



universität  
wien

# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Analysis of potential human pheromones“

verfasst von / submitted by

Isabella Maria Fischer BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of  
Master of Science (MSc)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt /  
degree programme code as it appears on  
the student record sheet:

A 066 862

Studienrichtung lt. Studienblatt /  
degree programme as it appears on  
the student record sheet:

Masterstudium Chemie

Betreut von / Supervisor:

Univ.-Prof. Dr. Christian Friedrich Wilhelm Becker

Mitbetreut von / Co-Supervisor:

## Table of contents

Abbreviations .....	4
Abstract .....	5
Zusammenfassung .....	6
1. Introduction .....	7
1.1 Pheromones .....	7
1.2 Odorant binding proteins (OBP) .....	7
1.3 Lipocalins .....	8
1.4 Olfactory receptors .....	9
1.5 Vomeronasal organ .....	10
1.6 Pheromonic communication in humans (homo sapiens) .....	10
1.6.1 h-SAL .....	13
1.7 Pheromonic communication in pig ( <i>sus scrofa</i> ) .....	14
1.7.1 SAL1 protein: .....	14
1.8 Pheromonic communication in panda ( <i>Ailuropoda melanoleuca</i> ) .....	14
1.8.1 Odorant binding protein 3 .....	15
1.9 Reverse chemical ecology .....	15
1.10 Objective .....	15
2. Materials and Methods .....	17
2.1 Materials .....	17
2.1.1 Chemicals .....	17
2.1.2 Competent cells .....	17
2.1.3 LB and LB plates .....	17
2.1.4 DNA .....	17
2.2 Methods .....	18
2.2.1 Agarose Gel electrophoresis .....	18
2.2.2 Cloning of H-Sal into pET30a .....	18
2.2.3 Mutagenesis .....	19
2.2.4 Protein expression .....	21
2.2.5 SDS-PAGE .....	22
2.2.6 Protein purification .....	23
2.2.7 Protein characterization .....	23
2.2.8 Calculation of the Dissociation constants from the binding curves .....	24
3. Results .....	26

3.1 Theoretical predictions about the H-SAL protein and comparison with OBP3 and SAL1 .....	26
3.2 Cloning of h-Sal into pET30a .....	27
3.3 Protein expression of h-SAL.....	27
3.3.4 Protein purification.....	28
3.3.5 h-SAL protein characterization.....	33
3.4 Protein expression of the h-SAL HEIE Mutant.....	45
3.4.1 Introduction of the HEIE mutation into hSAL.....	45
3.4.2 Protein expression.....	47
3.4.3 h-SAL HEIE mutant protein purification .....	47
3.4.4 Protein characterization .....	48
3.4.5 Fitting of Dissociation constants .....	52
4. Discussion.....	54
4.1 Research question .....	54
4.2 Improvements in h-SAL protein purification.....	54
4.3 Different approaches for introducing the HEIE mutation into h-SAL .....	55
4.4-NPN fluorescence quenching vs. tryptophan quenching .....	55
4.4.1 Theoretical predictions and docking experiments .....	55
4.4.2 Retinol.....	55
4.4.3 $\beta$ -Ionone .....	56
4.4.4 Conclusion.....	56
4.5 h-SAL protein ligands.....	56
4.6 h-SAL protein vs. h-SAL HEIE mutant protein .....	57
4.6.1 Comparison of protein models .....	57
4.6.2 Comparison of dissociation constants.....	57
4.6.3 Conclusion.....	57
4.7 Similarity to SAL1 ( <i>sus scrofa</i> ) and OBP3 ( <i>Ailuropoda melanoleuca</i> ).....	58
4.7.1 Comparison of protein models.....	58
4.7.2 Comparison of ligand binding experiments .....	58
4.7.3 Conclusion.....	58
4.8 Outlook .....	58
5. Supplementary Data.....	59
5.1 h-SAL HEIE protein sequence source.....	59
5.2 Physical and chemical protein parameters by ExPASy .....	59
5.2.1 Human SAL.....	59

5.2.2 Human SAL HEIE mutation .....	60
5.3 Sequencing results .....	60
5.3.1 h-Sal protein sequencing April 4, 2019 .....	60
5.3.2 h-SAL protein HEIE mutation sequencing July 18, 2019.....	61
5.4 Original fluorescence data .....	62
5.4.1 h-SAL protein .....	62
5.4.2 h-SAL protein HEIE mutation.....	69
6. References.....	73
7. Acknowledgement .....	80

## Abbreviations

E-3M2H	(E)-3-methyl-2-hexenoic acid
1-AMA	1-Aminoanthracene
ATP	Adenosine triphosphate
AS	Amino acid
ASOB	Apocrine secretion odour-binding proteins
DNA	Deoxyribonucleic acid
$K_{\text{diss}}$ , $K_{\text{d}}$	Dissociation constant
DDT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
FPLC	Fast protein liquid chromatography
GPCR	G-protein-coupled receptor
IC50	Half-maximal inhibitory concentration
h-SAL	Human salivary lipocalin
pI	Isoelectric point
IPTG	Isopropyl- $\beta$ -D-thiogalactopyranoside
LB	Lysogeny broth
MHC	Major histocompatibility complex
MUP	Major urinary protein
MUPP	Major urinary protein pseudogene
MUSCLE	Multiple Sequence Comparison by Log- Expectation
NCBI	National Center for Biotechnology Information
1-NPN	N-phenyl-1-naphthylamine
OBP	Odorant binding protein
OR	Odorant receptor
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PGP 9.5	Protein gene product 9.5
RNA	Ribonucleic acid
SAL	salivary lipocalin
SDS	Sodium dodecyl sulfate
SDS PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Tris	Tris(hydroxymethyl)aminomethane
TAE buffer	Tris-acetate-EDTA buffer
V1R	Vomer nasal receptor type 1
V2R	Vomer nasal receptor type 2

## Abstract

The question of whether humans communicate with pheromones and what they might look like has been in the focus of research for many years. Previous studies on this topic investigated the expression of pheromone receptors and the presence of a vomeronasal organ, the site where pheromone receptors are found in many other mammals or sought to discover the activity of specific pheromone candidates through behavioural studies.

The purpose of this thesis is to gain information about the structure of possible human pheromones, in case evolution had taken a different course, using the "reverse chemical ecology" approach, that has emerged in recent years and was particularly successful in the search for the pheromones of giant pandas. More specifically, this thesis investigated how different ligands fit into the binding pocket of the "odorant-binding protein" (here called h-SAL, two different sequences were found in literature), that transported pheromones from the glands, where they are synthesised into the environment and from the environment to the odorant-binding receptors in the common ancestor of humans and monkeys. Since the time of these ancestors, 400,000 to 350,000 years ago, this protein is no longer transcribed due to the gene sequence no longer being translated correctly because of a mutation at the splice donor site of the second intron.

Since these different sequences were found, both proteins were studied. The difference between them is four additional amino acids around the splicing site where the mutation occurred. For this project, the shorter gene sequence was ordered as a synthetic gene and the protein was expressed in *E. coli*. The purification was adapted to the specific properties of this protein and binding experiments based on fluorescence quenching with different ligands were performed. Thereafter, the 4 missing amino acids were inserted into the gene sequence by means of PCR and the experiments were repeated.

In the end it could be shown, that none of the tested ligands binds the protein with dissociation constants in the same range as other typical pheromone-odorant binding protein complexes. One reason for this could be that either suitable ligands were not tested, or the gene sequence has changed too much over the period, in which this protein was no longer functional, and the original pheromones can no longer bind.

## Zusammenfassung

Die Frage, ob Menschen mittels Pheromonen kommunizieren und wie diese aussehen könnten, beschäftigt die Forschung schon seit vielen Jahren. Bisherige Studien zu diesem Thema suchten entweder nach dem Pheromonrezeptor und dem Vomeronasalen Organ, dem Ort an dem die Pheromonrezeptoren in vielen anderen Säugetieren zu finden sind, oder versuchten die Aktivität spezieller Pheromonkandidaten durch Verhaltensstudien herauszufinden.

In dieser Arbeit soll Mithilfe der Strategie der „umgekehrten chemischen Ökologie“, die sich in den letzten Jahren besonders durch ihre Erfolge bei der Suche nach den Pheromonen von großen Pandas hervorgetan hat, versucht werden mehr über die Struktur möglicher menschlicher Pheromone herauszufinden. Genauer gesagt, sollte erforscht werden, wie verschiedene Liganden in die Bindetasche des „geruchsbindenden Proteins“ passen, das diese Pheromone im gemeinsamen Vorfahren von Menschen und Affen von der Drüse, in der diese Synthetisiert werden in die Umwelt und von der Umwelt zu den Pheromonrezeptoren transportierte (hier h-SAL genannt, es wurden in der Literatur zwei unterschiedliche Sequenzen gefunden). Seit der Zeit, in der diese Vorfahren lebten, vor 400.000 bis 350.000 Jahren, wird dieses Protein nicht mehr transkribiert, da die Gensequenz durch eine Mutation an der Spleißstelle des zweiten Introns nicht mehr korrekt übersetzt werden konnte.

Da sich die unterschiedliche Proteinsequenzen fanden, wurden beide Proteine bearbeitet. Der Unterschied dabei beträgt vier zusätzliche Aminosäuren rund um die Spleißstelle, bei der die Mutation aufgetreten ist. Für dieses Projekt wurde die kürzere Gensequenz als synthetisches Gen bestellt und das Protein in *E. coli* exprimiert. Es wurde die Aufreinigung an die Eigenschaften dieses Proteins angepasst und Bindungsexperimente, auf der Basis von Fluoreszenzquenching, mit verschiedenen Liganden wurden durchgeführt. Danach wurden, mittels PCR, die vier fehlenden Aminosäuren in die Gensequenz eingefügt und die Experimente wiederholt.

In dieser Arbeit konnte gezeigt werden, dass keiner der getesteten Liganden mit dem Protein eine Bindung eingeht, deren Dissoziationskonstanten in einem Bereich liegt, wie für andere Pheromon-bindende Proteinkomplexe üblich. Ein Grund hierfür könnte sein, dass entweder nicht die passenden Liganden getestet wurden oder sich die Gensequenz in den Jahren, die dieses Protein nicht mehr funktionstüchtig war, durch zufällige Mutation zu stark verändert hat und somit die ursprünglichen Pheromone nicht mehr binden kann.

# 1. Introduction

## 1.1 Pheromones

The name “pheromones” was proposed by Karlson and Lüscher from the Greek words “*pherein*”, which means “to transfer”, and “*hormón*”, which means “to excite”. They define this class as substances, released to the outside by an animal, that induces a specific reaction in an individual of the same species <sup>1</sup>. Pheromones can be composed of one or more chemicals in a blend <sup>2</sup>.

The original definition classified the pheromones into two categories: releaser and primer pheromones. The first induce a specific behaviour while the later influence developmental processes <sup>1</sup>.

On the one hand, releaser pheromones can further be classified according to their function <sup>2</sup>. Some examples include 5-androst-16-en-3-one as the pigs (*sus scrofa*) sex pheromone <sup>3</sup>, (+)-(1R,2S)-grandisol as the main component and (-)-(R)-terpinen-4-ol as the minor component of the bark beetle *Polygraphus punctifrons* aggregation pheromone <sup>4</sup>, (-)- $\alpha$ -pinene as the alarm pheromone of the aphid *Megoura viciae* <sup>5</sup>, glutamic acid as the host-marking pheromone of the fruit fly *Ceratitis rosa* <sup>6</sup> and (Z)-pentacos-12-ene as the oviposition-detering pheromone of the coccinellid beetle *Cheilomenes sexmaculata* <sup>7</sup>.

Primer pheromones on the other hand, typically affect endocrine or neuroendocrine systems. Ethyl oleate, for example, is a primer pheromone of the honey bee (*apis mellifera*), which delays the onset of foraging in young workers <sup>8</sup>.

## 1.2 Odorant binding proteins (OBP)

The excretion of pheromones, as well as the uptake for detection, is aided by odorant binding proteins. These proteins are typically small and can be present as monomers or dimers, with subunit molecular weights around 20 kDa <sup>9,10,11</sup>. The typical pI of odorant binding proteins is between 4-5 <sup>11</sup>. The sequence similarity of odorant binding proteins of different species is low. Bovine OBPs share only 30 % of their amino acids with rat OBPs, for example <sup>12</sup>.

Odorant binding proteins of vertebrates belong to the lipocalin family. These small, soluble proteins can reversibly bind volatile odorants and pheromones <sup>13</sup>. They can be present in biological fluids such as saliva (pig) <sup>14</sup>, urine (mouse) <sup>15</sup>, vaginal discharge (hamster) <sup>16</sup>, sweat (horse) <sup>17</sup> or tears (mouse) <sup>18</sup>, but are also expressed in the nasal area. They were first identified in the nasal mucosa <sup>10, 19, 20</sup>. It has been shown, that OBPs are loaded with pheromones when secreted outside the nose <sup>14,21</sup>.

The mechanism of action of OBPs has not yet been fully understood. However, Pelosi proposed three hypotheses. First, the odorant binding proteins could assist the hydrophobic pheromone by providing a hydrophobic space in the aqueous mucosa, from which the pheromone might bind and dissociate, and thereby increase the amount of pheromone in this phase. The pheromone might be released by a change in protein folding due to pH differences <sup>22</sup>. Secondly, the pheromone might only be transported to the olfactory receptors if aided by the OBPs. Thirdly, the olfactory receptors might recognize the OBP-pheromone complex but not the pheromone alone <sup>12</sup>.



### 1.3 Lipocalins

This diverse protein family consists of low molecular weight, soluble, extracellular proteins, binding for the most part hydrophobic molecules<sup>23</sup>. They are found to be transport proteins and bind to specific cell surface receptors<sup>24</sup>, but are also involved in transport of small ligands, like the neutrophil gelatinase associated lipocalin<sup>25</sup>, in olfaction, for example, the mouse major urinary protein<sup>21</sup>, in the regulation of the immune response, like lipocalin 2 in pulmonary mycobacterial infections<sup>26</sup>, in the cryptic colouration, for example, Crustacyanin subunit A and C<sup>27</sup>, in the enzymatic synthesis of prostaglandins like the prostaglandin D synthase<sup>28</sup> and the mediation of cell homeostasis, for example, 24p3R<sup>24, 29</sup>.

The structure of lipocalins is barrel-like, built by eight antiparallel  $\beta$ -sheets and a short  $\alpha$ -helical segment close to the C terminus<sup>30</sup>. In a typical lipocalin, like in bovine b-lactoglobulin, the ligand binding site has been found to be on the inside of the barrel<sup>31</sup>. Although the amino acid sequence similarity is quite low between members of the lipocalin family, there are three common motives found, that aid to assign a protein to this family<sup>32</sup> (see figure 1 and figure 2).

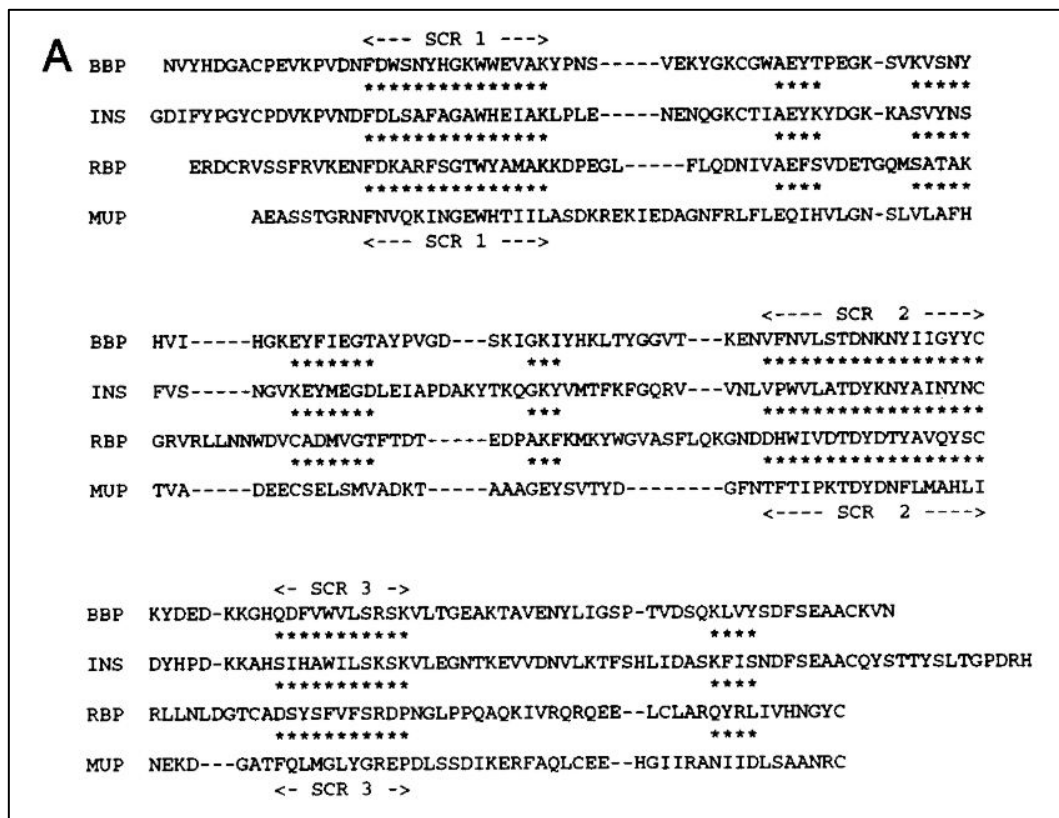


figure 1: Alignment of different lipocalins. The protein sequences used are bilin-binding protein (BBP), insecticyanin (INS), mouse major urinary protein (MUP) and retinol-binding protein (RBP). SCR1 -3 are the three structurally conserved regions. Source: (Flower, North, Attwood. 1993)<sup>32</sup>

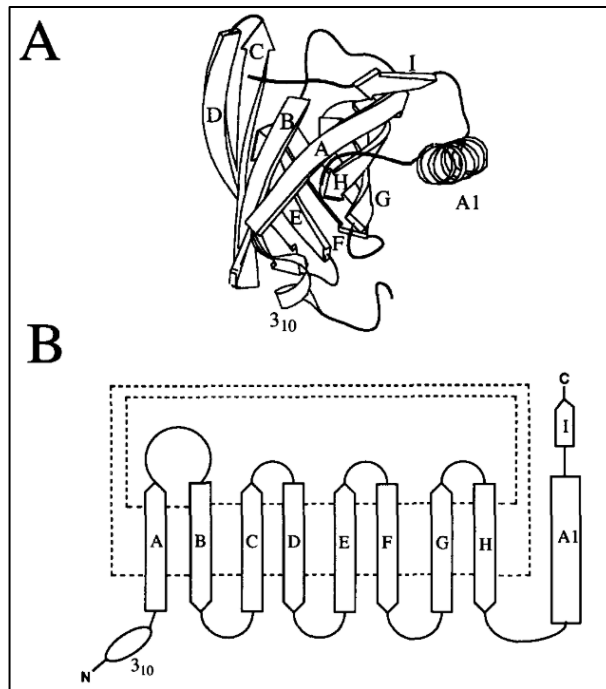


figure 2: A: ribbon drawing of the lipocalin fold (mouse major urinary protein). B: Schematic of the lipocalin fold (mouse major urinary protein). The strands of the  $\beta$ -barrel are labelled A to I. The helices are labelled 3<sub>10</sub> and A1. Source: (Flower, 1994)<sup>30</sup>

## 1.4 Olfactory receptors

The receptors used to convey general odorant information are significantly different from the ones used in pheromonic communication. The odorant receptors belong to the G-protein-coupled receptor (GPCR) superfamily and are expressed in olfactory sensory neurons of the main olfactory epithelium. They are seven-transmembrane domain proteins<sup>33</sup>. The receptors used for pheromone recognition belong to the V1R gene family, which contains no introns and the V2R gene family, which contains introns<sup>34</sup>. These receptors also belong to the G-protein-coupled-receptor superfamily but share no significant sequence identity with the odorant receptors. Also, they are expressed in a different location, namely the vomeronasal organ<sup>35</sup> (see figure 3).

