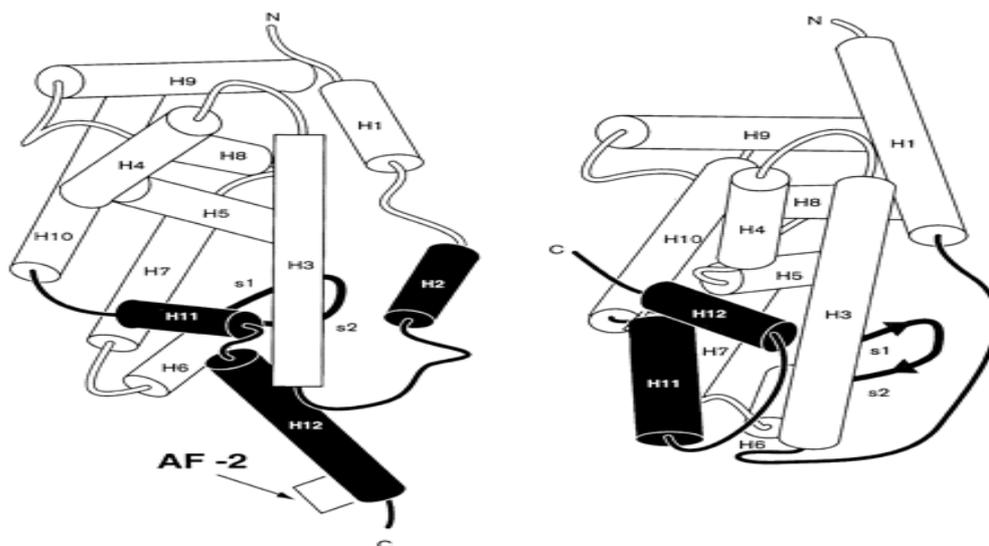
Figure 4 – Structure of the LBD in a nuclear receptor. Left; nuclear receptor in an open conformation unbound to a ligand (apo-form). Right; Nuclear receptor bound to an agonist in a closed conformation. Binding of an agonist changes the structural conformation of the receptor from an open form to a closed form (Aranda *et al.,* 2001)

1.2.1 Ecdysone receptor

The ecdysone receptor (EcR) is a ligand-dependent nuclear transcription factor found in arthropods. The receptor binds to ecdysteroid-hormones, which play a significant role in reproduction and control vital processes such as development, mounting and metamorphism (Kato *et al.,* 2007). A non-covalent heterodimerization of EcR and ultraspiracle protein (USP) is essential for activation of gene transcription and occur after binding of EcR to an agonist. USP is another member of the nuclear hormone family (Hill *et al.,* 2013).

The major active steroid hormone in insects and crustaceans is 20-hydroxyecdysone (20E) (Figure 5). The hormone is known to play an essential role in growth processes and studies have showed that 20E coordinates with juvenile hormone. Cross talk between these hormones has also been confirmed by experimental studies

Figure 5 –Structure of 20E and ponasterone A.
Top: The structure of the steroid hormone 20E. Bottom: The structure of the plant-derived steroid ponasterone A. Ponasterone A and 20E have three six membered and one five membered ring. The four membered ring structures are attached to an alkyl chain bound to functional groups. All steroid hormones possess the same skeleton with four rings structures and an alkyl group with various substitutions. Source picture: https://www.ebi.ac.uk/ chembl/compound/inspect/CdgfsdgdfgH EMBL224128, 2014.

in Arthropods (Mu *et al.,* 2004). In nature, many structural analogues of 20E exist. Ponasterone A is a phytoecdysteroid that have proved to be the most potent agonist for the EcR. Ponasterone A differs from 20E by the absence of a hydroxyl group at C-25 (Figure 5) (Gonsalves *et al.,* 2011).

The structure of the LBD of EcR is dynamic and structural rearrangements upon agonist binding ensure that the various ligands fit into the binding pocket (Hill *et al.,* 2013). The structural flexibility of the pocket makes it easy for compounds with structural similarity to the steroid hormones to mimic the endogenous hormones and disturb the hormone system. Many insecticides target growth by disrupting the hormonal system, but their effect on Arthropods such as *D.magna* has not been investigated (Kato *et al.,* 2007).

In total, 5 X-ray crystal structures of the EcR LBD from different species have been resolved and are available in Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do). The structures are obtained from three different species: *Bemisia tabaci*, *Tribolium castaneum* and *Heliothis virescens*. Two additional crystal structures of the EcR in *Heliothis virescens* are resolved, but co-crystallised with different agonists. Structural analysis of the X-ray crystal structures show that the steroids are fully trapped within the LBD and binds in a similar manner (the same position, orientation and interactions) despite variable volume of the pockets. The LBD within the resolved structures show the canonical tertiary structure of nuclear receptors with 12 α-helices and an antiparallel β-sheet between H5 and H6 (Hill *et al.,* 2013).

1.3 Daphnia as a model system

Most *Daphnia* species contains the molecular tools for production of haploid gametes and diploid eggs. Under normal conditions *Daphnia* reproduce asexually, but under external conditions such as high population density and depletion of food or extreme temperatures, *Daphnia* produce haploid resting eggs by meiosis that needs to be fertilized. The ability of cyclic parthenogenesis makes *Daphnia* an excellent subject for studies of the molecular mechanism of parthenogenesis (Stollewerk 2010).

Daphnia uses ecdysteroids and terpenoids as major endocrine signal molecules, which coordinates essential processes such as growth and development. *Daphnia* are very important in the aquatic ecosystem as previously explained (Section 1.1). The lack of knowledge concerning EDs effects on invertebrates has made *Daphnia* a model system for linking the mode of action and adverse outcomes of EDs. This is an approach for increasing the knowledge and improving the risk assessment of endocrine disrupting chemicals in invertebrates (Ashby *et al.,* 1997 and LeBlanc 2007).

1.4 In silico structural biology and drug discovery

Molecular modeling is defined as computational techniques used to mimic the behaviour of molecules. The technique is commonly used in fields of computational biology, drug design and lead optimisation.

Structure based drug-design techniques can serve as a guide for molecular modifications leading to increased potency and bioavailability of specific compounds. If a 3D structure or a model of the target is available, structure-based virtual screening can be used to identify potential binders. If the X-ray crystal structure or NMR structure of a protein is not available, a theoretical model can be build by homology modeling (Bohacek *et al.,* 1996).

An appropriate docking program can be used to place a small molecule (potential binder) into a target structure (e.g. receptor) in several different positions, conformations and orientations which are called binding poses. Each of these binding poses is given a score based on the fit into the binding pocket and the conformational energy of the small molecule in that particular pose. High-throughput docking is a rapid and inexpensive approach to predict the binding mode and affinity of many compounds towards a target, which can be selected for further experimental investigation (Kroemer 2007).

1.4.1 Molecular mechanics and Force fields

In computational chemistry, molecular modeling is used to construct models that mimic the behaviour of a molecular system. The model is a description of the inter- and intra-molecular forces that describe the 3D structure of the molecular system. The description may be quantum mechanically (QM), molecular mechanically (MM) or a combination of both (QM/MM) (Höltje *et al.,* 2008 and Gabrielsen *et al.,* 2011). For big molecular systems such as proteins and protein complexes, an MM description is most convenient due to the size of the molecular system. This approach in combination with docking can be a powerful tool for predicting the affinity and binding pose of environmental pollutants towards different proteins (Wu *et al.,* 2009). In MM, atoms are treated as individual particles and the atomic structure of a molecule is considered as a collection of masses interacting through harmonic forces (Gabrielsen *et al.,* 2011). Energies and interactions resulting from bond-stretching, angle-bending, torsional energies and non-bonding interactions are calculated without any consideration to electrons in the system (Höltje *et al.,* 2008).

Force fields describe intra and intermolecular forces of a molecular system and are used to calculate the total energy of the system. Deviations from a reference with

unstrained bonds, torsions and angles, in addition to non-bonding interactions are used in the calculations of the total energy. The collection of these unstrained values in addition to force field constants, which are empirically derived fit parameters, are the force fields (total potential energy) of a molecule (Höltje *et al.,* 2008). Force fields can be written as;

$$E_{tot} = E_{bonded} + E_{non\text{-}bonded},$$

$$E_{bonded} = (E_{bond} + E_{angle} + E_{dihedral}) + (E_{vdw} + E_{elec})$$

E_{tot} is the total potential energy calculated from energy deviations of bond-stretching, angle-bending and dihedral angles from their reference ($E_{bonded} = E_{bond} + E_{angle} + E_{dihedral}$) and non-bonded interactions that is electrostatic forces and Van Der Waals interactions ($E_{non\text{-}bonded} = E_{vdw} + E_{elec}$) (Höltje *et al.,* 2008).

1.4.2 Homology modeling

The 3D structure of a protein reveals a lot of information concerning structural and functional properties. Resolving 3D structures is vital in identification and analysis of the LBD, and essential for engineering protein properties and drug design. Homology modeling is a theoretical approach for predicting a 3D model of a protein with unknown 3D structure. The modeling technique takes advantage of structural conservation found in similar proteins that have evolved from a common ancestor. The amino acid sequence of the protein with unknown structure is often referred to as the target. The sequence of the target is used for a homology search to find similar sequences with resolved 3D structures to use as template for constructing a theoretical model of the target. Conserved regions of a protein are regions where the structural and sequence similarity between the template and the target are highest. These regions are easy to model, in contrast to non-
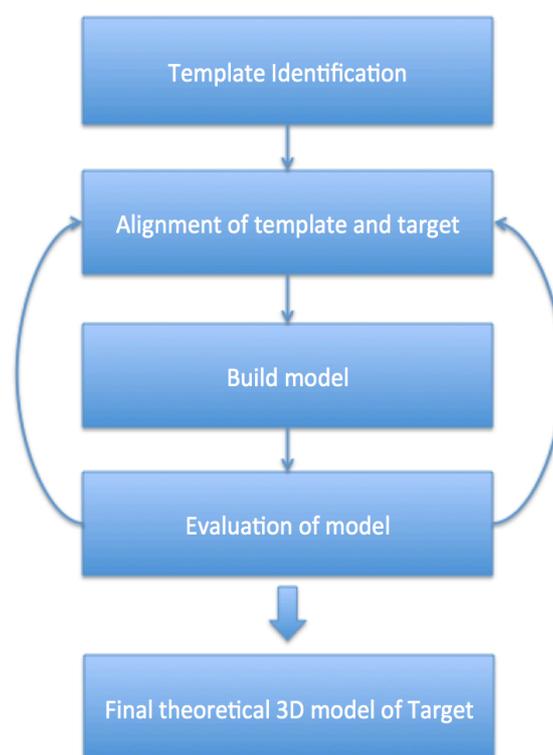
Figure 6 – Flowchart of the basic steps in homology modeling. The theoretical model of the target is built based on structural similarities between template and target.

10

conserved regions such as loops that connect important structural regions where the sequence can vary significantly among proteins (Krieger *et al.,* 2003).

Homology modeling consists of five main steps: (1) template identification, (2) amino acid alignment of template and target, (3) model building and (4) evaluation of model (Sylte *et al.,* 2012) (Figure 6).

Template identification

An appropriate template is identified by a sequence homology search in the PDB using the target sequence for obtaining similar sequences with resolved 3D structure (Bermann *et al.,* 2007).

Alignment

A sequence alignment between the template and the target is necessary for studying conservation and particular features such as catalytic or binding site residues etc. Aligning two sequences can be a difficult process if the sequence similarity is low. The corresponding positions in the two sequences must be matched for an optimal alignment. Experimental studies and analysis of proteins within the same family have shown that they are highly conserved with regards to residues and structure in important regions. In this case, the modeling process is simpler. A multiple sequence alignment of sequences with high similarity can be helpful for constructing a theoretical model of the target because it highlighting sequence conservation (Krieger *et al.,* 2003).

Build model and energy refinements

A theoretical 3D model can be constructed on the background of the alignment. This requires a suitable modeling program, which automatically constructs the model from the sequence alignment. The construction is a step-wise process that starts with modeling the core regions. The backbone conformation is transferred from the template to the target and conserved residues retain their side chain conformations. The next step is modeling of loop regions. This is a challenging step, since loops tend to be less conserved and of different lengths between the template and the target. A loop homology search can be performed for construction of a non-conserved loop region, where the sequence of the loop is used to search for loops with similar sequence and known conformation. The last step in building the theoretical model is rotation of non-conserved side chains and optimisation (Gabrielsen *et al.,* 2011).

Energy refinement of the constructed model is a method to increase the quality and optimize the energy of the model. Energy minimization is performed based on observations in nature, where stable state molecular systems correspond to low energy conformations of molecules (Sylte *et al.,* 2012). Refinement is a step-wise process with the purpose to remove close unnatural contact between amino acids and lower the energy of high-energy conformations added in the construction steps. A refinement can be performed using (1) energy minimization, (2) Monte Carlo simulations or (3) Molecular dynamics calculations (Höltje *et al.,* 2008). Energy minimization is based on iterations followed by energy calculation, which is used to refine the model towards an energy minimum. Monte Carlo simulations consist of random (stochastic) conformational moves followed by an energy minimization. The energy calculations from each round is stored and compared, and the conformation representing the lowest energy is saved. The aim of molecular dynamic calculations is to reproduce the behaviour of molecules during a period time. The atoms are moved at different time points followed by calculations of the new positions and velocity of the atoms. The new conformations are recorded and the procedure is repeated in a predefined number of times before the conformation representing the lowest energy is saved (kolinsky *et al.,* 1994).

Model evaluation

An evaluation of the constructed model is an important approach when the construction of theoretical models based on homologue proteins contains many elements of uncertainty. The Structural Analysis and Verification Server (SAVES; http://nihserver.mbi.ucla.edu/SAVES) is a commonly used service for evaluations and analysis of stereochemical quality of constructed models. A model is uploaded to the service, which provides different programs for evaluating the stereochemistry and geometry (Centeno *et al.,* 2005). Any errors in the model related to these features are reported to the user.

Docking of known ligands and mutational studies are other approaches that can be used to evaluate the quality of a constructed model. Data from mutagenesis experiments could support predictions from theoretical predictions proposed in docking studies concerning specific residues that are important for binding of a particular ligand or receptor activation

1.4.3 Semi-flexible Docking and scoring

Docking is defined as the positioning of a ligand in the active site of an enzyme or transporter protein or in the ligand binding site of a receptor. Scoring is a quality assessment of the docked ligands describing the interactions between the ligand and the target in terms of free energy to predict the binding affinity. Docking and scoring are important techniques used to predict the binding mode of known active ligands, predict binding affinities of compounds similar to the active ligands, and identify new ligands or chemicals that potentially bind to the target by using virtual screening (Leach *et al.,* 2006). Docking and scoring is a commonly used technique in drug discovery, but is also an excellent technique for predicting affinity of potentially toxic substances towards a target.

In an ideal docking-project, both the target and the ligand should be fully flexible to reflect the nature of structural flexibility in molecules, but the complexity of macromolecules make this very challenging and in many cases computationally unfeasible. Most available docking programs use a semi-flexible docking approach where the ligand is treated as flexible and the target as rigid (Leach *et al.,* 2006). There are several different approaches for including some degree of flexibility into the target. One approach is ensemble docking where flexibility is introduced to the target by docking the ligand in different conformations of the binding pocket. The various conformations of the pocket can be obtained from experimental crystal structures and/or computationally generated (Nabuurs *et al.,* 2007). Induced-fit docking is another approach that introduces flexibility to the receptor by performing a refinement of the side chains in the pocket with considerations to the ligand (with the ligand present) (Sherman *et al.,* 2006).

A scoring function can be used to rank the possible conformations/orientations of the ligands according to binding tightness in the pocket. Ideally, a scoring function will give the experimentally determined mode top rank. Scoring functions can be divided in three basic types; Force field, empirical and knowledge-based scoring function (Huang *et al.,* 2010). Force field is based on non-bonding interactions such as van der Waals (VDW) interactions, electrostatic interactions and stretching/bending/torsional forces. The empirical scoring function is based on a set of weighted energy terms such as entropy, desolvation and VDW for calculating binding affinities. Knowledge-based

scoring functions are based on energy potentials derived from structural information found in experimentally determined structures (Huang *et al.,* 2010).

1.5 *In vitro* testing to verify *in silico* predictions

Molecular docking can be used to predict binding affinities of small molecules towards a target. The predicted interactions should be experimentally verified to ensure consistency between computational predictions and actual ligand-binding specificity and affinity.

