

# Discovery of RAR-related Orphan Receptor $\gamma$ t Inverse Agonists

Master's Thesis

University of Turku

MSc Degree Programme in Biomedical Sciences

Drug Discovery and Development

November 2021

Miika Niemeläinen

## **Supervisors**

Professor Olli Pentikäinen

PhD Sanna Niinivehmas

MedChem.fi

University of Turku

Institute of Biomedicine

*The originality of this thesis has been verified in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service*

UNIVERSITY OF TURKU  
Institute of Biomedicine, Faculty of Medicine

NIEMELÄINEN, MIIKA: Discovery of RAR-related Orphan Receptor  $\gamma$ t Inverse Agonists

Master's Thesis, 48 pages  
MSc Degree Programme in Biomedical Sciences/Drug Discovery and Development  
November 2021

---

## ABSTRACT

**BACKGROUND** Technological advances have provided modern approaches for drug discovery and development. Its application, computer-aided drug design (CADD), enables solving complex drug design challenges practically unsolvable by traditional methods. Here, CADD was applied for a drug target retinoic acid receptor-related orphan receptor gamma t (ROR $\gamma$ t). ROR $\gamma$ t is a nuclear receptor that mediates inflammation and is connected to multiple chronic inflammatory diseases. ROR $\gamma$ t inhibition proposes anti-inflammatory effects and, as such, ROR $\gamma$ t has been identified as an appealing drug target. The highest inhibition is achieved with inverse agonist -inhibitors. Due to unspecificity and complexity of ROR $\gamma$ t, no effective drugs targeting ROR $\gamma$ t have been developed so far. Accordingly, CADD approach can be justified to address this challenge.

**UNMET MEDICAL NEED** Some patients suffering from inflammatory diseases do not respond sufficiently to the current therapies. Hence, search for novel therapies can be justified, and ROR $\gamma$ t inhibition is a promising option for this purpose.

**AIM OF THE STUDY** Modern CADD approaches were executed to discover novel ROR $\gamma$ t inverse agonists.

**MATERIALS AND METHODS** Several CADD methods were utilized including ligand- and structure-based drug design.

**RESULTS** 29 novel potential ROR $\gamma$ t inverse agonists were discovered.

**CONCLUSION** CADD protocol was demonstrated suitable for discovering novel potential ROR $\gamma$ t inverse agonists.

**Keywords:** ROR $\gamma$ t, CADD, inverse agonist

## TABLE OF CONTENTS

<b>1 INTRODUCTION</b> .....	1
<b>1.1 Retinoic acid receptor-related orphan receptor gamma t (ROR<math>\gamma</math>t)</b> .....	1
<i>1.1.1 The role in inflammatory diseases</i> .....	2
<i>1.1.2 Biomolecular and pharmacological characteristics</i> .....	2
<b>1.2 ROR<math>\gamma</math>t and drug discovery</b> .....	7
<b>1.3 Computer-aided drug design (CADD)</b> .....	9
<i>1.3.1 Structure-based drug design</i> .....	10
<i>1.3.2 Ligand-based drug design</i> .....	11
<i>1.3.3 Other CADD approaches</i> .....	13
<b>1.4 ROR<math>\gamma</math>t and CADD</b> .....	14
<b>1.5 Thesis summary</b> .....	15
<b>2 RESULTS</b> .....	16
<b>2.1 Structure-based drug design</b> .....	16
<i>2.1.1 Evaluation and validation of the docking- and scoring methods</i> .....	16
<i>2.1.2 Optimization of the final Negative image-based (NIB)-models</i> .....	17
<b>2.2 Ligand-based drug design</b> .....	19
<b>2.3 Acquiring compounds with structure- and ligand-based drug design screen</b> .....	19
<b>2.4 Final screening and obtaining the final compounds</b> .....	20
<b>3 DISCUSSION</b> .....	24
<b>3.1 CADD protocol</b> .....	24
<b>3.2 Limitations of the study</b> .....	26
<b>3.3 Significance of the study</b> .....	28
<b>3.4 Conclusions</b> .....	28
<b>4 MATERIALS AND METHODS</b> .....	29
<b>4.1 Acquiring proteins and ligands and their preparation</b> .....	29
<i>4.1.1 Obtaining the binding site models</i> .....	30
<i>4.1.2 Proteins preparation</i> .....	30
<i>4.1.3 Acquiring active- and inactive compounds</i> .....	31
<i>4.1.4 Ligand preparation</i> .....	31
<b>4.2 Binding site identification</b> .....	32
<b>4.3 Structure-based drug design</b> .....	32
<i>4.3.1 NIB-model generation with panther</i> .....	32

<i>4.3.2 Evaluation and validation of the docking- and scoring methods</i> .....	32
<i>4.3.3 Docking and scoring</i> .....	33
<i>4.3.4 Optimization of the final NIB-models</i> .....	34
<b>4.4 Ligand-based drug design</b> .....	35
<b>4.5 Acquiring compounds from structure- and ligand-based drug design screen</b> .....	37
<b>4.6 Final screening</b> .....	37
<b>4.7 Obtaining final compounds</b> .....	38
<b>5 ACKNOWLEDGEMENTS</b> .....	39
<b>6 ABBREVIATIONS</b> .....	40
<b>7 REFERENCES</b> .....	41

# 1 INTRODUCTION

## 1.1 Retinoic acid receptor-related orphan receptor gamma t (ROR $\gamma$ t)

Retinoic acid receptor -related orphan receptor gamma (ROR $\gamma$ , also known as NR1F3) is a nuclear receptor that has a role as a transcription factor in various cells. ROR $\gamma$ t is expressed as inactive in the cytosol and activates upon agonist binding. Activated ROR $\gamma$ t functions as a transcription factor and modulates (activates or silences) its target genes in the cell nucleus mediating a physiological response. ROR $\gamma$  originates from the protein-coding gene *RORC* (GeneCards, 2021). ROR $\gamma$  has multiple physiological roles, such as in the regulation of circadian rhythm, metabolism, and immunity (Jetten AM, 2009). It is also involved in the progression of inflammation-associated cancers (Yang et al., 2014).

An isoform of ROR $\gamma$ , known as ROR $\gamma$ t (or ROR $\gamma$ 2), is expressed in a variety of immune cells where it is involved in the regulation and activation of the immune system. Unlike ROR $\gamma$ , ROR $\gamma$ t is solely expressed in immune cells, and hence its effects are specifically limited to immune system. Not all ROR $\gamma$ t-mediated regulatory pathways are known due to their complexity, but there are several. Some regulatory effects include T helper 17 cell maturation and activation of the NLRP3 inflammasome, both leading to an inflammatory response (Billon et al., 2019; Weaver et al., 2013). ROR $\gamma$ t may also play a role in T cell receptor activation pathways aiding in the maturation of alpha/beta T-cells (GeneCards, 2021). In addition, the development of immune system components such as lymph nodes is orchestrated by ROR $\gamma$ t-mediated pathways (Jetten et al., 2018). ROR $\gamma$ t is therefore considered as a master regulator of the immune response (Unutmas D, 2009).

The interest in ROR $\gamma$ t-related research is increasing, as determined by PubMed search (search word 'ROR $\gamma$ t' in 2021). The first related ROR $\gamma$ t study was published in 1994, and ever since, the number of such studies has increased. Nowadays, the number of annually published such studies is around 300.

Previously, the typical aim in ROR-targeting drug design studies (for discovering anti-inflammatory agents) was to nonexplicitly inhibit ROR protein family members such as ROR $\alpha$ , ROR $\beta$ , and ROR $\gamma$ . However, the focus shifted specifically towards ROR $\gamma$  due to off-target effects (Huang et al., 2020). Over time, it was later discovered that the isoform of ROR $\gamma$ , known as ROR $\gamma$ t, could provide an even more explicit anti-inflammatory effect. The current focus of discovering ROR-targeting drugs has indeed been

set at ROR $\gamma$ t inhibition. Furthermore, ROR $\gamma$ t has become a more relevant and approachable target for drug discovery as its 3D protein structure was determined just a few years ago.

### ***1.1.1 The role in inflammatory diseases***

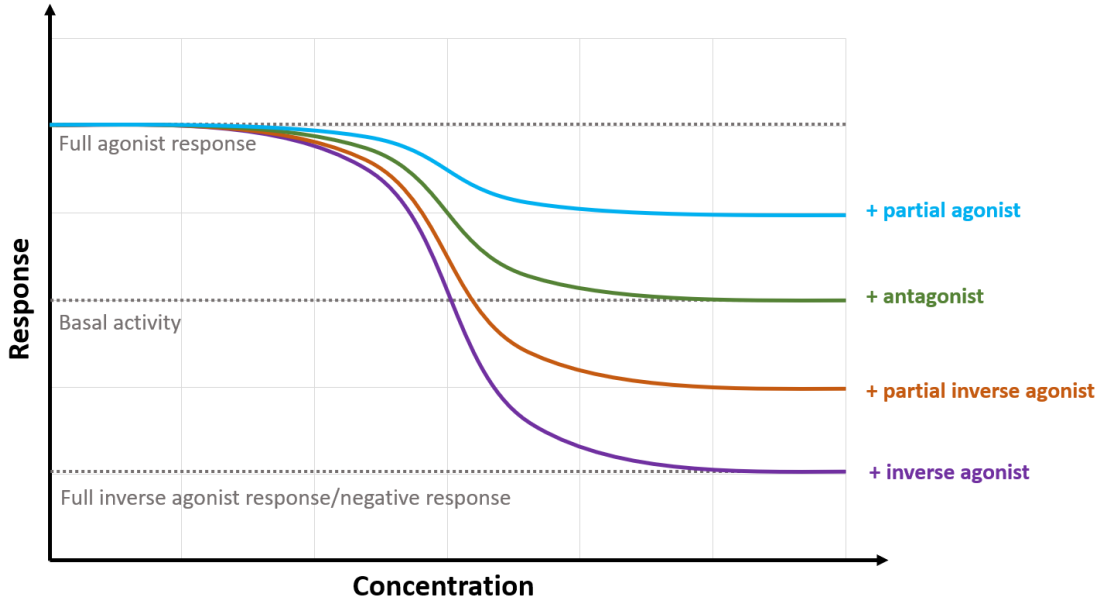
ROR $\gamma$ t is important in mediating an immune response to aid in fighting against pathogens or destruction of abnormal malignant cells in the body. However, constantly elevated expression of ROR $\gamma$ t has been connected to multiple inflammatory diseases such as arthritis, psoriasis, inflammatory respiratory diseases, autoimmune encephalomyelitis, multiple sclerosis, and inflammatory bowel disease (Weaver et al., 2013; Yang et al., 2014).

A typical characteristic of these diseases is persistent and chronic inflammation, which is partly due to constantly active ROR $\gamma$ t. As a master regulator of inflammation, ROR $\gamma$ t disruption has been hypothesized to neutralize its inflammatory effects and aid in directing the physiology of the body towards the normal condition.

### ***1.1.2 Biomolecular and pharmacological characteristics***

A compound exerts biological activity when it chemically interacts with a biological target, typically a receptor protein, in such a way that leads to a physiological response. This response is generally reached when a compound binds to the binding site of the protein and changes its biological activity. Such compounds are also called modulators which either increase or decrease the protein activity. Modulators increasing activity are called agonists while compounds decreasing activity are called inhibitors.

A group of inhibitors, antagonists, hamper the biological activity by inhibiting agonist binding and neutralizing the biological effect. However, protein bound to an antagonist may still have basal biological activity as receptor proteins can signal in the absence of agonists (figure 1). Greater inhibition can be achieved by blocking the basal activity of a receptor by inhibitors known as inverse agonists. Inverse agonists show negative intrinsic efficacy and provide opposite or ‘inverse’ effects to that of agonists. Modulators showing partially characteristic effect are called partial modulators. For example, partial inverse agonists do not completely block basal biological activity, like full inverse agonists (Kelly et al., 2018).



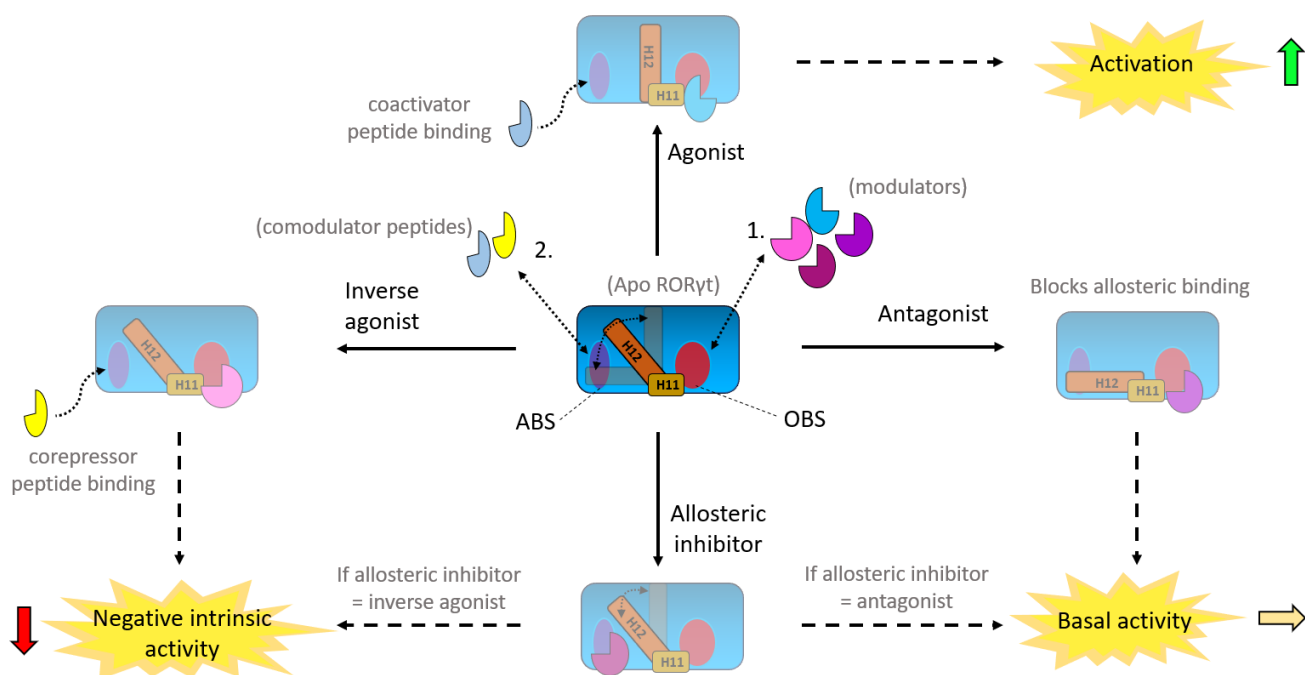
**Figure 1. Inhibitory response of a different group of inhibitors on active protein presented as half maximal inhibitory concentration ( $IC^{50}$ ) curve** (Berg et al., 2018). Inhibitory response is presented in the Y-axis while inhibitor concentration is presented in the X-axis. Inverse agonists provide the highest inhibitory effect. Figure has been adapted and modified from Berg et al., 2018.