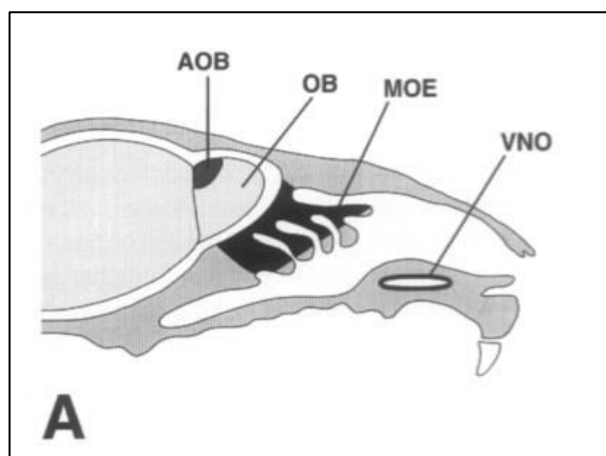


figure 3: The location of the vomeronasal organ (VNO), main olfactory system (MOE), main olfactory bulb (OB) and accessory olfactory bulb (AOB) in the skull of a rat. Source: (Dulac, Axel, 1995)<sup>35</sup>

## 1.5 Vomeronasal organ

The vomeronasal organ also called Jakobson's organ <sup>36</sup>, is part of the "vomeronasal organ complex" <sup>37</sup>. In a functional state, it consists of the epithelial tubular organ (vomeronasal organ), vomeronasal duct, seromucous glands, paravomeronasal ganglia, blood vessels, vomeronasal nerve bundles, cartilage and accessory olfactory bulbs <sup>38</sup>.

The vomeronasal organ has been found to be essential for certain pheromonic effects. A not exhaustive list of examples includes reproductive behaviour: Clancy et al. found, that an intact vomeronasal organ is necessary for the normal sexual behaviour and aggression in male mice <sup>39</sup>. Puberty acceleration in female mice was found to be dependent on this organ <sup>40</sup>. Another pheromonic effect is related to parental behaviour: Vomeronasal removal increased maternal behaviour in virgin female rats <sup>41</sup>. Also marking behaviour in some species is dependent on an intact vomeronasal organ, like in the male opossum, as differential scent marking was lost in male opossums after the vomeronasal nerve was cut <sup>42</sup>.

However, while the vomeronasal organ is necessary for some pheromonic detection and sexual behaviour, also the olfactory system plays a major role. As an example, Meredith could show, that the olfactory and the vomeronasal input converge into a shared circuit in the brain of male golden hamsters, controlling reproductive behaviour <sup>43</sup>. Halpern and Martínez-Marcos suggest, that this shared circuit is used to provide appropriate context for behaviours <sup>44</sup>.

In some cases, the vomeronasal organ is also not even necessary for pheromonic communication at all. In male rats, olfactory cues mediated via the olfactory system but not the vomeronasal system are necessary to display penile erection in the presence of inaccessible females in estrus. <sup>45</sup>. Additionally, mice were found to be able to discriminate between MHC-determined odour types without a functional vomeronasal organ <sup>46</sup>. Another case, where the pheromonal effect is mediated via the olfactory system is the nipple-search behaviour of new-born rabbits <sup>47</sup>.

However, some studies on sex pheromones have shown, that it makes a difference if the animals were sexually experienced before the vomeronasal organ was removed. Unfortunately, many studies do not indicate if this was the case with their experimental animals or they use exclusively experienced animals <sup>48, 45</sup>.

## 1.6 Pheromonic communication in humans (*homo sapiens*)

One of the most controversial questions is if humans possess pheromonic communication. While one side claims to have found said pheromones or at least behavioural evidence, the other side argues, that the physical requirements for pheromone detection are not met.

The first question and one of the main arguments against pheromonic communication in humans is the presence and functionality of the different parts of the vomeronasal organ complex. As stated in 1.5 Vomeronasal organ, Bhatnagar and Meisami found the whole complex to be necessary for functionality. Therefore, different parts of it will now be discussed.

The vomeronasal organ could be positively identified by Bhatnagar et al. in all their human samples, using serial histology <sup>49</sup>. The location of the vomeronasal organ can be seen in figure 4.

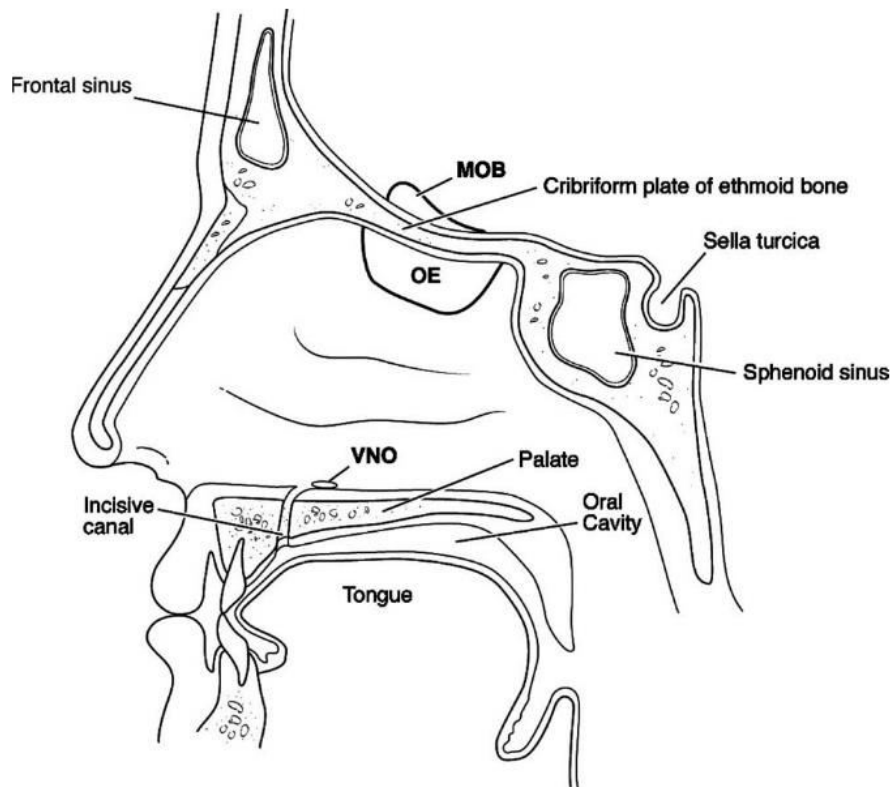


figure 4: The human skull with the approximate location of the human vomeronasal organ (VNO). Also shown are the main olfactory bulb (MOE) and the olfactory epithelium (OE). Source: (Halpern, Martinez-Marcos. 2003)<sup>44</sup>

The vomeronasal duct, on the other hand, was found by Witt et al. only in 15 of 23 samples, when using serial histology<sup>50</sup> and Knecht et al. stated in their review, that other groups got similar results to Witt et al.<sup>51</sup>

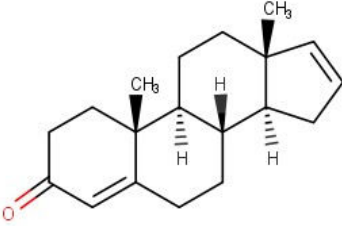
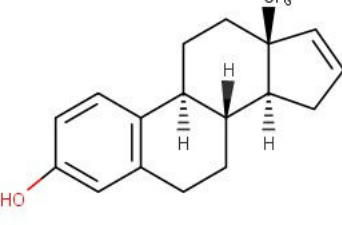
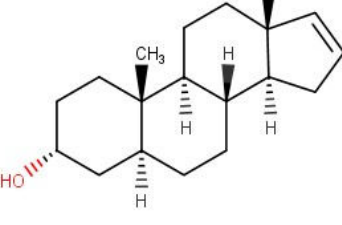
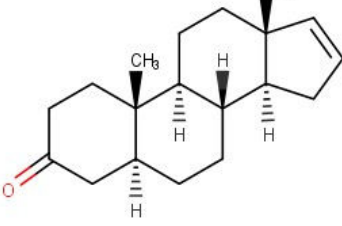
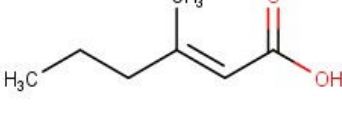
The nerve bundles from the vomeronasal organ to the accessory olfactory bulb could not be found by antibody staining against olfactory marker protein (expressed in vomeronasal receptor neurons of other species) or protein S100 (expressed in Schwann cells, which support neurons)<sup>52</sup> or the olfactory marker protein and PGP 9.5<sup>50</sup> (a marker for certain neurons<sup>53</sup>). This led to the conclusion, that there are no functional vomeronasal nerve bundles. Thus said, studying human embryos, Smith et al. concluded, that there is a vomeronasal organ developed in embryonic humans, but likely loses some functional components prenatally<sup>54</sup>.

In fact, no accessory olfactory bulb was found either<sup>55</sup>, so in summary, it can be said, that the vomeronasal organ appears to be non-functional in humans. However, since it has been demonstrated, that a vomeronasal organ is not necessary for pheromonic communication, the second question will be the presence and functionality of olfactory receptors of the V1R and V2R gene family.

On the one hand, only five V1r-like genes have been identified in humans and named V1RL1-5. All other human V1R-like genes are found to be pseudogenes<sup>56</sup>. From those five genes, only V1RL1 has been found in the olfactory mucosa<sup>57</sup>. However, the use of this receptor in pheromonic communication has yet to be proven. The V2R gene family on the other hand only comprises of pseudogenes in humans<sup>58</sup>.

The third question concerns the human pheromones or what is believed to be an effect of pheromonic communication. The main compounds, that might be pheromones in humans are summarised in table 1<sup>59</sup>.

table 1: Different compounds, described in the literature as putative human pheromones

Name	Structure	Additional information
4,16-androstadien-3-one		Activate the human hypothalamus <sup>60</sup> , effect on autonomous nervous system <sup>61</sup> , modulates mood <sup>62</sup>
Estra-1,3,5(10),16-tetraen-3-ol		Activate the human hypothalamus <sup>60</sup> , modulates mood <sup>62</sup>
5α-Androst-16-en-3α-ol		Detection threshold lower in period synchronized woman <sup>63</sup>
5α-Androst-16-en-3-one		Females have a stronger brain response to male body odour when able to smell 5α-androst-16-en-3-one <sup>64</sup>
(E)-3-Methyl-2-hexenoic acid		Found on the skin surface, carried by ASOB1 and apolipoprotein D, Apolipoprotein D is a lipocalin <sup>65</sup>

Though these putative pheromones are described in the literature, the findings must be taken with caution, since the studies all share the problem of small sample size or time scale. In the end, no bioassay guided thesis could prove to have found a true human pheromone to this date, so only speculations can be made<sup>66</sup>.

Nevertheless, some more effects are described, which are believed to be an effect of pheromonic communication:

- Menstruation cycle regulation<sup>67</sup>
- Mood effects<sup>68</sup>
- Effect on Luteinizing Hormone<sup>68</sup>

- Recognition of kin <sup>69</sup>
- Recognition of sex <sup>70</sup>
- MHC – dependent mate choice <sup>71, 72</sup>

If these are really the result of pheromonic communication and which pheromones these effects rely on is not proven up to this date. <sup>50, 73, 74, 66</sup>

### 1.6.1 h-SAL

The protein, this work is based on will be called human salivary lipocalin (h-SAL). This is to emphasize the relationship with the pig's salivary protein. The gene, this protein is translated from is called Major urinary protein pseudogene (MUPP), so the translated protein is also called MUP in the literature <sup>75</sup>. It is calculated to became a pseudogene in a hominid 400,000 to 350,000 years ago <sup>76</sup>. MUPP is found in NCBI under the Gene ID 100129193 and h-SAL under the GenBank accession number: EAW50553.1.

MUPP has no functional counterpart in the genome. It became a pseudogene by a splice junction mutation, preventing the RNA from being correctly spliced. This mutation is located at the splice donor site of the second intron <sup>75</sup>.

The h-SAL protein sequence (GenBank), see figure 5, is found to be 59 % identical to Chain A, odorant Binding Protein 3 of *Ailuropoda melanoleuca* (giant panda) (see figure 8) and 48 % identical to Chain A, Salivary Lipocalin of *Sus scrofa* (pig) (see figure 7) by using the BLASTP program <sup>77, 78</sup> on NCBI. These two have been chosen for multiple reasons. First, it is known, that pheromonic communication exists in panda and pig. Secondly, the ligands for the odorant binding proteins are also known. Thirdly, apart from the apes and monkeys, whose proteins were more similar, the blast revealed decent similarities with the h-SAL protein.

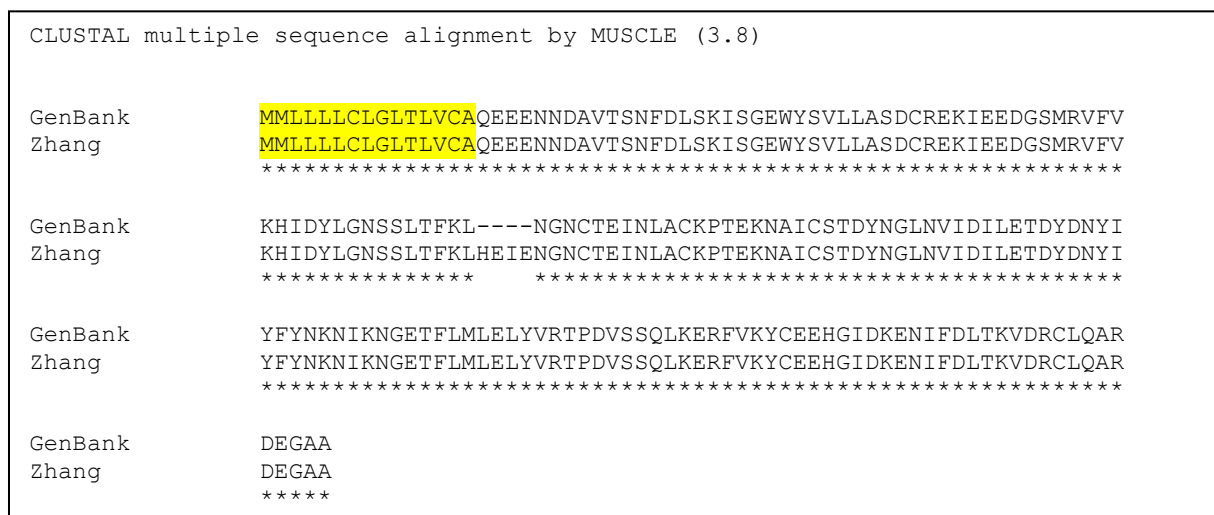


figure 5: h-SAL protein sequences and sources, the signal peptide (yellow) was identified using the SignalP 4.1 web server <sup>103, 104</sup> The alignment was made using the MUSCLE alignment tool version 3.8.31 <sup>105</sup>

There is, however, a second sequence for the h-SAL protein, which contains 4 additional amino acids. Zhang et al. reported in figure 3 (see additional information) of their paper, the sequence, were the amino acids HEI are located just before the second intron and E after it <sup>75</sup>. Since it is not clear, which one is the correct sequence, both were used in this project. When comparing protein models of the two h-SAL versions, the only clearly visible difference

is the length of one loop (see figure 6). In this work, the protein sequence from the GenBank will be further referred to as h-SAL and the sequence from Zhang et al. as h-SAL HEIE mutation.

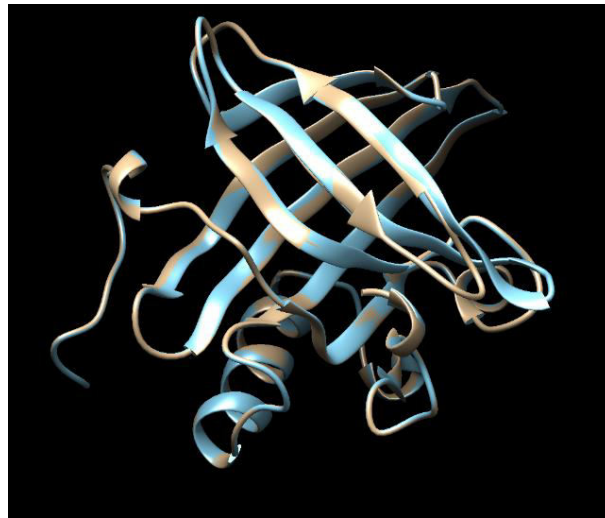


figure 6: overlap of the h-SAL protein from the gene bank (brown) and from Zhang et al. (blue) The protein models were built using swiss model<sup>79, 80, 81, 82, 83</sup> and the figure was made using the program UCSF Chimera<sup>84</sup>.

## 1.7 Pheromonic communication in pig (*sus scrofa*)

Pheromonic communication is mediated via the saliva in pigs. It contains the pheromones 5 $\alpha$ -androst-16-en-3-one and 5 $\alpha$ -androst-16-en-3 $\alpha$ -ol<sup>85, 14</sup>. The odorant binding protein associated with it is called salivary lipocalin (SAL1)<sup>14</sup>.

### 1.7.1 SAL1 protein:

The gene can be found on the NCBI webpage under the Gene ID 396739 and the protein sequence on UniProt with the number P816080. In figure 7 the protein sequence is shown. It is expressed in the boar's submaxillary glands, and in the nasal and vomeronasal area of both sexes<sup>86, 3</sup>. The ligands for the SAL1 protein are the pheromones 5 $\alpha$ -androst-16-en-3-one and 5 $\alpha$ -androst-16-en-3 $\alpha$ -ol<sup>14</sup>.

```
>P81608|SAL PIG Salivary lipocalin
MKLLLLLCLGLTLASSHKEAGQDVVTSNFDASKIAGEWYSILLASDAKENIEENGSMRVF
VEHIRVLDNSSLAFKFQRKVNGETDFYAVCDKVG DG VYTVAYYGENKFRLLLEVNYSDYV
ILHLVNVNGDKTFQLMEFYGRKPDVEPKLKDKFVEICQQYGI IKENIIDLTKIDRCFQLR
GSGGVQESSAE
```

figure 7: the sequence of the pigs salivary lipocalin, the signal peptide (yellow) was identified using the SignalP 5.0 web server<sup>104, 106</sup>

## 1.8 Pheromonic communication in panda (*Ailuropoda melanoleuca*)

Giant pandas possess a vomeronasal organ and mediate their reproductive behaviour via pheromonic cues. Male pandas can distinguish estrus females from non-estrus ones by investigating, displaying flehmen and licking female urine. Giant pandas also scent mark their territory with a secretion from their anogenital gland<sup>87</sup>. The structure of the pheromones they use, is, however, not found yet. Zhu et al. tried to use the reverse chemical ecology approach to form a hypothesis on pheromone candidates, by identifying the best ligands of the odorant

binding protein 3, which is similar to pheromone carrying proteins in rodents, pig and rabbits  
88

### 1.8.1 Odorant binding protein 3

The sequence of the protein can be found on UniProt under the number G1LK14 (see figure 8) and the three-dimensional structure of the OBP3 can be obtained from the Protein Data Bank under the number 5NGH\_A. It belongs to the lipocalin family and binds natural terpenoids and long-chain unsaturated aldehydes, such as Cedrol,  $\beta$ -Ionone, Citral, Safranal, Z11-16-Aldehyde, E2-10-Aldehyde, Z9-16-Aldehyde, and Z9-14-Aldehyde.

```
>G1LK14 AILME Lipocln_cytosolic_FA-bd_dom
MKLLVLCGLILVCAHEEGNDVRRNFDVSKIISGYWYSVLLASDVREKTEENSSMRVFNH
IEVLSNSSLLFNMHKVDGKCTEIALVSDKTEKDGEYSVEYDGYNVFRIVETDYTDYIIF
HLVNFKEKDSFQMMELSAREPDTSEEVRKRFVEYCYQKHGIVKENIFDLTEVDRCLOARGS
EKA
```

figure 8: the sequence of the *pandas* Odorant Binding Protein 3, the signal peptide (yellow) was identified using the SignalP 5.0 web server<sup>104, 106</sup>

## 1.9 Reverse chemical ecology

The classical chemical ecology approach aims to find the composition and chemical structure of a chemical signal by first extracting secretions from the senders of the chemical signal. Afterwards, these extracts are split into fractions and tested on the recipients of the chemical signal. This is to find the active compounds. Finally, the chemical structure is identified, and the signal can be imitated synthetically.