There are many experimental methods for testing theoretically predicted interactions between e.g. a receptor and an agonist or antagonist or between an enzyme and a substrate or inhibitor. Chemical tags like fluorescence labels can be used in binding studies and to reveal interaction, while radioactive labels can be used to measure binding affinity. Chemical tags are often objects to biases due to weak expressions of proteins (Whisenant *et al.,* 2010).

A two-hybrid assay is a powerful technique for detection of protein-protein interactions. The system can also be used for discovering molecules affecting these interactions, in addition to identification of residues or domains involved in the interactions (Miller *et al.,* 2004). The principle behind this technique is the usage of fusion proteins, where interaction between these fusions results in transcription of a detectable end product such as firefly luciferase (Figure 7). The assay can, among other applications, be used to determine the biological activity of a drug-related compound for specific biological targets and efficacy of a receptor towards pollutants. Many compounds can exhibit toxicity against cells that could lead to lack of response in many cases.

The cytotoxicity of compounds subject for testing can be evaluated by measuring changes in central cellular processes such as the metabolism and membrane integrity of cells exposed to different compounds (Bopp *et al.,* 2008).

14
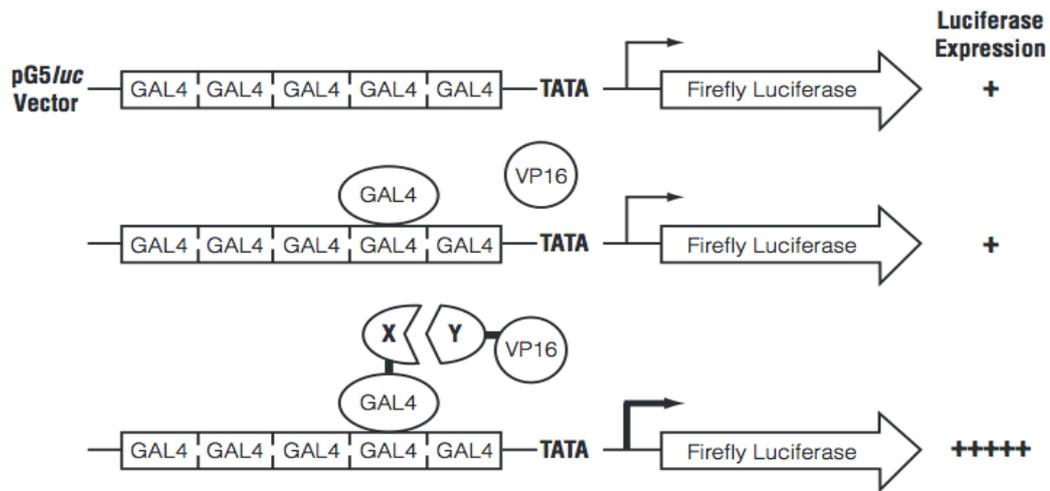
Figure 7 –Representation of the CheckMate^{TM} Mammalian two-hybrid system. The pG5*luc* vector contains five GAL4 binding sites upstream of the firefly luciferase gene. Interactions between protein X in fusion with GAL4 (pBIND) and protein Y fused with VP16 (pACT) results in expression of firefly luciferase. Source photo: https://www.promega.jp/resources/product-guides-and-selectors/protocols-and-applications-guide/protein-purification-and-analysis/

2 Aim

Several studies have shown that many environmental pollutants can bind to nuclear receptors and perturb their signalling. Such binding may initiate adverse outcomes. Knowledge about the relationships between the molecular structure, exposure and concentration of the pollutants on one side, and the interactions with cellular signalling pathways relevant for human and wild life on the other side, is extremely important for the risk-assessments of the particular pollutant. Disruption of hormone signalling by binding of pollutants to nuclear receptors can be verified experimentally, but this is often time consuming and associated with high costs. The need for rapid and cost-efficient testing approaches for high-throughput screening of a high number of chemicals have led to development of computational approaches using various prediction models.

A high quality 3D model of a protein can be used as a tool to predict binding affinities towards different compounds and thereby give an indication of signal pathways that can potentially be affected. This can contribute to lowering the costs and time associated with experimental testing since fewer pathways and targets need to be experimentally tested. In that way, homology modeling and docking can be used as an important supplement to experimental testing in e.g. risk assessments of pollutants.

This study focuses on exploring the possibility of 3D homology models to predict binding affinity of pollutants towards the EcR in *D. magna.*

The structure of the EcR in *D. magna* is unknown. The aims of the project were therefore to:

- Predict the 3D structure by using homology modeling.
- Predict putative interactions between a set of molecules/pollutants and the 3D homology model.
- Transfer a dual-luciferase reporter assay system from NIBB, Japan, to NIVA, Oslo, for testing agonist binding of compounds to the EcR.
- Support the 3D model by experimental (*in vitro*) testing.

Docking studies were used as a high throughput screening of potential binders and will in addition reveal information concerning molecular interactions and residues critical for interactions between pollutants and the EcR. .

3 Materials and Methods

3.1 Materials

3.1.1 Software

Molsoft Internal coordinate Mechanics software (v.3.7.3c)

The Internal Coordinate mechanics (ICM) software is a molecular modeling package that contains a variety of algorithms, prediction and analysis tools that can be used for operations such as homology modeling, structure prediction, docking (flexible and semi-flexible), pharmacophore modeling, calculations of electrostatic potentials at protein surfaces, sequence analysis and alignments. In this project the ICM software was used for multiple sequence alignments of the EcR sequence from *D. magna* with different template sequences from the PDB. Construction of homology models, docking and scoring procedures was also performed using the ICM software (Abagyan *et al.,* 2004).

Molcart in ICM

Molcart is a chemical management system integrated in the ICM software. The ICM cheminformatic tool is used to connect to the collection of compound databases before searching and analysing compounds of interest. Inactive ligands (decoys) were obtained from the ChemDiv database of the MolCart library (https://www.molsoft.com/molcart-compounds.html).

GraphPad Prism (v.6)

Prism is scientific graphing and statistics software providing features such as nonlinear regression with various options such as comparative models, comparative curves, nonparametric comparison, analysis of possibility tables etc. The obtained data from the two-hybrid assay was analysed using this software (http://www.graphpad.com/scientific-software/prism/).

BLASTP

Basic Local Alignment Search Tool for protein sequences (BLASTP) is a software package for performing alignment-based database queries using amino acid sequences. In this project, BLASTP was used to align two sequences in order to obtain the degree of sequence similarity between the EcR LBD form *D.magna* and from *D.melanogaster* (Altschul *et al.,* 1990).

19

3.1.2 Databases

ChEMBL

ChEMBL is a database containing information on bioactive molecules. The database provides experimental information concerning molecular interactions, drug-approvals and other clinical candidates (Gaulton *et al.,* 2012). The 19 active EcR ligands were found and downloaded from ChEMBL (Harada *et al.,* 2009) (https://www.ebi.ac.uk/chembldb/).

ChemDiv

The ChemDiv database specializes in drug-like compounds for drug discovery. The library contains 1,511,689 compounds (as of 01.02.14). The assumed non-binders (decoys) were obtained from this library by using each of the known active ligands to search for similar compounds. The database finds structures similar to the ligands by using specific characteristics of the ligands structure (fingerprints) to search for structures with similar features. Fingerprint characteristic can include number of ring structures, size, molecular weight and charge among other features (http://chemistryondemand.com/compound-library).

UniProtKB

UniProtKB is a database consisting of two different sections: manually annotated records (UniProtKB/Swiss-Prot) and computationally analysed records of protein sequences (UniProtKB/TrEMBL). The primary amino acid sequence of EcR in *D.magna* (target) was found and downloaded from UniProtKB/Swiss-Prot (accession: B0L4A2_9CRUS) (http://web.expasy.org/docs/relnotes/relstat.html).

Protein Data Bank (PDB)

The protein data bank is large structural database containing data on nucleic acids and protein structures including atomic coordinates obtained by X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy (Berman *et al.* 2007). The database allows the user to visualize and analyse structures and provide available annotations concerning the sequences. The crystal structures of the EcR co-crystallized with an ecdysteroid (1Z5X_e and 2R40_d, Table 1) were downloaded from PDB (http://www.rcsb.org/pdb/home/home.do).

Structural Analysis and Verification Server (SAVES)

SAVES is a metaserver that enable the users to perform structural evaluations of protein structures. In this project the PROCHECK, ERRAT and VERTIFY_3D programs were used to analyse the structure of the constructed homology models. PROCHECK performs a quality check of the stereochemistry by analysing the residue-by-residue geometry in addition to the overall geometry. The result of the analysis is represented by a Ramachandran plot that visualises the backbone dihedral angles (ϕ and Ψ) of the amino acids in the structure. ERRAT uses different algorithms to evaluate the statistics of non-bonded interactions between different types of atoms (Colovos *et al.,* 1993). VERTIFY_3D analyse the compatibility of the atomic model. Based on the environment and location of the residues, the structural class is decided for each residue before the results are compared to a collection of known structures as a reference for scoring (Bowie *et al.,* 1991). The database can be found at: http://nihserver.mbi.ucla.edu/SAVES/.

3.1.3 Chemicals

Ponasteron A and triclosan were purchased from Wako Pure Chemical Industries Ltd, Osaka, Japan. 20E was purchased from ICN (Costa Mesa, USA). Bisphenol A was purchased from Trademark (TCI) and emamectin benzoate from Flukar. TFOA, endosulfan and diethyl phthalate were purchased from Sigma-Aldrich, St.Louis, USA. The dual-luciferase reporter assay system was purchased from Promega, Madison, USA (E1960).

3.1.4 Detection system

The GloMax ®- Multi + Detection system from Promega (Madison, USA) was used for detection of luminescence when performing the assays at the University of Basic Biology, Japan. Victor[3] 1420 Multilabel counter with software version 3.00 from PerkinElmer (Massachusetts, USA) was used to measure the luminescence and fluorescence when performing the assays at NIVA, Oslo.

3.2 Method

3.2.1 Homology modeling

The Crystal structure of the EcR in *D. magna* has not been resolved. A homology modeling approach was therefore performed to construct theoretical 3D models of the receptor. Only two homology models were made of the target based on different templates. The reason for that was that a previous project at the Medical Pharmacology and Toxicology Research group (unpublished) identified two structures as more appropriate as templates than other available templates.

Amino acid sequence alignments

The complete amino acids sequence of the EcR ligand-binding domain (LBD) from *D. magna* was available in the UniProt database (http://www.uniprot.org/uniprot/B0L4A2). X-ray crystal structure coordinates of the two previously identified template structures were downloaded from the PDB ((ID: 1Z5X_E: *Tribolium castaneum* and 2R40_D: *Heliothis virescens*) and used as templates (Table 1). The target sequence and the template sequences were aligned and adjustments were made in the alignments to avoid gaps in important structural domains (helices and β-sheets). In addition, a multiple sequence alignment between the EcR LBD sequence and the top 8 ranked homologues sequences in the PDB was made with the purpose of investigate the structural conservation of EcR in different species (supplementary Figure S1) No adjustments were made in this alignment.

Construction of models

The ICM build model macro was used to construct the two models based on the obtained alignment between the two previous recommended templates and the target (default settings). The macro construct a model based on three main steps: 1) modeling of the core regions by transferring the backbone conformation from the template to the target, 2) Construction of non-conserved loops regions by a loop homology search in PDB, 3) placing the side chains and optimisation. In the last step, the conserved side chains are directly transferred from the template to the target and the non-conserved side chain are modelled or transferred without any reference to the template (Sylte *et al.,* 2012).

Model refinement

The ICM refineModel macro was used for energy optimisation of the constructed homology models. The macro performs side chain conformational sampling by using the Monte Carlo module integrated in the macro, with iterative annealing and a second side chain sampling (Gabrielse *et al.,* 2012). Five iterations were performed. The iterations consist of random movements of the side chains followed by a local energy minimization. The random movements generate an energy gradient and the side chains with energy above the gradient are selected for energy minimization. The complete energy is calculated and the iterations are accepted or rejected (Abagyan *et al.,* 2004)

Model evaluation

The constructed models were uploaded to the Structural Analysis and Verification server (SAVES) to check the stereochemical quality PROCHECK, ERRAT and Vertify_3D were used for this purpose (http://nihserver.mbi.ucla.edu/SAVES/).

icmPocketFinder

The icmPocketFinder macro was used to detect possible binding pockets in the constructed 3D models of the EcR. The algorithm does not require knowledge concerning potential ligands since it is based on a transformation of the Lennard-Jones potential calculated from the 3D structure of the receptor (An *et al.,* 2005). The tolerance level was set to 4.6 (default setting). The agonists in the X-ray crystal structure were displayed in the constructed homology models and the pocket corresponding to the position of the agonists was selected for the docking project.

3.2.2 Construction of test set of compounds

Selection of ligands

Ligands for the EcR were found in the ChEMBL database (https://www.ebi.ac.uk/chembl/). The database listed 19 known ligands for the EcR in *Drosophila melanogaster* with corresponding IC_{50} values (Harada et al. 2009). The EcR sequence of *D. melanogaster* and *D. magna* have an overall sequence similarity of 50%, but the identity in the ligand binding pocket is approximately 85,7% and the ligands of *D.melanogaster* are therefore believed to bind the EcR of *D. magna* (supplementary Figure S2).

Ligands with an IC_{50} <2600 nM (9 compounds) were considered active, and selected for the test set.

Screening of the Chemdiv database collection for decoys

The Chemdiv database in the Molsoft library was used to search for decoys (presumed non-binders) using the 9 active ligands as references. The decoys were selected based on physiochemical similarity with the 9 binders by using fingerprint similarity search. A fingerprint similarity search is based on unique characteristic within a structure of a compound and searches for compounds with similar characteristics, but with some degree of chemical dissimilarity so that they can be considered non-binders. The maximum distance value for the searches was set to 0,4 (default setting). The max distance value determines the degree of identity between the reference compounds and the target compounds. Lower values allow more similarity between the reference compound and the potential decoy.

Clustering of active ligands and decoys

The active ligands and the decoys were saved in one list and clustered using the TREE method with weighted pair group method with arithmetic mean (UPGMA) to see the diversity of the structures (Loewenstein *et al.,* 2008). The compounds were clustered with distance range set to 0.1 and one representative decoy within each cluster was manually chosen. A total of 155 decoys were selected for the docking project.

Pollutants dataset

A ligand dataset consisting of 655 pollutants with putative EDs were obtained from Dr. Lisa Bjørnsdatter Helgason working on environmental pollutants at UiT The Arctic University of Norway. The dataset was originally constructed by Howard and co-workers (Howard *et al.,* 2010) as an approach to identify commercial chemicals that might be persistent and bioaccumulative, but has not been included in contaminant measurement programs. The dataset included siloxanes, PCBs, PBDEs and many fluorinated compounds in addition to well-known chemicals like triclosan.

3.2.3 Semi-flexible docking

The dataset consisting of the 655 pollutants and the test set with binders and decoys were docked into the two constructed homology models. A semi-flexible docking approach allowed the compounds to be fully flexible, but the models were represented

as rigid structures. The ICM software represents the homology models as a set of rigid pre-calculated grid potential maps representing interacting terms such as hydrogen bonds, Van der Waals, hydrophobic and electrostatic forces.

A Monte Carlo global optimisation procedure was used to predict the binding pose of the compounds in the ICM software. A diverse set of conformations of the compounds was generated *in vacuo* by sampling of the torsional and rotational degrees of freedom (Abagyan *el al.* 2004). The conformations are placed into the binding pocket of the homology models as a starting point for global optimisation of the energy function. The global optimisation procedure performs iteration of random torsional and positional moves followed by a local energy minimization. Torsional moves are randomization of a single arbitrary chosen torsion angle and positional moves are pseudo-Brownian random translation or rotation of the whole structure of the compound. The conformation of the compounds are either accepted or rejected based on the energy (Bursulaya *et al.,* 2003). The low energy conformations are stacked, saved and ranked based on the docking energy.

Re-docking of compounds from X-ray complexes

In order to test the accuracy of the docking performance, the ligands were removed from the X-ray structure of 1Z5X (PDB ID) and 2R40 (PDB ID), and the ligand was re-docked into the structure. Ponasterone A was co-crystallised with 1Z5X and 20E was crystallised with 2r40. The Root mean square deviation (RMSD) between the ligand in the native X-ray complex and the re-docked ligand was calculated. This value describes the conformational differences between the predicted and the observed pose. The prediction is considered successful if RMSD <2.0 Å (Huang *et al.,* 2010).