However, the comprehensive model for pharmacological function of ROR $\gamma$ t would be far more complicated than the simplified model described above. Its biological complexity can be explained by its dynamic structural states and the second binding site, also known as allosteric binding site or ABS (Scheepstra et al., 2015; Sun et al., 2018). Additionally, novel discoveries suggest several regulatory factors for ROR $\gamma$ t, such as a third binding site in the hinge domain (HD) called hinge domain binding site (HD-BS). It is subject to SUMOylation which alters the biological function of ROR $\gamma$ t by specifically activating T cell regeneration but halting lymph node development (Lao et al., 2019; He et al., 2018).

To elucidate the actual mechanism of action of ROR $\gamma$ t inhibition, the biological function of ROR $\gamma$ t upon activation must be understood. ROR $\gamma$ t-mediated inflammation initiates when an agonist is bound to its binding site leading to a conformation change cascade. Consequently, one of the dynamic helices, known as helix 11, locks into a fixed position stabilizing adjacent dynamic helix 12 consolidating ABS. Resulting ABS stabilization enables a co-activator peptide binding to ABS which activates ROR $\gamma$ t (figure 2). The activated ROR $\gamma$ t-coactivator complex shifts to the cell nucleus, binds to the specific DNA sequences as a monomer, and activates genes responsible for inflammation while downregulating anti-inflammatory genes (Zhang et al., 2015). Some of the activated genes include Th17 cell genes *IL-17A*, *IL-17F* and *IL-*

23R (Castro et al., 2017). Since ROR $\gamma$ t is an orphan receptor, no endogenous ligands are known. However, oxysterols, such as desmosterol, have been hypothesized to be natural agonists of ROR $\gamma$ t, but this has not been confirmed so far (Hu et al., 2015).

Antagonists have been shown to block ROR $\gamma$ t activation by preventing a co-activator binding to the ABS, which neutralizes its effects. On the other hand, inverse agonists paradoxically tip ROR $\gamma$ t towards anti-inflammatory effects. Inverse agonist binding stabilizes ABS into such conformation that it binds a co-repressor peptide instead of a co-activator. Co-repressor binding modulates ROR $\gamma$ t function so that the physiological effects are the opposite of an agonist-bound ROR $\gamma$ t. This is since the genes modulated by an agonist-bound ROR $\gamma$ t are reversely modulated by an inverse agonist-bound ROR $\gamma$ t. Consequently, the genes responsible for inflammation are downregulated and anti-inflammatory genes upregulated resulting in anti-inflammatory effects.



**Figure 2. Molecular mechanism of ROR $\gamma$ t.** A modulator (inhibitor or agonist) binding to OBS (orthosteric site, 1.) stabilizes ABS (allosteric site) through stabilization of adjacent helices 11 (H11) and 12 (H12) enabling allosteric modulator binding (corepressor or coactivator peptide, 2.). This results in a physiological response leading to either activation (by agonists) or repression (by inverse agonists). Antagonists block both of these effects but retain basal activity. Figure has been adapted and modified from Scheepstra et al., 2015.

Inhibition of a biological target is typically evaluated and measured as half-maximal inhibitory concentration (IC<sub>50</sub>) values which indicates the concentration of a given compound (here inverse agonist) that



is needed to inhibit the function of a given biological target (here ROR $\gamma$ t) by 50 %. Inverse agonists provide the highest inhibitory effects, as indicated in figure 1.

ROR $\gamma$ t consists of several functional sequences. In short, the DNA-binding domain (DBD) is responsible for inflammatory regulation of the target genes while HD provides flexibility and is involved in the regulation of ROR $\gamma$ t function. On the other hand, the ligand-binding domain (LBD) binds a ligand which determines the physiological response in the end (Huang et al., 2020). The LBD comprises 14  $\alpha$ -helices, four short beta strands, and two turns (The UniProt Consortium, 2021). The main binding site, orthosteric binding site (OBS) lies in the core, while the ABS is adjacent to it.

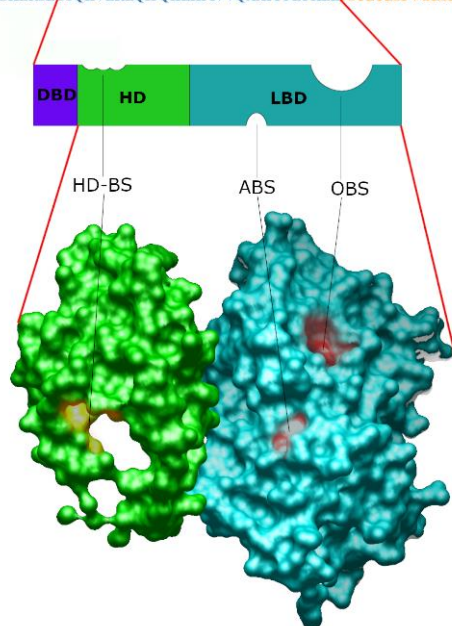
Regarding drug design, the most exciting domain is the LBD, as it is responsible for protein activation and can be chemically modulated by compounds. Consequently, the focus of this project is on the LBD. HD-BS appears also to be druggable but as the function of HD is largely unknown, it was excluded from the scope of this project.

The most important mechanism of action for ROR $\gamma$ t inverse agonism is the disruption of the H-bonds between amino acids tyrosine 502 (TYR502) and histidine 479 (HIS479) in LBD (Sun et al., 2018). Phenylalanine 506 (PHE506) may also form an H-bond with TYR502 or/and HIS479. These H-bonds are stabilized upon agonist binding and disrupted upon inverse agonist binding, which eventually determines co-activator/co-repressor binding to ABS and the physiological effect. Other important amino acids for inverse agonist binding are histidine 323 (HIS323), leucine 324 (LEU324), arginine 367 (ARG367), phenylalanine 377 (PHE377), phenylalanine 378 (PHE378), and phenylalanine 388 (PHE388). Additionally, a water molecule named WAT770 in proximity to PHE377 and HIS323 is considered important in mediating an inverse agonism response (Li et al., 2017; Kallen et al., 2017). The positions of the eight important amino acids for inverse agonist binding (and hence inverse agonism) in sequence are presented in figure 3. The binding sites are illustrated in figure 4.

MDRAPQRQHRSRELLAAKKTHTS**QIEVIP**CKICGDKSSGIHYGVITCEGCKGFFRRSQRC  
 NAAYSCTRQNCPIDRTSRNRCQHCRLLQKCLALGMSRDAVKFGRMSKKQRDSLHAEVQ  
 KQLQQRQQQQEPVVKTPPAGAQQADTLTYTLGLPDGQLPLGSSPDLPEASACPPGLLK  
 ASGSGPSYSNNLAKAGLNGASCHLEYSRPERGKAEGRESFYSTGSQLTDPDRCLRFEEHRH  
 PGLGELGQGPDSYGSPSFRSTPEAPYASLTEIEHLVQSVCKSYRETQCLRLEDLLRQRSNIF  
 SREEVTGYQRKSMWEMWERCAHHLTEAIQYVVEFAKRLSGFMELCQNDQIVLLKAGA  
 MEVVLVVMCRAYNADNRTV**FF**EGKYGGMEL**F**RALGCSELISSIFDFSHLSALHFSEDEIA  
 LYTALVLINAHRPGLQEKRRKVEQLQYNLELAFHHHLCKTHRQSILAKLPPKGLRSLCSQ  
**H**VERLQIFQHLHPIVVQAAFPPI**Y**KEL**FSTETESPVGLSK**

**Figure 3. The sequence of 518 amino acids for ROR $\gamma$ .** The sequence consists of N-terminal domain at 1-30 (red), DNA-binding domain (DBD) at 31-95 (purple), hinge domain (HD) at 96-265 (green), ligand binding domain (LBD) at 266-505 (blue), and a C-terminal domain at 506-518 (orange). ROR $\gamma$ t, an isoform of ROR $\gamma$ , differs only in the first 24 amino acids (amino acids 1-21 truncated, amino acids 22-24 mutated: HTS  $\rightarrow$  MRT) but is otherwise completely identical. The eight most important amino acids for inverse agonist binding are marked as red circles.

MDRAPQRQHRSRELLAAKKTHTS**QIEVIP**CKICGDKSSGIHYGVITCEGCKGFFRRSQRCNAAYSCTRQNCPIDRTSRNRCQHCRLLQKCLALGMSRDAVKFGRMSKKQRDSLHAEVQKQLQQRQQQQEPVVKTPPAGAQQADTLTYTLGLPDGQLPLGSSPDLPEASACPPGLLKA  
 SGSGPSYSNNLAKAGLNGASCHLEYSRPERGKAEGRESFYSTGSQLTDPDRCLRFEEHRHPGLGELGQGPDSYGSPSFRSTPEAPYASLTEIEHLVQSVCKSYRETQCLRLEDLLRQRSNIF  
 SREEVTGYQRKSMWEMWERCAHHLTEAIQYVVEFAKRLSGFMELCQNDQIVLLKAGAMEVVLVVMCRAYNADNRTV**FF**EGKYGGMEL**F**RALGCSELISSIFDFSHLSALHFSEDEIALYTALVLINAHRPGLQEKRRKVEQLQYNLELAFHHHLCKTHRQSILAKLPPKGLRSLCSQHVERLQIFQHLHPIVVQAAFPPI**Y**KEL**FSTETESPVGLSK**



**Figure 4. ROR $\gamma$ t binding sites.** Orthosteric binding site (OBS) is the main binding site. Allosteric binding site (ABS) lies next to it, which participates in a co-modulator peptide binding. OBS and ABS are located in the ligand-binding domain (LBD). Hinge domain binding site (HD-BS), presumably exists and is located in the hinge domain (HD). HD-BS supposedly regulates ROR $\gamma$ t function and is subject to modifications such as SUMOylation. ROR $\gamma$ t activation can be interrupted by interfering with any of these mechanisms, indicating several ways to manipulate ROR $\gamma$ t function by chemical agents. Importantly, this information is essential in a drug design process for ROR $\gamma$ t. NB: This figure just visualizes ROR $\gamma$ t binding sites and does not demonstrate their accurate locations or shape.

ROR $\gamma$ t modulators can be classified into several different classes. First, there exist both full and partial ROR $\gamma$ t inhibitors (inverse agonists and antagonists). Full inhibitors completely stabilize ROR $\gamma$ t and, as such, provide the maximal effect, while partial inhibitors only partly stabilize ABS resulting in poorer co-repressor binding leading to lesser inhibitory effects. Additionally, inhibitors can bind to either OBS or ABS. Inhibitors binding to OBS are called orthosteric inhibitors, while inhibitors binding to ABS are

called allosteric inhibitors. In conclusion, a total of at least eight (2 x 2 x 2) different inhibitory pharmacological states of ROR $\gamma$ t have been discovered. See Table 1 for more information.

**Table 1. The pharmacological characteristics of known ROR $\gamma$ t inhibitors.**

INHIBITION TYPE	COMMENT
Full orthosteric inverse agonist	Provides a high anti-inflammatory effect and OBS is a well-characterized binding site. This group of inhibitors seems the most appealing in drug discovery.
Full allosteric inverse agonist	Provides a high anti-inflammatory effect with less off-target effects, but ABS-targeting inhibitors are a less studied field and not well characterized. ABS is also significantly smaller than OBS, which might be harder to target.
Full orthosteric antagonist	Cease inflammatory effects but do not result in anti-inflammatory effects like inverse agonists do.
Full allosteric antagonist	Similar effects as full orthosteric antagonists but with less off-target effects and is harder to target.
Partial orthosteric inverse agonist	Provide partial anti-inflammatory effects.
Partial allosteric inverse agonist	Provide partial anti-inflammatory effects. Are not considered as a good ROR $\gamma$ t inhibitor group.
Partial orthosteric antagonist	Inhibitory effect is not significant.
Partial allosteric antagonist	Inhibitory effect is not significant.

In conclusion, ROR $\gamma$ t agonists aggregate inflammation, antagonists neutralize ROR $\gamma$ t effects, while inverse agonists propose anti-inflammatory effects. ROR $\gamma$ t inverse agonists have the highest anti-inflammatory effect and show the highest therapeutic value for treating inflammatory diseases with ROR $\gamma$ t modulators.

## 1.2 ROR $\gamma$ t and drug discovery

ROR $\gamma$ t is important in mediating an immune response and proposing inflammatory effects, as mentioned previously. Consequently, ROR $\gamma$ t is considered an appealing drug target for multiple inflammatory conditions. ROR $\gamma$ t inhibition has already been demonstrated *in vivo* to reduce the symptoms of psoriasis-like skin inflammation, arthritis, and colitis in mice (Xue et al., 2016; Chang et al., 2014; Igaki et al., 2019). Additionally, ROR $\gamma$ t-deficient mice have been shown to develop resistance for experimental autoimmune encephalomyelitis and multiple sclerosis (Kumar et al., 2012; Fukase et al., 2018).

Until today, some patients impaired by chronic inflammatory diseases do not respond correspondingly to the available treatment. Hence, the arsenal of treatment options for chronic inflammatory diseases

could be expanded. As ROR $\gamma$ -targeting drugs are not currently commercially available and ROR $\gamma$  has been shown as a potential drug target for treating chronic inflammatory diseases, ROR $\gamma$  inhibitors search can be justified. Indeed, several clinical trials (Phase I/II) for the treatment of psoriasis with ROR $\gamma$  inverse agonists have emerged, proving the potential of ROR $\gamma$  inhibitors for therapeutic human use (ClinicalTrials.gov, 2016).