If this classical chemical ecology is not a viable approach, because secretions of the sender are hard or impossible to come by, there is also the possibility to use the reverse chemical ecology approach to find the putative chemical signal. In this approach, the genome is mined for the odorant binding proteins of the species. The most promising are the proteins, which are most similar compared to the pig's salivary lipocalin or other known pheromone binding proteins. If proteomics can be used to detect these proteins in the receiving organ and secretions of the sender (if available), it gives good clues on having found the right proteins. Then the odorant binding proteins are recombinantly expressed and ligand binding is investigated. Natural compounds, that are found to bind well inside the binding pocket are good candidates for pheromones and can be further investigated in behavioural studies<sup>89</sup>.

### 1.10 Objective

The aim of this thesis was to use the process of reverse chemical ecology to study the human salivary lipocalin protein in order to gain insight about the pheromonic communication of the hominids before the gene was switched off 350,000 to 400,000 years ago, thereby studying how human pheromonic communication could have worked if this event had not happened. This approach could provide new insights on the search for the pheromones of humans and related species.

In order to study h-SAL, the process of reverse chemical ecology is used. There it is necessary to express the protein recombinantly in *E. coli* and characterise its ligand-binding properties. Since there is a difference of four amino acids between the sequence obtained from GenBank and the paper from Zhang et al., both proteins are subject of this work. The



ligand-binding experiments are carried out as a competitive fluorometric assay, which is a well-established method for this kind of proteins.

## 2. Materials and Methods

### 2.1 Materials

#### 2.1.1 Chemicals

All Restriction enzymes and the suitable buffers were ordered from New England Biolabs, except for the DpnI Enzyme and the Fast Digest buffer, which were ordered from Thermo Fisher. The Polymerases and Ligases, as well as the suitable buffers, were ordered from Thermo Fisher, except for the T4 Ligase, which was ordered from Promega and the Taq Polymerase, which was ordered from VWR. All other chemicals were ordered from Sigma Aldrich. All enzymes and chemicals were used without any further purification. The water used was purified by a MilliQ water purification system.

#### 2.1.2 Competent cells

Competent *E. coli* DH5 $\alpha$  and BL21 were prepared as RbCl - competent cells. The protocol was comparable to the one found on OpenWetWare<sup>90</sup>, with minor changes.

The main difference was the omission of CaCl<sub>2</sub> and glycerol in RF1, the pH of 7.0 in RF2 and the preparation of competent cells from a previous stock instead from plate-streaked colonies. The cells were frozen and stored at -80 °C directly.

#### 2.1.3 LB and LB plates

The Lysogeny broth (LB) was prepared after the Miller recipe. For one Liter, 10g Tryptone, 5g Yeast extract and 10g NaCl was dissolved in MilliQ water and sterilised at 121°C for 20 minutes. For LB plates, 1.5 wt.% bacterial Agar was added before sterilisation and then poured on sterile petri dishes. If necessary, antibiotics were added after sterilisation in the following concentrations: 50  $\mu$ M Ampicillin or 50  $\mu$ M Kanamycin final concentration.

#### 2.1.4 DNA

All DNA was stored at -20°C.

##### *Plasmids*

Plasmids were amplified by transforming DH5 $\alpha$  cells with the desired plasmid and growing them in the presence of a suitable antibiotic at 37 °C and shaking overnight. Then the plasmids were extracted using the OMEGA plasmid extraction kit. Sterile MilliQ water was used for elution of the plasmid. In this work, mainly kanamycin resistance was used.

##### *Primers*

The T7 and the pET30rv primer were bought from Eurofins. The hSAL-HEIE-fw (forward) and the hSAL-HEIE-rv (reverse) primers were bought from LGC Genomics.

*table 2: primers and their sequences*

Primer name	Primer sequence (5' → 3')
T7	TAA-TAC-GAC-TCA-CTA-TAG-GG
pET30rv	TTG-TCG-ACG-GAG-CTC
hSAL-HEIE-fw	CTG-CAT-GAG-ATT-GAG-AAT-GGC-AAT-TGC
hSAL-HEIE-rv	GCC-ATT-CTC-AAT-CTC-ATG-CAG-TTT-GAA

## 2.2 Methods

### 2.2.1 Agarose Gel electrophoresis

All agarose gels were prepared by dissolving 1 wt.% agarose in TAE buffer (see below) by heating. 0.004 % gel red was used as a precast gel stain. 1 kb DNA Ladder from New England Biolabs was used in a 1:10 dilution as the marker. DNA gel loading dye purple was used to track DNA during the electrophoresis. Gels for gel extraction were run at 100 V, analytical gels were run at 125 V. For visualisation, fluorescence imaging was performed.

10 times concentrated TAE Buffer composition: (48.4g Tris, 11.4 ml glacial acetic acid, 3.7g EDTA disodium salt to 1L with MilliQ water)

### 2.2.2 Cloning of H-Sal into pET30a

A new synthetic gene was bought from Eurofins in the pEX-K168 standard vector, with NdeI as 5' restriction site and EcoRI as 3' restriction site. This synthetic gene uses kanamycin resistance.

#### *Restriction digestion*

The pET30a, as well as the pEX-K168-hSAL plasmid, was amplified, and restriction digestion was performed at 37 °C for 2 hours.

*table 3: reagents and amounts used for restriction digestion*

	pET30a	pEX-K168-hSAL
Plasmid in sterile H <sub>2</sub> O	39 µl (approx. 4 µg)	25 µl (approx. 7,5 µg)
10 x CutSmart® Buffer	4,5 µl	3 µl
NdeI	1 µl (1 Unit)	1 µl
EcoRI	1 µl (1 Unit)	1 µl

Gel extraction was performed after agarose gel electrophoresis (see above) from a 1 % agarose gel. The gel slices were cut on a UV transilluminator and the extraction procedure was performed using the OMEGA gel extraction kit.

#### *Ligation*

The h-Sal gene was ligated into the pET30a vector at 4 °C for 16 - 64 hours.

*table 4: reagents and amounts used for ligation*

	pET30a-hSAL
10x Ligation Buffer	1 µl
Digested pET30a	1 µl (approx. 50 ng)
Digested h-SAL	2 µl (approx. 25 ng)
20mM ATP	1 µl
T4 Ligase	1 µl (1 Unit)
Sterile H <sub>2</sub> O	4 µl

#### *Transformation of DH5α*

*E. coli* DH5α competent cells were transformed using the ligation mix. The transformation procedure was performed using the heat shock method. For this, the *E. coli* cells were thawed on ice and the ligation mix was added. Then they were incubated on ice for 1 hour before they were heat-shocked for 1 minute at 42°C. Afterwards, 300 µl LB was added and

the cells were shaken at 37°C for one hour before they were spread on LB plates containing the suitable antibiotic and incubated at 37°C overnight.

### Colony PCR and sequencing

Colonies were picked and colony PCR was performed. An agarose gel was run to estimate the size of the insert and colonies with the insert of right length were grown in LB with the antibiotic overnight. The plasmid was extracted using the E.Z.N.A.® Plasmid Mini Kit I from OMEGA and sent for sequencing.

table 5: reagents and amounts used for colony PCR

Taq PCR Mix (Master Mix)	50 vol%
10 µM T7 Primer	1 vol%
10 µM pET30rv Primer	1 vol%
25 mM Mg <sup>2+</sup>	6 vol%
Sterile H <sub>2</sub> O	To 20 µl/colony

table 6: PCR program for colony PCR

Number of cycles	Step 1	Step 2	Step 3
1	95°C 5:00 min		
35	95°C 0:30 min	50 °C 0:30 min	72°C 1:00 min
1	72°C 10:00 min		

## 2.2.3 Mutagenesis

In order to mutate the amino acids HEIE into the h-SAL protein, two different strategies were tested (see figure 9). The first one was using overlapping primers, including the mutation in both primers according to Liu and Naismith<sup>91</sup> and the second one was the two-step approach with only one mutated primer. The mutated sequence lies in the middle of each the two primers hSAL-HEIE-fw and hSAL-HEIE-rv.

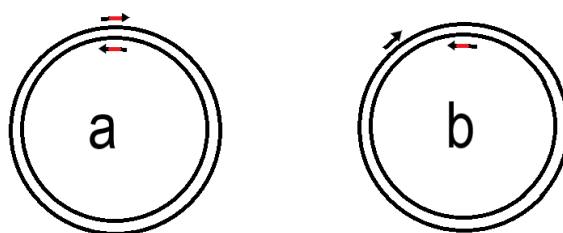


figure 9: primer positions for mutagenesis. The mutated parts of the primers are coloured red. A: approach with overlapping primers B: two-step approach using only one of the mutated primers

### Overlapping primers

The PCR was set up in accordance with Liu and Naismith<sup>91</sup>. With this method, the whole plasmid is amplified, and the mutation is inserted at the same time.

table 7: reagents and amounts used for PCR with overlapping primers

	1 <sup>st</sup> try	2 <sup>nd</sup> try
Sterile H <sub>2</sub> O	35.5 µl	32.5 µl
5x Phusion HF Buffer	10 µl	10 µl
10 mM dNTPs	2 µl	2 µl
10 µM hSAL-HEIE-fw	1 µl	2 µl
10 µM hSAL-HEIE-rv	1 µl	2 µl

Plasmid pET30a-hSAL diluted 1:100	1 µl (approx. 10 ng)	1 µl (approx. 10 ng)
Phusion Polymerase	0.5 µl (1 Unit)	0.5 µl (1 Unit)

*table 8: PCR program for mutation with overlapping primers*

Number of cycles	Step 1	Step 2	Step 3
1	95 °C 5:00 min		
12	95 °C 1:00 min	50 °C 1:00 min	72°C 15:00 min
1	50 °C 1:00 min	72 °C 30:00 min	

After PCR the mixture was digested with DpnI to remove the methylated template. Restriction digestion was carried out for 2 hours at 37 °C.

*table 9: reagents and amounts for digestion with DpnI*

PCR Product	17 µl
DpnI	1 µl (1 Unit)
10x Fast Digest Buffer	2 µl

5 µl of the digested PCR product was transformed into DH5α cells using the heat shock method and plated on kanamycin plates. The plates were incubated at 37 °C overnight.

#### *Two-step approach*

With this approach, the whole plasmid is amplified. Contrary to the approach with the two overlapping primers, it is done in two steps.

*table 10: amounts and reagents for the 1<sup>st</sup> PCR using the two-step approach*

Sterile H <sub>2</sub> O	32 µl
5x Phusion HF Buffer	10 µl
10 mM dNTPs	2 µl
10 µM Primer T7	2,5 µl
10 µM hSAL-HEIE-rv	2,5 µl
Plasmid pET30a-hSAL	0.5 µl (approx. 50 ng)
Phusion Polymerase	0.5 µl (1 Unit)

*table 11: 1<sup>st</sup> PCR program for mutation using the two-step approach*

Number of cycles	Step 1	Step 2	Step 3
1	95°C 5:00 min		
35	95°C 0:30 min	50 °C 0:30 min	72°C 1:00 min
1	72°C 10:00 min		

Gel extraction was performed after agarose gel electrophoresis (see above) from a 1 % agarose gel. The gel slices were cut on a UV transilluminator and the extraction procedure was performed using the E.Z.N.A.® Gel Extraction Kit from OMEGA. This served to purify the small amplified products, which were used as primers in the next step.

*table 12: amounts and reagents for the 2<sup>nd</sup> PCR using the two-step approach*

Sterile H <sub>2</sub> O	13,9 µl
--------------------------	---------

5x Phusion HF Buffer	10 $\mu$ l
10 mM dNTPs	2 $\mu$ l
Primers from gel extraction	22.6 $\mu$ l (approx. 280 ng)
Plasmid pET30a-hSAL diluted 1:100	1 $\mu$ l (approx. 10 ng)
Phusion Polymerase	0.5 $\mu$ l (1 Unit)

table 13: 2nd PCR program for mutation using the two-step approach

Number of cycles	Step 1	Step 2
1	98°C 5:00 min	
20	98°C 1:00 min	72°C 7:00 min
1	72°C 10:00 min	

After PCR the mixture was digested with DpnI to remove the methylated template. Restriction digestion was carried out for 5 hours at 37 °C.

table 14: reagents and amounts for DpnI digestion

PCR Product	17 $\mu$ l
DpnI	1 $\mu$ l (1 Unit)
10x Fast Digest Buffer	2 $\mu$ l

5  $\mu$ l of the digested PCR product was transformed into DH5 $\alpha$  cells using the heat shock method and plated on kanamycin plates. The plates were incubated at 37 °C overnight.

### Sequencing

Colonies were picked and grown in LB Kanamycin overnight. The plasmid was extracted using the E.Z.N.A.® Plasmid Mini Kit I from OMEGA and sent for sequencing.

### 2.2.4 Protein expression

*E. coli* BL21 were transformed with 2  $\mu$ l (100 – 200 ng) plasmid using the heat shock method and a preculture was grown overnight. 10 ml of this preculture was added to 1L of LB Kanamycin and grown at 230 rpm at 37°C to an OD<sub>600</sub> of 0.6.

Protein expression was induced with 400  $\mu$ l 1M IPTG for 2 hours. All the following steps were carried out on ice. The cells were harvested by centrifugation at 3500 rpm at 4°C for 15 min and the cell-pellet resuspended in 20 ml of 50 mM Tris buffer pH 7.4. No protease-inhibitors were added since they could influence the binding experiments.

Cell disruption was induced by sonication.

table 15: sonicator settings

4 Times – 5 min each
Output control 3
Duty cycle 50 %

The sample was kept on ice and there were 15 minutes break between each sonication to allow the sample to cool down.

Then the lysed cell mixture was centrifugated again at 12000 rpm at 4 °C for 1 hour. An SDS PAGE was run to find out if the protein was expressed in the supernatant or pellet. For storage, the supernatant was frozen at -20°C and was purified after another 12000 rpm at 4 °C for 1 h centrifugation step. The pellet was resuspended in 10 ml of 8 M Urea and incubated with 50 mM DDT for 1 hour at room temperature. It was then dialysed three times for 8-16 hours at 4°C and purified after another 12000 rpm at 4 °C for 1 h centrifugation step.

## 2.2.5 SDS-PAGE

The 13 % acrylamide/bisacrylamide SDS Gel was used in accordance to Laemmli <sup>92</sup>. 2,2,2-Trichloroethanol was added to the lower gel to enable fluorescence readout of proteins containing tryptophan. As a molecular weight marker, peqGOLD Protein Marker I from VWR and PageRuler™ Prestained Protein Ladder from Thermo Fisher were used to determinate the molecular weight of the protein samples.

To solubilise and denature the protein samples they were heated in boiling water. For cell cultures, 500 µl liquid culture was centrifugated at 12000 rpm for 1 minute at room temperature and the pellet was resuspended in 50 µl sample buffer and boiled for 10 minutes. For protein samples, after cell disruption 1-10 µl were boiled for 7 minutes with 10 µl sample buffer.

The gels were run at 200V and 20 mA/gel for 30-45 minutes.

### Gel readout

Immediately after electrophoresis, a fluorescence readout was performed. Then the gels were placed in stop-solution until the blue bands from the sample buffer de-stained and were stained again using the Coomassie staining solution. The gels were incubated in the staining solution overnight and de-stained by either boiling them in tap water or incubating them in stop solution. The gels were photographed.

### Buffers and Solutions for SDS-PAGE

table 16: reagents and amounts for 2 acrylamide/bisacrylamide gels

For 2 gels	Lower Gel	Upper Gel
Water	2.5 ml	1.5mL
30 % AA	3.5 ml	600µl
Lower gel buffer	2 ml	-
Upper gel buffer	-	500µl
3-Chloro-EtOH	40 µl	-
TEMED	10 µl	5µl
10 % APS	60 µl	30µl

table 17: composition of buffers and solutions

Lower gel buffer: 0.5 M Tris (12,1g), HCl to pH 8.8, SDS 0.4 % (0.8 g), MilliQ water to 200 ml
Upper gel buffer: 0.5 M Tris (12,1g), HCl to pH 6.8, SDS 0.4 % (0.8 g), MilliQ water to 200 ml
Running buffer: Tris 3.03 g, Glycine 14.4 g, SDS 1g, MilliQ water to 1 L
SDS sample buffer: 0.1 M Tris (0.6 g), HCl to pH 6.8, SDS 3 % (1.5 g), Mercaptoethanol 2.5 ml, Glycerol 5 mL, MilliQ water to 50 ml, Bromophenol blue a tiny spatula tip
Stop Solution: Acetic acid 100 ml, Ethanol 250 ml, MilliQ water to 1 L
Coomassie staining solution: Stop solution 100 ml, Coomassie blue R250 150 mg

## 2.2.6 Protein purification

Proteins were purified using FPLC with an anion exchange column as well as a size exclusion column. The program for purification of the h-SAL protein was optimised during this work. If the protein was purified more than once, the fractions of interest were dialysed against 50 mM Tris buffer, pH 7.4 for 16 hours before loading again.

### *Anion-exchange chromatography*

The column used was a HiPrep™ Q HP 16/10 from GE Healthcare, packed with Q Sepharose High Performance anion exchange resin.

Buffer A was 50 mM Tris, pH 7.4 and Buffer B was 50 mM Tris, pH 7.4 with additional 0.5 M NaCl. A gradient program was run, which was modified during the course of this work and also Buffer B was switched to 50 mM Tris, pH 7.4 with 1 M NaCl. The flow rate was 2 ml/min.

### *Size-exclusion chromatography*

The column used was Superose® 12 10/300 GL from GE Healthcare. It was run with an isocratic program with 50mM Tris, pH 7.4 and 0.5 M NaCl as buffer.

### *Ammonium sulphate precipitation*

This experiment was done in accordance with Wingfield “Protein Precipitation using Ammonium Sulphate”<sup>93</sup>. To check if the h-SAL protein precipitates from the once purified supernatant at a certain ammonium sulphate saturation, different amounts of ammonium sulphate were added to the sample. The aliquots were centrifuged at 4°C for 1h at 12000 rpm and the pellet resuspended in 100 µl 50 mM Tris pH 7.4. Then an SDS PAGE was run to check the presence of the protein.

## 2.2.7 Protein characterization

The binding affinities of the purified proteins were characterised using fluorescence spectroscopy. Quenching of the signal was measured and used to calculate dissociation constants. Two different fluorimeters were used during the course of this work. The first one was the Perkin Elmer LS 55 Fluorescence Spectrometer, which was then upgraded to the Perkin Elmer Fluorescence Spectrometer FL6500. Both fluorimeters have been tested to produce equal results after normalisation. The data was automatically background corrected. For the experiments 2 µM protein concentration was used in 50 mM Tris pH 7.4. 1-NPN, 1-AMA and the different ligands were used as 1mM solutions in methanol. The data were normalised to 100 % fluorescence intensity for the protein only if not otherwise stated.



### Competitive Binding assay

In the competitive binding assay, either 1-AMA or 1-NPN were used as a probe to bind in the binding pocket. The fluorescence of this binding probes is significantly higher when bound to the protein compared to in solution. The potential ligands quench the signal by competing for the binding site. The intensity of the free fluorescence probe in solution (without protein present) was deducted as additional background manually. The instrument settings for these experiments can be found in table 18.