Scoring

The ICM virtual ligand screening (VLS) scoring function was used to score, evaluate and compare the binding energy of the test set of ligands and decoys and dataset of pollutants. This is an empirical scoring function that uses steric, entropic, hydrophobic and electrostatic terms to calculate the score (Huang *et al.,* 2010). A correction term proportional to the number of atoms was included in the score calculations in order to avoid biases towards larger pollutants and ligands/decoys (Schapira *et al.,* 2003).

ICM batch docking was performed in three parallel runs for each of the two docking projects (the binder/decoy test set and the dataset of pollutants). Batch docking places and scores all compounds automatically. Alternative conformations of the pollutants, ligands and decoys from each run were scored, and a hit list was made to select the top ranked conformations for each ligand.

Docking and scoring of test set

The constructed homology models were evaluated according to their ability to separate binders from decoys i.e. the selectivity of the receptor, by making Receiver Characteristics Operator (ROC) curves. ROC-curves use the scoring values of the ligands and the decoys to compare the number of ligands that was predicted as binders (true positive) against decoys predicted to bind (false positives) (Lindin *et al., 2013*).

Selection of compounds for experimental testing

Docking scores of the dataset consisting of 655 pollutants were used as a criterion to select compounds for experimental testing. A total of 8 compounds were supposed to be selected. Two of the active ligands were chosen as positive control and two of the pollutants with poor score were selected as theoretical negative control. The 4 remaining compounds were selected based on 2 qualities in addition to docking score above or close to the threshold set by the active ligands: (1) structural similarity with active ligands and (2) commercial availability. Since one of the initially selected compounds was commercial unavailable, another compound was selected. Another of the 4 compounds did not arrive in time for the experiments and was therefore replaced by another compound with unknown docking score. This compound was docked after the experimental testing in the luciferase reporter assay

3.2.4 Experimental analysis

Two-hybrid testing system for EcR activity

A two-hybrid assay was used for detection of activity of the EcR in *D. magna* after exposure to selected potential ED. The system was applied for identification of interactions between EcR and environmental pollutants, where EcR dimerize with USP upon binding of an agonist and activate transcription of firefly luciferase (Fields *et al.,* 1994). The principle of this technique is that the binding of an agonist to EcR

and dimerization with USP activates transcription of a reporter gene, which is translated into an enzyme for witch the catalytic activity can be determined by measuring a luminescence signal.

Construction plasmid vectors

Three vectors of the CheckMate^{TM} Mammalian two-hybrid system Kit (Promega E2440, Madison, USA) were used: (1) pBIND vector, (2) pACT vector (3) pG5*luc* vector. The pBIND vector contained the yeast GAL4 DNA-binding domain upstream of a multiple cloning site where USP was inserted. The vector also contained *Renilla reniformis* luciferase, which was controlled by the VP16 promotor. The pACT vector contains the VP16 activation domain upstream of the cloning region where EcR was inserted. The *pG5luc* vector contained five biding sites for the GAL4 DNA-binding domain, which were upstream of the reporter gene coding for the firefly luciferase. FuGENE (Roche Diagnostics, Basel, Switzerland) was used as transfection reagent. This is a nonliposomal formulation designed to transfect cell lines. A fourth vector, pACT-droTaiman (LXXLL), was used with an insertion of the LxxLL domain of *D. melanogaster*. This is known to be co-factor for the USP/EcR dimerization (Kato *et al.,* 2007 and Zhu *et al.,* 2006).

The clones were obtained from National Institute for environmental Studies, Japan (NIES: Tsukuba, Japan:Tatarazako *et al.,* 2003) and the vectors were prepared by Prof. Taisen Iguchi, the National institute of basic biology (NIBB), Okazaki, Japan, prior to the project start (Kato *et al.,* 2007).

Cloning of vectors

To ensure sufficient amount of the vectors, competent *E.coli* (One Shot® TOP10 Chemically Competent *E. coli*, Invitrogen^{TM}, Carlsbad, USA) were transformed with: (1) pBIND-EcR, (2) the pACT-USP, (3) pG5*luc* and (4) pACT-droTaiman (LXXLL). The GenElute^{TM} HP Plasmid Midiprep Kit (NA0200 SIGMA) was used to purify the cloned vectors according to manufactures protocol.

Cell culture

Chinese hamster ovary (CHO-K1) cells were purchased from public health agency cultural collection, Microbiology Services (Cat. No. EC85051005). The cell line is a

sub-clone of the parental CHO cell line derived from the ovary of an adult Chinese hamster.

The cells were cultured on petri dishes with 10 mL Dulbecco´s modified Eagle´s medium (DMEM) (Gibco, Life Technologies, Carlsbad, USA) containing 10% fetal bovine serum (FBS) (Sigma-Aldrich, St.Louis, USA) at 37°C and 5% CO_2.

Transfection

One day prior to transfection, the cells were washed with phosphate buffered saline (PBS) (Gibco, Life Technologies) and detached from the petri dish by using 1 mL 0.25% trypsin ethylenediaminetetraacetic acid (EDTA) (Life Technologies). After removing the trypsin EDTA, 5 ml fresh media (DMEM+FBS) was added to the petri dish and the solution was transferred to a 50 mL falcon tube. The cells were centrifuged for 5 min at 1000 rpm before they were resuspended in 2 mL fresh media. The cells were counted using a hemocytometer, before 10^4 cells were transferred to the wells of a 24-well plate containing 900 μL fresh medium. The 24-well plate was incubated for 24 hours at 37°C and 5% CO_2.

After 24 hours incubation, the transfection was performed using 1μl FuGENE HD (Roche diagnostics) according to manufacturers protocol. Each of the 24 wells received 30 ng pBIND-dapEcR (LBD) vector, 30 ng pACT-dapUSP (LBD) vector, 100 ng aACT-droTaiman (LXXLL) vector, 300 ng pG5Luc vector and 100 μL FBS.

The 24-well plate was incubated for 4 hours at 37°C and 5% CO_2.

Hormones and chemicals selected for testing were diluted to obtain concentrations in the range of 10^{-2}-10^{-9}M. The compounds were additionally diluted 10x in growth media without FBS, before 10 μL were transferred to the wells with 3 technical replicates (Figure 8). The 24-well plate was incubated for 40 hours at 37°C and 5% CO_2.
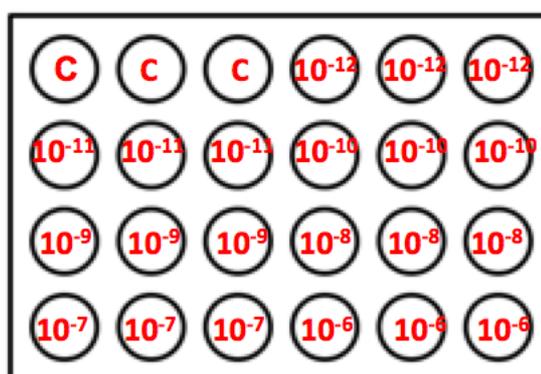
Figure 8 - 24-well plate with applied concentrations of the positive controls. All concentrations were tested in three technical replicates. C= control (DMSO). Concentrations are given in M.

28

The assay was conducted 3 times to obtain data that were independent of each other for each tested compound.

Detection

The cells were incubated for 40 hours with hormones and chemicals, before they were washed with 1 mL PBS. Cell lysis was performed by adding 100 μL passive lysis buffer (PLB) (Promega) diluted 1:5 in MQ water to each of the wells before shaking for 20 min on an orbital shaker (450 rpm). The lysate from each well were transferred to a 96-well plate to obtain a volume of 8 μL pr. well.

The amount of *Renilla* luciferase and firefly luciferase expressed in the cells was quantified using the Dual-Luciferase® Reporter assay system kit (Promega E1910) according to the manufactures protocol.

Cytotoxicity test

The viability of the cells was measured at the end of the exposure period. Almar blue (AB) was used as a marker for the metabolic integrity of the cells because the probe is taken up by the mitochondria, where it becomes reduced into a detectable fluorescence product. The cell membrane integrity was analysed by using 5-carboxyfluorescein diacetoxymethyl ester (CFDA). CFDA is taken up in the cytosol and hydrolysed to another detectable fluorescence product (Peters *et al.,* 2007).

A stock solution of CFDA-AM was pre-made using 5 mg CFDA-AM (5mg CFDA (Molecular probes VWR C-1354, Radnor, USA) and 2350 μL DMSO (Sigma-Aldrich D-8779)) to obtain a final concentration of 4 mmol/L. A Tris buffer was pre-made to obtain a concentration of 50mM Tris with pH 7.5 (0.97 g Trizma base (Sigma-Aldrich), 6.61 g Trizma HCl (Sigma-Aldrich) in 1L of distilled water (made by Karina Petersen, NIVA).

The incubation media was prepared by adding 11.6 μL CFDA-AM stock solution, 579 μL AB (Canadian life technol, Medprobe DAL1100) and 11 mL Tris buffer to an Erlenmeyer flask to obtain final concentrations of 4 μM CFDA-AM and 5% AB.
Preparation of the cells for the cytotoxicity test was performed by the same procedure as the cells cultured for transfection and detection as previously described, although without performing the cell lysis. After 40 hours incubation, the growth medium was

removed and 400 µL incubation media were added to each well. The plates were incubated for 30 minutes in the dark at room temperature before fluorescence readings were performed using the Victor[3] multilabel counter with excitation and emission wavelength of 530-590 (AB) and 485-530 (CFDA-AM).

A cobber standard was also made to represent the maximum cell death (positive control), which was used to normalize the data (solvent control 100% viability and positive control 0% viability). The cobber standard was made in concentrations from 1mg/mL to 0.0156 mg/mL on a 24-well plate with 3 technical replicates.

Statistical and graphical treatment

The expression of *Renilla* was normalized to the expressed firefly luciferase, before the total expression was normalized to unexposed cells. The median of each technical replicate was calculated. The assay was conducted 3 times for each compound, which gave 3 medians for each of the selected concentrations.

The measurements from the cytotoxicity test were analysed by dividing the data obtained from each concentration to the average of the control (100% viability) and the cobber standard for 0% viability. The test was only conducted 1 time for each compound.

The data obtained the two-hybrid assay and the cytotoxicity test from for were uploaded to GraphPad Prism, where the mean of the medians were plotted against the measured change in signal obtained at different concentrations. Error bars were applied to indicate the standard deviation of the mean within the measurements obtained from every concentration of compound.

Difference in mean expression at different concentrations of compounds were analysed by a one-way analysis of variance (ANOVA) followed by a Dunnetts´ test for multiple comparison of concentrations against the controls. The tests were performed with default settings and the alpha for the Dunnett´s test was set to 0.05.

4 Results

One of the objectives of this project was to predict the 3D structure of EcR in *D. magna.* Homology modeling was used to construct two theoretical 3D models of the receptor structure. Structural analysis and molecular docking were performed to evaluate the models. Docking studies were also performed to predict the binding affinity of selected compounds towards the receptor and to examine the receptor-ligand interactions. A second goal was to support the theoretical model of the EcR by experimental testing of the predicted agonists to study binding to the EcR by using a ligand dependent two-hybrid luciferase reporter assay.

4.1 Homology modeling

4.1.1 Identification of template and sequence alignment

In order to identify templates, the protein sequence of the EcR LBD was uploaded to PDB for a homology search. Two previously identified template structures (PDB ID: 1Z5X and 2R40) were selected as templates and downloaded from the database (Table 1). The template sequences and the target sequence were aligned and adjusted to avoid gaps in structurally conserved regions (Figure 9). A multiple sequence alignment between EcR LBD of *D. magna* and the top 8 most similar sequences in PDB was also made and displayed with their corresponding primary and secondary structures to highlight sequence and structural conservation (supplementary Figure S1). The alignments showed that helical segments were highly conserved at the structure level and less conserved at the sequence level. Three β-sheets were found in all species and possessed high sequence conservation in addition to structural conservation. Loop domains were less conserved between the species, being dissimilar in both length and sequence.

Table 1 –**Templates selected for alignment and construction of models.** The top ranked X-ray crystal structure from the homology search in Protein Data Bank with the amino acid sequence of the EcR in *D. magna*. Identity states the amino acid identity of the target compared to corresponding template in percent. Model states the numbering of the homology models constructed from the different templates. LBD = Ligand binding domain.

PDB code	Species	Function	Co-crystallised agonist	Identity (%)	Resolution (Å)	Modell
1Z5X_E	*Bemisia tabacil*	EcR LBD	Ponasteron A	71	3.1	I
2R40_D	*Heliothis virescens*	EcR LBD	20-hydroxyecdysone	58	2.4	II

```
ID=71% pP=36.7       .P#.PEQEELI.RLVYFQ.E#-.PS.ED#++I....#.E......-.+F+HITE#TILTVQL.VEFSKRLPGFD.L#REDQ
B0L4A2_9CRUS1    1   KPLSPEQEELINRLVYFQEEFDQPSEEDLRKISTSGIHESDA---DAKFKHITEMTILTVQLTVEFSKRLPGFDTLLREDQ
1z5x_e           1   -PITPEQEELIHRLVYFQNEYEHPSPEDIKRIVNAAPEEE--NVAEERFRHITEITILTVQLIVEFSKRLPGFDKLIREDQ
1z5x_e
                           H1              H2                      H3            H4

                     I.LLKACSSEVMM#R#ARRYDA.TDSI#FA.N.PYTRESY.#AG#GDT#-.L#RF.+.M.#MKVDNAEYALLTAIVIFSER
B0L4A2_9CRUS1  257   ITLLKACSSEVMMLRCARRYDANTDSIVFANNLPYTRESYNMAGVGDTADSLFRFGKTMSLMKVDNAEYALLTAIVIFSER
1z5x_e          79   IALLKACSSEVMMFRMARRYDAETDSILFATNQPYTRESYTVAGMGDTVEDLLRFCRHMCAMKVDNAEYALLTAIVIFSER
1z5x_e
                           H5               H6              H7            H8

                     P.L.E..KVEKIQEIY#EAL.AYV.N.......#.T.FAKLLSVLTELRTLGN#NSE.CFSLKLKN+K#P.FL.EIWDV#
B0L4A2_9CRUS1  338   PGLVEARKVEKIQEIYLEALQAYVMNHRV---RPMTEFAKLLSVLTELRTLGNLNSEMCFSLKLKNKKLPPFLAEIWDVH
1z5x_e         160   PSLSEGWKVEKIQEIYIEALKAYVENR--RKPYATTIFAKLLSVLTELRTLGNMNSETCFSLKLKNRKVPSFLEEIWDVV
1z5x_e
                           H9                      H10                   H11


ID=58%               .V.PL.#.Q..LI.RLV##QE.#-QPSEEDL++#.............#.F+.ITEMTILTVQL.VEF.K.LPGF..#...DQI
2r40_d           1   -VPPLTANQKSLIARLVWYQEGYEQPSEEDLKRVTQTD--------MPFRQITEMTILTVQLIVEFAKGLPGFAKISQSDQI
B0L4A2_9CRUS2    1   PVKPLSPEQEELINRLVYFQEEFDQPSEEDLRKIST--SGIHESDADAKFKHITEMTILTVQLTVEFSKRLPGFDTLLREDQI
2r40_d
                           H1              H2              H3            H4

                     TLLKACSSEVMMLR#ARRYDA.TDS##FANN.#YTR-.Y..AG#....#-.L#.F.+.M..M.#DN#.YALLTAIVIFS-RPGL
2r40_d          74   TLLKACSSEVMMLRVARRYDAATDSVLFANNQAYTRDNYRKAGMAYVIEDLLHFCRCMYSMMMDNVHYALLTAIVIFSDRPGL
B0L4A2_9CRUS2  370   TLLKACSSEVMMLRAARRYDANTDSIVFANNLPYTRESYNMAGVGDTADSLFRFGKTMSLMKVDNAEYALLTAIVIFSERPGL
2r40_d
                           H5               H6            H7            H8

                     ..#..VE.IQ.#YL..L.#Y##N.........##..F.K#L.#LTE#RTLG..NS.MC#SLKLKN+KLPPFL.EIWDV#.
2r40_d         157   EQPLLVEEIQRYYLNTLRVYILNQNSASPR--CAVIFGKILGILTEIRTLGMQNSNMCISLKLKNRKLPPFLEEIWDVA-
B0L4A2_9CRUS2  453   VEARKVEKIQEIYLEALQAYVMNHR-----VRPMTEFAKLLSVLTELRTLGNLNSEMCFSLKLKNKKLPPFLAEIWDVHS
2r40_d
                           H9                      H10                   H11
```

Figure 9 – Alignement of ecdysone receptor sequence from *Daphnia magna* and template sequences found in the Protein Data Bank. Secondary structures are indicated below the sequences. The red cylinders represents helices and green arrows represents β-sheets. The black squeres indicate residues within 4 Å sphere radius from the agonist in the binding pocket of the target and the red squares indicate residues within 4 Å sphere radius from the agonist in the binding pocket of the template. Top: Alignment of the EcR sequence from *D.magna* and 1Z5X (*B. tabacil.*) The sequence identity was 71%. Bottom: Alignment of the EcR sequence from *D.magna* and 2R40 (*H. virescens*). Sequence identity between the target and the template was 58%

4.1.2 Construction of homology model

Two models of the EcR LBD were constructed based on the sequence alignments between the target and the templates of the two crystal structures obtained from the PDB (ID: 1Z5X_E: *T. castaneum* and 2R40_D: *H. virescens*). The icmPocketFinder macro was used to identify the binding pocket in the constructed homology models. The macro identified more than one pocket in each of the homology models and the crystal structures of the templates with the co-crystallised ligands were superimposed with the constructed models to ensure correct selection of the LBD. The volume of the pockets differed between model I and model II (Table 3).