However, the ongoing clinical trials have already demonstrated some issues with ROR $\gamma$  inhibitors implying higher potency and efficiency might be required to reach a significant therapeutic value (Huang et al., 2020). Another issue with ROR $\gamma$  is its unspecificity. As ROR $\gamma$  mediates several physiological pathways, it is not considered as a specific target. Consequently, ROR $\gamma$  inhibition might lead to several adverse effects, and due to its complexity and unspecificity, it is challenging to predict the potential undesired effects. Hypothetically, one expected adverse effect might include immune system deprivation.

Due to its relatively recent discovery as a promising drug target, ROR $\gamma$ -related research has many potentially undiscovered applications in the pharmaceutical industry for treating inflammatory conditions. Especially full ROR $\gamma$  inverse agonists appear to provide the most potent anti-inflammatory effect (table 1).

ROR $\gamma$  druggability has been demonstrated. According to the latest studies, ABS might be a more druggable target due to its higher specificity and corresponding lesser number of off-target effects (Huang et al., 2020). This is because similar OBS (as for ROR $\gamma$ ) can also be present in other receptors of ROR-family (such as ROR $\alpha$ , ROR $\beta$ , and ROR $\gamma$ ) while ABS are typically unique to each protein subtype. However, ABS is not as well characterized as OBS since significantly less research has been conducted on ABS. It is also substantially smaller in volume than OBS suggesting ABS might be more challenging to target due to less surface area for binding. The scope of this study was limited to discovering one class of inhibitors targeting one specific site. Here, the focus was set at discovering inverse agonists for OBS mainly due to its better characterization.

### 1.3 Computer-aided drug design (CADD)

Drug design is a process of developing a novel medication by rational design. A similar term, drug discovery, refers to the process of how new medications or drug candidates are discovered. An ancient form of drug discovery is dated to thousands of years ago and was based on experiments conducted by consuming either processed or unprocessed plants or other extracts originating from living organisms. Generation after generation, the lore of natural medicine accumulated thanks to serendipitous discoveries over time. Herbal medicine could be considered the oldest form of medicine.

Modern medicine commenced when the effect of a medication was conceptualized to be mediated by a specific chemical compound, which interacts with a biological target (usually receptor protein) and modulates its bioactivity. In the case of medication, this modulation is considered favorable as disturbing the function of a protein by either inhibiting or activating it can alleviate or even eliminate symptoms of a specific disease. For example, ROR $\gamma$ t is a biological target that mediates inflammation. If its function is inhibited, it proposes anti-inflammatory effects, and would be beneficial for treating inflammatory diseases.

Nowadays, computer tools are applied to model the complex interaction between compound and target. Additionally, other computer approaches are also used to boost the discovery of novel medications. Such approaches can be put under the concept called Computer-Aided Drug Design (CADD). Indeed, recent technological advances have enabled the development of sophisticated CADD protocols to provide superior aid in discovering novel pharmacologically active compounds.

Computing tools for drug discovery and development have existed as long as there have been computers. However, early CADD approaches were relatively simple. Initially, computers stored data originating from experimental studies, which was analyzed with statistical methods. Over time, when computing power increased, more sophisticated tools were developed solely for drug design purposes.

The modern concept for CADD was compiled in the early 1980s, but no single comprehensive description for CADD exists. In short, CADD can be defined as a computational approach to discover and analyze biologically active compounds against the target of interest (Odilia et al., 2016). The field is rapidly growing, which can be determined by PubMed search. For example, there are 391 relevant articles published in 1980 with the search word 'computer-aided drug design' (all publications), while in 2000, there

are 6,373 publications and 24,326 publications in 2020 with the same search parameters. Indeed, the number of relevant publications steadily increases each year.

CADD is typically divided into ligand- and structure-based drug design, even though several other approaches exist (Amy et al., 2003; Acharya et al., 2011). Ligand-based drug design utilizes the structural data of known bioactive compounds towards the target of interest. Structural characteristics of these compounds, such as shape and charge, responsible for bioactivity can be modeled. Furthermore, the derived model can be used to discover or design novel compounds that contain similar characteristics and may subsequently show the desired activity towards the target of interest. On the other hand, structure-based drug design utilizes a three-dimensional (3D) structure of the biological target. Based on the structural data, it is possible to predict whether a given compound binds to the biological target or not.

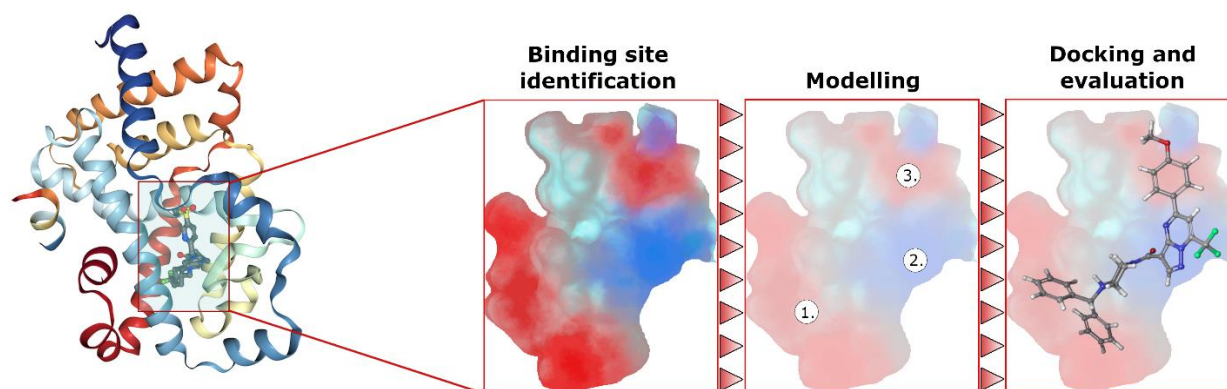
Thanks to the technological advances, more and more complex CADD approaches can be developed in the future. One such promising approach is called molecular dynamic simulation, where the dynamic state of chemical interactions can be modeled (Liu et al., 2018). The movement of both the compound and target are predicted over time in the compound binding process, which enables higher prediction accuracy as the information of dynamic state is utilized. These simulations are especially applicable when the target is highly dynamic. However, such simulations require an immense amount of computing power due to the complexity of the models. Anyhow, CADD models will keep on evolving, which will enable the discovery of evermore complex drugs in the future.

### ***1.3.1 Structure-based drug design***

Protein 3D structures has several applications in drug design. The first protein 3D structure to be solved was myoglobin in 1958 and over the following years, more and more protein 3D structures were deciphered (Kendrew et al., 1958). Structural data of proteins became publicly available upon foundation of the Protein Data Bank (PDB) in 1971, accelerating the utilization of structural data (Berman et al., 2000). A critical application of structural data was discovered in drug design, which led to the concept of structure-based drug design.

Structure-based drug design utilizes the 3D structure of the biological target. Structural data enables prediction of what kind of compounds might bind to the biological target by modeling chemical interactions between a given compound and binding site of the biological target. Generally, the compounds that

are complementary both in shape and polarity/charge to the binding site show strong chemical interactions and hence provide high binding affinity (figure 5).



**Figure 5. Principle of structure-based drug design.** Protein 3D structure can be utilized to model its binding site. Here, red color describes hydrophobic and blue color polar binding site regions. Based on the structure, it is possible to design such compounds that are complementary in shape and charge in respect to the binding site and thus might show high binding affinity. Here, three (1. hydrophobic end, 2. hydrophilic area, 3. hydrophobic area) important regions are defined (1-3). If a given compound placed/docked in the binding site (final frame) fits without steric obstruction and shows favoring polarity/charges upon binding, it probably has a good binding affinity. With this approach, structural data can be utilized to estimate whether a given compound binds to the target or not.

3D protein structures are determined by a protein crystallization process followed by X-ray crystallography imaging. However, structural state of the protein may alter during the crystallization process which causes flaws to the model. Additionally, 3D protein structure represents only one structural state even though especially dynamic proteins have actually several different conformations. Considering only one structural state has indeed limitations. This challenge can be partly overcome by running computer simulations that consider the dynamic movement of the protein, and consequently, different structural states can be predicted.

### ***1.3.2 Ligand-based drug design***

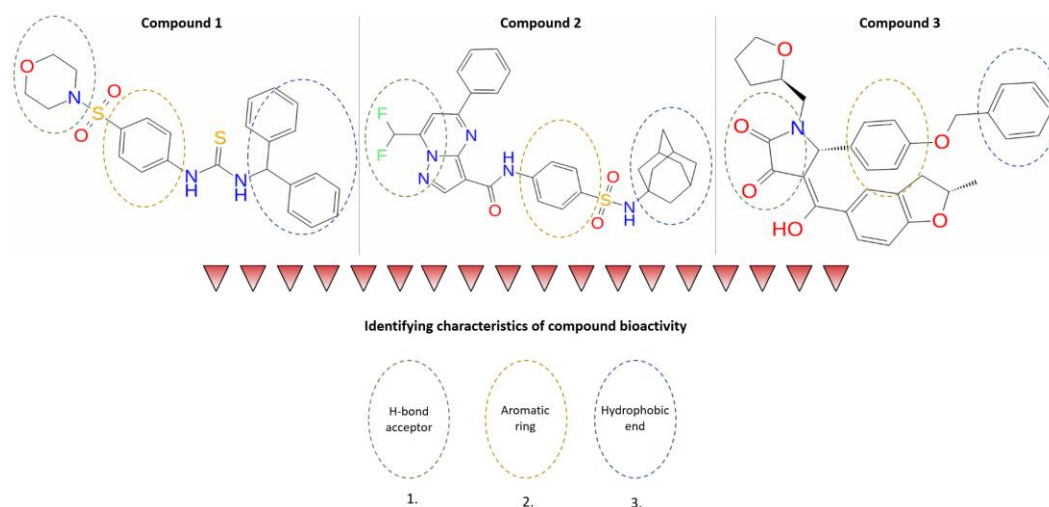
If a set of active compounds towards the biological target are known, it is possible to identify those structural characteristics that make them bioactive towards the target. This abstract description of molecular features is called pharmacophore. The pharmacophore concept dates back to the end of the nineteenth century when methylene blue was discovered to selectively bind to nerve fibers, leading to the idea that there must be such chemical properties that make a compound bind to a specific biological

target. However, the pharmacophore concept itself was introduced far later via publication in 1971 and its widely accepted definition was introduced a few years later in 1977. Pharmacophore is currently defined as “a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule's biological activity” (Daisy et al., 2011).

Ligand-based drug design relies on pharmacophore modeling, which defines those characteristics that are required for compound binding to the target. These characteristics involve structural properties of the compound, such as shape, charge or functional group. To evaluate this, quantitative structure-activity relationship (QSAR) is defined to model the relationship between structural features of compounds and their known biological activities towards the target, such as inhibition potency  $IC_{50}$ . Furthermore, the derived model can be used to discover or design novel compounds with similar characteristics which, as such, may be pharmacologically active towards the same target. If any given compound does not meet the required characteristics for binding (as defined by the QSAR model), it is eliminated in the ligand-based drug design process (Verma et al., 2010).

The basic idea of pharmacophore is presented in figure 6, where three known bioactive compounds (towards the same biological target) are compared. In this example, each compound has three similar structural properties which include an H-bond acceptor end, an aromatic ring in the middle, and a hydrophobic end, each in the same approximate positions. These three structural properties can be hypothesized to be crucial for bioactivity, and a pharmacophore model can be constructed. This information can be used in CADD to exclude those compounds that do not show the required three structural properties in similar positions. However, no reliable pharmacophore model can be constructed solely based on three compounds. Instead, tens or preferably hundreds of bioactive compounds are looked upon to discover the common characteristics. The more compounds are included, the more detailed and reliable QSAR model can be constructed.





**Figure 6. Principle of pharmacophore, demonstrated by a quantitative structure-activity relationship (QSAR) model generation.** Three common characteristics (H-bond acceptor end, aromatic ring in the middle, and hydrophobic end) of the three known bioactive compounds are identified for constructing a simple QSAR model. Compounds that do not show the three required bioactive characteristics presented by QSAR are excluded.

One limitation with pharmacophore includes lack of discovery of novel bioactive structural features as QSAR model favors compounds similar to already known bioactive compounds. This basically leads to a less innovative drug discovery approach as compounds that are remarkably different from the defined characteristics are excluded even though they might have desired biological activity towards the target.

### 1.3.3 Other CADD approaches

Other relevant CADD methods are shortly described here. In general, there are hundreds of different programs that can be utilized in the CADD protocol to conduct a single virtual screen (VS). Here, VS refers to a set of CADD methods that are applied to search the most potential bioactive compounds against the biological target from a compound library containing hundreds of thousands of molecules. As hundreds of different programs exist and different program settings may be used, there are practically unlimited ways to conduct a single VS.

Generally, CADD tools are used to predict which compounds have bioactivity against the target of interest. Some programs can directly predict the binding affinity of the compound, while other tools are solely used for optimization purposes. Additionally, chemical properties like lipophilicity of a given compound can be predicted which might provide useful information as highly hydrophobic compounds make poor

drugs due to poor solubility. Compounds with high logP values are typically discarded early on. Thus, it is possible to predict before any experimental test whether a given compound has such chemical properties that makes it a good drug or not. Other similarly challenging compounds include pan-assay interference compounds (PAINS). These are such compounds that react with numerous biological targets and disturb experimental tests. Such compounds can be identified with computational tools and discarded (Sun et al., 2018).

Lately, molecular dynamics tools have been developed for CADD (Liu et al., 2018). Molecular mechanics with generalized Born and surface area solvation (MM/GBSA) is a tool used to calculate binding affinity of compounds to the biological target of interest (Virtanen et al., 2015). It essentially calculates dynamic compound binding energies to the target by running molecular dynamics simulations.