### Tryptophan quenching

The fluorescence intensity, the wavelength at maximum intensity and the quantum yield of tryptophan strongly depend on the polarity of the surrounding microenvironment. This is used to distinguish tryptophan in the binding pocket of proteins from those facing the solvent. In this experiment the fluorescence signal of the tryptophan in the binding pocket is measured. It gets quenched when ligands bind<sup>94</sup>. The instrument setting can be found in table 18.

table 18: fluorometer settings for different experiments. A: also includes 1-NPN measurements for background subtraction. B: includes 1-NPN measurements to find the dissociation constant and for background subtraction

	NPN binding	Protein binding h-SAL <sup>a</sup>	Protein binding h-SAL <sup>b</sup>	1-AMA binding	Protein binding h-SAL HEIE mutation <sup>b</sup>	Tryptophan quenching
Instrument	LS 55	LS 55	FL6500	FL6500	FL6500	LS 55
Emission start [nm]	380	380	380	460	380	310
Emission end [nm]	450	450	450	550	450	380
Excitation [nm]	337	337	337	375	337	295
Scan rate [nm/min]	200	200	240	240	240	200
Photomultiplier voltage (gain)	low	medium	low 400	low 400	Low 400	medium
Excitation Slit [nm]	5	5	5	5	5	7
Emission Slit [nm]	5	5	5	5	5	7
Peaks measured at [nm]	410	410	418	516	416	337
Accumulation number	-	-	1	1	1	-
Emission correction	-	-	on	On	On	-
Gain PMT	-	-	*1	*1	*1	-

### 2.2.8 Calculation of the Dissociation constants from the binding curves

For the graphs and the curve fittings the program Origin 2019 and Origin 2020 were used. The calculations were done in Microsoft Excel.

### *The dissociation constant of h-SAL and h-SAL HEIE mutation protein and 1-NPN and 1-AMA*

For the calculation of the fluorescence probes dissociation constants the model of a protein, with one binding site was used. The formula used was:  $y = \frac{B_{max} * x}{k_1 + x}$ .  $B_{max}$  is the abbreviation for the total receptor number, expressed in the number of fluorescence intensity and  $k_1$  the equilibrium dissociation constant, expressed in  $\mu\text{M}$ .

### *The IC<sub>50</sub> of different ligands*

For the calculation of the IC<sub>50</sub>, the normalized fluorescence intensity was plotted against the ligand concentration. Then, assuming a simple dose-response relationship, the formula  $y = A_1 + \frac{A_2 - A_1}{1 + 10^{(\log(x_0) - x) * p}}$  was used as a fit function.  $A_1$  is the abbreviation for the top asymptote,  $A_2$  is the bottom asymptote and  $\log(x_0)$  is the centre of the curve.  $A_1$  was fixed to 100 % for all fittings since the fluorescence without ligand present was normalised to be 100 %.  $p$  is the parameter for the hill slope.

### *The dissociation constants of protein and different ligands*

The formula obtained from literature<sup>95</sup>. The formula is  $K_{diss} = \frac{IC_{50}}{1 + [L]/K_d}$ , where  $[L]$  is the concentration of the free fluorescent probe and  $K_d$  is the measured dissociation constant of the protein-fluorescence probe complex.

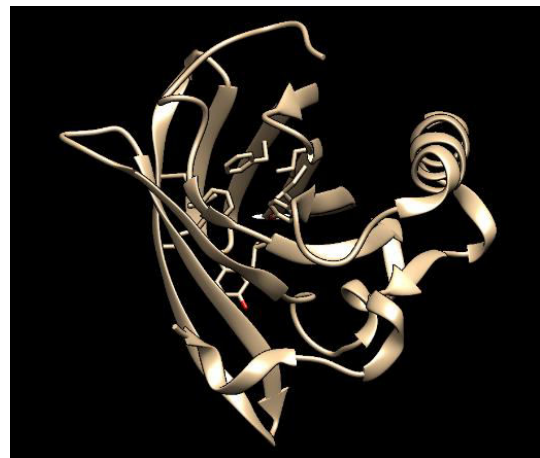
### 3. Results

#### 3.1 Theoretical predictions about the H-SAL protein and comparison with OBP3 and SAL1

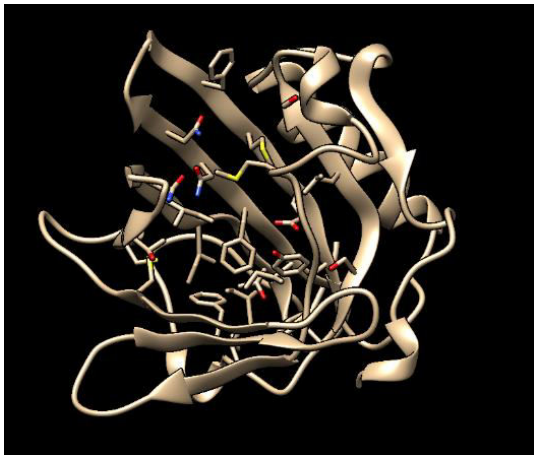
A model for the h-SAL protein was built using Swiss Model on-line software<sup>79, 80, 81, 82, 83</sup>. The amino acids in the binding pocket were examined manually using the program UCSF Chimera<sup>84</sup>. The digital protein models for giant panda OBP3<sup>88</sup> and pig SAL<sup>96</sup> were obtained from the PDB Database. figure 10- figure 12 were made using the program UCSF Chimera in order to visualise the protein folding and gain first insights into the properties of the binding pocket. However, this approach proved to be unsuccessful in finding possible pheromone candidates, as the binding pocket of h-SAL differs significantly from both giant panda OBP3 and pig SAL.



*figure 10: model of the giant pandas OBP 3. AS in the binding pocket are shown.*



*figure 12: the model of the pig SAL1. AS in the binding pocket are shown.*



*figure 11: the model of human h-SAL. AS in the binding pocket are shown*

### 3.2 Cloning of h-Sal into pET30a

The h-SAL gene was cloned into pET30a for expression. This was done by first performing restriction digestion on the pET30a vector and on the pEX-K168-hSAL vector with the restriction enzymes NdeI and EcoRI. The DNA was then cleaned up via agarose gel electrophoresis and gel extraction and the insert was ligated into the digested vector.

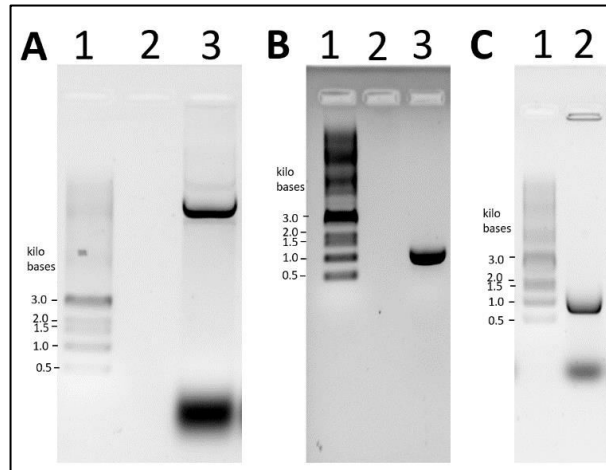


figure 13: agarose gel electrophoresis A: digested pET30a. B: digested h-SAL insert. C: colony PCR of the colony with the right insert.

The ligation mixture was used to transform *E. coli* DH5 $\alpha$ , which were then plated. To check the result of cloning, colonies were picked, and colony PCR was performed. The sequence of one colony, containing an insert of the right size, was confirmed by sequencing. The sequencing result can be found in the supplementary data.

### 3.3 Protein expression of h-SAL

The h-SAL protein was expressed as described in chapter 2.2.4. It was carried out multiple times since purification proved to be challenging. In figure 14 two exemplary SDS Pages can be seen. In the first three expressions, the protein was found mainly in the supernatant. However, in the other three expressions, there was approximately half of the h-SAL protein found in the pellet, which made purification from the pellet possible. The protein was expressed without any tag, as it would need to be removed before the ligand binding experiments. Overexpression at the expected molecular weight was taken as indication, that the desired protein was expressed.

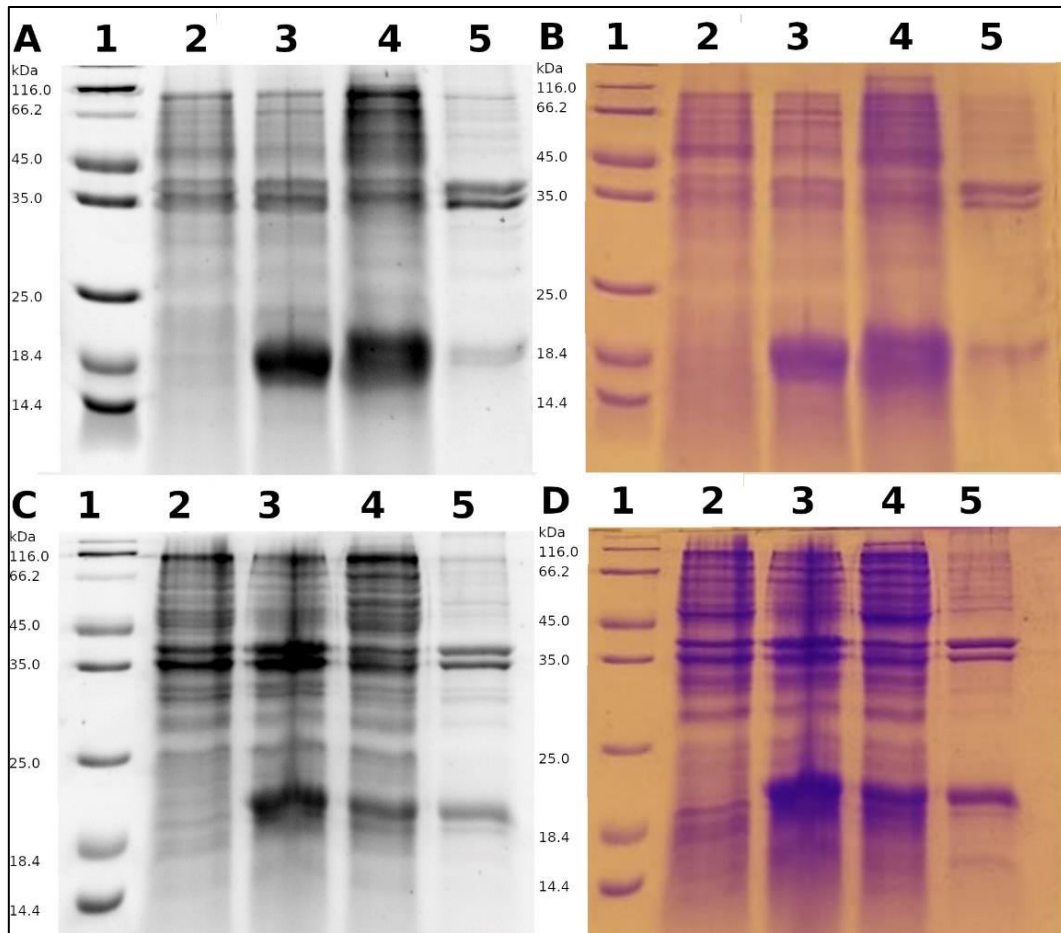


figure 14: SDS page after expression of h-SAL. A: fluorescence readout of the third expression. B: Coomassie-stained gel of the third expression. C: fluorescence readout of the fourth expression. D: Coomassie-stained gel of the fourth expression. 1: Marker. 2: bacterial sample before inducing. 3: bacterial sample after inducing. 4: supernatant after sonication and centrifugation. 5: pellet after sonication and centrifugation.

### 3.3.4 Protein purification

#### *The initial purifications*

The protein was initially purified from the supernatant. An exemplary chromatographic trace can be seen in figure 15. There were some main issues with this purification: first, there were lots of high molecular weight impurities still found in the sample, even after three purification steps with the anion exchange column. Also, as seen in figure 19, purification with a size exclusion column does not lead to increased purity. Furthermore, the protein peak tailed, which lead to a substantial loss of protein over the course of multiple purifications.

#### *Anion-exchange chromatography*

The initial purification program can be seen in table 19. Buffer concentrations were 50 mM Tris, pH 7.4 for Buffer A and 50 mM Tris, pH 7.4 and 0.5 M NaCl for Buffer B. The chromatogram of the first purification can be seen in figure 15.

table 19: the initial anion exchange program

ml	%B
10	0
200	0-50
50	50-100
100	100

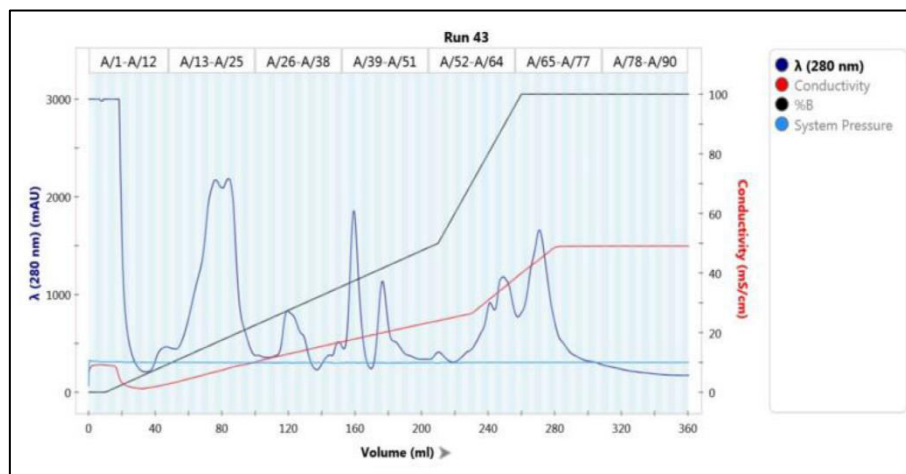


figure 15: chromatogram of the first purification run by anion exchange chromatography

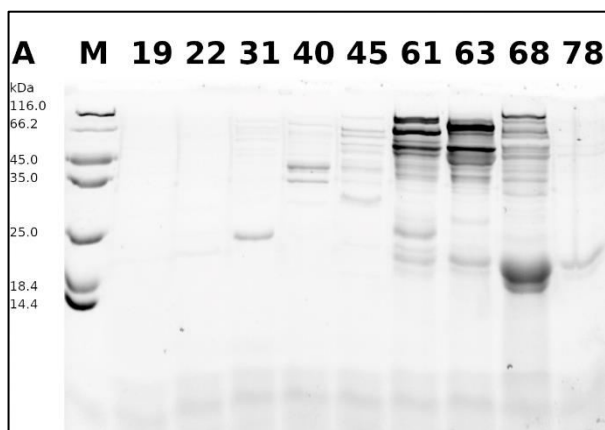


figure 16: fluorescence readout of the fractions from the first purification. The fractions are labelled with fraction number. M: Marker

In figure 16 it can be seen, that, starting from fraction 61, the protein is eluted up until fraction 78 and in a small amount probably further. The amount of high molecular weight impurities decreases with fraction number and the fractions with the most h-SAL protein are 67-70 in this case (see figure 17 and figure 18).

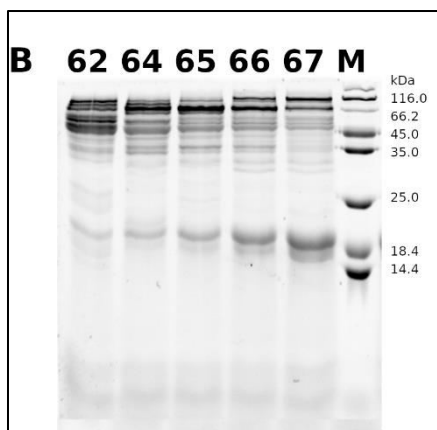


figure 17: fluorescence readout of fraction 62-67 from the first purification. The fractions are labelled with fraction number. M: Marker. The first half of the same gel as figure 18.

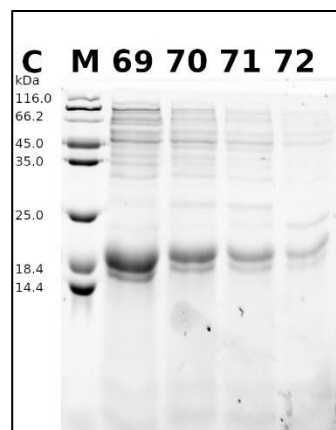


figure 18: fluorescence readout of fraction 69-72 from the first purification. The fractions are labelled with fraction number. M: Marker. The second half of the same gel as figure 17.

### Size-exclusion chromatography

Size exclusion chromatography was used after anion exchange chromatography in order to separate the high molecular weight impurities from the h-SAL protein. The program run was isocratic with a 50mM Tris, pH 7.4 and 0.5 M NaCl buffer (Buffer B).

For this purification step, no chromatogram can be shown, due to an air bubble in the detection chamber.

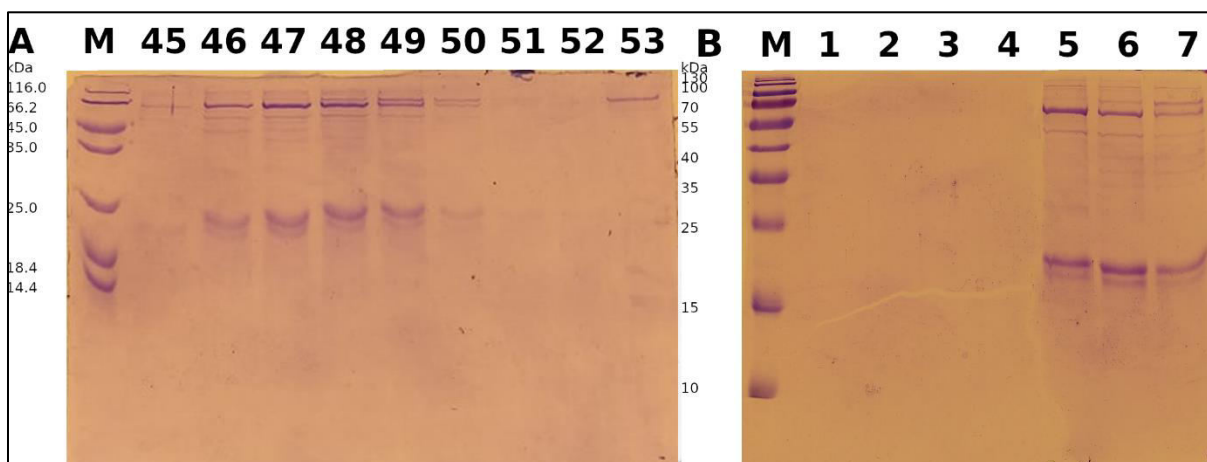


figure 19: Coomassie-stained gels. The lanes are labelled with fraction number. M: Marker A: fractions after anion exchange. 46-50 were pooled and used for size exclusion chromatography. B: fractions after size exclusion chromatography.

Fractions 46-50 after anion exchange chromatography were pooled and concentrated using the ultrafiltration device on ice (see figure 19). Then the fractions were freeze-dried and dissolved in 700  $\mu$ l MilliQ water. This solution was centrifuged at 12000 rpm for 15 min at 4°C before loading on the column. Protein was only found in fractions 5-7. This was verified by running fractions 5-13 on another gel (not shown). It can be clearly seen, that size-exclusion chromatography lead to no significantly improved purity of the protein fractions.

### Ammonium sulphate precipitation

A 1:10 times diluted protein fraction, obtained after an autosampler failure was used as the sample. This fraction was pre-concentrated by ultrafiltration at room temperature for 12 h and contained 0.5 M NaCl. 1 ml was used for each ammonium sulphate concentration.

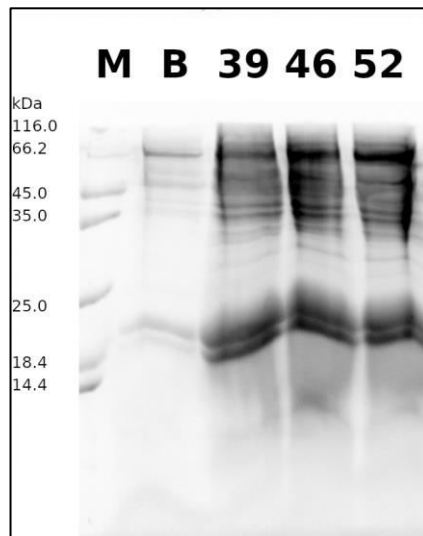


figure 20: fluorescence readout of SDS PAGE with redissolved pellets obtained at different Ammonium Sulfate concentration. The Lanes are Labelled with Percentage of Ammonium Sulfate Saturation. M: Marker, B: diluted protein fraction before the experiment. At 33% Ammonium Sulfate Saturation no pellet was formed.