Residues within a 4 Å sphere radius of the template ligands were selected and considered as the binding site in the constructed homology models. The residues were displayed in the homology models and compared to the residues within a 4 Å sphere radius around the template ligand in the crystal structures of the templates. The conservation of these residues was relatively high (Table 2 and Figure 10). The constructed homology models had the same number of residues (28) in close proximity (<4 Å) to the ligand in the pocket. The majority of the residues within 4 Å sphere radius from the agonists in the binding pockets were located in helix 3, helix 5 or in loop domains. Two residues were located in the β-sheets (Ile467 and Val468) (Table 2 and Figure 11).

Table 2 –The table shows the residues within a 4 Å sphere radius of the ligand in homology model I and II of the ecdysone receptor in *Daphnia magna*. The residues marked in red indicate non-conserved residues in the models compared to its respective template. The agonists co-crystallised with the templates were used for selection of the residues within 4 Å sphere radius. Segment column describes the location of the residues in the models. Model I was constructed with 1Z5X as template and the displayed residues are within 4 Å sphere radius from ponasterone A. Model II was constructed with 2R40 as a template and the residues listed are within 4 Å sphere radius from 20E. "*"Indicates that the corresponding residues in the template structure were forming hydrogen bond with the ligand.

Model I (Ponasterone A)	Model II (20E)	Segment
D384*	D384*	Loop
Q385	Q385	Loop
P386	P386	Loop
H410		H3
I411	I411	H3
T412	T412	H3
M414	M414	H3
T415	T415	H3
T418*	T418	H3
L421	L421	H3
M452	M452	H5
	M453	H5
R455*	R455*	H5
C456	A456	H5
R459	R459	Loop
I467	I467	B-sheet
V468	V468	B-sheet
F469	F469	Loop
A470*	A470*	Loop
N471	N471	Loop
Y480*	Y480*	H6
	T488	H7
L492	L492	H7
N573	N573*	H10
M576	M576	H10
C577	C577	H10
L580	L580	H10
L587		Loop
L591	L591	H11
W595	W595	H11

4.1.3 Structural evaluation of homology models

Evaluation of stereochemistry and geometry

PROCHECK, ERRAT and Vertify_3D were used to evaluate the stereochemical quality of the models. No significant deviations were reported in the evaluation. The Ramachandran plot provided by PROCHECK showed that the majority of the residues were in the favourable regions and a minority of residues were found in additional allowed regions. The ERRAT output scored the models as very reliable (97.3 and 98.7) and the compatibility of the homology models was in an acceptable range of 84.6-89.5 (Table 3).

Model evaluation by molecular docking and scoring of active ligands

The ligands co-crystallised with the template were re-docked into the crystal structures and obtained a docking score of -29.9 for 1Z5X (ponasterone A) and -47.9 for 2R40 (20E). These scoring values were used as a reference in evaluation of the scoring values for the active ligands found in ChEMBL. The 9 active ligands were docked into the constructed models using ICM docking batch method with a semi-flexible docking approach. The average docking score for model I was -23.4 and -42.6 for model II (Table 4). The docked ligands were superimposed in order to analyse and compare the binding pose, which were found to be similar for all the ligands (Figure 11).

The binding pocket was examined with respect to interactions between ligands and amino acid residues. As both ponasterone A and 20E are uncharged, ionic interactions were not present. Both models had 5 hydrogen bonds between their respective ligands and residues in the pocket, and 4 of these bonds were found in the both models (Table 2). The main chain of D384 formed a bond with one of the hydroxyl groups on ring A of the steroids (Carbon 3) and the backbone of R455 formed a hydrogen bond with the other hydroxyl group on ring A (Carbon 2). The main chain of A470 formed a hydrogen bond with the ketone on ring B (Carbon 6). The side chain (amide group) of Y480 formed a hydrogen bond with the hydroxyl group of the alkyl side chain of ring D (Carbon 20) (Figure 12).

The RMSD between the docked conformation of ponasterone A in model I and in the template (PDB id: Iz5x) was calculated to 0.28 Å, while the RMSD between the

docked conformation of 20E in model II and the template (PDB id: 2R40) was 0.14
Å.

Table 3 – Structural validation of constructed homology models. The stereochemistry and compatibility of the ecdysone receptor homology models were analysed. Model I was constructed with 1Z5X as template and model II was constructed with 2R40 as template. PROCHECK displays a summary of the residue geometry with a Ramachandran plot: most favoured/ additional allowed/generously allowed. ERRAT shows the confident limit for verification and VERTIFY_3D show the over all compatibility of the structure in percent. The volume of the ligand-binding pocket in model I was 94.6 Å^3 smaller than in model II.

Model	PROCHECK	ERRAT	Vertify_3D	Pocket (Å^3)
I	89.5/10/0.5	97.3	84.6	318.2
II	93.7/6.3/0	98.7	89.5	412.8

Table 4 – Scoring value of the known binders. Nine active ligands considered as active were docked in two constructed homology models of the ecdysone receptor ligand binding domain (LBD) in *D. magna*. The more negative scoring value, the better the ligand fit into the LBD. VLS score show negative values. VLS = Virtual ligand screening

Ligand (name or Id)	VLS score Model I	VLS score Model II
Ponasterone A	-29.9	-47.4
20-hydroksyecdysone	-29.9	-47.9
Makisterone	-26.8	-53.7
Inkosterone	-18.7	-36.1
Ecdysone	-17.3	-32.7
Cyasterone	-24.9	-47.9
CHEMBL559048	-26.4	-48.2
CHEMBL564892	-20.8	-37.6
CHEMBL559941	-15.5	-31.5
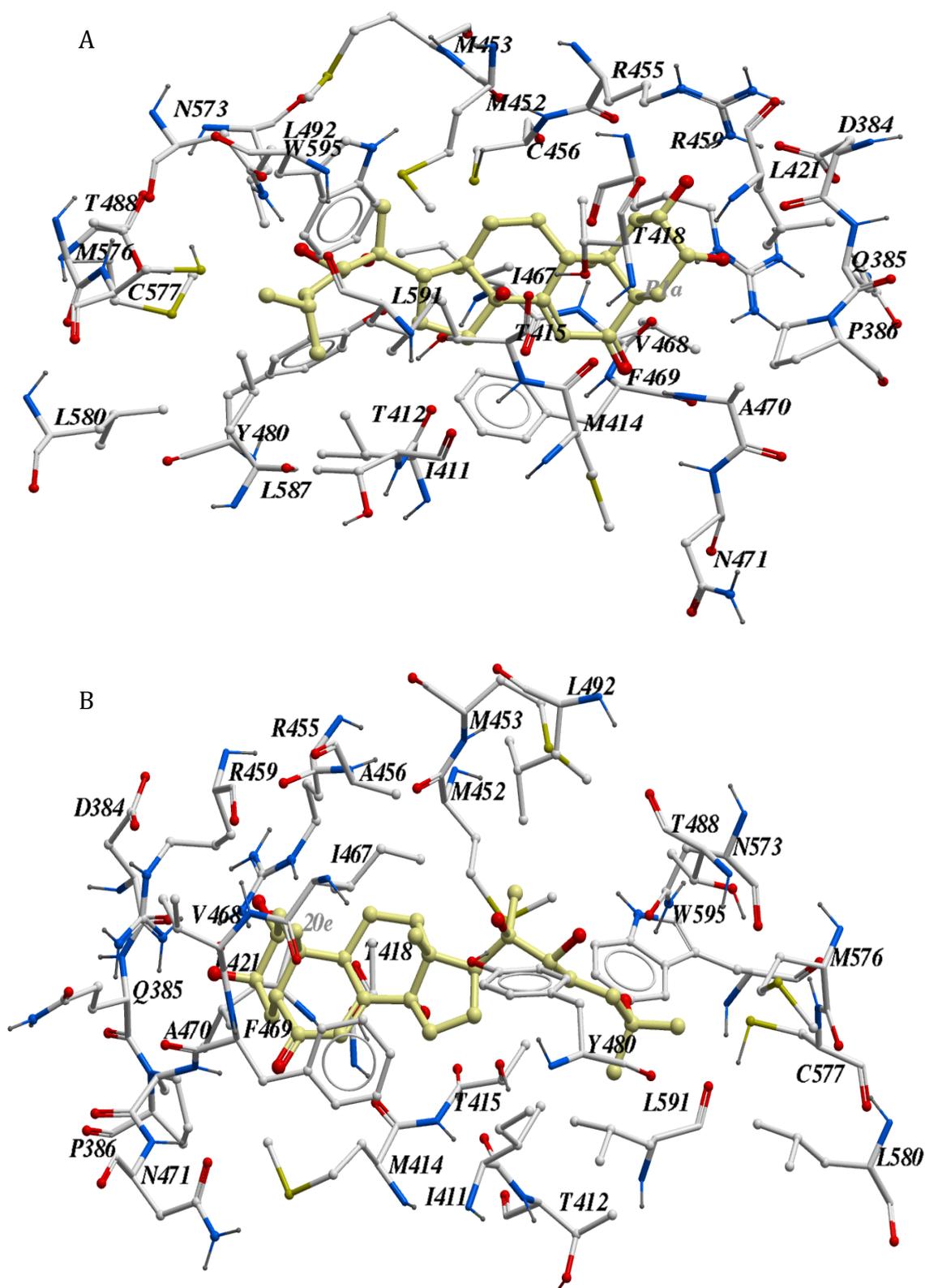
Figure 10 – The residues in the ligand binding pocket 4 Å sphere radius from the ligand. There was 28 residues within 4 Å sphere radius from the ligand in model I and model II. Five hydrogen bonds can be seen in each of the models as coloured dotted lines. (A) LBD of model I with ponasterone A displayed in the pocket, which was in the crystal structure of the template. (B) LBD of model II with 20-hydroxyecdysone displayed in the pocket, which was co-crystallised with the template
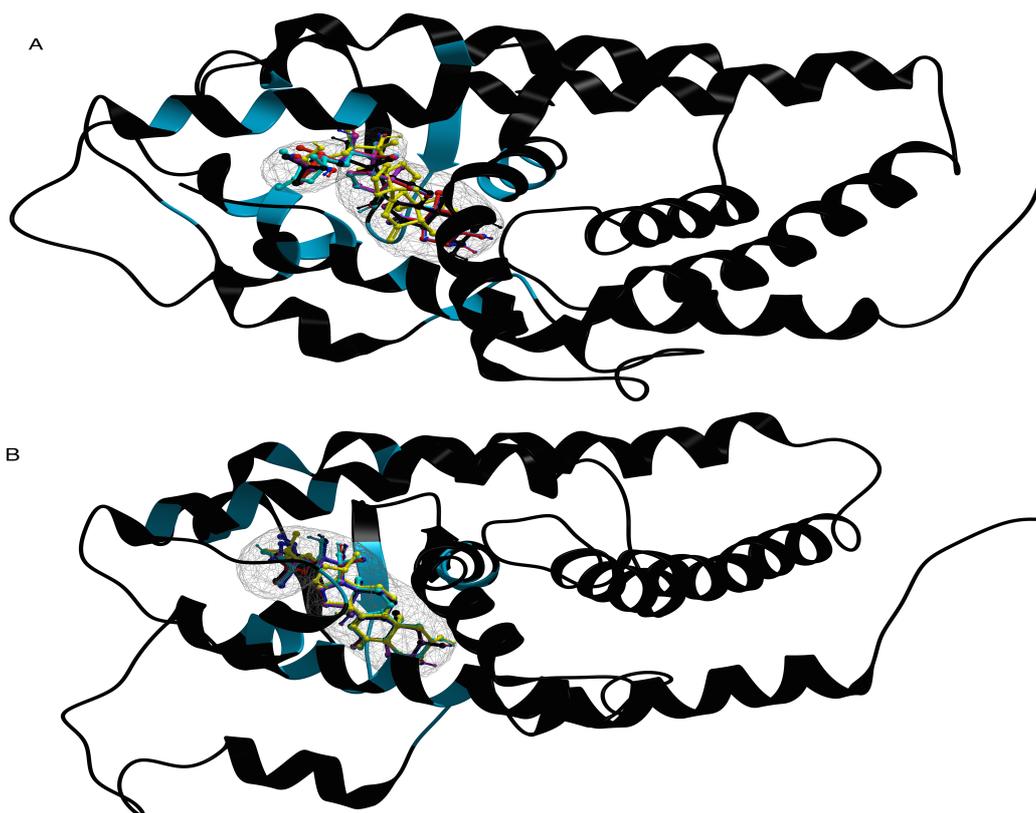
Figure 11 –Ligand binding pocket in two homology models of the ecdysone receptor from *Daphnia magna* with superimposed ligands. The ligand-binding pocket is displayed as a light grey mesh. The known active ligands are superimposed in the ligand binding domain with different colours. Residues within 4Å sphere radius from the agonist are displayed in blue. (A) Homology model I was constructed with 1Z5X as template (B) Homology model II was constructed with 2R40 as template.
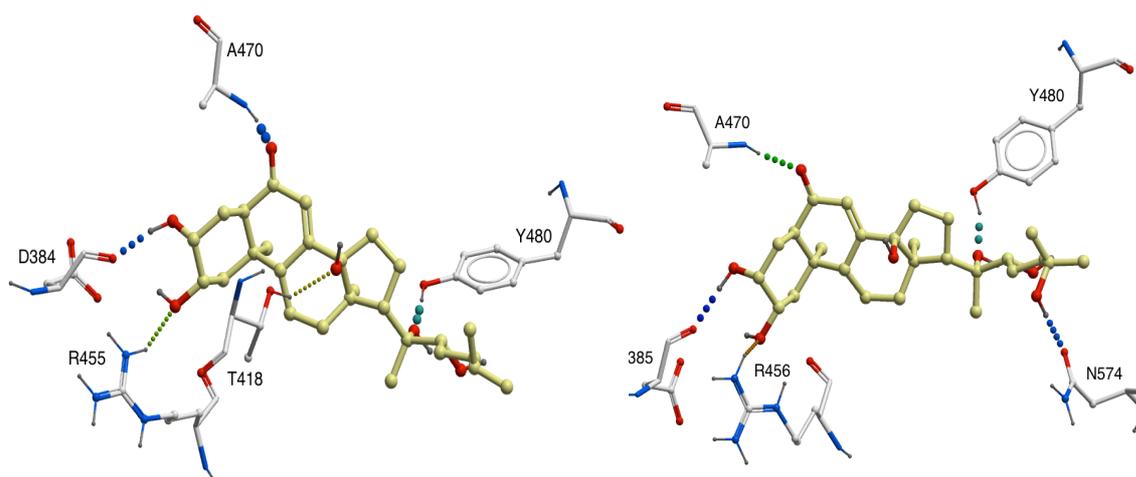
Figure 12– Hydrogen bonds between ligand and residues in the ligand binding domain. A total of 5 HBs between the ligand and residues in the ligand binding domain were observed in both homology models and 4 of these were identical in the models. Left: hydrogen bonds (HB) between ponasterone A and residues in the LBD of model I. Model I was constructed with 1Z5X as a template and the x–ray structure of 1Z5X was co-crystallised with ponasterone A. Right: HB between 20E and residues in the LBD of model II. Model II was constructed with 2R40 as a template and the x–ray structure of 2R40 was co-crystallised with 20E.