After CADD protocol has been executed, compounds (originating from a compound library) showing the highest potential are selected. Finally, these compounds may be inspected individually by using biomolecular visualization tools to estimate reliability of compound binding. This process is called visual screening. Thanks to VS, only a few (typically tens or hundreds) compounds need to be inspected instead of the whole compound library. Furthermore, the most potential compounds based on visual screening are manually selected for *in vitro* or *in vivo* experiments to discover its actual bioactivity. Therefore, CADD is nowadays considered as a crucial part of drug discovery and development process.

#### **1.4 ROR $\gamma$ t and CADD**

As previously mentioned, ROR $\gamma$ t is a rather challenging drug target. This is mainly due to its complex function and numerous roles in regulating inflammation in the body. Additionally, its structure has been considered complex and dynamic which proposes additional challenges for drug design projects. To overcome these challenges, powerful CADD approaches can be justified (Liu et al., 2018).

There is a sufficiently large amount of data for both ligand- and structure-based drug design approaches for discovering novel ROR $\gamma$ t modulators. Until today (as in 2021), there are around 1,700 known ROR $\gamma$ t ligands in ChEMBL database, a few hundred of which are ROR $\gamma$ t inhibitors and around half of those inhibitors are inverse agonists (Gaulton et al., 2012). Additionally, there are over 100 ROR $\gamma$ t structures currently available in PDB, most of which are inhibitor-ROR $\gamma$ t complexes. The first inverse agonist -

bound structure was published in 2015 while there are already 78 inverse agonist -bound structures available as in 2021.

ROR $\gamma$ t inhibition has demonstrated its opportunities in treating inflammatory conditions but because of its complexity and its relatively recent discovery as a potential drug target, no actual commercially available ROR $\gamma$ t inhibitors exist so far. Hence, the discovery of novel ROR $\gamma$ t inhibitors can be justified and recent CADD methods provide useful and modern approach for this challenge. Additionally, there are plenty of new ligand- and structure data for ROR $\gamma$ t which enables a powerful VS approach for discovering potent ROR $\gamma$ t inhibitors.

## 1.5 Thesis summary

The purpose of the study is to discover novel promising full ROR $\gamma$ t inverse agonists by utilizing CADD methods. Here, the applied methods are hypothesized to be applicable for discovering ROR $\gamma$ t inhibitors by utilizing available ROR $\gamma$ t protein structures and known ROR $\gamma$ t inhibitors. Additionally, ROR $\gamma$ t is assumed to be a feasible drug target justifying this study (Huang et al., 2020).

The designed CADD protocol includes several methods. The two main methods include ligand- and structure-based drug design which both were designed with the aid of two ‘in-house methods’ known as SDFCONF (for ligand-based drug design), and brute force negative image-based optimization (BR-NiB, for structure-based drug design) developed by MedChem group ([www.medchem.fi/](http://www.medchem.fi/)) (Kurkinen et al., Submitted). Note that ligand-based drug design approach with SDFCONF is not considered as a traditional ligand-based drug design. Instead, SDFCONF is a tool that can be applied for filtering compounds that do not meet the pharmacophore-points characteristics set by SDFCONF. For simplification, SDFCONF-based approaches are from now on referred as ligand-based drug design as SDFCONF was only used for ligand-based drug design -like applications in this project.

Additional tools were utilized for developing CADD models such as protein structure optimization, ligand preparation, ligand docking and NIB-modeling, to name a few. In the end, other applicable CADD methods were also utilized including PAINS- and logP-filtering, MM-GBSA and visual screening. In the end, a successful CADD protocol was designed to run a ROR $\gamma$ t inverse agonist VS. Details are provided in chapter 4.

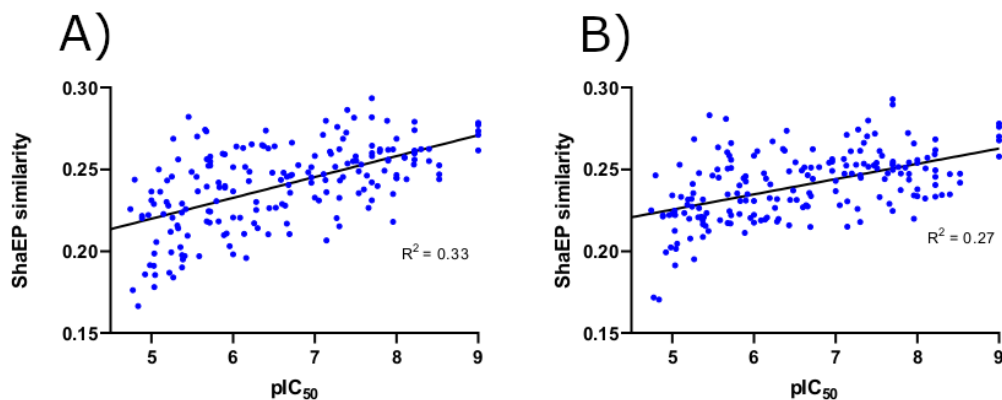
## 2 RESULTS

### 2.1 Structure-based drug design

Two 3D protein structures (presented as 5NTW and 5VB6) were successfully utilized for constructing models for structure-based drug design. The two protein structures were optimized which were also utilized in ligand-based drug design. Structure-based drug design approach proved to be a powerful method for identifying inverse agonist candidates as described in this chapter.

#### 2.1.1 Evaluation and validation of the docking- and scoring methods

Pearson correlation coefficients were calculated for known experimental data (as  $IC_{50}$  values) of known 191 full orthosteric inverse agonists which was proportionated with the predicted VS score for each compound. The best  $R^2$  score was achieved when the compounds were docked with PLANTS and scored with ShaEP. The  $R^2$  values and graphs for 5NTW and 5VB6 models (figure 7) were managed with GraphPad Prism (GraphPad Prism version 8.0.0 for Windows, GraphPad Software, San Diego, California USA, [www.graphpad.com](http://www.graphpad.com)).  $R^2$  values of 0.33 and 0.27 were calculated for 5NTW and 5VB6, respectively.



**Figure 7.** Correlations of the activity data of the 191 CHEMBL compounds against ShaEP scores for 5NTW (A) and 5VB6 (B). The highest correlation coefficient ( $R^2$ ) was achieved when the compounds were docked with PLANTS and scored with ShaEP. This yielded  $R^2$  values of 0.33 for 5NTW and 0.27 for 5VB6 model. Potency of the compounds is presented as  $pIC_{50}$  in the X-axis and the corresponding ShaEP score in the Y-axis.

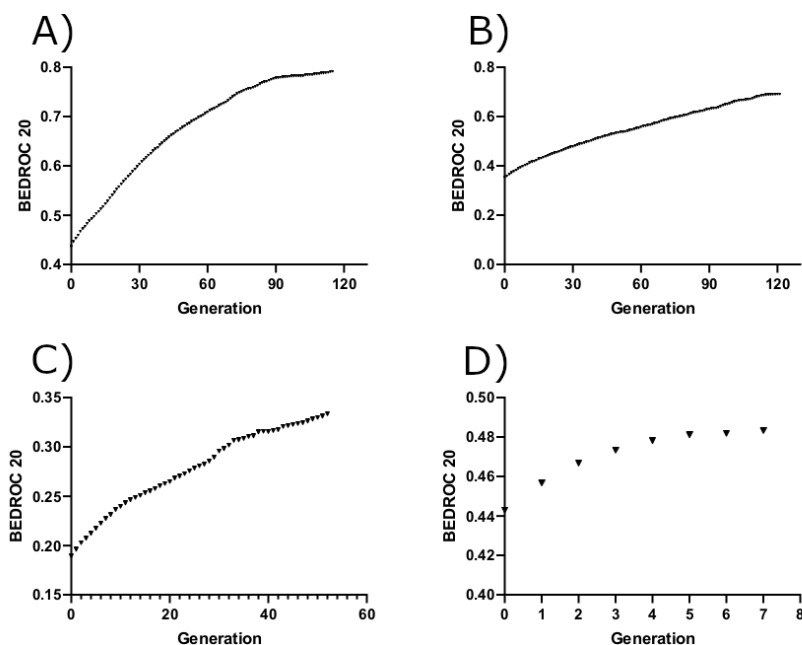
### ***2.1.2 Optimization of the final Negative image-based (NIB)-models***

Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic 20 (BEDROC20 or BR20) got higher generation by generation as seen in figure 8. Here, each new generation is created upon elimination of a single NIB-point from the previous NIB-model generation. The values improved quickly in the beginning but the steepness of the curve got milder in the end. The NIB-models were generated by BR-NiB, but manual modifications were also conducted.

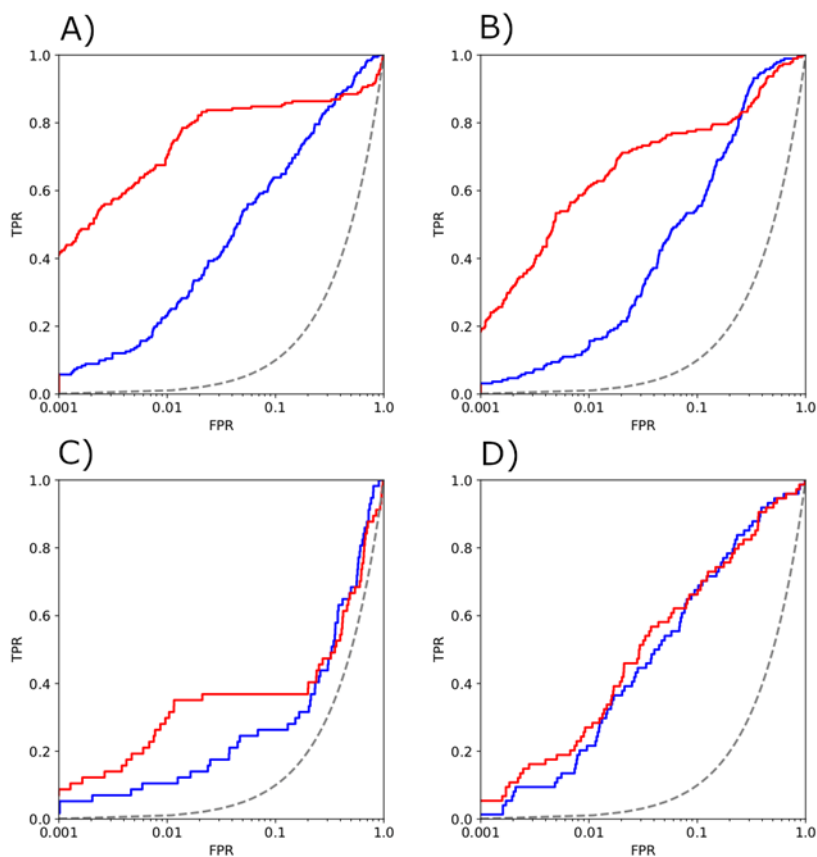
5NTW model with a full set of active compounds was restarted after generation 101 as BR20 did not improve in generation 102. BR-NiB was continued despite of the tiny decline in the BR20 score. Eventually, BR20 score improved until 114<sup>th</sup> generation (figure 8A). 5VB6 NIB-model optimization with a full set of active compounds halted at the 51<sup>st</sup> generation due to failure in BR20 improvement. However, the optimization process was continued and BR20 improved until 121<sup>st</sup> generation (figure 8B).

5NTW model with the lowest-ranked active compounds present was similarly restarted after 39<sup>th</sup> generation as BR20 did not improve in 40<sup>th</sup> generation and BR20 steadily improved until 52<sup>nd</sup> generation (figure 8C). 5VB6 NIB-model with the lowest-ranked active compounds present was generated without any modifications (figure 8D).

Rocker graphs were obtained for each four NIB-models to characterize the enrichment of the active compounds. NIB-model optimization significantly improved enrichment of the active compounds in each of the models as illustrated in the figure 9.



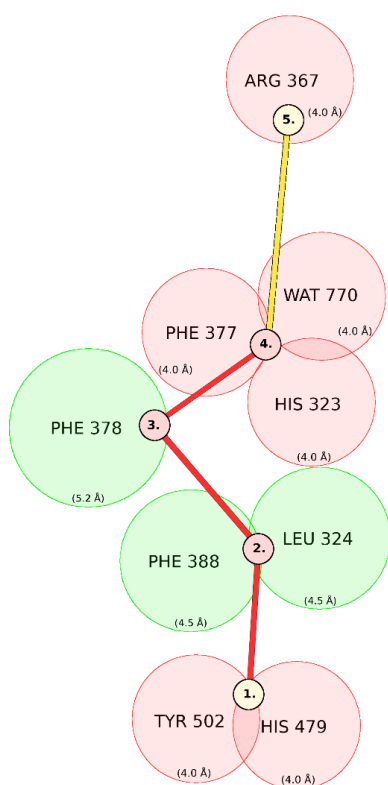
**Figure 8. Improvement of BR20 values generation by generation in each of the four NIB-models.** Generations of the optimized NIB-models by BR-NiB are presented in the X-axis and BR20 values in the Y-axis. As seen from the graphs, BR-NiB improved the BR20 values of the NIB-models remarkably. A) 5NTW model with the full set of active compounds B) 5VB6 model with the full set of active compounds C) 5NTW model with the lowest-ranked active compounds included only D) 5VB6 model with the lowest-ranked active compounds included only. The graphs were generated with GraphPad Prism (version 8.0.0).



**Figure 9. Roc curves of the four different models illustrating enrichment of the active compounds against one fourth of randomly selected inactive compounds from Specs.** The blue roc curves present the enrichment of the active compounds based on the original NIB-models while the red graphs present the enrichment based on the optimized, final NIB-models. True positives or active compounds are presented in the Y-axis while false positives or Specs compounds are presented in the X-axis. As seen from the graphs, BR-NiB improved the enrichment of the active compounds. The roc graphs presented herein describe the following models for enrichment: A) 5NTW model with the full set of active compounds B) 5VB6 model with the full set of active compounds C) 5NTW model with the lowest-ranked active compounds included only D) 5VB6 model with the lowest-ranked active compounds included only. The graphs were generated with Roc software.

## 2.2 Ligand-based drug design

Crucial amino acids for ligand binding were determined for constructing a ligand-based drug design model with 'in house' SDFCONF script. Ligand-based drug design screen separated active compounds from inactive ones. With the optimized settings for the nine coordinate points with corresponding radiuses (figure 10), SDFCONF-screen excluded 87 % of all the remaining Specs compounds while 30-40 % (percentage depended on which of the four models were used) of the active compounds were discarded in the screening.