It was found that the h-SAL protein precipitates between 33 – 39 % ammonium sulphate saturation (see figure 20), so the fractions were pooled and brought to 39 % ammonium sulphate saturation. After centrifugation at 4 °C for 1h at 12000 rpm, the pellet was resuspended in 4 ml of 50 mM Tris pH 7.4. and dialysed 3 times against 50 mM Tris pH 7.4 for 8-16 hours each before loading on the anion-exchange column. The chromatogram can be found in figure 21.

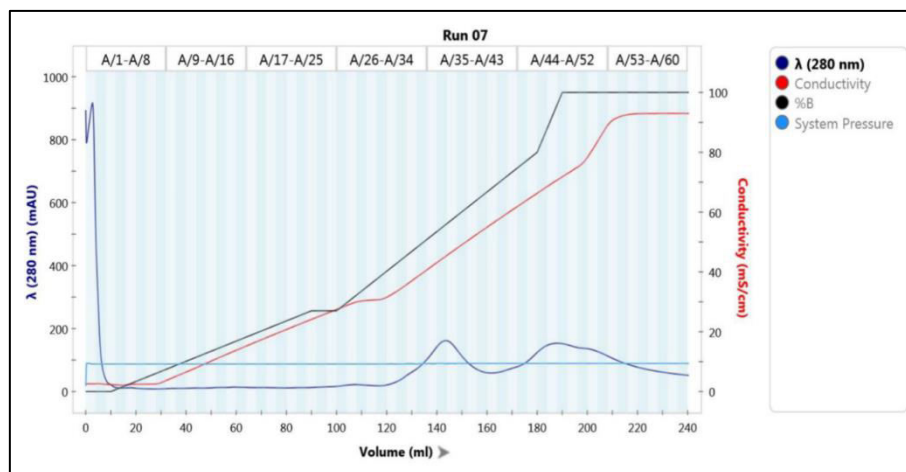


figure 21: chromatogram of the purification by anion exchange chromatography after ammonium sulphate precipitation



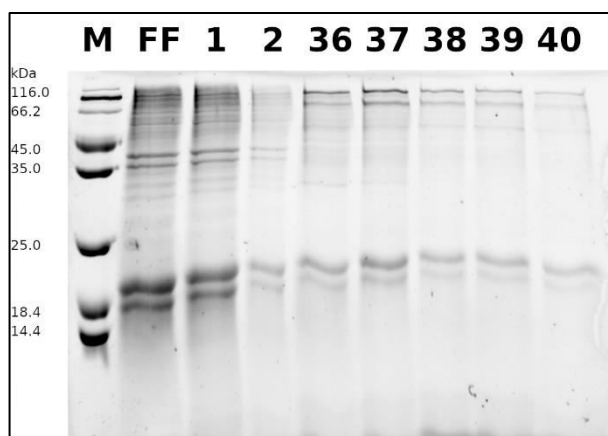


figure 22: fluorescence readout of the SDS PAGE after anion exchange chromatography. The lanes are labelled with fraction number. M: Marker. FF: follow-through while loading the column.

However, since there was a substantial amount of degradation of the protein found (see double bands just above 18,4 kDa in figure 20 and figure 22), most likely due to the long exposure to room temperature of these fractions, they were discarded.

#### The new purification-approach

The main difference to the initial purifications was the possibility to purify the protein from the pellet, which was already containing fewer impurities at the start. The proteins were solubilised in 8M Urea and the disulphide bridges of the proteins were reduced with 50 mM DDT. The solution was then dialysed 3 times for 8-16 hours against 50 mM Tris pH 7.4. Then the solution was centrifuged at 12000 rpm, at 4°C for 1 hour. Consequently, the protein solution was purified by anion exchange chromatography. The main changes in the chromatographic program were the switch to higher salt concentrations in buffer B. Initially it was 0.5 M NaCl and was changed to 1M NaCl. The chromatographic program was shortened between 0 and 0.27 M NaCl concentration and prolonged between 0.27 and 0.8 M NaCl concentration to allow the h-SAL protein peak to get separated from the DNA peak adjacent to it (see table 20). This lead, however, to more diluted protein fractions. A sample chromatogram can be seen in figure 23 and the corresponding SDS PAGE in figure 24.

table 20: the new anion exchange program

ml	%B
10	0
85	0-27
10	27
80	27-80
10	80-100
50	100

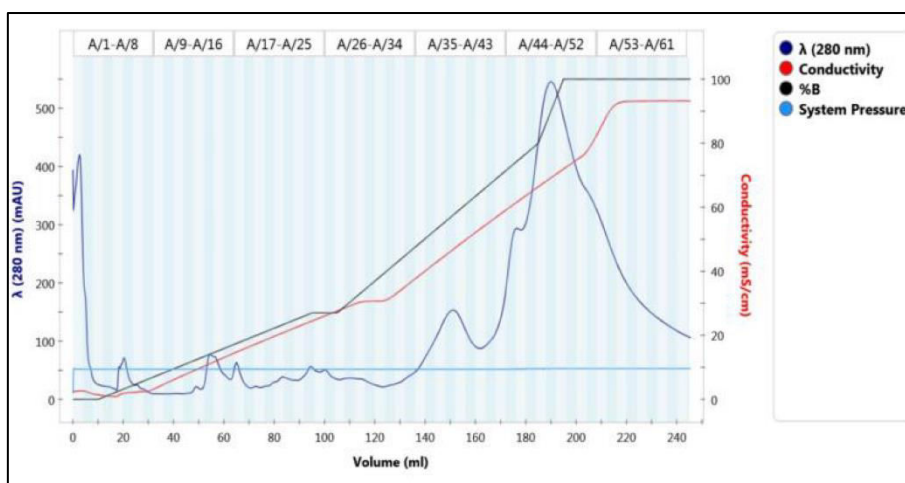


figure 23: chromatogram after the first purification step with the new purification approach

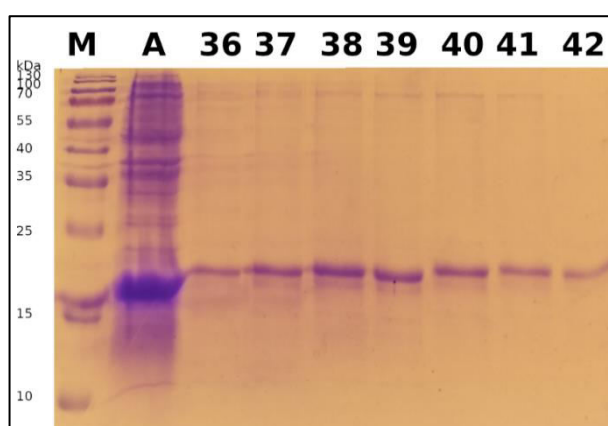


figure 24: coomassie-stained SDS PAGE after purification with the new anion exchange program. The lanes are labelled with fraction numbers. M: Marker. A: corresponding cell culture after inducing with IPTG

The main advantage of this chromatographic program is, that it yields up to 4 fractions of pure protein, without substantial quantities of DNA (see table 21).

### 3.3.5 h-SAL protein characterization

#### Amount of protein obtained

The amount of protein in the fractions was calculated using the extinction coefficient obtained from EXPASy<sup>97</sup>. Given that all pairs of cysteines form cystines, the extinction coefficient at 280 nm is  $19,285 \text{ M}^{-1} \text{ cm}^{-1}$  when measured in water.

The absorbance at 260 nm was also measured, to evaluate if the amount of DNA would significantly influence the measurement at 280. The results of the different fractions can be found in table 21.

table 21: available fractions after purification, absorption at different wavelengths and amount of protein calculated. Fraction volume 4 ml each for non-extracted fractions. Extracted fractions volume 1ml each taken from the previously mentioned fraction.

Fraction	Abs at 280	Abs at 260	[mol/l]
First purification from pellet			
38	0.400	0.287	0.020
39	0.371	0.305	0.019
38 extracted pH 7.4	0.374	0.269	0.019
39 extracted pH 4	0.234	0.149	0.012
Third purification from pellet			
36	0.155	0.113	0.008
37	0.226	0.154	0.011
38	0.285	0.186	0.015
39	0.269	0.183	0.014
40	0.199	0.162	0.010

#### Finding a suitable fluorescence probe

In order to evaluate the binding affinity of the fluorescence probes to the h-SAL protein, different amounts of 1-NPN or AMA were added to a 2  $\mu$ M solution of h-SAL in 50 mM Tris, pH 7.4. Then the protein was extracted at pH 7.4 and at pH 4 with Dichloromethane to investigate if there was a ligand already bound after bacterial expression, which would influence the proteins binding capabilities. The results are summarised in figure 25 and table 22, as well as figure 26. For the 1-NPN measurements, the  $K_d$  was calculated as a mean of three measurements, while for the 1-AMA measurements it was done from a single measurement. This also applies to the background measurements for the respective fluorescence probes.

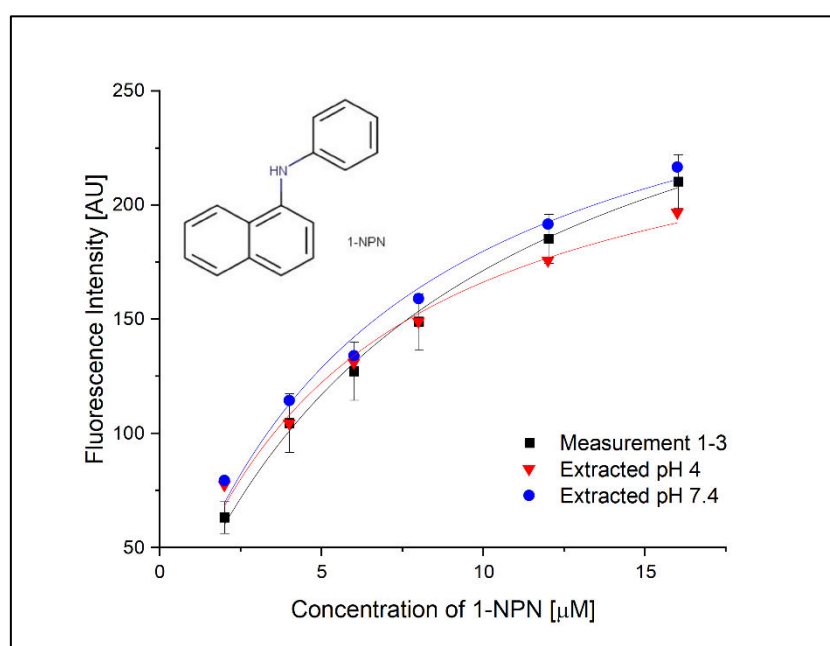


figure 25: the 1-NPN binding experiment using the different h-SAL fractions

table 22: fitted parameters from the 1-NPN binding experiments (see figure 25)

Measurement	Mean 1-3	Extracted pH 7.4	Extracted pH 4
$B_{\max}$ [AU]	320	298	260
$\sigma B_{\max}$ [AU]	14	20	14
$k_1$ [ $\mu\text{M}$ ]	8,7	6,6	5,6
$\sigma k_1$ [ $\mu\text{M}$ ]	0,8	1,0	0,8

As table 22 demonstrates, that the dissociation constants only vary within the same order of magnitude, which can be due to the typical data fluctuations with this kind of experiments. Therefore, it can be said, that there is no already bound ligand, which would disturb further binding experiments.

Concerning the binding curve with 1-AMA (see figure 26) it should be noted, that the dissociation constant and also the standard deviation is significantly higher than the one obtained with NPN binding (see table 22). Therefore, NPN was chosen as the fluorescence probe for all further binding experiments.

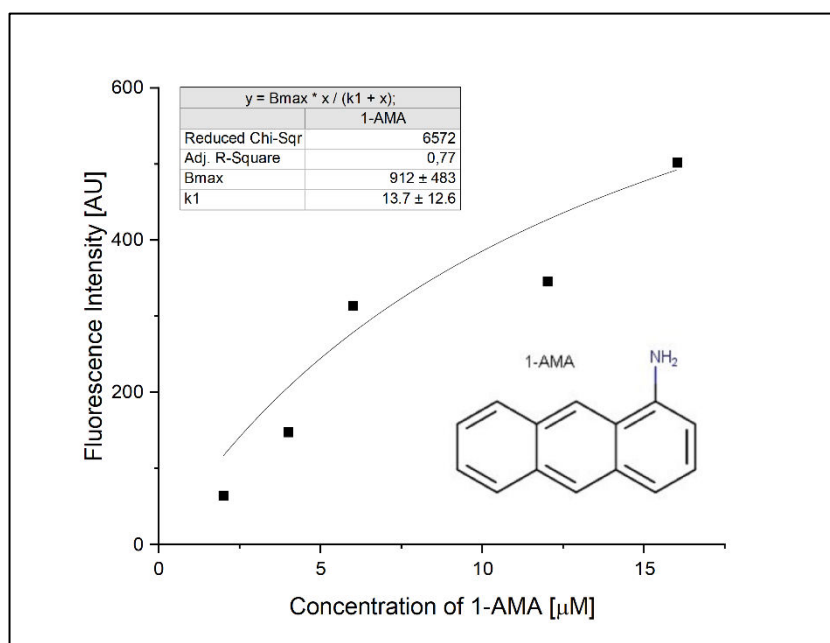


figure 26: the 1-AMA binding experiment and fitted parameters

### Ligand-binding experiments

For the ligand-binding experiments 2  $\mu\text{M}$  1-NPN was used in a competitive assay with different concentrations of ligand. The initial measurements were done only once. If the fluorescence intensity went below 70 % at 16  $\mu\text{M}$  ligand concentration, the measurements were repeated and the mean of three measurements was plotted in the following graphs. The exceptions to this are the ligands Piperonyl alcohol, which was measured only once, Quercetin, which was measured four times instead of three times and Vanillylamine, which was also measured three times. If the measurement has been done three or more times, error bars corresponding to  $\pm$  one standard derivation have been reported. The curves obtained from the initial measurements were not fitted, since they were measured only once.

In figure 27, the quenching curves with the ligands Retinol,  $\beta$ -Ionone, Farnesol and Safranal are shown. All three structures are similar, with  $\beta$ -Ionone and Farnesol being like different

parts of Retinol. Safranal is also a terpenoid.  $\beta$ -Ionone, Farnesol and Safranal bind to some extent but Retinol binds the h-SAL protein well. This is somehow to be expected since, upon many others, human lipocalins have been shown to bind retinoids<sup>98</sup>.

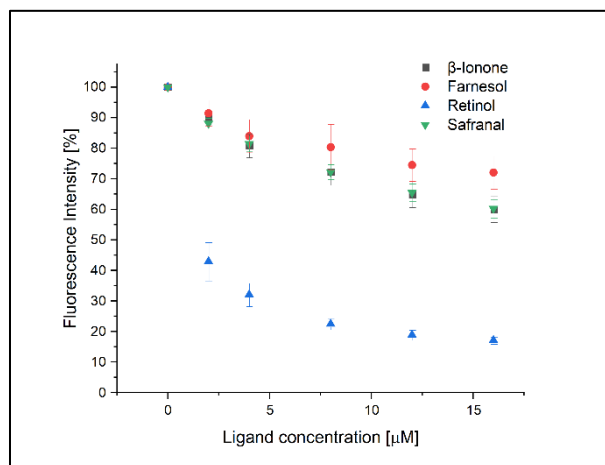


figure 27: quenching curves for Retinol,  $\beta$ -Ionone, Farnesol and Safranal

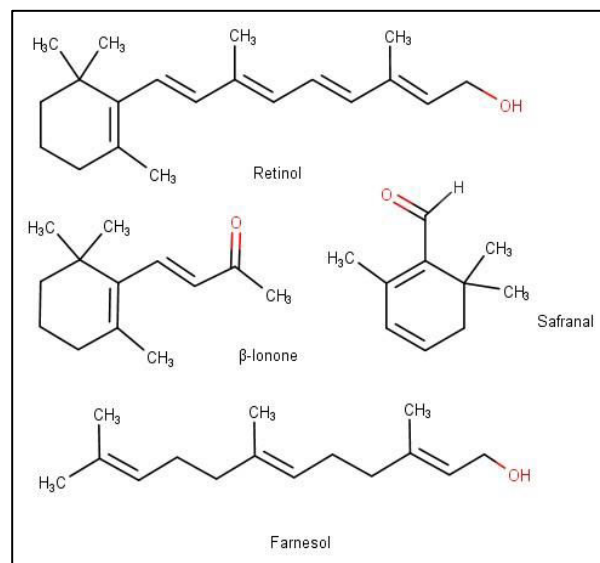


figure 28: structures of Retinol,  $\beta$ -Ionone, Farnesol and Safranal

In figure 29, the quenching curves with the ligands Butyl benzoate, Hexyl benzoate, Octyl benzoate and 3,7-Dimethyloctyl benzoate are displayed. The trend shows, that the longer sidechains fit better into the binding pocket.

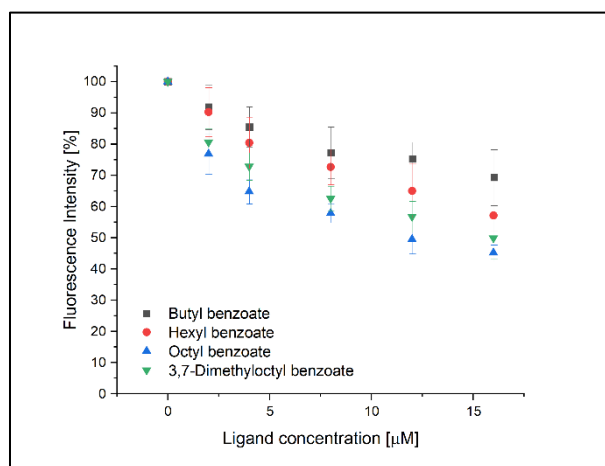


figure 29: quenching curves for Butyl benzoate, Hexyl benzoate, Octyl benzoate and 3,7-Dimethyloctyl benzoate

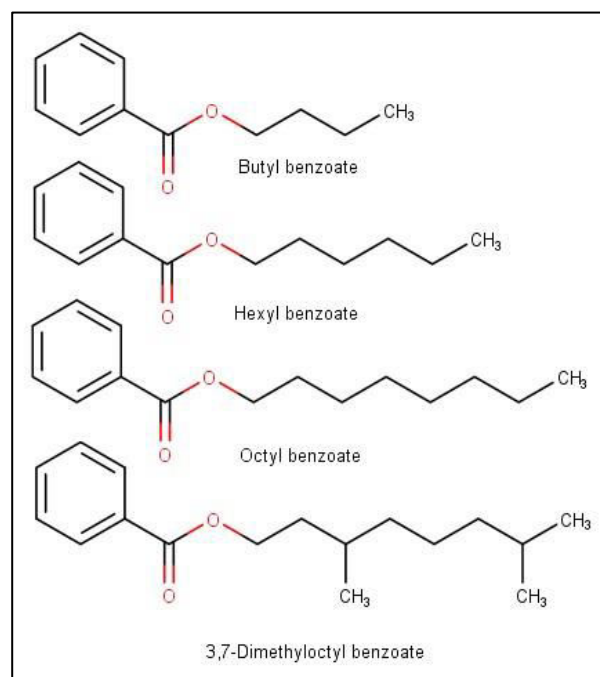


figure 30: chemical structures of Butyl benzoate, Hexyl benzoate, Octyl benzoate and 3,7-Dimethyloctyl benzoate

In figure 31, the quenching curves with the ligands (+) Butyl fenchol, (-) Butyl fenchol,  $\alpha$ -Santalol, 5 $\alpha$ -Androstan-17 $\beta$ -ol-3-one, Androstenone and Quercetin are shown. Only Quercetin exhibits considerable probe displacement. The steroid compounds 5 $\alpha$ -Androstane-

17 $\beta$ -ol-3-one and Androstenone do not bind the protein at all and therefore make other steroids as putative pheromones unlikely. In the binding curve for Androstenone an increase in fluorescence for ligand concentrations larger than 4  $\mu$ M can be seen. This might be due to micellar formation, which increases the fluorescence of trapped 1-NPN<sup>99, 88, 100</sup>. The terpenoids Butyl fenchol and  $\alpha$ -Santalol might not bind well due to their bridged ring structure.