Docking and scoring of the test set consisting of binders and decoys

The Chemdiv database identified 155 decoys based on finger print similarities with each of the 9 active ligands. A test set consisting of the binders and the decoys was generated and docked into the homology models before ROC-curves were generated for the models. The models ability to differentiate between binders and decoys was measured as the area under the curve (AUC). AUC of 100 implies that the models have a very high specificity towards its targets and are able to differentiate between binders and decoys. AUC <50 indicates an insignificant test where the LBD of the models do not separate decoys from binders. The AUC of model I was 86, indicating that the model have a moderately to high ability to separate between the compounds in the test set (Figure 13). AUC of model II was 98, which signifies that the model has an excellent ability to separate binders from decoys because of high specificity towards the binders (Figure 14).
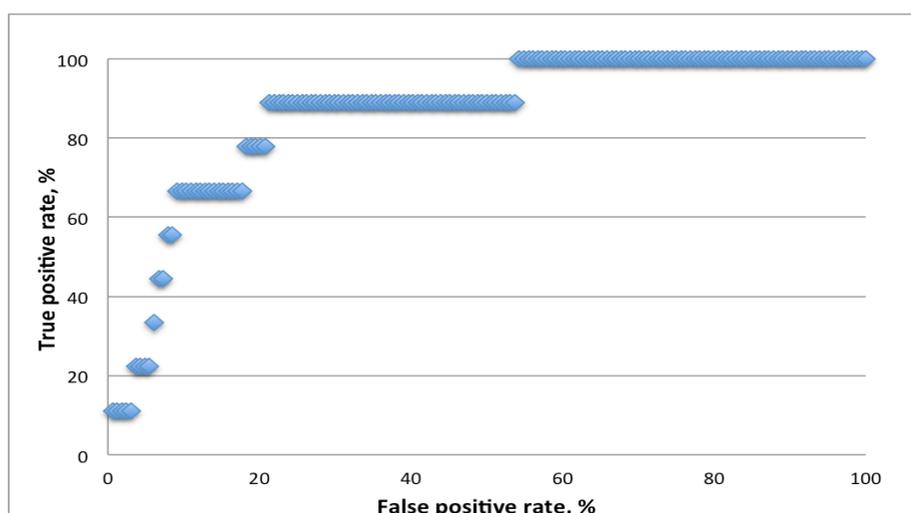
Figure 13 –ROC curve for model I. The Receiver characteristics operator (ROC) curve shows the false positive rate (%) plotted against the true positive rate (%) based on docking scores of binders and decoys in model I. The AUC of the curve was 86.
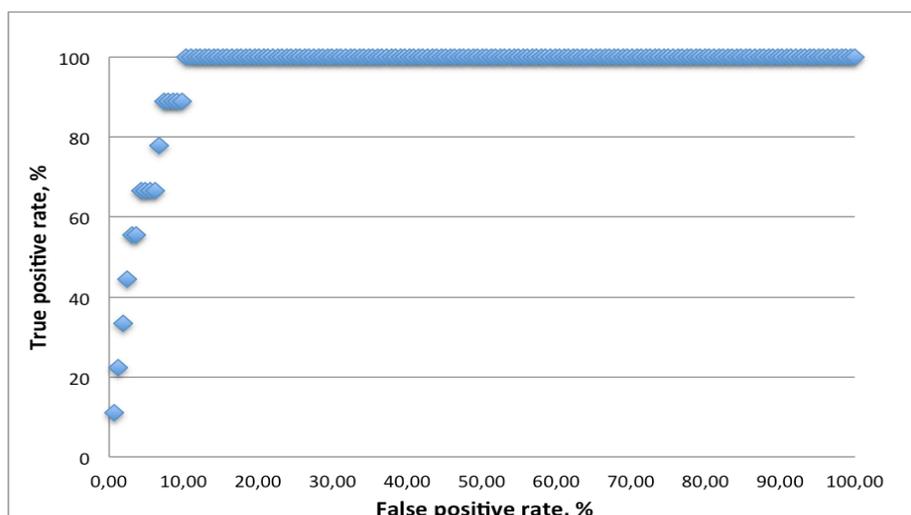
**Figure 14 –ROC curve for model II.** The Receiver characteristics operator (ROC) curve shows the false positive rate (%) vs. true positive rate (%) of model II based on docking scores of known active ligands and generated decoys. The AUC was 98.

39

4.2 Docking and scoring of pollutants

The dataset of 655 pollutants from Howard and Muir (Howard *et al.,* 2010) was docked and scored in both homology models. A threshold at -23.4 for model I and -42.9 for model II, based on the average docking score of the 9 active ligands, were used to analyse the screening of the 655 pollutants (Table 4). In model I, 13.4% of the pollutants gave better scores than threshold. None scored better than the threshold in model II. The binding pose of the compounds that scored above threshold in model I was displayed in the LBD for investigation. This showed that the compounds had a similar binding pose as the active ligands.

4.2.1 Selection of pollutants for experimental verification of 3D models

A total of 8 compounds were selected for experimental verification of the predicted 3D models of EcR. Ponasterone A and 20E were used as positive controls since these hormones got good docking scores and are known agonists of the receptor (Hill *et al.,* 2013). Bisphenol A and emamectin-benzoate were used as negative controls because of low scoring values. The 4 additionally tested compounds chosen from the pollutant dataset were (1) 1H, 1H, 2H, 2H-perfluorooctyl acrylate (TFOA), (2) triclosan, (3) diethyl phthalate and (4) endosulfan (Table 5). All the compounds were superimposed in the LBD of both the homology models and revealed a good fitting within the binding cavity (except endosulfan which was selected during the experimental testing).

The originally 4 selected compounds for experimental verification were: (1) TFOA, (2) triclosan, (3) PubChem ID 93253 (no name) and (4) phenol, 3-[2-chloro-4-(trifluoromethyl)phenoxy]-, acetate (supplementary table S2). One of these compounds was commercially unavailable (PubChem ID 93253) and phenol, 3-[2-chloro-4-(trifluoromethyl)phenoxy]-, acetate arrived too late.

Table 5 –Compounds selected for *in vitro* verification of the theoretically predicted structure of ecdysone receptor in *Daphnia magna*. The first column shows the names of compounds selected for experimental testing to support the theoretically homology models of ecdysone receptor. The second column shows the docking scores obtained for model I and model II respectively. The third column shows the structure of the selected compounds

Name	Score Model I/ Modell II	Structure
20-hydroxyecdysone	-30.4 / -48.7	
Ponasterone A	- 29.7 / -47.4	
Bisphenol A	-11.5/-22.5	
Emamectin-benzoate	-1.6 / 0.26	
1H, 1H, 2H, 2H-perfluorooctyl acrylate (TFOA)	-23.5 / -28	
Diethyl phthalate	-30.8 / -22	
Endosulfan	-7.84 / -11.25	
Triclosan	-20.5 / -35.8	
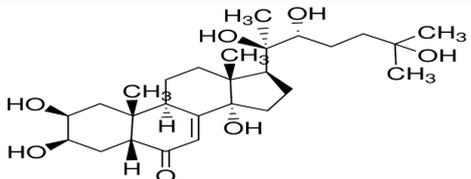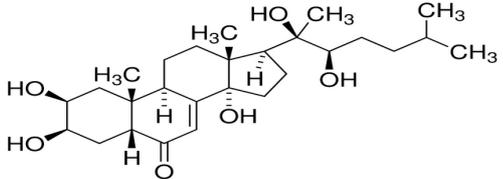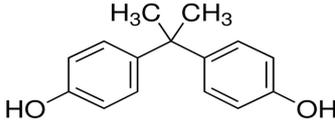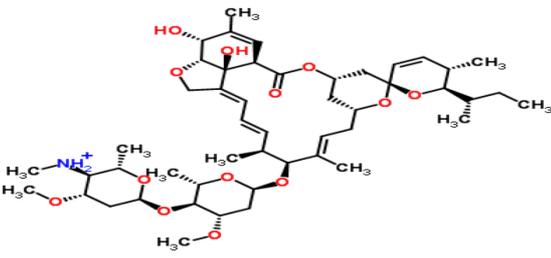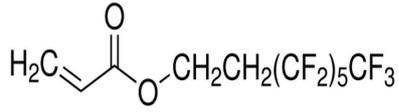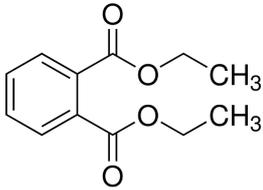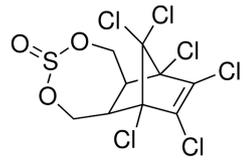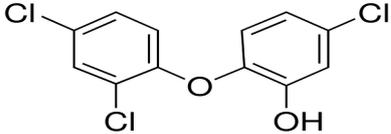
4.3 Two-hybrid assay

A two-hybrid assay was constructed as a reporter system for the agonist-dependent transcriptional activation of the EcR in *D. magna*. Bisphenol A and emamectin-benzoate was used as negative controls. Endosulfan was tested without knowledge of docking score, which later turned out to be bad (Table 5). The remaining 5 selected compounds were tested to determine if model predictions could be confirmed by *in vitro* transcriptional activation.

The results of luminescence detection revealed an activation of firefly luciferase after stimulation with the hormones, ponasterone A and 20E. This showed that the hormones are able to bind to the EcR and cause the conformational change necessary for the EcR and USP to dimerize. Ponasterone A was able to cause a concentration-dependent increase in luciferase activity at nM-concentrations, whereas 20E was only causing a partial concentration-response curve at roughly 20 times higher concentrations (Figure 15). Ponasterone A also showed a higher efficacy by inducing a 2.5-fold higher increase in the luciferase activity compared to control than 20E. The EC50 values for ponasterone A and 20E were 6.38×10^{-9} nM and 3.57×10^{-6} nM respectively. A comparison of these EC50 values illustrate that the potency of ponasterone A to activate the EcR is 559.6 fold greater than for 20E. The statistical analysis calculated a significant increase in expression at the highest concentration (10^6Mol/L) of 20E. A significant increase in expression levels of luciferase after stimulation with ponasterone A, were also found to be when the cells were stimulated with the highest concentration.

Bisphenol A and emamectin benzoate were used as negative controls and did not induce any response in the cells, as expected based on the docking studies. No response was observed after stimulation with TFOA, triclosan, diethyl phthalate or endosulfan and the results from the ANOVA and Dunnett`s test showed none significant difference in the expression levels at the different concentrations compared to control. These results indicate that the TFOA, triclosan and diethyl phthalate are not able to activate the EcR and is contradictory to what the results from the docking studies predicted (Figure 15).
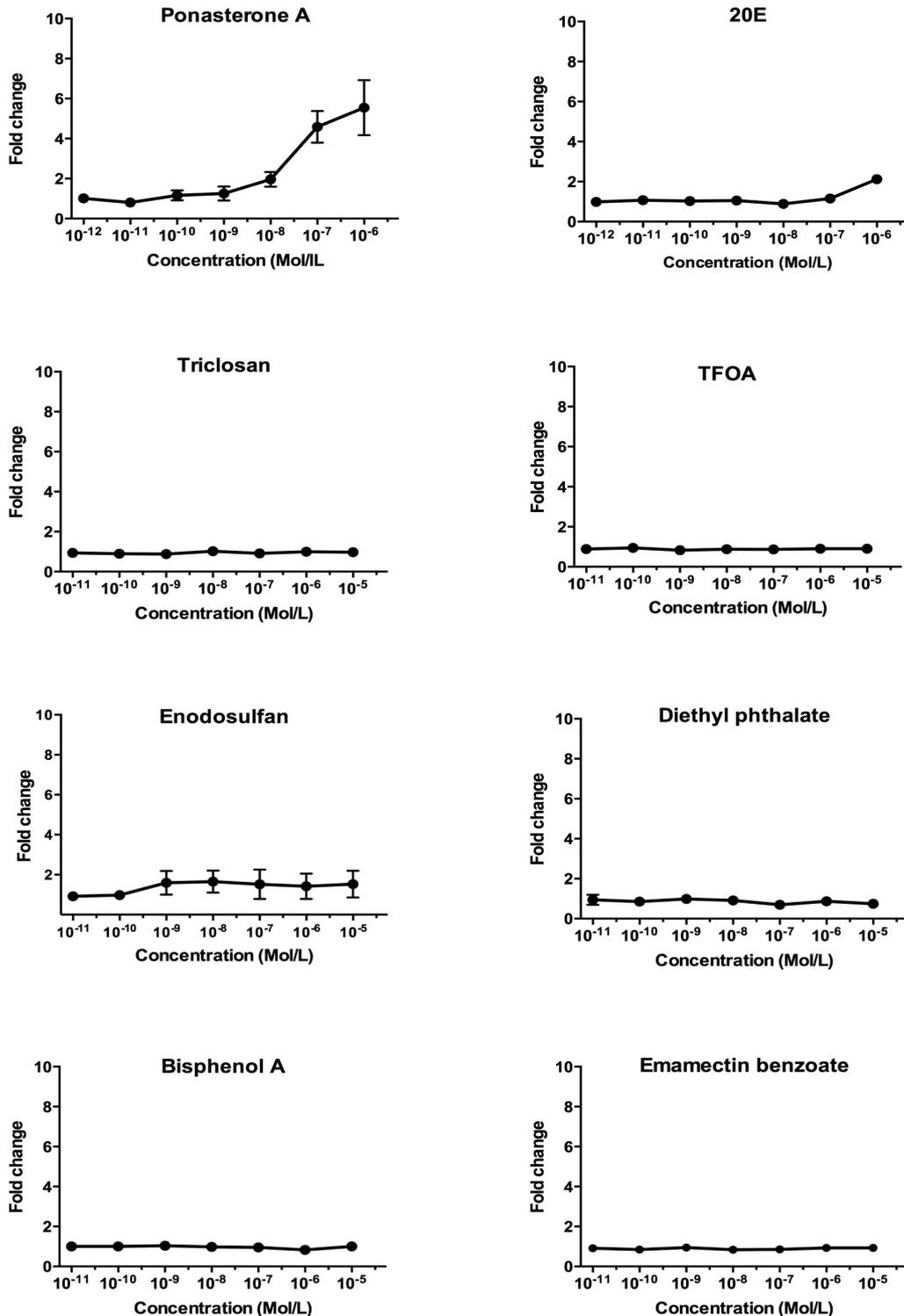
Figure 15 – Response of the ecdysone receptor reporter system after exposure to different chemicals. A total of 8 different compounds were tested for their ability to activate to the EcR and induce the agonist dependent expression of firefly luciferase. Ponasterone A and 20E were able to induce a concentration-dependent increase in the firefly luciferase. Bisphenol A and Emamectin benzoate were used as theoretically negative controls and did not activate the EcR. Triclosan, TFOA, β-endosulfan and diethyl phthalate did not induce expression of firefly luciferase. The x-axis show the concentration in Mol/L of chemical and the y-axis show the fold changes. The bars indicate SEM of median.

4.4 Cytotoxicity test

The potential cytotoxicity of the experimental tested compounds were determined by analyzing the changes in metabolic activity and membrane integrity of the CHO cells. The results obtained for triclosan, diethyl phthalate and endosulfan indicate a small decrease in both metabolism and membrane integrity at the highest concentrations (data not shown). None of the other chemicals showed any clear changes in viability at the concentrations tested. The test was only performed in one out of totally three biological replicates, and can thus only be regarded as indicative of any cytotoxicity occurring.

5 Discussion

Homology modeling was used to construct two models of the EcR from the sea flea *D. magna*. This technique is used to build a 3D model of a protein with unknown structure based on the structure of a protein homolog with resolved 3D structure. The homology models were validated based on folding and stereochemistry, by docking a set of presumed binders and decoys and by experimental testing.