**Figure 10. SDFCONF-based (3D) QSAR pharmacophore model illustrated as 2D.** 3D coordinates of each of the selected amino acids were acquired from the optimized protein structures of 5NTW and 5VB6 and their position was used to identify the most potential compounds presenting ROR $\gamma$ t-inverse agonism by setting criterions for screening. The applied criterions included: (1) N, O, S, F, Cl, Br or I present within 4.0 Å radius from either HIS479 or TYR502, (2) any aromatic atom within 4.5 Å radius from either LEU324 or PHE388, (3) any aromatic atom within 5.2 Å radius from PHE378, (4) N, O, S, F, Cl, Br or I present within 4.0 Å radius from either HIS323, PHE377 or WAT770, (5) N, O, S, F, Cl, Br or I present within 4.0 Å radius from ARG367. Compounds that fulfilled the criterions 1, 2, 3 and 4 or fulfilled the criterions 2, 3, 4 and 5 but also had any heavy atom within 4Å radius from either HIS479 or TYR502 passed the SDFCONF screen. Red circles demonstrate radiuses for hydrophilic groups and green circles radiuses for hydrophobic groups in respect to the selected amino acids.


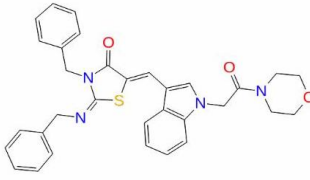
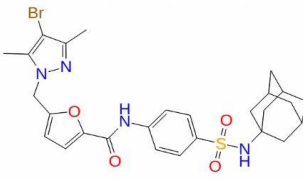
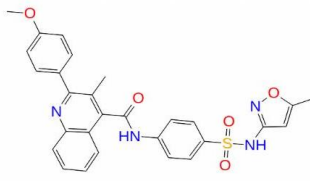
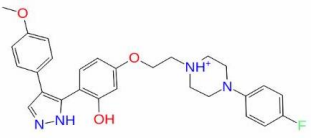
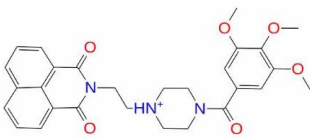
## 2.3 Acquiring compounds with structure- and ligand-based drug design screen

A total of around 6,800 unique (duplicates removed) compounds remained after the structure-based drug design screen (from each of the four NIB-models) originating from the total number of 1,900,000 Specs compounds. Furthermore, the number of 6,800 compounds was reduced to a total of 865 compounds after SDFCONF ligand-based drug design screen. The remaining 865 compounds were acquired for the final screening.

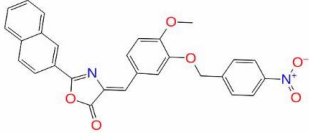
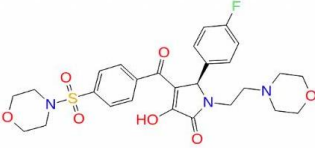
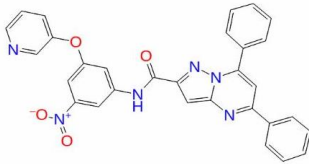
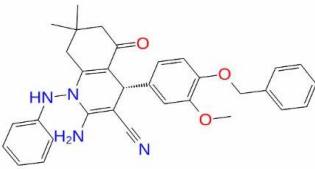
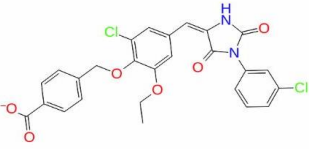
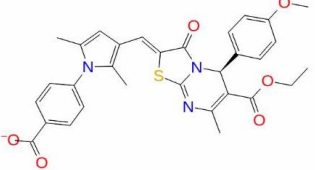
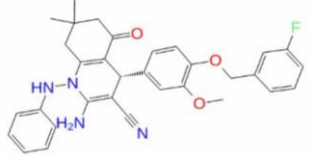
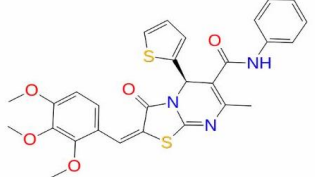
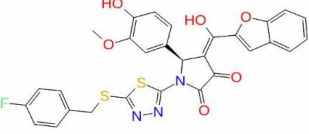
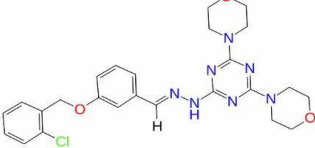
## 2.4 Final screening and obtaining the final compounds

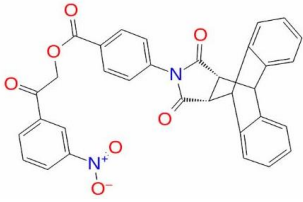
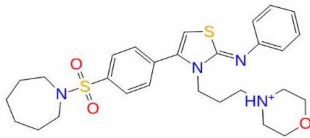
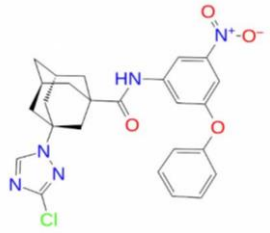
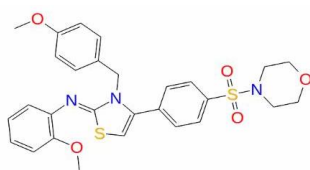
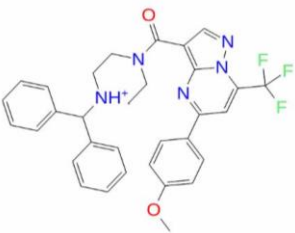
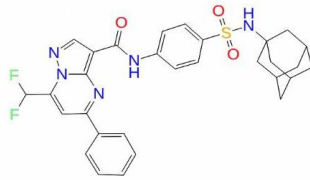
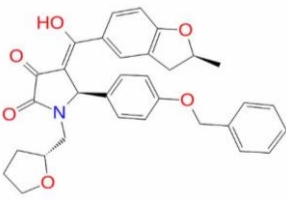
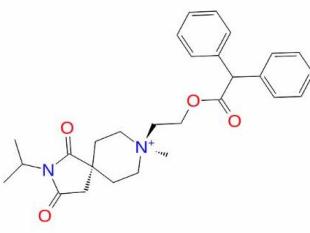
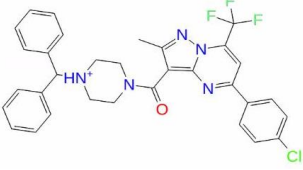
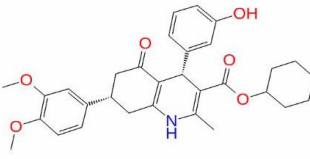
The last step consisted of four separate screens described herein. After the first three screens (logP-, PAINS- and MM/GBSA-screening), a total of 230 compounds remained (5NTW 1<sup>st</sup> model: 63 compounds, 5NTW 2<sup>nd</sup> model: 62 compounds; 5VB6 1<sup>st</sup> model: 58 compounds, 5VB6 2<sup>nd</sup> model: 47 compounds). Hundred compounds possessing the highest MM/GBSA-scores, which were considered as the most potent compounds, were visually inspected by including the 25 highest scored compounds from each of the four models. Out of the hundred compounds, 29 compounds were eventually selected as the promising inverse agonists. The compounds are shown in table 2 with comments on LBD goodness of fit.

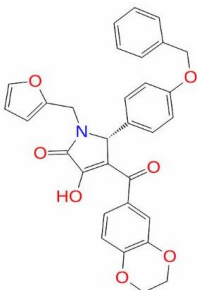
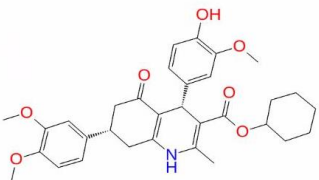
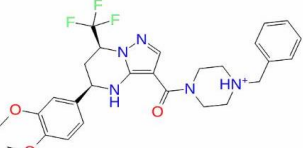
**Table 2. The selected 29 compounds of the final screen shown with compound ID, compound structure and comment on inverse agonist goodness of fit binding based on the visual screen. The general fit, H-bonding with ARG367 and TYR502-HIS479-PHE506 agonist lock disruption was evaluated. Nitro groups were also described as they may show toxicity issues in clinical usage.**

COMP OUND ID	COMPOUND STRUCTURE	COMMENT	COMP OUND ID	COMPOUND STRUCTURE	COMMENT
AH- 487/40 936254		Good fit to OBS. Strong H-bond with ARG367 and moderately blocks TYR502-HIS479-PHE506 agonist lock formation. Includes nitro group which may show toxicity issues in clinical usage.	AO- 081/15 386957		Good fit to OBS. No H-bond with ARG367 and moderately blocks TYR502-HIS479-PHE506 agonist lock formation.
AK- 968/15 608930		Good fit to OBS. Moderately strong H-bond with ARG367 and moderately blocks TYR502-HIS479-PHE506 agonist lock formation.	AK- 968/15 253414		Excellent fit to OBS. Strong H-bond with ARG367 and moderately blocks TYR502-HIS479-PHE506 agonist lock formation.
AO- 022/43 452979		Good fit to OBS. Strong H-bond with ARG367 and significantly blocks TYR502-HIS479-	AF- 399/15 599217		Good fit to OBS. No H-bond with ARG367 and moderately blocks



		PHE506 agonist lock formation.			TYR502-HIS479-PHE506 agonist lock formation.
AK-968/12 162001		Good fit to OBS. No H-bond with ARG367 and moderately blocks TYR502-HIS479-PHE506 agonist lock formation. Includes nitro group which may show toxicity issues in clinical usage.	AF-399/41 318865		Excellent fit to OBS. Strong H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation.
AG-690/11 570086		Good fit to OBS. No H-bond with ARG367 and moderately blocks TYR502-HIS479-PHE506 agonist lock formation. Includes nitro group which may show toxicity issues in clinical usage.	AN-989/41 838397		Good fit to OBS. No H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation.
AH-487/41 654264		Good fit to OBS. Moderately strong H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation.	AN-648/15 596220		Excellent fit to OBS. No H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation.
AN-989/41 838307		Excellent fit to OBS. No H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation.	AO-081/15 386326		Good fit to OBS. No H-bond with ARG367 and excellently blocks TYR502-HIS479-PHE506 agonist lock formation.
AF-399/42 017933		Excellent fit to OBS. No H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation.	AG-205/36 940103		Good fit to OBS. Strong H-bond with ARG367 and significantly blocks TYR502-

					HIS479-PHE506 agonist lock formation.
AK-918/11 909161		Good fit to OBS. No H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation. Includes nitro group which may show toxicity issues in clinical usage.	AF-399/41 895766		Good fit to OBS. Strong H-bond with ARG367 and moderately blocks TYR502-HIS479-PHE506 agonist lock formation.
AG-205/33 688028		Excellent fit to OBS. No H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation. Includes nitro group which may show toxicity issues in clinical usage.	AK-968/15 607331		Good fit to OBS. No H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation.
AG-219/11 789371		Good fit to OBS. Strong H-bond with ARG367 and moderately blocks TYR502-HIS479-PHE506 agonist lock formation.	AK-968/41 022069		Excellent fit to OBS. No H-bond with ARG367 and moderately blocks TYR502-HIS479-PHE506 agonist lock formation.
AF-399/42 017292		Good fit to OBS. No H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation.	AE-641/30 156032		Good fit to OBS. No H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation.
AG-219/12 748006		Good fit to OBS. Moderately strong H-bond with ARG367 and significantly blocks TYR502-HIS479-PHE506 agonist lock formation.	AG-690/40 753951		Good fit to OBS. No H-bond with ARG367 and excellently blocks TYR502-HIS479-

					PHE506 agonist lock formation.
AF-399/42 017398		Excellent fit to OBS. No H-bond with ARG367 and excellently blocks TYR502-HIS479-PHE506 agonist lock formation.	AG-690/40 752975		Good fit to OBS. No H-bond with ARG367 and excellently blocks TYR502-HIS479-PHE506 agonist lock formation.
AK-968/15 359742		Good fit to OBS. No H-bond with ARG367 and excellently blocks TYR502-HIS479-PHE506 agonist lock formation.			

## 3 DISCUSSION

### 3.1 CADD protocol

Functionality of the CADD protocol was demonstrated in theory as several potential ROR $\gamma$ t inverse agonists were discovered based on the promising CADD scores and visual screening. Even though the project was considered successful, some issues were faced. The overall project methodologies with goods and limitations are discussed herein.

The selected docking (PLANTS) and scoring (Panther/ShaEP) methods yielded a correlation coefficient  $R^2$  value of 0.33 for 5NTW model and 0.27 for 5VB6 model. This is not considered a high correlation value but it could be explained by the heterogenous nature of the activity data as it originated from multiple laboratory measurements. This issue was confirmed as the activity data (ROR $\gamma$ t inhibition) of the ChEMBL compounds was acquired by several different experiments. Indeed, there was a lot of dispersion reported among the  $IC_{50}$  values for the same ROR $\gamma$ t-inverse agonists and the  $IC_{50}$  values typically varied multi-fold for the same compound (Gaulton et al., 2012). A better correlation would probably have been achieved if the active compounds would have been selected more systematically. However, this would have resulted in the exclusion of several active compounds and thus decreasing the amount of utilizable data. Considering heterogeneous data origin, the selected docking- and scoring methods were demonstrated to sufficiently predict compound activity.

NIB-model optimization was considered successful. As seen in the figure 8, BR20 got higher generation by generation proving the NIB-models improved with BR-NiB. However, 5VB6 models were not optimized as decently as 5NTW models. Especially, 5VB6 model with the lowest-ranked active compounds improved only by eight generations which is a modest number when compared to other models. No comprehensive explanation was identified for this issue.

BR-NiB was validated as a functional method in the CADD protocol as demonstrated by the remarkable improvement of the optimized NIB-models to discern active compounds from inactive ones (figure 9). However, one rather impractical feature of BR-NiB was identified, as it automatically halts the program when BR20 do not increase in the next generation. This may not always serve the user. Here, BR-NiB was manually rerun a total of three times (once for both of the 5NTW models and once for the 5VB6 model with the lowest-ranked active compounds) after the first decline in BR20 value but still the final

obtained BR20 value eventually got higher. Such a setting decreases the amount of time required for one run but a more practical approach would be to run the program until to the very last generation/NIB point and select the NIB-model with the highest BR20 score.