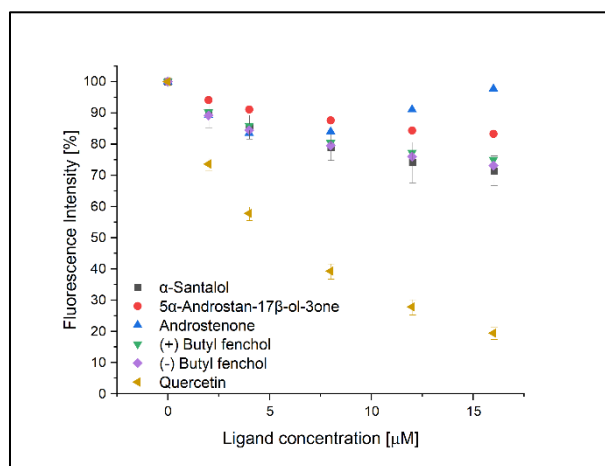


figure 31: quenching curves for (+) Butyl fenchol, (-) Butyl fenchol,  $\alpha$ -Santalol, 5 $\alpha$ -Androstan-17 $\beta$ -ol-3-one, Androstenone and Quercetin

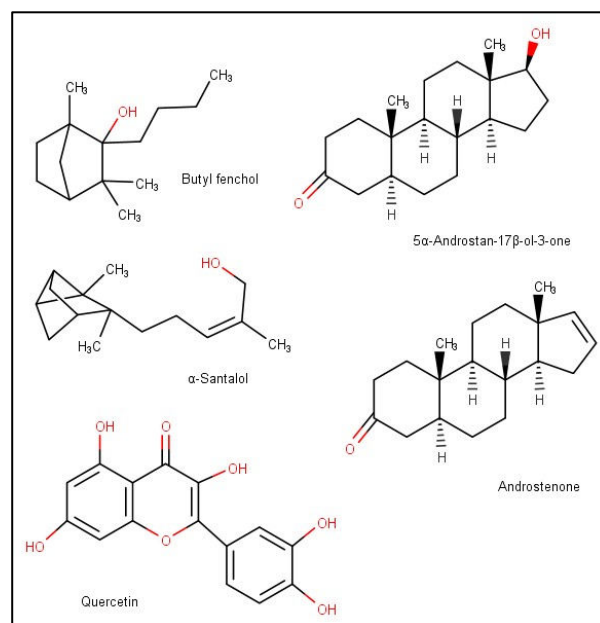


figure 32: chemical structures of (+) Butyl fenchol, (-) Butyl fenchol,  $\alpha$ -Santalol, 5 $\alpha$ -Androstan-17 $\beta$ -ol-3-one, Androstenone and Quercetin

In figure 33, the quenching curves with the ligands 1-Dodecanol, (Z)-11-Hexadecen-1-ol and 1-Hexadecanol are displayed. The linear alcohols bind the protein to some extent, but the unsaturated (Z)-11-Hexadecan-1-ol displaces the probe even less.

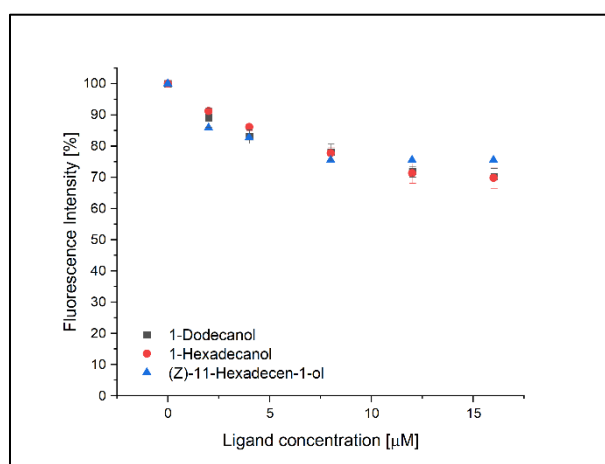


figure 33: quenching curves for 1-Dodecanol, (Z)-11-Hexadecen-1-ol and 1-Hexadecanol

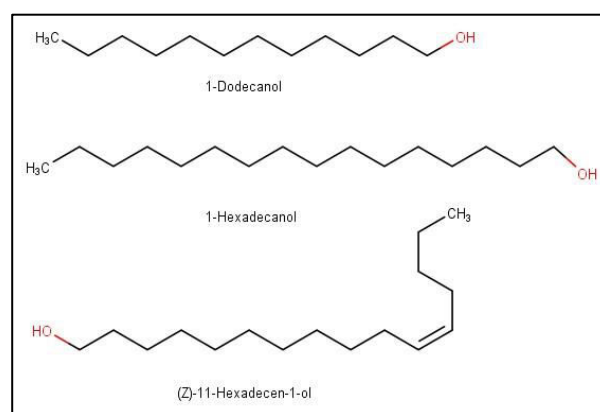


figure 34: chemical structures of 1-Dodecanol, (Z)-11-Hexadecen-1-ol and 1-Hexadecanol

In figure 35, the quenching curves with the ligands (Z)-13-Octadecenal, (Z)-9-Tetradecenal, (Z)-9-Hexadecenal, Dodecanal, 3,7,7-Trimetyloctanal and (Z)-11-Hexadecenal are shown.

None of these aldehydes can be considered similar to the putative pheromone since they don't displace the fluorescence probe enough.

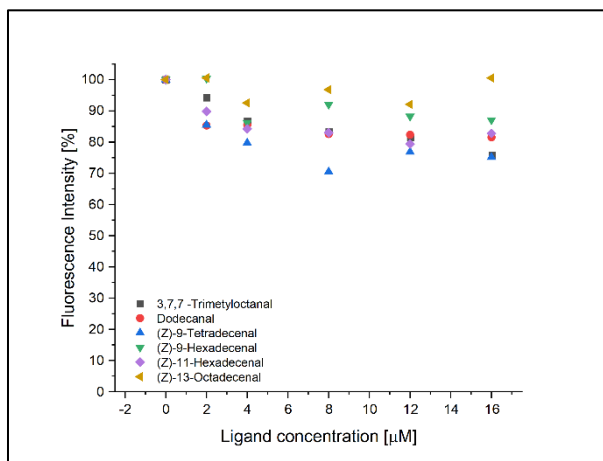


figure 35: quenching curves for (Z)-13-Octadecenal, (Z)-9-Tetradecenal, (Z)-9-Hexadecenal, Dodecanal, 3,7,7-Trimetyloctanal and (Z)-11-Hexadecenal

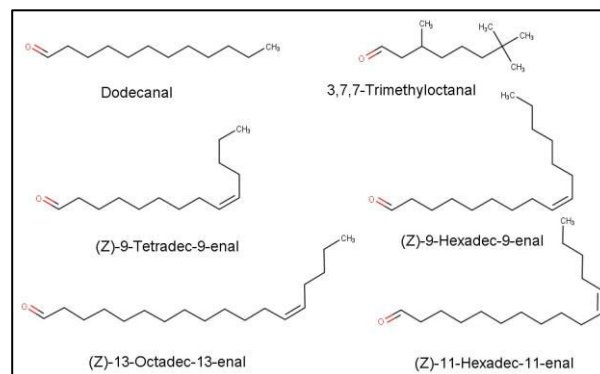


figure 36: chemical structures of (Z)-13-Octadecenal, (Z)-9-Tetradecenal, (Z)-9-Hexadecenal, Dodecanal, 3,7,7-Trimetyloctanal and (Z)-11-Hexadecenal

In figure 37, the quenching curves with the ligands Piperonyl alcohol, Homovanillic acid, Coniferyl aldehyde and Vanillin are displayed. Coniferyl aldehyde shows good binding affinities, vanillin binds a bit but Piperonyl alcohol and Homovanillic acid do not bind at all.

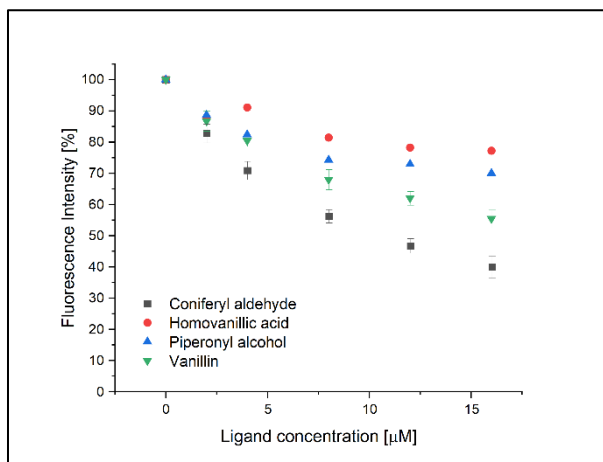


figure 37: quenching curves for Piperonyl alcohol, Homovanillic acid, Coniferyl aldehyde and Vanillin

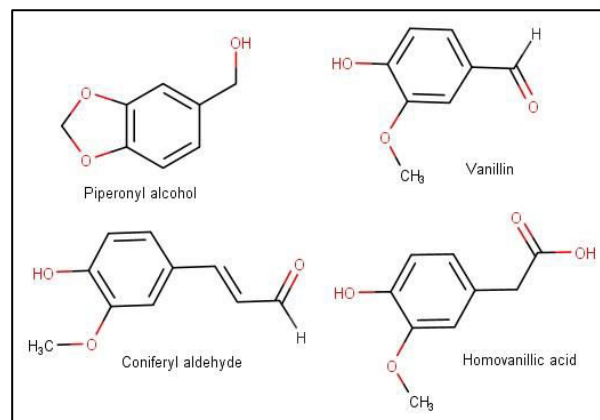


figure 38: chemical structures of Piperonyl alcohol, Homovanillic acid, Coniferyl aldehyde and Vanillin

In figure 39, the quenching curves with the ligands Vanillylamine, Eugenol and Nonivamide are shown. None of them displaces the probe significantly

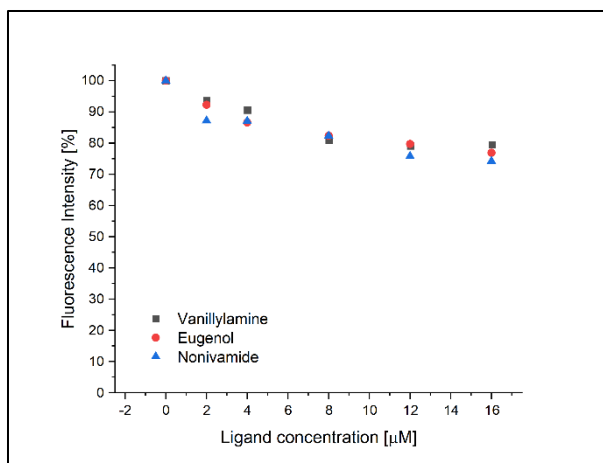


figure 39: quenching curves for Vanillyllamine, Eugenol and Nonivamide

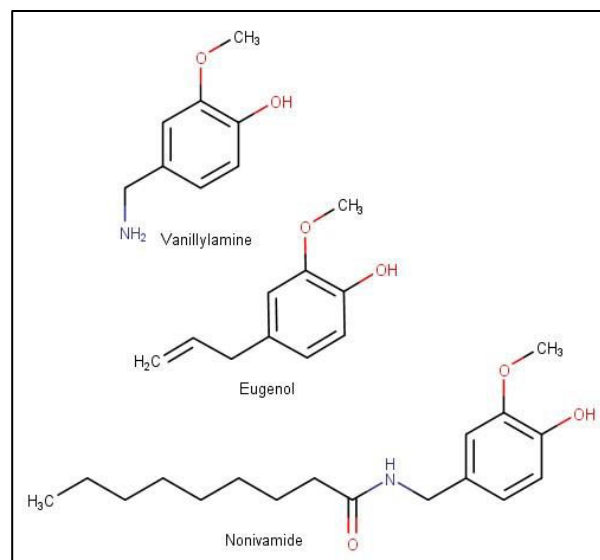


figure 40: chemical structures of Vanillyllamine, Eugenol and Nonivamide

In figure 41, the quenching curves with the ligands  $\gamma$ -Undecalactone, Amylcinnamaldehyde, 2-Methoxycinnamaldehyde and Nonanoic acid are displayed. Only Amylcinnamaldehyde binds the protein significantly.

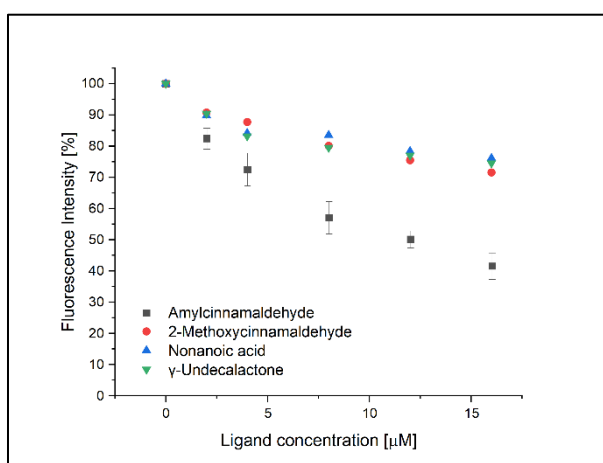


figure 41: quenching curves for  $\gamma$ -Undecalactone, Amylcinnamaldehyde, 2-Methoxycinnamaldehyde and Nonanoic acid

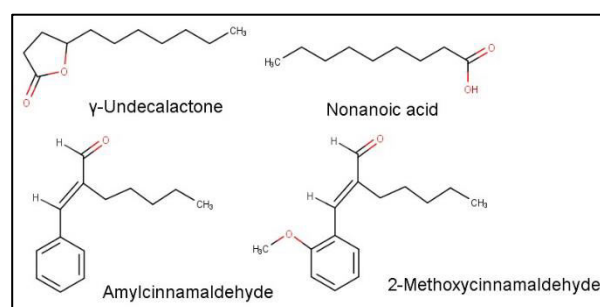


figure 42: chemical structures of  $\gamma$ -Undecalactone, Amylcinnamaldehyde, 2-Methoxycinnamaldehyde and Nonanoic acid

In figure 43, the quenching curves with the ligands Disparlure, Cyclamen aldehyde, Ethyl laureate, Dodecanoic acid are shown. None of these ligands can be considered similar to the structure of the putative pheromone. The fluorescence increases when Disparlure is added. This is most likely due to the hydrophobicity and low critical micellar concentration of Disparlure.



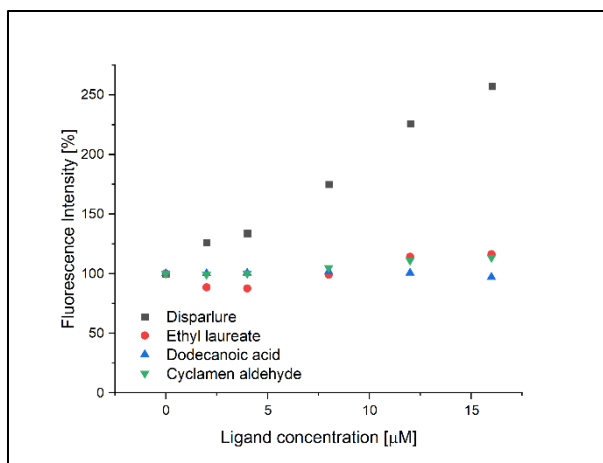


figure 43: quenching curves for Disparlure, Cyclamen aldehyde, Ethyl laureate, Dodecanoic acid

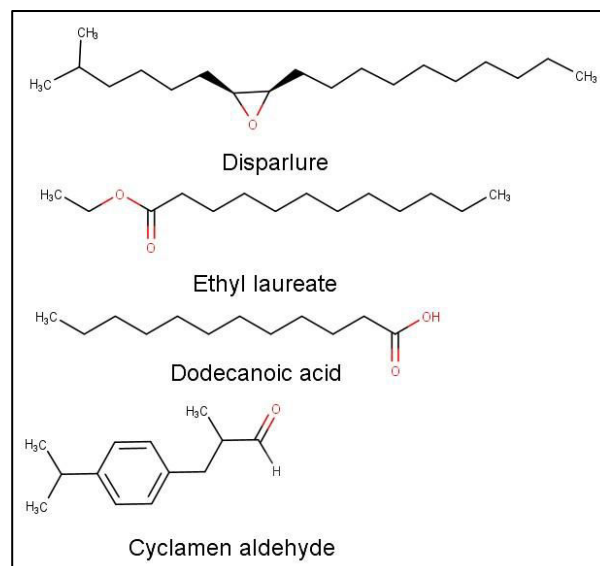


figure 44: chemical structures of Disparlure, Cyclamen aldehyde, Ethyl laureate, Dodecanoic acid

In figure 45, the quenching curves with the ligands Citral, (+) Carvone and Geranyl acetone are displayed. None of these ligands shows good affinities to the binding pocket of the protein.

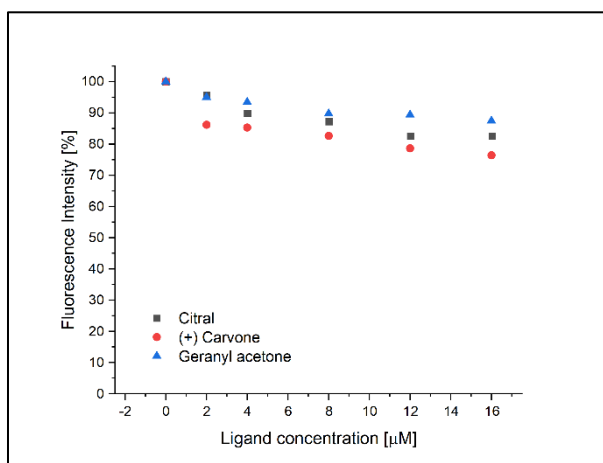


figure 45: quenching curves for Citral, (+) Carvone and Geranyl acetone

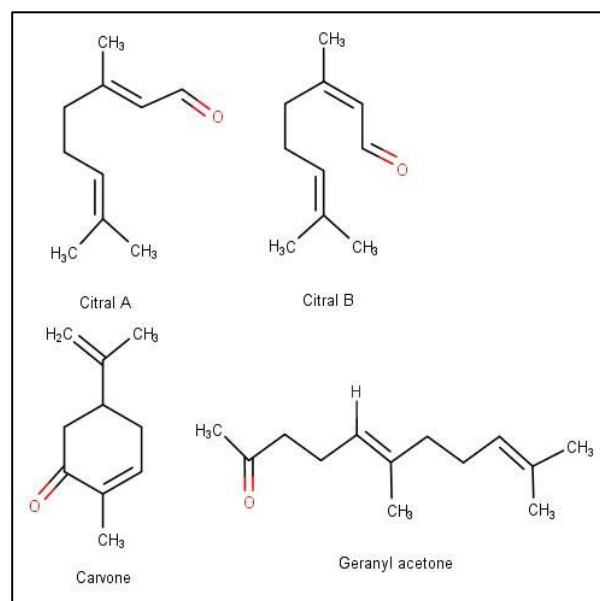


figure 46: chemical structures of Citral A and Citral B, (+) Carvone and Geranyl acetone

### Dissociation constants

For the Ligands, that reached below 50 % fluorescence intensity at 16  $\mu\text{M}$  ligand, the Dissociation constants were calculated. The graphs used for these calculations are the mean of three measurements. The error bars represent  $\pm$  one standard derivation.