Homology modeling and docking studies are theoretical approaches for predicting the 3D model of a protein and binding affinities of ligands to the protein. Docking studies were also used to study interactions between pollutants and EcR. The results from this docking were validated by *in* vitro testing of selected pollutants predicted to bind to the EcR in addition to theoretically negative controls (not predicted to bind). *In vitro* testing is necessary due to the fact that modeling and docking are theoretical approaches with some limitations and inaccuracies.

A putative drawback with the procedure is that the protein commonly is treated as a rigid structure, which does not take into account protein dynamics. Another problem is that the docking places the compound in the LBD without any considerations regarding the compounds capability of passing biological barriers, such as cell membranes, and reaching the target. Based on these drawbacks, among others, the method should only be used as a screening for possible binders, which then needs to be tested experimentally.

5.1 Alignment and structural analysis of the homology models

Visual investigation of the multiple sequence alignment between EcR LBD in *D. magna* and the top 8 ranked similar sequences from PDB (with a known 3D structure) showed high structure conservation. Secondary structure elements were to a large extent intact, but helices were of variable lengths and the number of helices varied from 10 to 13. The sheets were conserved in all 8 proteins, while the loops varied in length. A paper by Hill *et al.,* (2012) presented structural features of the EcR-USP based on a general characterization of the 5 known 3D crystal structures obtained from *Bemisia tabaci*, *Tribolium castaneum* and *Heliothis virescens* (3 crystal structures from *H. virescens* co-crystallised with different steroids). The analysis from the paper showed that EcR contains approximately 12 α-helices.

Normally, a deviation in the number of helices between template and target is caused by fusion of helices in the alignment due to differences in their sequences. Both our constructed homology models consisted of 11 α-helices, which was the same as in the crystal structures of the templates. Some of the EcR LBD structures in the multiple alignments contained more than 12 helices. This may be caused by differences in the sequences leading to helical kinks and thereby a higher number of helices.

The volumes of the LBD were different for the two homology models, 318.2 Å^3 for model I and 412.8 Å^3 for model II. A paper published by Wurtz *et al.,* (2000) describing homology models of the EcR LBD in *Chironomus tentans* constructed by using different templates, found that the volume and the shape of the EcR LBD could diverge drastically between models when different templates were used. Even though the receptor is well conserved among species, this could be caused sequence-dependent variation in topography. The templates was also co-crystallised with different ligands, which may also contribute to the volume differences in the LBD.

5.2 Docking studies

Docking studies in model II with the known active ligands showed that ponasterone A and 20E gave a similar score and the inactive form of the ecdysone scored significantly lower in comparison (Table 4). In model I, ponasterone A and 20E gave the same score and the active form of ecdysone gave better score (more negative) than the inactive form (Table 4). These results follow the trend as stated in the literature (Kato *et al.,* 2007), where experimental studies have shown that ponasterone A has a higher affinity to EcR than 20E. Ecdysone is the inactive form of 20E and is therefore expected to have a lower affinity to EcR LBD compared to its active form (Gonsalves *et al.,* 2011 and Harada *et al.,* 2009) (structure ecdysone see supplementary Table 3).

The results from the docking studies with the active ligands together with the calculated AUC of the ROC curves confirm the specificity of the LBD and supports that the constructed models are a valid 3D representation of the EcR in *D. magna*. Based on the ROC curves, the AUC was calculated to be 86 for model I and 98 for model II, respectively. Model II is therefore considered to be more accurate than model I.

Docking and scoring of the dataset consisting of 655 pollutants showed that none of the compounds scored better than threshold in model II in comparison to model I where 13.4% scored better than threshold. These results together with the results obtained from docking of binders and decoys, could strengthen the argument that model II is more accurate than model I. The docking score of the pollutants predicted to be agonists and selected for experimental verification, were close to the threshold value for model I, but significantly lower for most of the compounds for model II.

The RMSD between the ligands in the template crystal structure and the same ligands docked into the corresponding models was 0.28 for model I and 0.14 for model II. An RMSD < 2 Å between an X-ray structure pose and a docking pose of the same ligand is considered as successful docking (Thomsen *et al.,* 2006). This indicates that the binding pose of the steroid molecule after docking is very similar to the binding pose of the steroid molecule in the EcR structures used as templates, which shows that the ICM docking program is reliable in the present study.

5.2.1 Interacting residues in the LBD

There were some differences in the residues within 4 Å sphere radius of the ligand in the homology models compared to the templates (Table 2). Investigation of the binding pose of the steroids in the LBD of the homology models and the template structures indicate that the steroid binding sites have an evolutionary conserved topography, since the binding poses of the steroids were similar.

The binding pose of ponasterone A was studied in model I since this model was based on a template co-crystallised with this ligand. Model II was studied with 20E in the pocket because this ligand was co-crystallised with the template used for constructing this homology model. Both models featured 5 hydrogen bonds (HBs) between EcR and their respective steroids. D384, R455, A470 and Y480 were implicated in HB interactions with the steroid molecule in both models (Table 2 and Figure 12).

A comprehensive study published by Billas *et al.,* (2009) describes the crystal structure of the EcR/USP in three different species: *Heliothis virescens* (PDB ID: 2R40*), Bemisia tabaci* (PDB ID*:* 1Z5X) and *Tribolium castaneum* (PDB ID: 2NXX). Billas *et al.,* (2009) found that ponasterone A formed the same 6 HBs with residues in

all three EcR structures (numbered according to *H. virescens:* E309, T343, T346, R384, A398, Y408). When comparing these 6 HB forming residues in the three X-ray structures with the residues contributing to HB formation in our models, 4 out of these 6 residues were identical and formed HB with the steroid in model I (T418, R455, A470 and Y480) and 3 were identical in model II (R455, A470 and Y480). The E309 HB (*H. virescens* numbering) observed in the X-ray structures was changed to a HB forming aspartic acid in the sequence of *D. magna*. Model II has an HB between N574 and the hydroxyl group at the C25 position of 20E (Figure 2), which is not present in ponasterone A (the only difference between the steroids is this hydroxyl group). In the three crystal structures described in the paper (Billas *et al.,* 2009), a HB was also formed by T346 (*H. virescens* numbering), which is replaced by a lysine in the sequence of *D. magna.* This lysine is not in sufficiently close proximity for HB formation in our modelled complexes.

Billas *et al.,* (2009) also compared amino acids within 4.5 Å sphere radius around the steroid molecule in an EcR from a lepidoptera (*H. virescens)* to corresponding amino acids in the X-ray structure from *T. castaneum* and *B. tabaci.* They showed that Val384 (*H. virescens* numbering) was replaced with methionine, glycine or alanine in the X-ray complexes in other species. In addition, V395 was replaced with isoleucine, V416 was replaced by a threonine, asparagine or serine, and M342 was replaced with a valine or isoleucine in the species other then a lepidoptera. In the EcR of D. *magna*, Val384 is replaced with a threonine and V395 is replaced with an isoleucine as expected for other orders than Lepidoptera. The sequence EcR of *D.magna* shared the same conservation of methionine (M414 in model I and M415 in model II) as the Lepidoptera M342. However, in *D.magna* V416 is replaced with a cysteine. When docking studies were performed in model II, this cysteine was mutated to an alanine to avoid interferences between the steroid and this residue.

These findings indicate that the *D. magna* EcR may have some differences in conservation compared to the Lepidoptera, Hemiptera and Coleopteran investigated in the study by Billas *et al.,* (2009). It is likely that residues in the LBD with identical conservation across all the investigated orders are critical in binding of the ecdysteroids. Variations of residues in the LBD are probably contributing to a difference in affinity of EcR towards other compounds across animal orders.

5. 3 *In vitro* testing of selected pollutants

Ponasterone A, 20E, Bisphenol A and Emamectin benzoate were tested during the visit at NIBB, Japan. To verify the assay performance, ponasterone A and 20E were re-tested during the stay at NIVA. The results of the tests at NIVA were to a large extent similar to those obtained in Japan for both the Ponasterone A and 20E. This showed that we were able to successfully transfer the technique from the NIBB, Japan, to the lab at NIVA and that all results are highly comparable despite obtained from different laboratories.

Pollutants are expected to have lower affinity towards NRs than its natural ligands and the docking results obtained from model II are consistent with this theory (Strunck *et al.,* 2000). The negative controls and the positive controls were tested at NIBB, Japan, with concentrations ranging from 10^{-6}M to 10^{-12}M. The remaining 4 compounds selected from the docking studies were tested at NIVA, and the concentrations were changed to 10^{-5}M to 10^{-11}M due to the expectation of lower affinity towards the receptor. The negative controls should have been tested with concentration within this range because they were also pollutants.

The upper concentration limit was set to 10nM to avoid cytotoxicity. A cytotoxicity test was also performed with selected pollutants, to verify that the used dosages were not lethal. The test was only done in one biological replicate due to limited time and the achieved results are therefore not validated. If some of the chemicals were toxic for the cells, it would have been indicated in the assay as a decrease in the signal when the measured luminescence was normalized to unexposed cells (not treated with chemicals).

5.3.1 Physiochemical properties

Many studies have showed a connection between biological responses and physiochemical properties such as lipophilicity and volatility of compounds tested *in vitro* (Nynke *et al.,* 2012). Riedl *et al.,* (2007) investigated how physiochemical properties of different compounds affected the toxicity of the compounds towards algal. The EC_{50} values of different chemicals were calculated after performing a microplate assay and comparative bioassays in glass vessels. The results showed that the chemicals with a $K_{ow} > 3$ and log $K_{AW} > -4$ were less toxic in microtiter plates.

The article suggested two possible reasons for the observation (1) volatility of the substances into the headspace and (2) absorption of the chemicals to the walls of the plate. The cellular response and the EC_{50} value of a chemical are dependent on the concentration of freely dissolved chemical. If a chemical is bound to other plasma protein, the plastic wells or has evaporated, less is available for the cells. As a result, toxicity of compounds tends to be underestimated (Seibert *et al.,* 2002).

Schreiber *et al.,* (2007), investigated the loss of phenanthrene (log K_{OW} 4.46) and phenanthridine (log K_{OW} 3.48) after 48h exposure of zebra fish embryos cultivated in microtiter plates of polystyrene. The results reveal that concentration of phenanthrene decreased by 99% and the concentration of phenanthridine decreased by 17%.

Triclosan has a log K_{OW} of 4.2-4.8 (Chen *et al.,* 2011), TFOA has a log K_{OW} of 6.96, and endosulfan has a log K_{OW} of 3.66 (Montgomery, 2007). For comparison, ponasterone A has a log K_{OW} of 2.1 and 20E has a log K_{OW} of 0.5 (chemspider, 2014). Endosulfan was docked after the experimental testing and showed a high docking score in both homology models (Table 5) and is therefore not expected to bind to EcR in the assay. The EC_{50} values for these compounds could not be calculated, as the compounds did not show measurable responses in the ligand-dependent EcR binding assay. Since the lipophilic characteristics for these compounds are relatively high, especially for TFOA, the lack of expression may be explained by binding of the compounds to the polystyrene wells of the microtiter plate. If sufficient amounts of compound were bound to the wells, the remaining concentration of chemicals could have been too low to induce any detectable expression of the reporter gene. Bisphenol A and Emamectin benzoate have a log $K_{OW} > 3$ and it is therefor necessary to consider if some of the compounds could have attached to the wells, despite that they were not expected to induce any response (Crane *et al.,* 2009). Diethyl phthalate have a log K_{OW} of 2.38, (Montgomery 2007) and is not considered to be highly volatile (Sekizawa *et al.,* 2003). The docking studies predicted diethyl phthalate to be a good binder in model I, but the lack of equivalent results in the assay does not support this result.

5.4 Comparison of *in silico* and *in vitro* results

Molecular docking was used to predict the binding affinity of 655 pollutants. Based on these results, 4 pollutants were selected for experimental verification of the docking results. Ponasterone A and 20E were selected as positive controls in addition to two theoretical negative controls. A two-hybrid luciferase assay was used for the verification, where the expression of firefly luciferase is ligand dependent and the expression of *Renilla* luciferase is proportional to the amount of transfected cells. The docking scores of the two steroids and the negative controls (bisphenol A and emamectin-benzoate) were confirmed by the experimental results (Table 5, Figure 15). Both models also predicted endosulfan to be a weak binder, which was also confirmed by the experimental testing (Figure 15). If we compare the scoring values of the presumed binders with the tested compunds (Table 4), diethyl phthalate had very good scoring in model I and relative bad scoring in model II, while TOFA had quite good scoring in model I, but lower than the presumed binders in model II. The experimental testing indicated that none of them were agonists, and that indicates that model II is more realistic than model I. Contrary to the experimental results both models predicted that triclosan is a strong binder. In summary, the results from the experimental verification were to some extend in agreement with the docking predictions, but contradictory for some of the compounds (especially model I predictions). There are many possible reasons for this as the source of error can be in the models as well as in the experiment.

The templates were co-crystallised with agonists. This causes the models to best differentiate between agonists and non-binders. When a NR bind to a ligand, a folding of the H12 moiety closes the structure (Holo-form) (Figure 4). Binding of antagonists can induce misfolding or inhibit the folding of H12. If compounds do not fit into the pocket generated based on an agonist bound LBD, the model is not optimal for discriminating between a non-binder and an antagonist. For identification of EcR antagonists, it is better to use a template co-crystallised with an antagonist.

Studies by Hashimoto *el al.,* (2005) on nuclear receptor antagonists described how compounds that binds in the ligand binding pocket, but interfere with the folding of helix 12 should be considered as antagonists for that corresponding NR. This is called "the helix-folding inhibition hypothesis" and could potentially be used to identify EcR

antagonists. In theory, this could imply that compounds interfering with H12 when docked into a homology model based on a structure co-crystallised with an agonist should be considered as potential antagonists. It is possible for an antagonist to fit in the LBD, but the binding can interrupt binding of essential co-activators. Models obtained from homology modeling introduce a lot of uncertainties in addition to inaccurate predicting and the "helix-folding inhibition hypothesis" would therefore be difficult to use for our purpose. Discovery of antagonist would be a lot simpler if any crystal structures co-crystallised with antagonists existed.

A docking study with the antagonist cucurbitacin B was performed in both the crystal structures of the templates and in the homology models (Duportets *et al.,* 2013). The purpose of this study was to see if an antagonist could fit into the binding pocket generated from a template completed with an agonist. Cucurbitacin B gave a high score in both homology models and both crystal structures (supplementary table 3) indicating that it binds to another conformation of EcR than the conformations of model I and II. These results could support the argument that the 3D structures of the homology models reflect the 3D structure of the EcR found in certain invertebrates in nature even though no observable steric hindrance was present. The study also shows that an antagonist bound template should be used in the search for receptor antagonists.

Kato *et al.,* (2007) demonstrated a dose-dependent response of the pesticide tebufenozide in an EcR/USP reporter system. The max response was approximately 1,5 times lower than the measured response for ponasterone A in the same assay. This result indicated that tebufenozide can bind to the EcR/USP complex, but exhibit a lower efficacy towards the receptor compared to ponasterone A.

On the background of the some conflicting theoretical and experimental results, tebufenozide was docked in the homology models to see if this could help to verify the constructed models. The result showed a good score in model I (-20.86), based on the threshold value of -23.4. Tebufenozide got a high score at -20.87 (the more negative the better) in model II compared to the threshold at -42.9 (supplementary table 3). To compare these scores, tebufenozide was docked into the crystal structures of the templates. The trends in the scoring values were the same, but tebufenozide had

an even worse score (-20) when docked into the template of model II (2R40). These results are quite interesting taken in consideration that model II is based on an EcR template of a lepidoptera. The EcR sequence of lepidoptera is strictly conserved, but since some of the residues in the pocket are replaced with other residues in other arthropods it is reasonable to assume that it would affect the binding affinity.

The study published by Billas *et al.,* (2009) revealed that the binding affinity of different ligands varied across taxonomic orders, but the affinity of ecdysteroids towards EcR did not diverge much. The paper further states that this is likely caused by the requirement of ecdysteroids to bind effectively to the receptor. The differences in the affinity of ecdysteroids compared to other binders are presumably caused by variation in the sequences other than in residues critical for binding of steroids. This introduces another uncertainty in the homology modeling and docking approach compared to the experimental approach, when only small changes in the position of residues could have a large influence on the binding affinity of compounds towards the EcR.

Another possible source for the some deviating results could be that the docking approach only places the compounds in the pocket of the EcR without considerations to the environment. Before a compound can bind the EcR, it has to be transported across cell membranes. The compounds can possess unfavourable characteristics like size, charge and solubility, prohibiting it from membrane passage.