Ligand-based drug design screen with SDFCONF was determined being successful. The final settings for SDFCONF-screen (figure 10) were manually and experimentally selected and excluded 87 % of the Specs compounds while only around 30-40 % of the ChEMBL compounds were excluded with the same settings. Additionally, the 30-40 % discarded active compounds were mostly the least potent compounds showing the highest IC50 values (IC50 values were obtained from ChEMBL database). This demonstrated the capacity of SDFCONF-screen to separate compounds entailing characteristics of especially potent ROR $\gamma$ t inverse agonism.

In the final screen, the selected high logP threshold value of 5.5 was justified by highly lipophilic LBD as indicated in the generated NIB-models (figure 12). The final screen reduced the number of remaining compounds to 230 which were looked upon in the visualization window of Maestro. The most promising compounds were noticed to be the compounds entailing the highest MM/GBSA scores. This is an expected outcome since ROR $\gamma$ t is a dynamic protein and, as such, molecular dynamics approach with MM/GBSA should provide accurate predictions. Consequently, compounds with the lowest MM/GBSA scores were excluded and the top 25 compounds for each model (totaling one hundred compounds) were selected for the visual screening. The most promising inverse agonist candidates were manually selected by visual screening from those compounds. The final screening was considered successful as the number of promising compounds was reduced to a reasonable number of 29.

Each of the 29 compounds has multiple benzene rings but also hydrophobic and hydrophilic areas (table 2). Almost every compound has a highly hydrophilic end. The most common functional groups include, in addition to benzene group, amine-, ketone and ether. Additionally, some compounds contained halogen-, sulfide- and nitro side groups. However, as each of the compounds were considered rather unique, no single new structural component responsible for inverse agonism was identified.

The docked 29 selected compounds were evaluated with maestro visualization tool to predict which of the compounds appeared the most promising inverse agonists (table 2). A total of nine compounds were estimated as very promising which included compounds AN-648/15596220, AN-989/41838307, AO-081/15386326, AG-690/40753951, AF-399/42017398, AG-690/40752975, AK-968/15359742, AK-968/15253414 and AF-399/41318865.

Compounds AK-968/15253414 and AF-399/41318865 form an H-bond with ARG367 but also show excellent OBS fit. Both compounds also demonstrate moderate or good agonist lock disruption. Other compounds were shown to have either good or excellent OBS fit or agonist lock disruption. AF-399/42017398 was the only compound that has both excellent OBS fit and agonist lock disruption and hence it was considered the most promising inverse agonist. Additionally, especially AK-968/15253414 and AF-399/41318865 are expected to show robust inverse agonism response as they form an H-bond with ARG367 in addition to good OBS fit and agonist lock disruption.

### **3.2 Limitations of the study**

General limitations and potential continuation of the study topic are discussed herein. First, the background of the study is discussed and this is summarized with considerations.

ROR $\gamma$ t is an extremely complex protein which interacts with numerous biological machineries and targets. This comes with two main challenges in respect to drug design which can be summarized into words ‘specificity’ and ‘complexity’ (Huang et al., 2020; GeneCards, 2021). First of all, physiological effects of ROR $\gamma$ t are versatile and as such not specific. On the other hand, it is a complex and dynamic protein and as such a difficult drug target. Additionally, there are many different biological mechanisms that modify ROR $\gamma$ t function, including regulatory effects of HD-BS SUMOylation (He et al., 2018). These two challenges could be overcome by conducting more research on ROR $\gamma$ t.

In this study, novel potential orthosteric ROR $\gamma$ t inverse agonists were discovered. However, the obtained potential inverse agonists were not tested for inhibition and hence the results were not confirmed in practice. Additionally, this study could have been broadened to include other inverse agonist binding mechanisms for solving the issue with unspecificity. As previously mentioned, there are other binding sites such as ABS and HD-BS demonstrated in figure 4 (Huang et al., 2020). Several ABS-targeting inhibitors already exist but no HD-BS-targeting drugs are known as the binding site was just discovered and its biological mechanism is largely unknown. Theoretically, it is possible to develop HD-targeting drugs for minimizing off-target effects. However, its druggability remains a challenge unless otherwise proven.

Another approach for minimizing off-target effects could be achieved by dual inhibition. As orthosteric- and allosteric binding pockets are constantly interacting, it is hypothesized whether combinatory drug

treatment could be possible. In such case, both ABS and OBS would be stabilized with bitopic inhibitors, which show binding affinity towards both the binding sites. Indeed, such dual inhibition has been demonstrated but more research would be needed to discover its potential in practice (Meijer et al., 2021). The same outcome should be achieved by administering orthosteric- and allosteric inhibitors separately to the patient at once. Here, an experimental study could have been conducted in such a way that known potent allosteric inhibitors would have been administered together with each herein discovered potential orthosteric inverse agonist for discovering potential dual inhibition effects.

ABS is specifically expressed in ROR $\gamma$ t and as such is a more specific binding site than OBS (Huang et al., 2020). This is since OBS is similar among ROR family receptors such as ROR $\alpha$ , ROR $\beta$ , and ROR $\gamma$  increasing off-target effects. Consequently, targeting ROR $\gamma$ t could also show effects on other ROR receptors increasing the adverse effects. This implies ABS could be a more druggable binding site than OBS. However, OBS was selected here for the main focus due to its better characterization as more studies have been conducted on it. It has also larger binding area suggesting that designing novel inverse agonists for OBS would be an easier task over ABS.

The complexity issue with ROR $\gamma$ t could have been partially overcome by conducting a more robust CADD protocol with comprehensive molecular dynamics modelling. Several ROR $\gamma$ t structures could have been utilized in this study in addition to a larger number of inactive compounds. However, such approach would be rather time-consuming and such workload would exceed the requirement for Master's thesis project.

In conclusion, ROR $\gamma$ t inhibitors, especially inverse agonists, have been shown as promising agents for treating chronic inflammatory diseases. However, as ROR $\gamma$ t inhibition comes with off-target effects, adverse effects are likely to follow. The most common approach to increase specificity is achieved with highly potent and selective compounds (compounds with very low IC<sub>50</sub> values). However, this might not always provide satisfactory results. More specific effects could also be achieved by targeting ABS or HD-BS. Targeting ABS or HD-BS either separately or combinatory should provide more specific effects on ROR $\gamma$ t with less off-target effects. Dual inhibition of both OBS and ABS could also increase specificity and potency of drug treatment (Meijer et al., 2021). Due to the potential adverse effects of ROR $\gamma$ t inhibition, systematic administration of inverse agonists might be challenging. Instead, topical administration of ROR $\gamma$ t inverse agonists on the skin for treating psoriasis, for example, could provide a more specific approach for treating inflammatory conditions with less adverse effects. With these aspects in

mind, more specific treatment for chronic inflammatory diseases by ROR $\gamma$ t inhibition could be designed but it would require an immense amount of research due to the challenges mentioned above.

### **3.3 Significance of the study**

New information of the structural state of inverse agonists was produced as the arsenal of known ROR $\gamma$ t potential inverse agonists was expanded by 29. These compounds could be used for treating chronic inflammatory diseases if shown high potency and efficiency without significant adverse effects in practice. However, this is not probable as ROR $\gamma$ t inhibition was not tested in practice and also several ROR $\gamma$ t inverse agonists have previously failed as a drug treatment. Additionally, the developed CADD protocol with two 'in house' methods was shown promising for discovering ROR $\gamma$ t inverse agonists and it could be applied for similar drug discovery projects in the future.

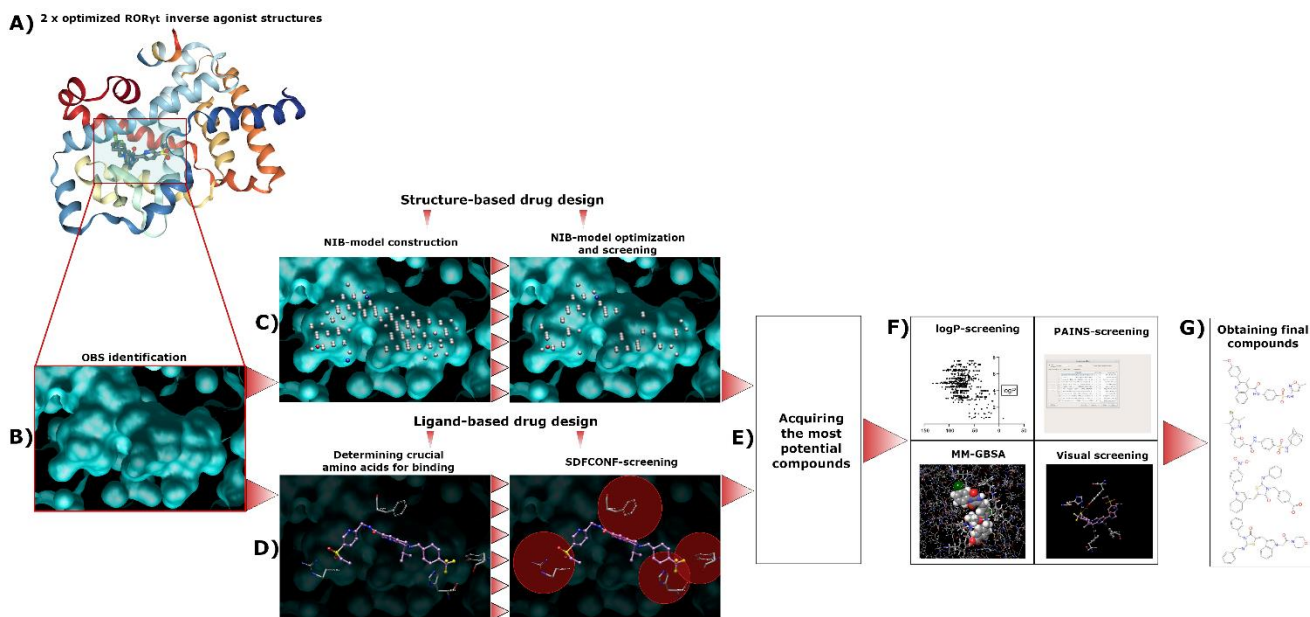
### **3.4 Conclusions**

The study expanded the arsenal of known, potential ROR $\gamma$ t inhibitors for designing novel treatments for chronic inflammatory diseases. The developed CADD protocol was demonstrated promising for drug search.



## 4 MATERIALS AND METHODS

The CADD protocol was conducted as described in figure 11. Both structure- and ligand-based drug design protocols were used in addition to four other CADD approaches (logP-, PAINS-, MM-GBSA- and visual screening). This was followed by selection of the most promising inverse agonists. Detailed description of the methodology is provided herein with all the relevant programs and settings included.



**Figure 11. CADD protocol for discovery of novel potential ROR $\gamma$ t inverse agonists.** A) Optimization of two selected ROR $\gamma$ t inverse agonism structures B) Identification of the OBS C) Structure-based drug design; screening based on the optimized NIB-models of OBS D) Ligand-based drug design; SDFCONF-based pharmacophore screening based on the model constructed by known inverse agonists E) Obtaining the highest scored, most promising compounds from Specs compound library F) Final screening. Compounds presenting logP > 5.5 or PAINS-activity were excluded. The remaining compounds were screened with MM-GBSA to obtain the highest ranked molecules which were visually screened for selecting the most promising compounds. G) Obtaining the promising inverse agonists.

### 4.1 Acquiring proteins and ligands and their preparation

Protein structures were acquired from the Protein Data Bank (PDB) and ligands from and Specs database (Specs, The Netherlands; [www.specs.net](http://www.specs.net); 10 mg compound library). Both proteins and ligands were processed for optimizing the CADD process. Details are provided in this chapter.

#### ***4.1.1 Obtaining the binding site models***

All the available 99 X-ray structures of ROR $\gamma$ t were obtained from PDB at the time of the search (18 July 2019). The selection criteria for the two protein structures: (1) Must have a full inverse agonist bound to OBS. (2) Structures possessing resolutions above 2.3 Å (the lesser is better quality) were excluded and the R-Value Free could not exceed the resolution/10 by more than 0.05. Additionally, R-free had to be approximately equal with the R-value (The Protein Databank; R-value and R-free). (3) The structures could not miss any structural data (such as amino acids).

The structures fulfilling these conditions were superimposed by chain A and analyzed thoroughly with molecular modeling tools *bodil 3D* and *Coot* to determine two most suitable ROR $\gamma$ t structures for the VS (Lehtonen et al., 2004; Emsley et al., 2010). Eventually, structures named as 5NTW (Kallen et al., 2017) and 5VB6 (Li et al., 2017) in PDB were concluded to be the two representable models for ROR $\gamma$ t inverse agonism for VS. 5NTW was selected based on the great resolution and its structure was identified being almost identical with most of the other full inverse agonism structures hence presenting the average conformation state of ROR $\gamma$ t full inverse agonism.

On the other hand, 5VB6 was selected based on the good resolution and its exceptional conformation state of the LBD which is explained by the unique X-ray crystallography process. As a part of the crystallography process, ROR $\gamma$ t was covalently tethered with a cofactor peptide stabilizing its structure thermodynamically providing higher conformational flexibility. Hence, such a conformational state of ROR $\gamma$ t is emphasized in 5VB6 that may not be feasible to crystallize in regular conditions providing new atomic insight for the ROR $\gamma$ t inverse agonism (Li et al., 2017).

#### ***4.1.2 Proteins preparation***

5NTW and 5VB6 structures were downloaded as PDB format and prepared with Protein Preparation Wizard included in Maestro 2018-1 (Maestro, Schrödinger, LLC, New York, NY, 2018) with the following options: (1) import and process: delete waters, preprocess; (2) review and modify: delete chain B, C and D (only for 5VB6), delete A: NA601 (sodium; only for 5VB6). Option ‘Create zero-order bonds to metals’ was deselected; options ‘Convert selenomethiones to methiones’, ‘Fill in missing side chains using Prime’ and ‘Fill in missing loops using Prime’ were selected; (3) refine: use PROPKA pH: set 7.4 +/- 0, optimize. H<sub>2</sub>O:s with < 3 hydrogen-bonds were deleted. (4) Restrained minimization: hydrogens

only. Conformations of LEU324 and CYS393 (only for 5NTW) were changed to the other state not suggested by the default settings. Default settings were otherwise employed. (5) Ligands 98N and 927 were removed; 5NTW and 5VB6 were exported as MOL2 format.