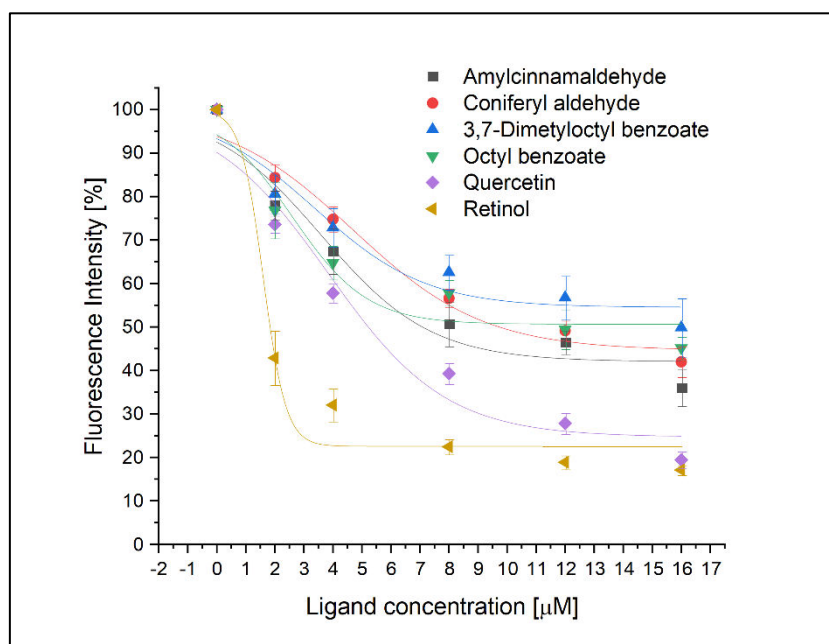


figure 47: plot of the best quenching curves for the h-SAL protein. The parameters of the fitted curves can be found in table 23. The values plotted are the mean of three measurements.

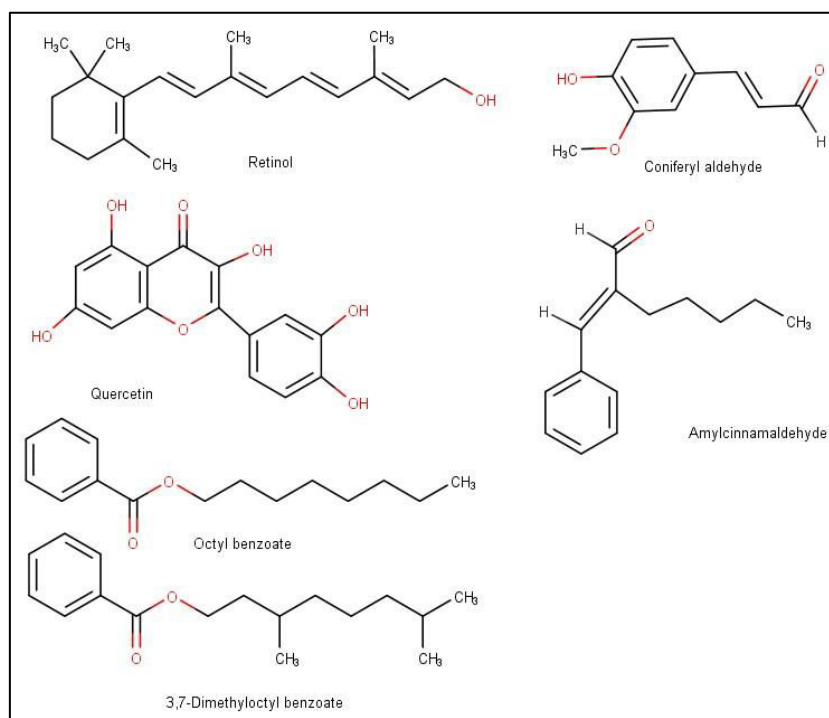


figure 48: chemical structures of the best ligands for the h-SAL protein

table 23: parameters of the fitted curves in figure 47. The parameter A2 was fixed to a value of 100 % for all curves since these were the measurements without ligand present.

	Quercetin	Octyl benzoate	Coniferyl aldehyde	Amyl-cinnamaldehyde	3,7-Dimethyloctyl benzoate	Retinol
A1 [%]	25	51	45	42	54	23
$\sigma$ A1 [%]	6	4	47	5	4	3
Logx0 [ $\mu$ M]	3.8	2.6	4.7	3.6	3.4	1.6
$\sigma$ logx0 [ $\mu$ M]	0.8	0.6	0.8	0.8	0.9	0.4
p	-0.2	-0.3	-0.2	-0.2	-0.2	-1.1
$\sigma$ p	0.1	0.2	0.1	0.1	0.1	1.2

table 24: calculated values for the dissociation constants.  $K_d = 8.7 \mu\text{M}$  (NPN) see table 22,  $[L] = 2 \mu\text{M}$

	Logx0 = IC50 [ $\mu$ M]	$1+[L]/K_d$	$K_{diss}$ [ $\mu$ M]
Quercetin	3.8	1.2	3.1
Octyl benzoate	2.6	1.2	2.1
Coniferyl aldehyde	4.7	1.2	3.8
Amylcinnamaldehyde	3.6	1.2	2.9
3,7-Dimethyloctyl benzoate	3.4	1.2	2.8
Retinol	1.6	1.2	1.3

### Theoretical predictions

Docking experiments using the h-SAL protein model and 1-NPN, Retinol,  $\beta$ -Ionone and Quercetin were performed using SwissDock<sup>101, 102</sup>. Only  $\beta$ -Ionone was found to bind inside the binding pocket with these docking experiments. The binding pocket of the protein model is either smaller or the expressed protein changes conformation upon binding. Either way, it can be concluded that there is a substantial difference between the theoretical model and the folding of the expressed protein during binding.

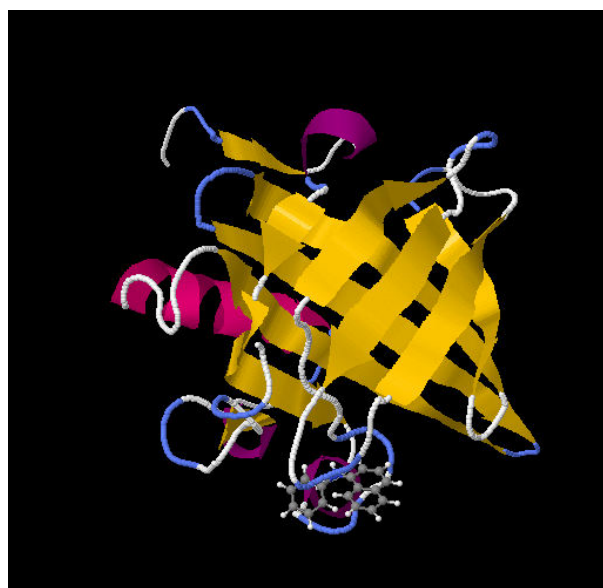


figure 49: docking of the h-SAL protein with 1-NPN



figure 50: docking of the h-SAL protein with Retinol

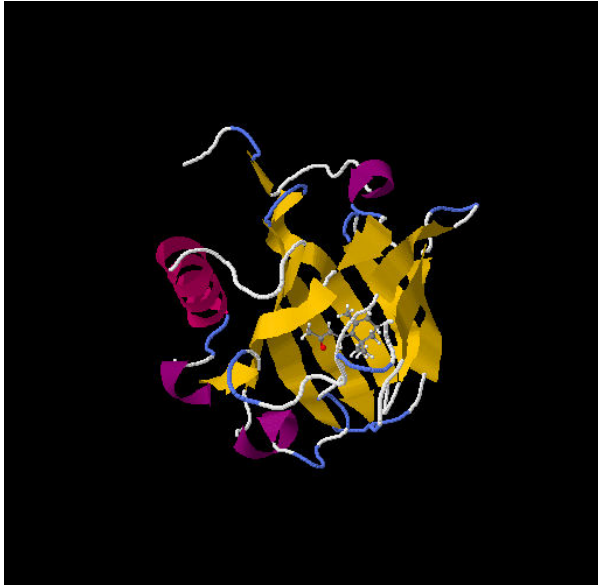


figure 51: docking of the h-SAL protein with  $\beta$ -Ionone

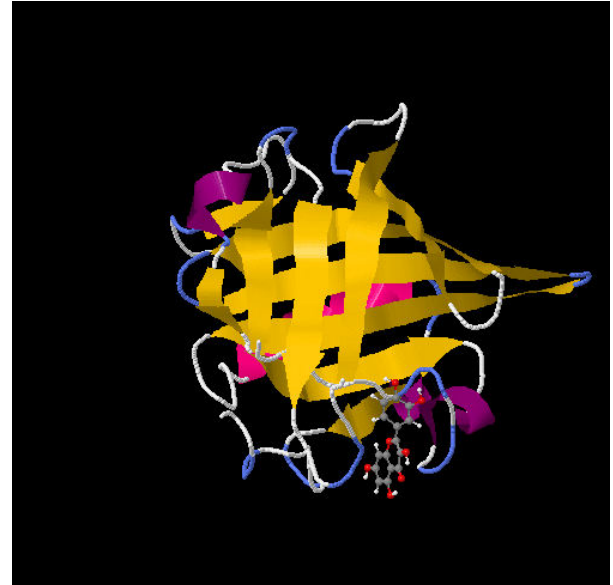


figure 52: docking of the h-SAL protein with Quercetin

### Tryptophan quenching

Tryptophan quenching experiments were done with Retinol and  $\beta$ -Ionone to investigate if the 1-NPN binding was influencing the ligand binding results. Both experiments were done as a triple measurement.

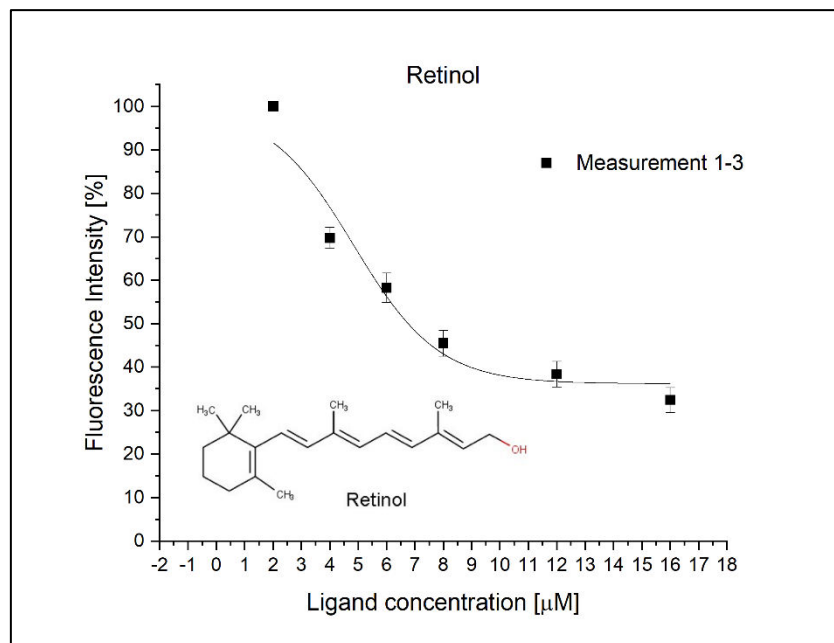


figure 53: tryptophan quenching experiments with Retinol. The parameters for the fitted curves can be found in table 25

table 25: parameters of the fitted curves in Fehler! Verweisquelle konnte nicht gefunden werden.. The parameter A2 was fixed to a value of 100 % for all curves since these were the measurements with the lowest ligand concentration

	Retinol Measurement 1-3
A1 [%]	36

$\sigma_{A1}$ [%]	5
$\text{Log}x_0$ [ $\mu\text{M}$ ]	4.8
$\sigma_{\text{log}x_0}$ [ $\mu\text{M}$ ]	0.6
$\rho$	-0.3
$\sigma_\rho$	0.1

table 26: calculated values for the dissociation constants.  $[L] = 0$

	$\text{log}x_0 = \text{IC}_{50} = K_d$ [ $\mu\text{M}$ ]
Measurement 1-3	4.8
$\sigma$	0.6

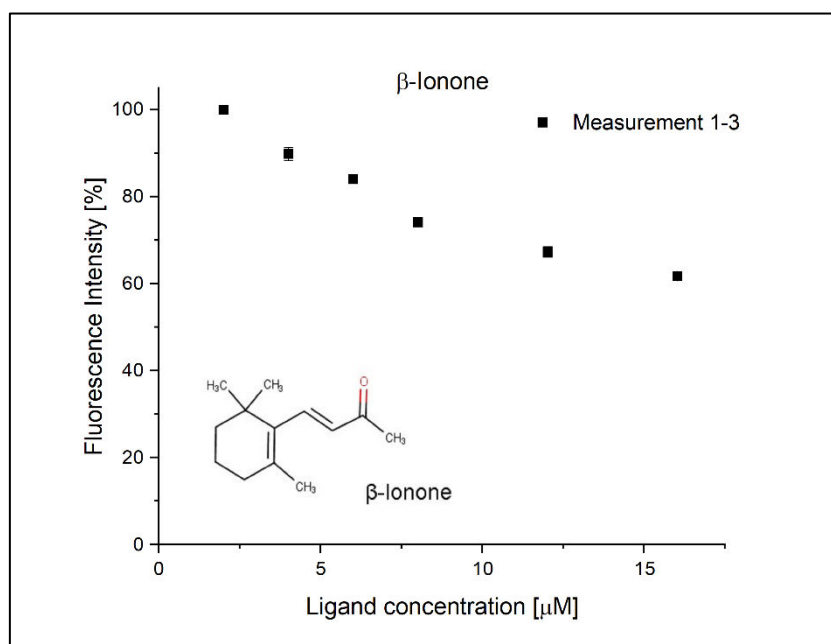


figure 54: tryptophan quenching experiments with  $\beta$ -ionone

Additionally, the data from the 1-NPN binding experiments for Retinol and  $\beta$ -ionone was plotted and evaluated without using the datapoint for 0  $\mu\text{M}$  ligand concentration, since this is how the tryptophan quenching experiments were analysed.

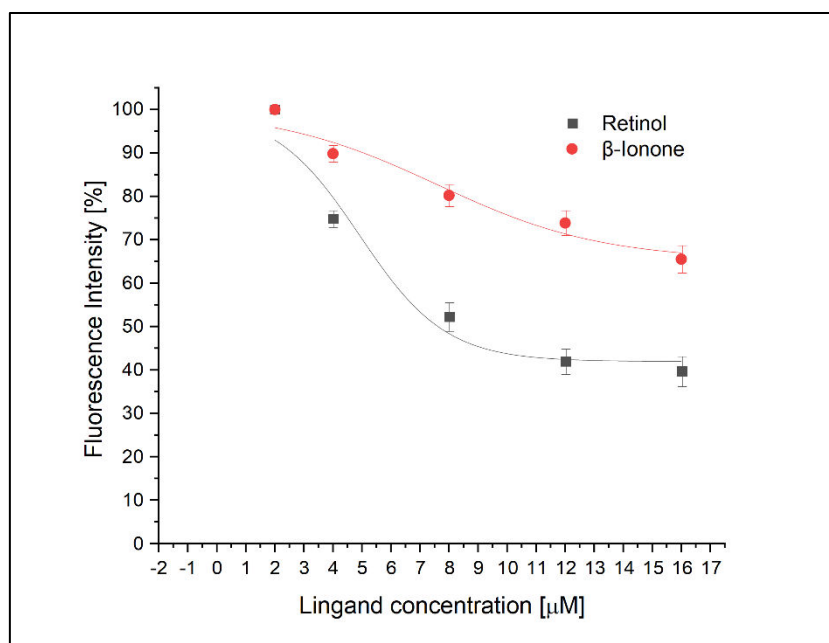


figure 55: 1-NPN quenching experiments for Retinol and  $\beta$ -Ionone. The datapoint for 0  $\mu$ M Ligand has been omitted and 2  $\mu$ M Ligand has been taken as 100 % fluorescence intensity in order to be comparable to the tryptophan quenching data

table 27: parameters of the fitted curves in figure 55. The parameter A2 was fixed to a value of 100 % for all curves since this is how the tryptophan quenching experiments were evaluated

	Retinol	B-Ionone
A1 [%]	42	65
$\sigma$ A1 [%]	5	6
Logx0 [ $\mu$ M]	4.9	7.6
$\sigma$ logx0 [ $\mu$ M]	0.9	1.8
p	-0.3	-0.2
$\sigma$ p	0.1	0.1

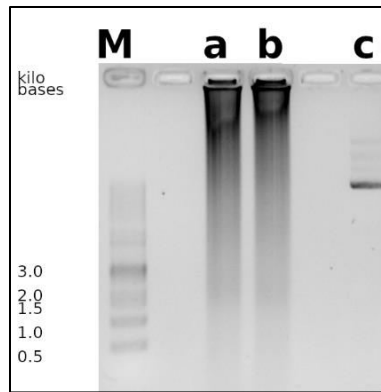
The calculation of the dissociation constant for Retinol was  $IC_{50} = K_{diss} = \log_{x_0} = 4.9 \mu M$ .

### 3.4 Protein expression of the h-SAL HEIE Mutant

#### 3.4.1 Introduction of the HEIE mutation into hSAL

##### Overlapping primers:

This approach proved to be not successful. The PCR yielded products of various length, which lead to a smeared band and some precipitate in the pockets, when run on an agarose gel and did not transform into *E. coli* DH5 $\alpha$ . This can be seen in figure 56 and did not improve by using double the amount of primers.



*figure 56: agarose gel electrophoresis M: Marker. a: 1st try before DpnI digestion. (see table 7) b: 1st try after DpnI digestion. c: pET30a h-SAL*

### *Two-step approach*

This method could also be done in one step, without gel extraction of the elongated primers, although the result proved to be more favourable if divided. This approach led to two colonies, which were both grown and sent for sequencing. The sequencing results can be seen in additional information. Both colonies contained the desired mutation.

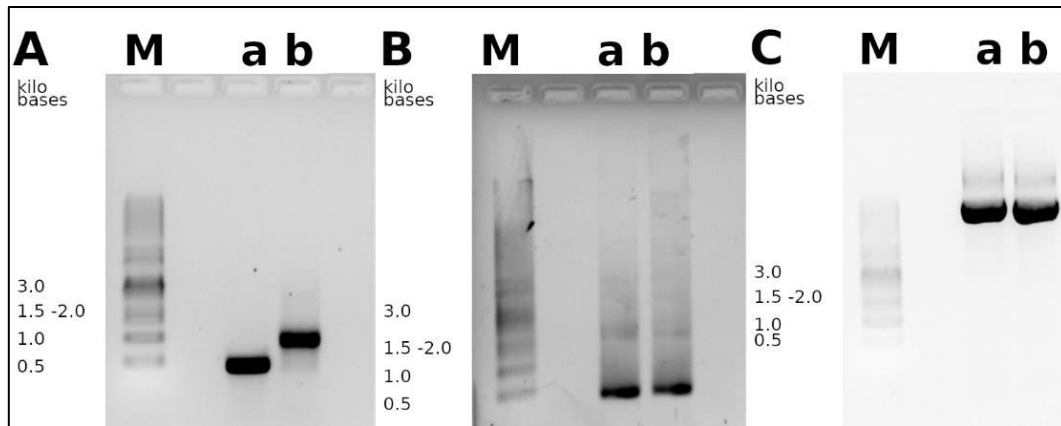


figure 57: agarose gel electrophoresis. M: Marker. A: after the first step and gel extraction. a: PCR tube 1 (used further) b: PCR tube 2 (discarded) B: after the second step. a: before DpnI digestion b: after DpnI digestion. C: Plasmids grown from colonies a: Colony 5 (not used) b: Colony 6 (used further)

### 3.4.2 Protein expression

The h-SAL protein HEIE mutant was expressed as described in “2.2.4 Protein expression”. More protein was found in the supernatant than the pellet (see figure 58). Still, the pellet was purified since this approach led to more pure protein in the h-SAL expressions.

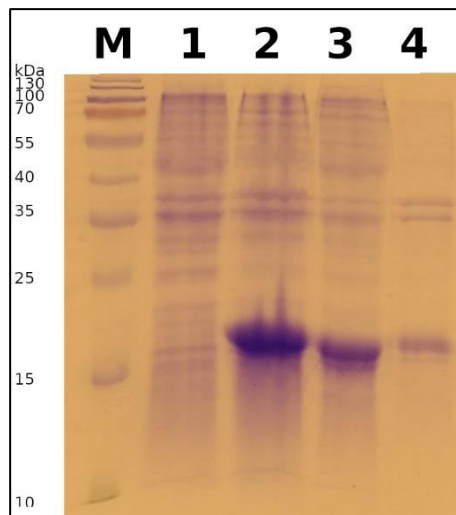


figure 58: SDS page after expression of h-SAL HEIE mutation. Coomassie-stained gel. M: Marker 1: bacterial sample before inducing. 2: bacterial sample after inducing. 3: supernatant after sonication and centrifugation. 4: pellet after sonication and centrifugation

### 3.4.3 h-SAL HEIE mutant protein purification

The proteins were solubilised in 8M Urea and the disulphide bridges of the proteins were reduced with 50 mM DDT. The solution was then dialysed 3 times for 8-16 hours against 50 mM Tris pH 7.4. Then it was centrifugated at 12000 rpm, at 4°C for 1 hour. Consequently, the protein solution was purified by anion exchange chromatography as described in 3.3.4 as “The new purification approach”.