The predicted affinities of compounds towards the EcR attained by docking studies cannot be used as a direct guide of expected response in the agonist assay. The scoring values obtained from docking are a prediction of the binding affinity, which is correlated to the free energy of binding (ΔG) and the binding constant K_i. The binding constant describes the strength of an interaction between a protein and a ligand and can be found experimentally. The purpose of the assay was to measure expression of firefly luciferase as an endpoint for interaction. The assay is not able to differentiate or evaluate the strength of binding between a receptor and a ligand.

The knowledge on how residues are differently conserved among species can be used to make a "consensus" receptor, which can easily be used to represent more species

than *D.magna.* This consensus can then be used for high throughput screening of pollutants to find potential binders and to link these interactions to the network of adverse outcome pathways (AOP) upon exposure. An AOP describes organism's response to a toxicant by outlining the series of molecular/biological events from chemical properties and receptor-ligand interactions to production of an adverse outcome. Knowledge concerning AOP can be used in predictive risk assessment. The homology models can contribute to this approach by predicting putative EDs and determine the binding affinity of chemicals.

This project is a cornerstone on the road to discover the true 3D structure of EcR in *D.magna.* The results obtained in this project provide useful insight into how theoretical studies and experimental studies must be combined in order to achieve successful and reliable results. The experimental studies done in this project did not include sufficient number of potential EcR ligands to fully verify the predicted 3D structure of the receptor. Based on the knowledge this thesis provides it is necessary to test more compounds from the docking study experimentally in order to verify the structure. The selection of pollutants for experimental verification was not optimal because many of the originally selected compounds were unavailable for purchase.

For future studies, a more accurate approach for selection of compounds for experimentally verification would be to select compounds based on physical and chemical similarities to the ligands. Inappropriate compounds can be removed from the test set based on different filtering applications such as Lipinski´s rule of five or Absorption-distribution-metabolism-excretion-toxicity (ADMET) (Lipinski 2004 and Van de Waterbeemd *et al.,* 2003). This was not done in this project due to the limited time. It would also be wise to test some of the compounds for antagonist activity by using e.g. concentration equivalent to EC_{50} of a hormone with increasing concentration of a pollutant.

6 Conclusion

A homology modeling approach was used to build 3D models of the EcR receptor in *D. magna*. Two models were constructed based on different templates. Structural analysis and docking studies with calculation of AUC revealed that both predicted models were highly accurate, model II being more specific than model I. A comparison of the LBD in the homology models and findings in the literature suggested that residues expected to be critical for binding of ligands were present in both homology models. Docking studies were also used for high throughput screening of pollutants to identify potential binders. Some of the pollutants with predicted affinity were tested experimentally by a two-hybrid assay in order to verify the predictions and thereby support the 3D the homology models. The results achieved experimentally for triclosan, diethyl phthalate and TFOA were conflicting with the results obtained by docking studies, especially for model I, possibly due to chemical properties of the tested compounds. The results of ponasterone A and 20E suggested that the technique is applicable for predicting protein interactions and to some extent supported the predicted 3D models. A future experimental verification of the models can benefit from the approach suggested by this study, by selecting chemicals with physiochemical properties more similar to the identified binders.

References

Abagyan, R., Totrov, M., Kuznetsov, D. (2004) ” ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation”. Journal of Computational Chemistry, 15 (5): 488-506. Doi: 10.1002/jcc.540150503.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman,D.J. (1990) Basic local alignment search tool, Journal of Molecular Biology, 215(3): 403-410.

An, J., Tostrov, M., Abagyan, R., (2005) ”Pocketome via comprehensive identification and classification og ligand binding envelopes”, Mol Cell Proteomics 4:752-761.

Aranda, A. Pascual, A., (2001) “Nuclear Hormone Receptors and Gene Expression”, *Physiol Rev*, 81: 1269-1304.

Ashby, J., Houthoff, E., Kennedy, S.J., Stevens, J., Bars, R., Jekat, F.W., Campbell, P., Van Miller, J., Carpanini, F.M., Randall, G.L., (1997), ”The challenge posed by endocrine-disrupting chemicals”, Environ. Health Perspect.105: 164-169.

Berman, H., Henrick, K., Nakamura, H., Markley, J.L. (2007) ”The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data” Nucleic Acids Res., 35(Database issue); D301–D303. Doi: 10.1093/nar/gkl971.

Billas, I. M. L., Browning, C., Lawrence, M. C., Graham, L. D., Moras, D., Hill, R. J., (2009), ”The Structure and Function of the Ecdysone Receptor”, Ecdysone: Structures and Functions, Smagghe, Guy (Ed.), pp 335-360.

Bohacek, R. S., McMartin, M., Guida, W.C., (1996), ””The Art and Practice of Structure-Based: Drug Design: A Molecular Modeling Perspective”, Medicinal Research Reviews, Vol. 16 (1): 3-50.

Bopp, S. K., Lettieri, T., (2008), “Comparison of four different colorimetric and fluorometric cytotoxicity assays in a zebrafish liver cell line”, *BMC Pharmacology*, 8:8. Doi:10.1186/1471-2210-8-8.

Bowie, J.U., Lüthy, R., Eisenberg, D., ”A method to identify protein sequences that fold into a known three-dimensional structure” (1991), Science, Jul 12;253(5016):164-70.

Breivik, K., Wania, F., (2002), Evaluating a model of the historical behavior of two hexachlorocyclohexanes in the Baltic Sea environment. Environmental Science and Technology 36 (5): 1014–1023.

Buck, C. R., Franklin, J., Berger, U., Conder, M. J., Cousins, I. T., de Voogt, P., Jensen, A. A., Kannan, K., Mabury, S. S., van Leeuwen, S. P., (2011), “Perfluoroalkyl and Polyfluoroalkyl Substances in the Environment: Terminology, Classification, and Origins”, Integr. Environ. Assess. Manag., 7(4): 513–54.1

Bursulaya, B.D., Totrov, M., Abagyan, R., Brooks III, C.L., (2003), "Comparative study of several algorithms for flexible ligand docking", Journal of Computer-Aided Molecular Design 17: 755–763.

Centeno, B.N., Planas-Inglesias, J., Baldomero, O., (2005), " Comparative modelling of protein structure and its impact on microbial cell factories", Microbial Cell Factories, **4**:20. Doi:10.1186/1475-2859-4-20.

Chemspider:http://www.chemspider.com/Chemical-Structure.4573597.html?rid=08402a6f-6782-4594-ba85-8fa7605d5f6a.

Chen, X., Nielsen, L. J., Furgal, K., Liu, Y., Lolas, I., B., Beser, K., (2011), "Biodegradation of triclosan and formation of methyl-triclosan in activated sludge under aerobic conditions", Chemosphere 84: 452–456.

Clavton, EM., Todd, M., Dowd, JB. Aiello, AE., (2011), "The impact of bisphenol A and triclosan on immune parameters in the U.S. population, NHANES 2003-2006", Environ. Health Perspect., 119(3):390-6. Doi: 10.1289/ehp.1002883.

Colovos, C., Yeates, T.O., " Verification of protein structures: patterns of nonbonded atomic interactions" (1993), Protein Sci. 2(9): 1511-9.

Crane, M., Boxall, A. B. A., Barret, K., (2009), "Veterinary Medicines in the Environment", SETAC, Pensacola, Florida

Darnerud, P. O., Eriksen, G. S., Jóhannesson, T., Larsen, P. B., Viluksela, M., (2001), "Polybrominated Diphenyl Ethers: Occurrence, Dietary Exposure, and Toxicology", Environ Health Perspect., 109(Suppl 1): 49–68

Duportets, L., Maria, A., Vitecek, S., Gadenne, ., Debernard, S., (2013), "Steroid hormone signalling is involved in the age-dependent behavioural response to sex pheromone in the adult male moth *Argotis ipsilonI*", general and comparative Endocrinology 186:58-66

Fields, S., Sternglanz, R., (1994), "The two-hybrid system: and assay for protein-protein interactions", Trends Genet, 10(8): 286-92.

Gabrielsen, M., Kurczab, R., Ravna, A.W., Kufareva, I., Abagyan, R., Chilmonczyk, Z., Bojarski, A.J., Sylte, I., " Molecular mechanism of serotonin transporter inhibition elucidated by a new flexible docking protocol" (2012), European Journal of Med. Chem., 47; 24-37.

Gabrielsen, M., Ravna, A.W., Kristiansen, K., Sylte, I., (2011), "Substrate binding and translocation of the serotonin transporter studies by docking and molecular dynamics simulations", J Mol Model. Doi: 10.1007/s00894-011-1133-1.

Gaulton, A., Bellis, L.J., Bento, P.A., Chambers, J., Davis, M., Hersay, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., Overington, J.P. (2012) "ChEMBL: a large-scale bioactivity database for drug discovery". Nucleic Acids Res., 40(D1): D1100–D1107. Doi: 10.1093/nar/gkr777.

Germain, P., Staels, B., Dacquet, C., Spedding, M., Laudet, V., (2006), "Overview of nomenclature of nuclear Receptors", Pharmacol. Rev., 58(4):685-704.

Gonsalves, S.E., Neal, J.S., Kehoe, A.S., Westwood, J.T., (2011), "Genome-wide examination of the transcriptional response to ecdysteroids 20-hydroxyecdysone and Ponasterone A in *Drosophila melanogaster*", BMC genomics,, 12:475.

Hansen, K.M., Christensen, J.H., Brandt, J., Frohn, L.M., Geels, C., (2004) "Modelling atmospheric transport of α-hexachlorocyclohexane in the Northern Hemisphere with a 3-D dynamical model: DEHM-POP", Atmos. Chem. Phys., 4:1125–1137.

Harada, T., Nakagawa, Y., Akamatsu, M., Miyagawa, H., (2009), "Evaluation of hydrogen bonds of ecdysteroids in the ligand-receptor interactions using a protein modeling system", Bioorg. Med. Chem., 17: 5868-5873.

Hashimoto, Y., Miyachi, H., (2005), "Nuclear receptor antagonists designed based on the helix-folding inhibition hypothesis", Bioorganic & Medical Chemistry, 13: 5080-5093.

Hawker, D. W., Connell, D. W., (1998), " Octanol-Water Partition Coefficients of Polychlorinated Biphenyl Congeners", Environ. Sci. Technol., 22, 382-387

Hill, R.J., Billas, M.L.I., Bonnetin, F., Graham, L.D., Lawrence, M.C., (2013), "Ecdysone receptors: From the Ashburner Model to Structural Biology", Annu. Rev. Entomol., 58:251-271. Doi: 1 0.1146/annurev-ento-120811-153610.

Howard, P.H., Muir, D.C.G., (2010), "Identifying New Persistent and Bioaccumulative Organics Among Chemicals in Commerce", Environ. Sci. Technol., 44(7):2277-2285.

Huang, S.Y., Grinter, S.Z., Zou, X., (2010), "Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions", Phys. Chem. Chem. Phys. 12:12899-12908. Doi: 10.1039/c0cp00151a.

Höltje, H.D., Sippl, W., Rognan, D., Folkers, G., (2008), "Molecular modeling. Basic principles and applications", 3th ed. WILEY-VCH Verlag GmbH & Co KGaA.

Kato, Y., Kobayashi, K., Oda, S., Tatarazako, N., Watanabe, H., Iguchi, T., (2007), "Cloning and characterization of the ecdysone receptor and ultraspiracle protein from the water flea *Daphnia magna*", J. Endocrinology, 193:183-194. Doi: 10.1677/JOE-06-0228.

Krieger, E., Nabuurs, S.B., Vriend, G., (2003), "Homology modeling", Methods Biochem. Anal., 44:509–523.

Kroemer, R. T., (2007), "Structure-Based Drug Design: Docking and Scoring", Current Protein and Peptide Science, 8: 312-328.

Kolinski, A., Skolnick, J., (1994), "Monte Carlo simulations of protein folding. I. Lattice Model and Interaction Scheme", Proteins: Structure, Function, and Genetics 18:338-352.

Leach, A.R., Schichet, B.K., Peishoff, C.E., (2006), "Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gabs", J. Med. Chem., 49(20).

LeBlanc, G.A., (2007), "Crustacean endocrine toxicology: a review." Ecotoxicol. 16(1): 61-81.

Lidin, I., Wuxiuer, Y., Kufareva, I., Abagyan, R., Moens, U., Sylte, I., (2013), "Homology modeling and ligand docking of Mitogen-activated protein kinase-activated protein kinase 5 (MK5)", Theoretical Biology and Medical Modelling, 10:56.

Lipinski, C. A., (2004), "Lead- and drug-like compounds: the rule-of-five revolution", Drug Discovery Today: Technologies, 1(4): 337–341.

Loewenstein, Y., Portugaly, E., Fromer, M., Linaial, M., (2008) " Efficient algorithms for accurate hierarchical clustering of huge datasets: tackling the entire protein space" oxford journals, Bioinformatics, 24 (13): i41-i49. Doi: 10.1093/bioinformatics/btn174.

Marty, MS., Carney, W.E., Rowlands, JC., "Endocrine disruption: Historical Perspectives and Its Impact on the Future of Toxicology Testing" (2011), Toxicological Sciences, 120(S1): S93-S108. Doi:10.1093/toxsci/kfq329.

Miller, J., Stagliar, I., (2004), "Using the two-hybrid system to identify interacting proteins", Methods Mol. Biol. 261:247-62.

Meylan, W. M., Howard, P. H., (2005), "Estimating octanol–air partition coefficients with octanol–water partition coefficients and Henrys law constants", Chemosphere (61): 640–644.

Montgomery, J. F., (2007), "Ground water chemicals", 4ed, CRC Press, Tylor & Francis group, Florida, US.

Mu, X., Leblanc, G.A., (2004), "Cross Communication Between Signaling Pathways: Juvenoid Hormones Modulate Ecdysteroid Activity in a Crustacean", J. Experimental zoology, 301:793-801.

Nabuurs, S.B., Wagener, M., de Vlieg, J., (2007), "A flexible Approach to Induced Fit Docking", J. Med. Chem., 50(26):6507-18.

Nynke, I. K., Rico-Rico, A., Blaauboer, J. B., Hermens, J. L. M., (2012) "Quantifying Processes Determining the Free Concentration of Phenanthrene in Basal Cytotoxicity Assays", Chem. Res. Toxicol., 25, 436−445.

Peters, Z.J., Nykamp, J.A., Passaperuma, K., Carlson, J.C., DeWitte-Orr, S.J., Greenberg, B.M., Bols, N.C., (2007), "Effect of copper on the cytotoxicity of

phenanthrene and 9,10-phenanthrenequinone to the human placental cell line JEG-3", Reprof. Toxicol., 23(4):513-20.

Rajapakse, N., Silva, E., Kortenkamp, A., (2002), "Combining Xenoestrogens at Levels below Individual No-Observed-Effect Concentrations Dramatically Enhances Steroid Hormone Action", Environmental health perspect, 110: 917-21.

Riedl, J., Altenburger, R., 2007. Physicochemical substance properties as indicators for unreliable exposure in microplate-based bioassays. Chemosphere 67, 2210-2220.

Robinson-Rechavi, M., Garcia, E.H., Laudet, V., (2003), "The nuclear receptor superfamily", Journal of Cell Science 116, 585-586. Doi: 10.1242/jcs.00247.

Safe, S., (1990);" Polychlorinated Biphenyls (PCBs), Dibenzo-p-Dioxins (PCDDs), Dibenzofurans (PCDFs), and Related Compounds: Environmental and Mechanistic Considerations Which Support the Development of Toxic Equivalency Factors (TEFs)", Critical Reviews in Toxicology, 21(1): 51-88.

Schapira, M., Abagyan, R., Totrov, M., "Nuclear hormone receptor targeted Virtual Screening" (2003), Journal of Med. Chem. 46:3045-3059

Schreiber, R., Altenburger, R., Paschke, A., Küster, E., 2008. How to deal with lipophilic and volatile organic substances in microtiter plate assays. Environmental Toxicology and Chemistry 27, 1676-1682.

Seibert, H., Mörchel, S., Gülden, M., (2002), "Factors influencing nominal effective concentrations of chemical compounds in vitro: medium protein concentration", Toxicology in Vitro 16: 289–297.