#### ***4.1.3 Acquiring active- and inactive compounds***

A set of ROR $\gamma$ t inverse agonists were acquired from ChEMBL-database with the target ID of ChEMBL1741186 (10 June 2019) (Gaulton et al., 2012). Only those full orthosteric inverse agonists with a reported IC<sub>50</sub> value (marked as =) which had traceable data origin were included. Whether several activity measurements existed for the same compound, the most potent IC<sub>50</sub> value was selected. The selected 191 active and unique compounds were downloaded as SMILES format (Simplified Molecular-Input Line-Entry System) and IC<sub>50</sub> values were transformed to a logarithmic scale (pIC<sub>50</sub>) for producing linear reference values. ChEMBL molecules were converted to 3D structures with LIGPREP (LigPrep, Schrödinger, LLC, New York, NY, 2018).

Inactive compounds were obtained by downloading all the available molecules from Specs as MOL2 format yielding totally around 170,000 unique compounds (28 June 2019) for VS. They served as decoy compounds in VS but also as a data set for discovering ROR $\gamma$ t inverse agonists.

#### ***4.1.4 Ligand preparation***

All the molecules from ChEMBL and Specs were processed with LIGPREP (LigPrep, Schrödinger, LLC, New York, NY, 2018) to generate possible tautomers and enantiomers at OPLS3 charges (OPLS 2005 force field) at pH 7.4 (Harder et al., 2016). Next, ten low-energy conformations were generated for each of the compounds with ConfGen (ConfGen, Schrödinger, LLC, New York, NY, 2018) (Watts et al., 2010). The prepared ligands were converted to MOL2 format with MOL2CONVERT in Maestro. Eventually 4,400 ChEMBL compounds and 1,900,000 Specs compounds were generated.

## 4.2 Binding site identification

The OBS of both the structures 5NTW and 5VB6 were identified and looked upon with bodil 3D. The electron density map resolution in the OBS among the structures were confirmed to be proper with Coot.

## 4.3 Structure-based drug design

The two selected protein structures were utilized for structure-based drug design approaches. Negative image -based (NIB)-models were generated and furthermore optimized for estimating binding affinity of the compounds of interest. The optimal docking- and scoring approach was discovered by calculating correlations.

### 4.3.1 NIB-model generation with panther

NIB-models of the LBD of prepared 5NTW and 5VB6 were constructed with Panther -multipurpose docking tool (version 0.18.19) (Niinivehmas et al., 2015). The NIB-models were scrutinized with bodil 3D to determine whether the models were correctly generated. In the end, three neutral NIB-points were manually changed to two negative charges and one positive charge in both 5NTW and 5VB6 NIB-models.

NIBs are models that consider shape and charge of OBS, which are needed as a part of the scoring process to estimate binding affinity of each compound. NIB-models were optimized as explained later in this chapter.

### 4.3.2 Evaluation and validation of the docking- and scoring methods

A set of molecular docking- and scoring methods were executed to discover the most optimal computational methods for identification of active ligands in VS. Docking means predicting binding of the compound of interest to the binding site. The prediction accuracy was determined by calculating correlation coefficient  $R^2$  for each of the conducted method.

The generated 4,400 active compounds were docked to the prepared LBD of 5NTW and 5VB6 with various docking tools. The docking tools included Glide (Glide, Schrödinger, LLC, New York, NY,

2018), Prime (Prime, Schrödinger, LLC, New York, NY, 2018), PLANTS (PLANTS<sub>PLP</sub>; version 1.2), and the docking pose -optimization tool of molecular superimposition/similarity analysis program ShaEP (version 1.3.0) (Korb et al., 2009; Vainio et al., 2009). ShaEP docking was based on the NIB-models generated by Panther.

Subsequently, several scoring methods were applied to yield VS-based predictions of the binding affinity for each of the docked compound. The docking tools Glide, Prime, PLANTS and ShaEP were applied to calculate docking scores. In addition, ShaEP was applied for scoring the active compounds docked with PLANTS yielding a total of five methods for predicting binding affinity (table 3). ShaEP scoring was based on the constructed NIB-models by Panther.

Next, the best docking score for each individual compound was extracted and proportionated with the highest pIC<sub>50</sub> value reported for that active compound in ChEMBL to yield a correlation coefficient. The aim was to discover the best suitable predictive docking- and scoring method for the VS which was evaluated based on Pearson correlation coefficient (R<sup>2</sup>) score. PLANTS was selected as the docking tool and ShaEP as the scoring tool for VS since the best correlation was achieved with these methods.

**Table 3. The five experimentally tested docking- and scoring methods presented as A) – E).** The methods were tested to select the method that had the most accurate VS-based prediction of the binding affinity of the active compounds. The VS-based prediction was evaluated as Pearson correlation coefficient R<sup>2</sup> scores.

	DOCKING	SCORING
A)	Glide	Glide
B)	Prime	Prime
C)	Panther/ShaEP	Panther/ShaEP
D)	PLANTS	PLANTS
E)	PLANTS	Panther/ShaEP

#### 4.3.3 Docking and scoring

All the 1,900,000 Specs- and 4,400 ChEMBL compounds were docked to the prepared ligand-free structures 5NTW and 5VB6 with PLANTS. The docking scores were calculated for 5NTW and 5VB6 separately with ShaEP utilizing the generated NIB-models. Shape/electrostatic complementary scores were

obtained for each of the docked ligand 3D enantiomers and tautomers including their conformers. Only the highest ShaEP score was obtained for each of the compounds in VS.

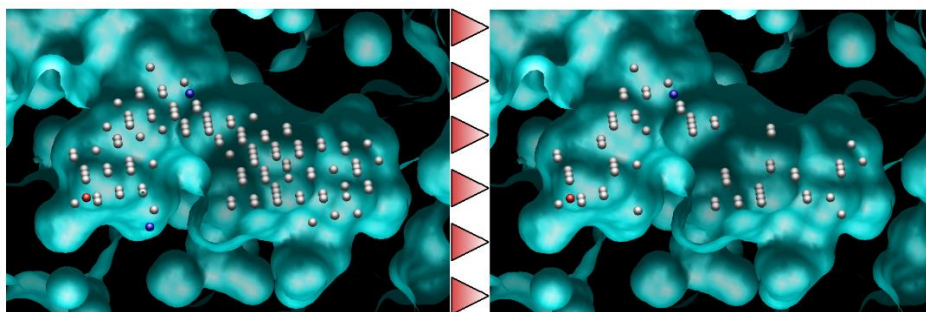
#### ***4.3.4 Optimization of the final NIB-models***

The generated NIB-models of LBD of 5NTW (184 NIB-points) and 5VB6 (198 NIB-points) were optimized by BR-NiB approach with the ‘in-house’ Brute Force Image-Based Rescoring -software to improve the capacity of ShaEP to discern active compounds from a random set of compounds based on the utilized NIB-models as seen in figure 12 (Kurkinen et al., Submitted). Here, the capacity was estimated based on BR20 value as the software was commanded to optimize BR20 values (Truchon et al., 2007). BR20 values can be considered as a measurement for evaluating the ability of the model to correctly discern active compounds from inactive compounds.

BR-NiB utilizes Rocker and ShaEP to optimize the NIB-models to enhance the capacity of ShaEP to score the active compounds (= ChEMBL compounds) higher on average compared to the decoy compounds (= Specs compounds) improving the capacity of the NIB-model to rank higher compounds possessing RORyt inverse agonist -like chemical structures (Lätti et al., 2016; Niinivehmas et al., 2015). BR-NiB deletes individual NIB-points one by one (the NIB-point that improves BR20 value the most) until the maximal BR20 has been reached. Here, BR-NiB halted optimization process if deletion of any NIB-point did not improve enrichment demonstrated as BR20.

Randomly selected one fourth of the generated around 1,900,000 ( $\approx$  475,000) compounds from Specs were chosen as the set of decoy compounds and all the generated 4,400 ChEMBL compounds were selected as the active compounds. This resulted in around 475,000 decoy compounds and 4,400 active compounds (incl. 10x conformations, enantiomers and tautomers) for NIB-model optimization process. Optimized NIB-models were generated for both 5NTW and 5VB6 structures.

Additionally, another two NIB-models were generated by excluding the highest-ranked ChEMBL molecules from the set of active compounds. Those ChEMBL compounds that ranked among the best 1 % of decoy compounds based on the final optimized NIB-model were separately excluded for models 5NTW and 5VB6. Then, NIB-models were optimized second time with the new set of active compounds. Subsequently, additional two NIB-models were separately generated for 5NTW and 5VB6 resulting in four optimized NIB-models as in total.



**Figure 12. NIB-model optimization with BR-NiB software.** Optimization was executed by deleting the excessive NIB-points to improve the scoring and enrichment of those compounds entailing characteristics of active compounds.

#### 4.4 Ligand-based drug design

The amino acids of ROR $\gamma$ t being responsible for ligand binding were identified based on the literature. The most relevant amino acids for ligand binding include HIS323, LEU324, ARG367, PHE377, PHE378, PHE388, HIS479 AND HIS502. Especially, hydrophobic-/stacking effects with LEU324 or PHE377 stabilize inverse agonism conformation as does the disruption of the H-bond formation between HIS479 and TYR502 by a ligand interaction. The latter could be achieved most effectively by formation of a hydrogen bond (H-bond) with either HIS479 or TYR502. Additionally, H<sub>2</sub>O (WAT770) in LBD may participate in ligand binding.

Consequently, these eight amino acids and the H<sub>2</sub>O (nine coordinate points) were used for construction of pharmacophore points for ligand-based drug design approach with the in-house algorithm SDFCONF (Lätti et al. manuscript in preparation). SDFCONF-based pharmacophore points screening as a part of CADD process has been demonstrated before (Rauhämäki et al., 2018). The selected 3D coordinates of the amino acids of the prepared 5NTW and 5VB6 were utilized to construct a pharmacophore model (figure 13). The 3D coordinates were carefully selected to yield a set of pharmacophore filter points for each of the amino acids to exclude those docked compounds that are not in proximity of the selected amino acids and hence probably do not interact with those amino acids.

Five important binding regions were identified in LBD of ROR $\gamma$ t. Next, such pharmacophore filter points were generated with SDFCONF to exclude the compounds not present within the defined range among the required regions.

The first region (1.) included HIS479 and TYR502 which can be either H-bond acceptors or donors. Two pharmacophore filter points were created to set a criterion that hydrophilic groups must be present near HIS479 or TYR502. The screen was executed by determination of possible H-bond acceptors or donors presented as heavy atoms N, O, S, F, Cl, Br or I near HIS479 or TYR502.

The second region (2.) included LEU324 and PHE388 which are important for forming hydrophobic- or stacking effect with ligands. Compounds not possessing any aromatic atom near LEU324 or PHE388 were excluded. Similarly, the third region (3.) included PHE378 and compounds not possessing any aromatic atom near PHE378 were excluded.

The fourth region (4.) included amino acids HIS323 and PHE377 but also WAT770 which can be either H-bond acceptors or donors. Pharmacophore filter points were generated for each of them to exclude those compounds that did not show hydrophilic groups (simplified as N, O, S, F, Cl, Br or I) near HIS323, PHE377 or WAT770.

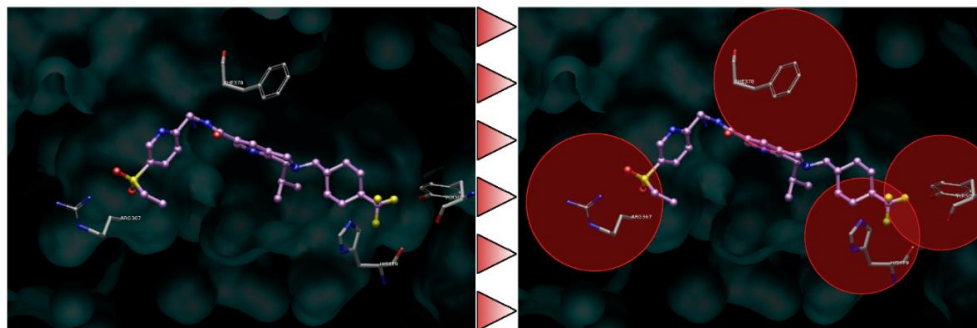
The fifth region (5.) included amino acid ARG367 which acts as a H-bond donor. A pharmacophore filter point was created to set a criterion that hydrophilic groups must be present near ARG367 demonstrated as heavy atoms N, O, S, F, Cl, Br or I near ARG367.

Next, SDFCONF-based pharmacophore screening was specified by setting the following requirements: the compound passed the screening if any of its tautomer, enantiomer or conformer fulfilled the criterions 1, 2, 3 and 4 or the criterions 2, 3, 4 and 5 but also had any heavy atom (any other than H) within 4Å radius from either TYR502 or HIS479.

Then, the 3D coordinates of individual heavy atoms of the docked compounds were obtained for each of the generated tautomers, enantiomers and their conformations. The 3D coordinates of all the compounds were then compared against the selected 3D coordinates of the amino acids HIS323, LEU324, ARG367, PHE377, PHE378, PHE388, HIS479 AND HIS502. The 3D coordinate of the center of oxygen of WAT770 was included for the pharmacophore model. SDFCONF-screening was executed separately to all the four groups of Specs compounds that were ranked among the best 1 % based on the ShaEP-score. SDFCONF-screening was experimentally validated by adjusting the radiuses of the nine selected 3D coordinates and determining the optimal radiuses which would discard as many decoy compounds as possible with relatively low number of active compounds being discarded. No more than 50 % of the



active compounds discarded by SDFCONF-screen was accepted. Eventually, SDFCONF-based pharmacophore screening radii were determined to be 4.0 Å for criterion 1, 4.5 Å for criterion 2, 5.2 Å for criterion 3, 4.0 Å for criterion 4 and 4.0 Å for criterion 5.



**Figure 13. SDFCONF pharmacophore model construction.** The amino acids being crucial for ligand binding were first identified and the corresponding 3D coordinates were selected for each of the amino acids. The model was used to exclude those docked compounds that do not present desired chemical groups within the selected radii (marked as red circles on the right) from the amino acids and hence do not likely form molecular interaction with those amino acids.

#### 4.5 Acquiring compounds from structure- and ligand-based drug design screen

All the 1,900,000 compounds, including generated enantiomers, tautomers and conformations, from Specs 10 mg compound collection (retrieved: 28 June 2019; [www.specs.net](http://www.specs.net)) were docked with PLANTS to the prepared LBD of 5NTW and 5VB6 and scored with ShaEP based on the four optimized NIB-models separately. 1 % of the highest ranked compounds were included from each of the four structure-based drug design models. The remaining compounds were SDFCONF-screened to exclude those compounds that do not meet the required structural characteristics set by pharmacophore points. Duplicates were discarded and a set of potentially activate compounds were obtained.

#### 4.6 Final screening

In the final CADD step, logP values were calculated with Maestro for the remaining compounds. The compounds possessing logP values of  $> 5.5$  were excluded as highly lipophilic compounds have poor solubility. PAINS compounds were removed with PAINS-filtering (PAINS1, PAINS2, PAINS3) with Canvas program (Canvas, Schrödinger, LLC, New York, NY, 2018) (Duan et al., 2010). Additionally, the compounds possessing reactive side groups such as aldehyde groups were excluded.

Free energy binding of the remaining docked compounds was calculated with Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) -tool in Maestro to exclude the compounds presenting weaker predicted binding energies (Genheden et al., 2015; Virtanen et al., 2015). MM/GBSA were generated individually to the compounds docked with PLANTS to 5NTW and 5VB6 structures. The following options were used: (1) VSGB solvation model and OPLS3 force field were applied; (2) use input partial charges ON, use implicit membrane OFF; (3) distance from ligand = 4 Å, sampling method: minimize.

The compounds presenting MM/GBSA-binding energies (MMGBSA dG Bind) of > -95 were excluded. The remaining molecules were visually evaluated with Maestro based on the generated MM-GBSA models. In visual selection, the goodness of fit of the docked compounds to the OBS was evaluated and compared to other known inverse agonist structures from PDB. Additionally, visual evaluation was conducted whether a given compound would disrupt TYR502-HIS479-PHE506 agonist lock formation and interaction with ARG367, both of which are known to be important upon inverse agonist binding. Each of these three properties were ranked from one to three to yield semi-quantitative evaluation of the goodness of fit.

#### **4.7 Obtaining final compounds**

The most promising inverse agonists were selected based on the overall CADD protocol and the visual screening. The data of the selected compounds was obtained for considerations.

## **5 ACKNOWLEDGEMENTS**

The thesis was executed in MedChem research group in the university of Turku. I sincerely thank my supervisors Olli Pentikäinen and Sanna Niinivehmas for guiding through the thesis process. Additionally, I acknowledge each MedChem group member for providing me advice and tools required for this work. Obviously, this thesis would not have been possible without your help. Finally, I thank all the close people for supporting me all these years.

## 6 ABBREVIATIONS

3D	Three-dimensional
ABS	Allosteric Binding Site
BR20	Boltzmann-Enhanced Discrimination of Receiver Operating Characteristic 20
BR-NiB	Brute Force Negative Image-Based Optimization
CADD	Computer-aided drug design
DBD	DNA-Binding Domain
HD	Hinge Domain
HD-BS	Hinge Domain Binding Site
H-bond	Hydrogen Bond
IC <sub>50</sub>	Half Maximal Inhibitory Concentration
LBD	Ligand Binding Domain
MM/GBSA	Molecular Mechanics/Generalized Born Surface Area
NIB	Negative Image-Based
OBS	Orthosteric Binding Site
QSAR	Quantitative Structure-Activity Relationship
ROR $\gamma$ t	Retinoic Acid-related Orphan Receptor Gamma t
VS	Virtual Screening

## 7 REFERENCES

- Acharya C, Coop A, Polli JE, Mackerell AD Jr. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr Comput Aided Drug Des.* 2011;7(1):10-22.
- Amy C. Anderson. The Process of Structure-Based Drug Design, *Chemistry & Biology*, Volume 10, Issue 9, 2003, Pages 787-797, ISSN 1074-5521.
- Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. Vol 28.; 2000.
- Billon C, Murray MH, Avdagic A, Burris TP. ROR regulates the NLRP3 inflammasome. *Journal of Biological Chemistry.* 2019;294(1):10-19.
- Castro G, Liu X, Ngo K, et al. ROR $\gamma$ t and ROR $\alpha$  signature genes in human Th17 cells. *PLoS One.* 2017;12(8):e0181868. Published 2017 Aug 1.
- Chang MR, Lyda B, Kamenecka TM, Griffin PR. Pharmacologic repression of retinoic acid receptor-related orphan nuclear receptor  $\gamma$  is therapeutic in the collagen-induced arthritis experimental model. *Arthritis and Rheumatology.* 2014;66(3):579-588.
- ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2000 Feb 29 - . Identifier NCT03004846, A Study of GSK2981278 Ointment in Subjects With Plaque Psoriasis; 2016 Dec 29. Available from: <https://clinicaltrials.gov/ct2/show/NCT03004846>.
- Daisy P, Singh SK, Vijayalakshmi P, Selvaraj C, Rajalakshmi M, Suveena S. A database for the predicted pharmacophoric features of medicinal compounds. *Bioinformation.* 2011;6(4):167-168. Published 2011 May 7.
- Duan J, Dixon SL, Lowrie JF, Sherman W. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *Journal of Molecular Graphics and Modelling.* 2010.
- Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography.* 2010;66(4):486-501.
- Fukase Y, Sato A, Tomata Y, et al. Identification of novel quinazolinone derivatives as ROR $\gamma$ t inverse agonist. *Bioorganic and Medicinal Chemistry.* 2018.
- Gaulton A, Bellis LJ, Bento AP, et al. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Research.* 2012;40(D1).
- GeneCards – the human gene database ([www.genecards.org](http://www.genecards.org)). RORC Gene - RAR Related Orphan Receptor C. Accessed October 27, 2021. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=RORC>.
- Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery.* 2015;10(5):449-461.

Harder E, Damm W, Maple J, et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation*. 2016;12(1):281-296.

He Z, Zhang J, Huang Z, et al. Sumoylation of ROR $\gamma$ t regulates TH17 differentiation and thymocyte development. *Nat Commun*. 2018;9(1):4870. Published 2018 Nov 19.

Huang M, Bolin S, Miller H, Ng HL. ROR $\gamma$  Structural Plasticity and Druggability. *Int J Mol Sci*. 2020;21(15):5329. Published 2020 Jul 27.

Hu X, Wang Y, Hao LY, Liu X, Lesch CA, Sanchez BM, Wendling JM, Morgan RW, Aicher TD, Carter LL, Toogood PL, Glick GD. Sterol metabolism controls T(H)17 differentiation by generating endogenous ROR $\gamma$  agonists. *Nat Chem Biol*. 2015 Feb;11(2):141-7.

Igaki K, Nakamura Y, Komoike Y, et al. Pharmacological Evaluation of TAK-828F, a Novel Orally Available ROR $\gamma$ t Inverse Agonist, on Murine Colitis Model. *Inflammation*. 2019;42(1):91-102.

Jetten AM. Retinoid-related orphan receptors (RORs): critical roles in development, immunity, circadian rhythm, and cellular metabolism. *Nuclear receptor signaling*. 2009;7.

Jetten AM, Takeda Y, Slominski A, Kang HS. Retinoic acid-related orphan receptor  $\gamma$  (ROR $\gamma$ ): Connecting sterol metabolism to regulation of the immune system and autoimmune disease. *Current Opinion in Toxicology*. 2018;8:66-80.

Kallen J, Izaac A, Be C, et al. Structural States of ROR gamma t: X-ray Elucidation of Molecular Mechanisms and Binding Interactions for Natural and Synthetic Compounds. *ChemMedChem*. 2017;12:1014-1021. 1014–1021.

Kelly A Berg, William P Clarke, Making Sense of Pharmacology: Inverse Agonism and Functional Selectivity, *International Journal of Neuropsychopharmacology*, Volume 21, Issue 10, October 2018, Pages 962–977.

Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC (March 1958). "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis". *Nature*. 181 (4610): 662–66.

Korb O, Stützle T, Exner TE. Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. *Journal of Chemical Information and Modeling*. 2009;49(1):84-96.

Kumar N, Lyda B, Chang MR, et al. Identification of SR2211: A potent synthetic ROR $\gamma$ -selective modulator. *ACS Chemical Biology*. 2012;7(4):672-677.

Kurkinen Sami T, Lehtonen Jukka V, Pentikäinen Olli T, Postila Pekka A. Optimization of cavity-based negative images to boost docking enrichment in virtual screening. Submitted.

Lao C, Zhou X, Chen H, Wei F, Huang Z, Bai C. 5,6,7,8-Tetrahydrobenzo[4,5]thieno[2,3-d]pyrimidine derivatives as inhibitors of full-length ROR $\gamma$ t. *Bioorg Chem*. 2019 Sep;90:103077.

Lätti S, Niinivehmas S, Pentikäinen OT. Rocker: Open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization. *Journal of Cheminformatics*. 2016;8(1).

Lehtonen J v, Still D-J, Rantanen V-V, et al. BODIL: A Molecular Modeling Environment for Structure-Function Analysis and Drug Design. <http://www.abo.fi/fak/mnf/bkf/research/johnson/bodil.html>. 2004

Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X. Molecular dynamics simulations and novel drug discovery. *Expert Opin Drug Discov.* 2018 Jan;13(1):23-37.

Li X, Anderson M, Collin D, et al. Structural studies unravel the active conformation of apo ROR gamma t nuclear receptor and a common inverse agonism of two diverse classes of ROR gamma t inhibitors. *J Biol Chem.* 2017;292:11618-11630. 11618–11630.

Meijer FA, Oerlemans GJM, Brunsveld L. Orthosteric and Allosteric Dual Targeting of the Nuclear Receptor ROR $\gamma$ t with a Bitopic Ligand. *ACS Chem Biol.* 2021;16(3):510-519.

Niinivehmas SP, Salokas K, Lätti S, Raunio H, Pentikäinen OT. Ultrafast protein structure-based virtual screening with Panther. *Journal of Computer-Aided Molecular Design.* 2015;29(10):989-1006.

Odilia Osakwe, Chapter 5 - The Significance of Discovery Screening and Structure Optimization Studies, Editor(s): Odilia Osakwe, Syed A.A. Rizvi, *Social Aspects of Drug Discovery, Development and Commercialization*, Academic Press, 2016, Pages 109-128, ISBN 9780128022207.

Rauhämäki S, Postila PA, Lätti S, et al. Discovery of Retinoic Acid-Related Orphan Receptor  $\gamma$ t Inverse Agonists via Docking and Negative Image-Based Screening. *ACS Omega.* 2018;3(6):6259-6266.

Scheepstra M, Leysen S, van Almen GC, et al. Identification of an allosteric binding site for ROR $\gamma$ t inhibition. *Nat Commun.* 2015;6:8833. Published 2015 Dec 7.

Sun N, Yuan C, Ma X, Wang Y, Gu X, Fu W. Molecular mechanism of action of ROR $\gamma$ t agonists and inverse agonists: Insights from molecular dynamics simulation. *Molecules.* 2018;23(12).

The Protein Databank; R-value and R-free; PDB-101 [website]; Available from: <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/r-value-and-r-free>.

The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49:D1 (2021). UniProtKB - P51449 (RORG\_HUMAN). Accessed October 27, 2021. <https://www.uniprot.org/uniprot/P51449>.

Truchon JF, Bayly CI. Evaluating virtual screening methods: Good and bad metrics for the “early recognition” problem. *Journal of Chemical Information and Modeling.* 2007;47(2):488-508.

Unutmas D. RORC2: The master of human Th17 cell programming. *European Journal of Immunology.* 2009;39(6):1452-1455.

Vainio MJ, Puranen JS, Johnson MS. ShaEP: Molecular overlay based on shape and electrostatic potential. *Journal of Chemical Information and Modeling.* 2009;49(2):492-502.

Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design--a review. *Curr Top Med Chem.* 2010;10(1):95-115.

Virtanen SI, Niinivehmas SP, Pentikäinen OT. Case-specific performance of MM-PBSA, MM-GBSA, and SIE in virtual screening. *Journal of Molecular Graphics and Modelling*. 2015.

Watts KS, Dalal P, Murphy RB, Sherman W, Friesner RA, Shelley JC. ConfGen: A conformational search method for efficient generation of bioactive conformers. *Journal of Chemical Information and Modeling*. 2010;50(4):534-546.

Weaver CT, Elson CO, Fouser LA, Kolls JK. The Th17 Pathway and Inflammatory Diseases of the Intestines, Lungs, and Skin. *Annual Review of Pathology: Mechanisms of Disease*. 2013;8(1):477-512.

Xue X, Soroosh P, de Leon-Tabaldo A, et al. Pharmacologic modulation of ROR $\gamma$ t translates to efficacy in preclinical and translational models of psoriasis and inflammatory arthritis. *Scientific Reports*. 2016;6.

Yang J, Sundrud MS, Skepner J, Yamagata T. Targeting Th17 cells in autoimmune diseases. *Trends Pharmacol Sci*. 2014 Oct;35(10):493-500.

Zhang Y, Luo XY, Wu DH, Xu Y. ROR nuclear receptors: Structures, related diseases, and drug discovery. *Acta Pharmacologica Sinica*. 2015;36(1):71-87.

Figures 1 and 2 were created with Microsoft PowerPoint. Figures 3, 4, 5, 6, 10, 11, 12 and 13 were created with GIMP image manipulation program (The GIMP Development Team. (2019). GIMP. Retrieved from <https://www.gimp.org>). 5NTW structure from PDB was obtained for creation of figures 5 and 11. Bodil 3D was used for creation of figures 4, 5, 11, 12 and 13.