Sekizawa, J., Dobson, S., Touch III, R., J., (2003), Concise International Chemical Assessment Document 52: DIETHYL PHTHALATE, World Health Organization, viewed 13 may, http://www.inchem.org/documents/cicads/cicads/cicad52.htm#5.1.

Sherman, W., Day, T., Jacobson, M.P., Friesner, R.A., Farid, R., (2006), "Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects", J. Med. Chem., 49:534-553.

Stollewerk, A., (2010), "The water flea *Daphnia* - a 'new' model system for ecology and evolution? ", Journal of Biology 9 (2):2. Doi: 10.1186/jbiol212.

Strunck, E., Stemmann, N., Hopert, A., Wünsche, W., Frank, K., Vollmer, G., (2000), "Relative binding affinity does not predict biological response to xenoestrogens in rat endometrial adenocarcinoma cells", J Steroid Biochem Mol Biol. 74(3):73-81.

Sylte, I., Ravna, R.W., (2012) "Homology modeling of transporter proteins (carriers and ion channels", Methods Mol.Biol. 857:281-299. Doi: 10.1007/978-1-61779-588-6_12.

Tancréde, M. V., Yanagisawa, Y., (1990), " An analytical method to determine Henry's law constant for selected volatile organic compounds at concentrations and temperatures corresponding to tap water use", J. Air Waste Manage Assoc., 40(12): 1658-63.

Tatarazako, N., Oda, S., Watanabe, H., Morita, M., Iguchi, T., (2003), "Juvenile hormone agonists affect the occurrence of male *Daphnia*", Chemosphere 53:827-833.

Thomsen, R., Christensen, M.H., (2006), "MolDock: A New Technique for High-Accuracy Molecular Docking", J. Med. Chem., 49:3315-3321.

Tyler, CR., Jobling, S., Sumpter, JP. (1998), "Endocrine disruption in wildlife: a critical review of the evidence", Crit. Rev. Toxicology, 28: 319-61.

Van de Waterbeemd, H., Gifford, E., (2003) "ADMET in silico modelling: towards prediction paradise?" Nature Reviews Drug Discovery 2, 192-204. Doi:10.1038/nrd1032.

Verreault, J., Shahmiri, S., Gabrielsen, G. W., Letcher, R. J., (2007), "Organohalogen and metabolically-derived contaminants and associations with whole body constituents in Norway Arctic glucous gulls", Environment International 33: 823-830.

Wang, MS., Chen, SJ., Lai,. YC., Huang, KL., C, GP., (2010), "Characterization of Persistent Organic Pollutants in Ash Collected from Different Facilities of a Municipal Solid Waste Incinerator", Aerosol and Air Quality Research, 10: 391–402.

Whisenant, T. C., Ho, D. T., Benz, R. W., Rogers, J. S., Kaake, R. M., Gordon, E. A., Huang, L., Baldi, P., Bardwell, L., (2010) "Computational prediction and experimental verification of new MAP kinase docking sites and substrates including Gli transcription factors", PLoS Comput Biol 6(8): e1000908. Doi:10.1371/journal.pcbi.1000908

Wu, B., Zhang, Y., Kong, J., Zhang, X., Cheng, S., (2009), "In silico predication of nuclear hormone receptors for organic pollutants by homology modeling and molecular docking", Toxicology Letters, 191(1):69-73

Wurtz, JM., Guillot, B., Moras, D., Tietjen, K., Schindler, M.,(2000) "A new model for 20-hydroxyecdysone and dibenzoylhydrazine binding: A homology modeling and docking approach" Protein Sci., 9: 1073-84.

Zhu, J., Chen, L., Sun, G., Raikhel, A.S., (2006), "The competence factor βFtz-F1 potentiates ecdysone receptor activity via recruiting a p160/SRC coactivator", Mol. Cell. Biol. Doi:10.1128/MCB.01318-06

Supplementary material

Supplementary Table S1 –Identified active ligands of EcR from ChEMBL. The table show the 9 active ligands with IC50 <2600 nM.

Supplementary figure S1 – Multiple sequence alignment. A multiple sequence alignment between the EcR LBD sequence in *D.magna* and the top 8 ranked sequences based on similarity found in the PDB. The alignment show conserved secondary structure elements

Supplementary Figure S2 – Alignment of EcR LBD. Sequence alignment between the EcR LBD sequence of *D.magna* (Query) and *D.melanogaster* (Sbjct). The red boxes show the residues in the LBD found in model II. The blue boxes show the corresponding residues in the sequence of *D.melanogaster*. The arrows indicate residues that are different in the LBD of the two sequences

Supplementary Table S2- Structure of the originally chosen compounds for experimental verification. The table show the pollutants with the best score in both in both the models. Since none of the pollutants scored above threshold in model II, the ones with highest score in this model were used if the corresponding score in model I was good.

**Supplementary Table S3- Structure of the agonist tebufenozide and the antagonist cucurbitacine B.**

Supplementary Table S1 –Identified active ligands of EcR from ChEMBL. The
table show the 9 active ligands with IC50 <2600 nM.

Ligand (name or id)	Structure
Ponasterone A	
20E	
CHEMBL559048	
Cyasterone	
Inkosterone	
Makisterone	
CHEMBL564892	
CHEMBL559941	
Ecdysone	

```
         .........##.+##.#...#......  ..................#..#..#....#..#V.#AK.#P.F..L...DQ#.LL+....E####....R..  .....-.#.##..#.#.+..#..AG#...#...#..#...M..#.#D..E#.#L.AI.##..D...#.....VE.#
3ipu_a      204 PQLSPEQLGMIEKLVAAQQQCNRRSFS-DRLRVTPWPPHSREARQQRFAHFTELAIVSVQBIVDFAKQLPGFLQLSREDQIALLKTSAIEVMLLETSRR----YNPGSESITFLKDFSYNREDFAKAGLQVEFINPIFEFSRAMNELQLNDAEFALLIAISIFSADRPNVQDQLQVERL
1p8d_a      218 VQLTAAQELMIQQLVAAQLQCNKRSFS-DQPKVTPWPLQSRDARQQRFAHFTELAIISVQBIVDFAKQVPGFLQLGREDQIALLKASTIEIMLLETARR----YNHETECITFLKDFTYSKDDFHRAGLQVEFINPIFEFSRAMRRLGLDDAEYALLIAINIFSADRPNVQEPGRVEAL
2r40_d      288 PPLTANQKSLIARLVWYQEGYEQPSEE-DLKRV-------TQT-DMPFRQITEMTILTVQLIVEFAKGLPGFAKISQSDQITLLKACSSEVMMLRVARR----YDAATDSVLFANNQAYTRDNYRKAGMA-YVIEDLLHFCRCMYSMMMDNVHYALLTAIVIFS-DRPGLEQPLLVEEI
B0L4A2_9CRUS 363 KPLSPEQEELINRLVYFQEEFDQPSEE-DLRKISTSGIHESDA-DAKFKHITEMTILTVQLTVEFSKRLPGFDTLLREDQITLLKACSSEVMMLRCARR----YDANTDSIVFANNLPYTRESYNMAGVG-DTADSLFRFGKTMSLMKVDNAEYALLTAIVIFS-ERPGLVEARKVEKI
3ixp_a      206 ------QELSIERLLEMESLVADPSEEFQFLRVGP-DSNVPPKFRAPVSSLCQIGNKQIAALVVWARDIPHFSQLEMEDQILLIKGSWNELLLFAIAWRSMEFLTSPPQLMCLMPGMTLHRNSALQAGVGQIFDRVLSELSLKMRTLRVDQAEYVALKAIILLNPDVKGLKNRQEVEVL
3fc6_a      225 ---SANEDMPVERILEAELAV-----------------------DPVTNICQAADKQLFTLVEWAKRIPHFSELPLDDQVILLRAGWNELLIASFSHRSI----AVKDGILLATGLHVHRNSAHSAGVGAIFDRVLTELVSKMRDMQMDKTELGCLRAIVLFNPDSKGLSNPAEVEAL
2acl_a      225 ---SANEDMPVERILEAELAVE----------------------DPVTNICQAADKQLFTLVEWAKRIPHFSELPLDDQVILLRAGWNELLIASFSHRSI----AVKDGILLATGLHVHRNSAHSAGVGAIFDRVLTELVSKMRDMQMDKTELGCLRAIVLFNPDSKGLSNPAEVEAL
1uhl_a      298 -------EMPVDRILEAELAVEQSP--------------------NDPVTNICQAADKQLFTLVEWAKRIPHFSSLPLDDQVILLRAGWNELLIASFSHRSI----DVRDGILLATGLHVHRNSAHSAGVGAIFDRVLTELVSKMRDMRMDKTELGCLRAIILFNPDAKGLSNPSEVEVL
2nxx_a      180 -------DMPLERIIEAEKRVECNDPL---------VALVVNENNTTVNNICQATHKQLFQLVQWAKLVPHFTSLPLTDQVQLLRAGWNELLIAAFSHRSM----QAQDAIVIATGLTVNKSTAHAVGVGNIYDRVLSELVNKMKEMKMDKTELGCLRAIILYNPDVRGIKSVQEVEML
```

```
         .##..L..Y..............F#+#L..L..LR.#.....E.#F.#+L.....#..#L.E##-...
3ipu_a      384 TYVEALHAYV---SIHHPHDRLMFPPRMLMKLVSLRTLSSVHSEQVFALRL-QDKKLPPLLSEIWDV-----
1p8d_a      398 PYVEALLSYT---RIKRPQDQLRFPPRMLMKLVSLRTLSSVHSEQVFALRL-QDKKLPPLLSEIWDVHE---
2r40_d      464 YYLNTLRVYILNQNSASPRCAVIFGKILGILTEIRTLGMQNSNMCISLKL-KNRKLPPFLEEIWDVA----
B0L4A2_9CRUS 536 IYLEALQAYVMNHRV---RPMTEFAKLLSVLTELRTLGNLNSEMCFSLKL-KNKKLPPFLAEIWDVHS---
3ixp_a      394 KMFLCLDEYC---RRSRSSEEGRFAALLLRLPALRSISLKSFEHLFFFHLVADTSIAGYIRDALRNHAPPI
3fc6_a      395 KVYASLEAYC---KHKYPEQPGRFAKLLLRLPALRSIGLKCLEHLFFFKL------DTFLMEMLEA-----
2acl_a      395 KVYASLEAYC---KHKYPEQPGRFAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA-----
1uhl_a      466 KVYASLETYC---KQKYPEQQGRFAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEA-----
2nxx_a      341 KIYGVLEEYT---RTTHPNEPGRFAKLLLRLPALRSIGLKCLEHLFFFKLIGDVPIDTFLMEMLEG-----
```


Supplementary figure S1 – Multiple sequence alignment. A multiple sequence alignment between the EcR LBD sequence in *D.magna* and the top 8 ranked sequences based on similarity found in the PDB. The alignment show conserved secondary structure elements.

Score	Expect	Method	Identities	Positives	Gaps
491 bits(1263)	9e-167	Compositional matrix adjust.	260/470(55%)	315/470(67%)	58/470(12%)

```
Query  163  L---------------GREDLSPPGSVNGGGGYSVDSFTDAKK-KKGPVPRQQEELCLV  205
            L               GR+DLSP  S+N   GYS +    DAKK KKGP PR QEELCLV
Sbjct  210  LGMQHTPRSDSVNSISSGRDDLSPSSSLN---GYSANESCDAKKSKKGPAPRVQEELCLV  266

Query  206  CGDRASGYHYNALTCEGCKGFFRRSITKNAVYQCKYGNGCDIDMYMRRKCQECRLKKCLT  265
            CGDRASGYHYNALTCEGCKGFFRRS+TK+AVY CK+G  C++DMYMRRKCQECRLKKCL
Sbjct  267  CGDRASGYHYNALTCEGCKGFFRRSVTKSAVYCCKFGRACEMDMYMRRKCQECRLKKCLA  326

Query  266  VGMRPECVVPEYQCAVKREAKKAQKEKDRPNSS-----------SLGSPDMKEPEIKAI  313
            VGMRPECVVPE QCA+KR  KKAQKEKD+  +S          S G  D  + EI  +
Sbjct  327  VGMRPECVVPENQCAMKRREKKAQKEKDKMTTSPSSQHGGNGSLASGGGQDFVKKEILDL  386

Query  314  HQQQQQQQQQMQQQQQQQQQLQQHIHVLDEKPMIVCGPANGGVSNGSPVKPLSPEQEELI  373
               +  Q                I +L ++ +  C   N       +  L+  Q  +I
Sbjct  387  MTCEPPQHAT--------------IPLLPDEILAKCQARN--------IPSLTYNQLAVI  424

Query  374  NRLVYFQEEFDQPSEEDLRKI-STSGIHESDADAKFKHITEMTILTVQLTVEFSKRLPGF  432
            +L+++Q+ ++QPSEEDLR+I S     +ES  D  F+HITE+TILTVQL VEF+K LP F
Sbjct  425  YKLIWYQDGYEQPSEEDLRRIMSQPDENESQTDVSFRHLTELTILTVQLIVEFAKGLPAF  484

Query  433  DTLLREDQITLLKACSSEVMMLRCARRYDANTDSIVFANNLPYTRESYNMAGVGDTADSL  492
             + +EDQITLLKACSSEVMMLR ARRYD ++DSI FANN  YTR+SY MAG+ D  + L
Sbjct  485  TKIPQEDQITLLKACSSEVMMLRMARRYDHSSDSIFFANNRSYTRDSYKMAGMADNIEDL  544

Query  493  FRFGKTMSLMKVDNAEYALLTAIVIFSERPGLVEARKVEKIQEIYLEALQAYVMNHRV--  550
            F + M  MKVDN EYALLTAIVIFS+RPGL +A+ VE IQ  Y++ L+ Y++N
Sbjct  545  LHFCRQMFSMKVDNVEYALLTAIVIFSDRPGLEKAQLVEAIQSYYIDTLRIYILNRHCGD  604

Query  551  -RPMTEFAKLLSVLTELRTLGNLNSEMCFSLKLKNKKLPPFLAEIWDVHS  599
             +   +AKLLS+LTELRTLGN N+EMCFSLKLKN+KLP FL EIWDVH+
Sbjct  605  SMSLVFYAKLLSILTELRTLGNQNAEMCFSLKLKNRKLPKFLEEIWDVHA  654
```

Supplementary Figure S2 – Alignment of EcR LBD. Sequence alignment between the EcR LBD sequence of *D.magna* (Query) and *D.melanogaster* (Sbjct). The red boxes show the residues in the LBD found in model II. The blue boxes show the corresponding residues in the sequence of *D.melanogaster*. The arrows indicate residues that are different in the LBD of the two sequences.

Supplementary Table S2- Structure of the originally chosen compounds for experimental verification. The table show the pollutants with the best score in both in both the models. Since none of the pollutants scored above threshold in model II, the ones with highest score in this model were used if the corresponding score in model I was good.

Chemical (name or ID)	Docking Score	Structure
Triclosan	-20.5 / -35.8	
phenol, 3-[2-chloro-4-(trifluoromethyl)phenoxy]-, acetate	-26 / -29	
PubChem ID 93253	-31 / -31.5	
TFOA	-23.5/ -28	
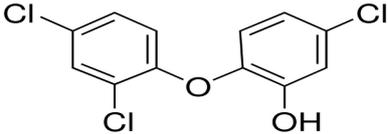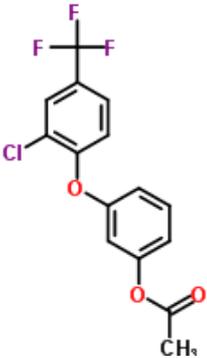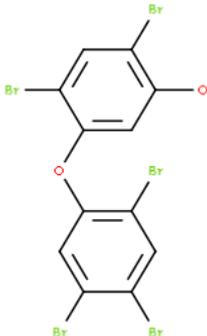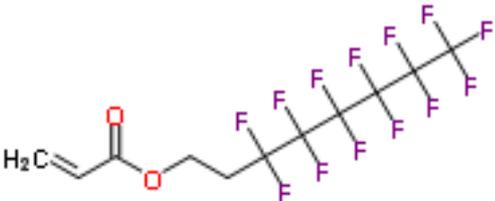
Supplementary Table S3- Structure of the agonist tebufenozide and the antagonist cucurbitacine B.

Name	Score	Structure
Cucurbitacin B	0.26/ -1.6	
Tebufenozide	20.86/ 20.87	