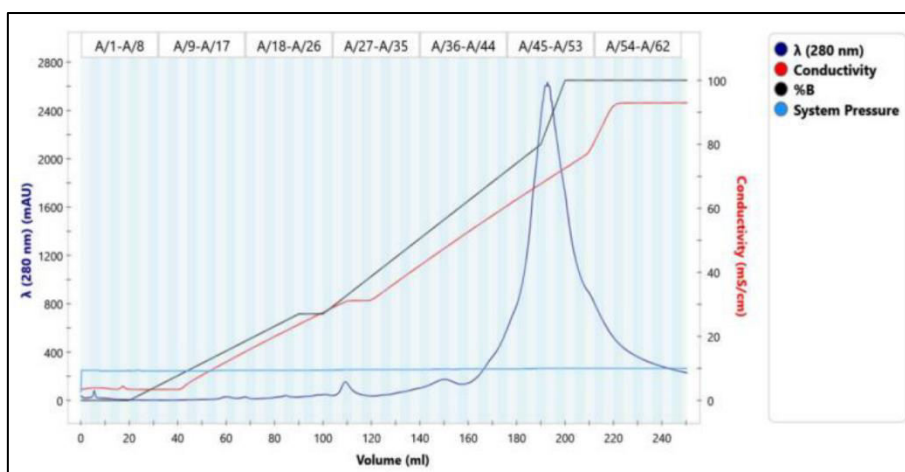


figure 59: chromatogram of the purification run for the h-SAL protein HEIE mutation by the improved anion exchange chromatography

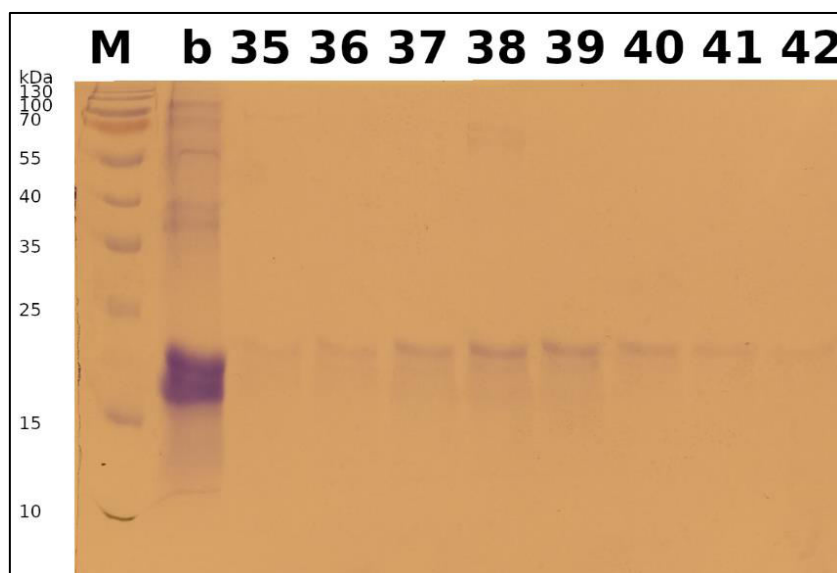


figure 60: coomassie-stained SDS PAGE after purification with the improved anion exchange program. The lanes are labelled with fraction numbers. M: Marker. b: corresponding cell culture after inducing with IPTG

### 3.4.4 Protein characterization

#### Amount of protein obtained

The amount of protein in the fractions was calculated using the extinction coefficient obtained from ExPASy<sup>97</sup>. Given that all pairs of cysteines form cystines, the extinction coefficient at 280 nm is  $19,285 \text{ M}^{-1} \text{ cm}^{-1}$  when measured in water, which is the same extinction coefficient as the non-mutated h-SAL protein.

The absorbance at 260 nm was also measured, to evaluate if the amount of DNA would significantly influence the measurement at 280. Three sufficiently pure fractions were obtained and mixed, in order to minimise errors associated with differences from the different fractions. The result can be found in table 28.

table 28: available mixed fraction after purification, absorption at different wavelengths and amount of protein calculated. Fraction volume of 12 ml in total.

Fraction	Abs at 280	Abs at 260	[mol/l]
36-39	0.267	0.201	0.014

### Finding a suitable fluorescence probe:

The fluorescence probe 1-NPN was used in accordance with the previous experiments.

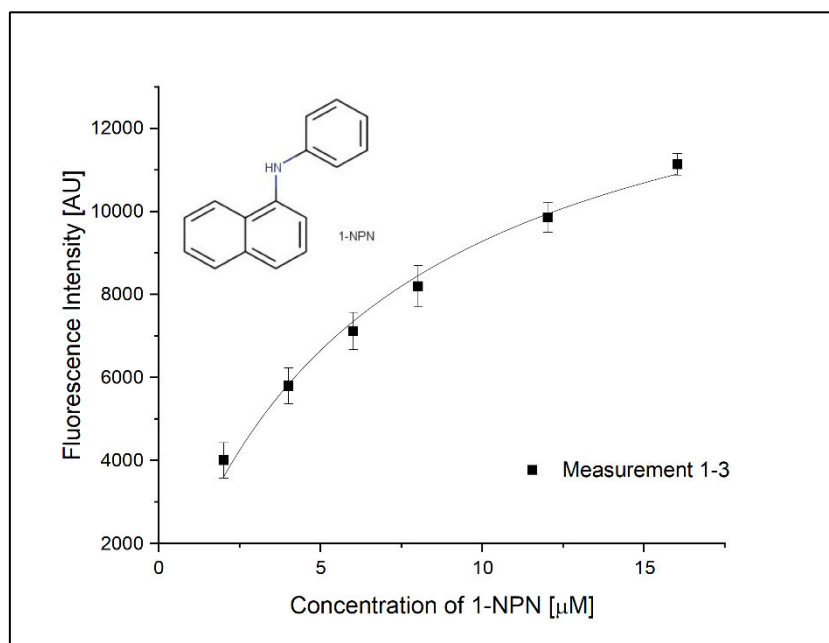


figure 61: the 1-NPN binding experiment using the h-SAL HEIE mutation fraction

table 29: fitted parameters from the 1-NPN binding experiment using the h-SAL HEIE mutation fraction (see figure 61)

Measurement	Measurement 1-3
$B_{\max}$ [AU]	15356
$\sigma B_{\max}$ [AU]	798
$k_1$ [ $\mu\text{M}$ ]	6.6
$\sigma k_1$ [ $\mu\text{M}$ ]	0.8

It can be noted, that the dissociation constant is comparable to the one obtained from the h-SAL measurements (see table 22).

### Ligand-binding experiments

For the ligand-binding experiments 2  $\mu\text{M}$  1-NPN was used in a competitive assay with different concentrations of ligand. Only ligands, that displayed binding with the h-SAL protein were tested. The best ligands, namely Quercetin, Octyl benzoate, Coniferyl aldehyde, Amyl cinnamaldehyde, 3,7-Dimethyloctyl benzoate and Retinol were measured in triplets. In the following graphs, the mean of the three measurements is plotted. All other ligands were only measured once, due to time constraints. An exception to this rule is Safranal and Hexyl benzoate, which were also measured three times. If the measurement has been done three or more times, error bars corresponding to  $\pm$  one standard derivation have been reported. The curves obtained from the initial measurements were not fitted, since they were measured only once.

In figure 62, the quenching curves with the ligands Retinol,  $\beta$ -Ionone, Farnesol and Safranal are displayed. As in the binding curves with h-SAL, Retinol is the best ligand.

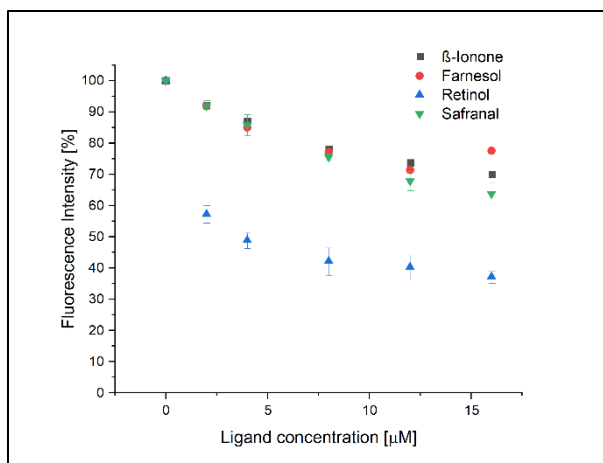


figure 62: quenching curves for Retinol,  $\beta$ -Ionone, Farnesol and Safranal

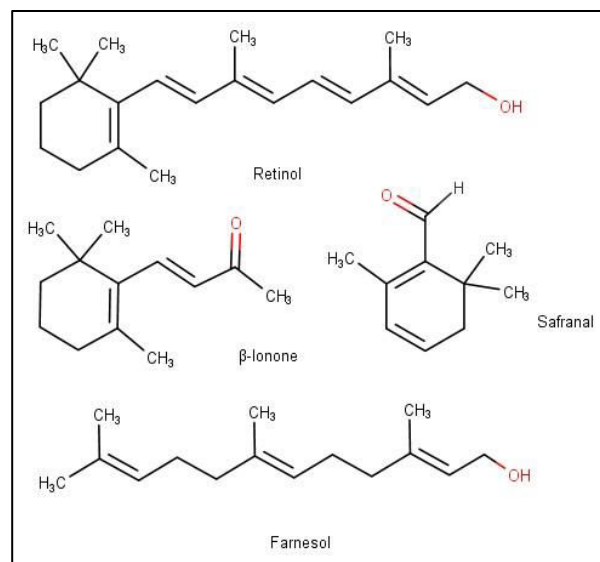


figure 63: chemical structures of Retinol,  $\beta$ -Ionone, Farnesol and Safranal

In figure 64, the quenching curves with the ligands Butyl benzoate, Hexyl benzoate, Octyl benzoate and 3,7-Dimethyloctyl benzoate are shown. The same trend as for h-SAL, longer sidechains fitting better into the h-SAL HEIE mutation protein, can be observed.

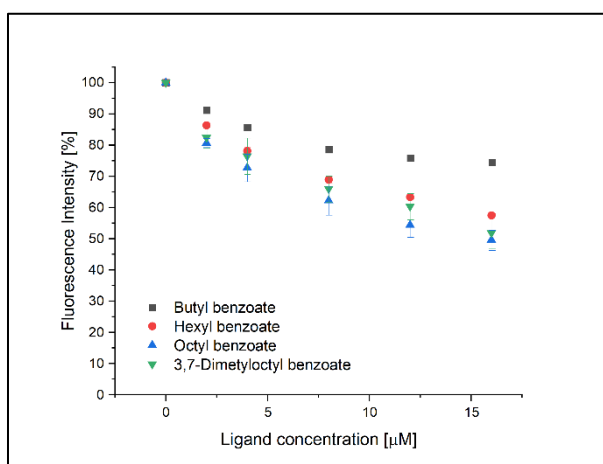


figure 64: quenching curves for Butyl benzoate, Hexyl benzoate, Octyl benzoate and 3,7-Dimethyloctyl benzoate

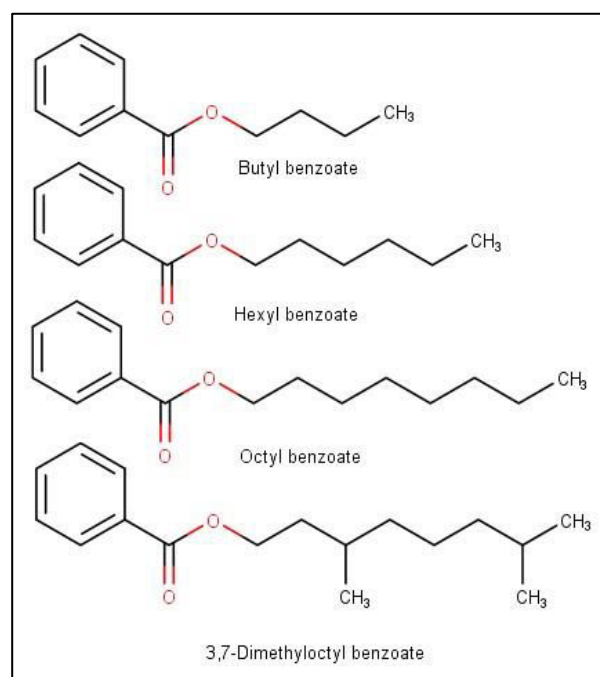


figure 65: chemical structures of Butyl benzoate, Hexyl benzoate, Octyl benzoate and 3,7-Dimethyloctyl benzoate

In figure 66, the quenching curves with the ligands  $\alpha$ -Santalol, Quercetin, 1-Dodecanol and 1-Hexadecanol are shown. Of these ligands, only Quercetin displays considerable probe displacement. The other compounds bind the h-SAL HEIE mutation protein weaker.

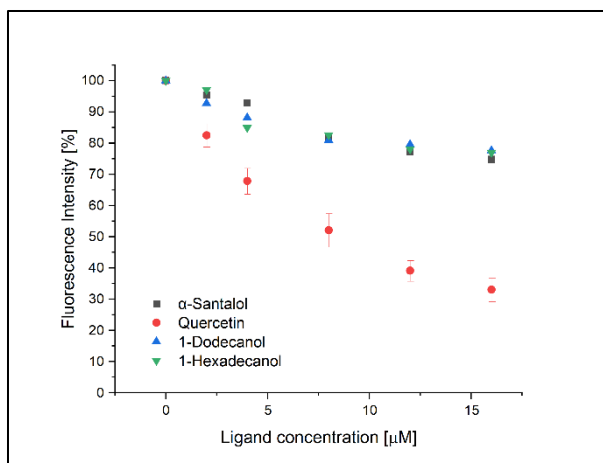


figure 66: quenching curves for  $\alpha$ -Santalol, Quercetin, 1-Dodecanol and 1-Hexadecanol

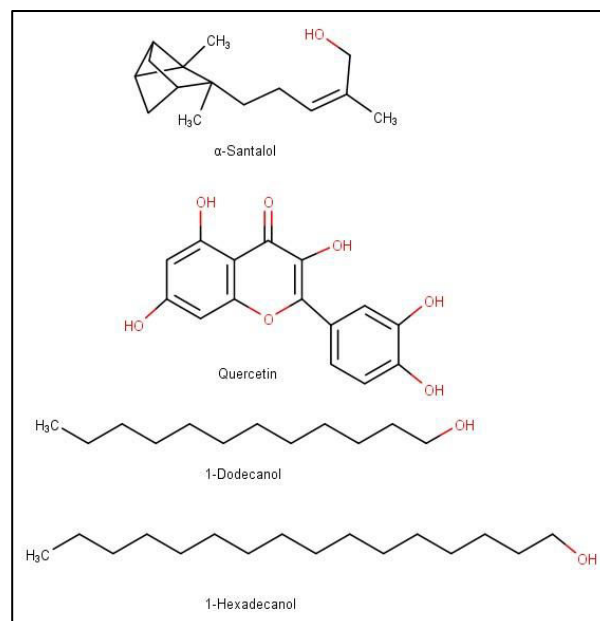


figure 67: chemical structures of  $\alpha$ -Santalol, Quercetin, 1-Dodecanol and 1-Hexadecanol

In figure 68, the quenching curves with the ligands Amylcinnamaldehyde, Coniferyl aldehyde and Vanillin are shown. Of the three ligands, only Vanillin doesn't reach below 50 % fluorescence intensity at 16  $\mu$ M ligand concentration.

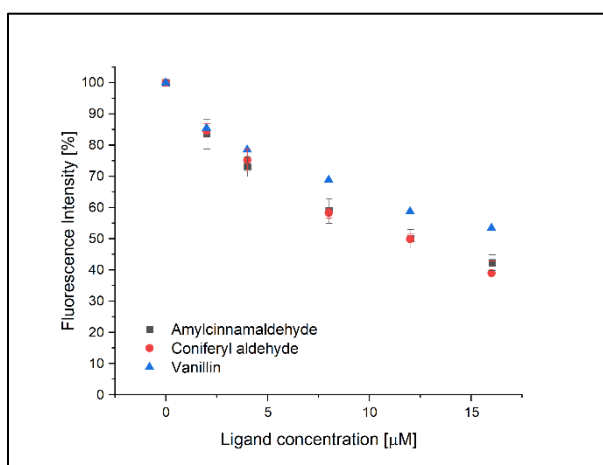


figure 68: quenching curves for Amylcinnamaldehyde, Coniferyl aldehyde and Vanillin

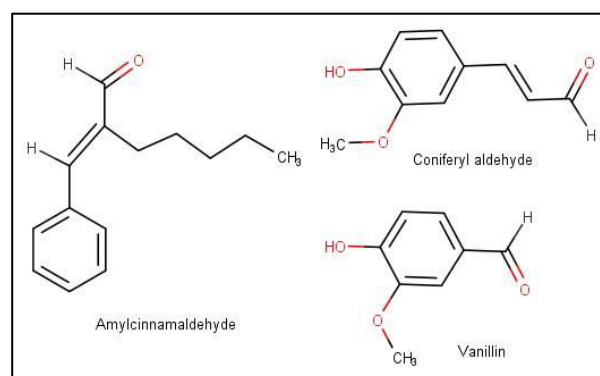


figure 69: chemical structures of Amylcinnamaldehyde, Coniferyl aldehyde and Vanillin

### 3.4.5 Fitting of Dissociation constants

For the Ligands, that reached below 50 % fluorescence intensity at 16  $\mu\text{M}$  ligand, the Dissociation constants were calculated. The graphs used for these calculations are the mean of three measurements. The error bars represent  $\pm$  one standard derivation.

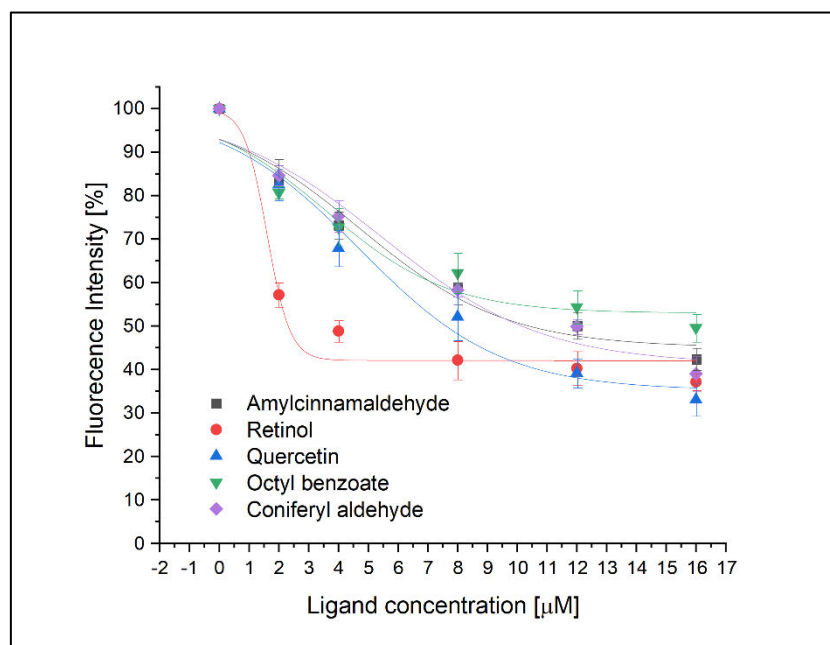


figure 70: plot of the best quenching curves for the h-SAL protein HEIE mutation. The parameters of the fitted curves can be found in table 30. The values plotted are the mean of three measurements.

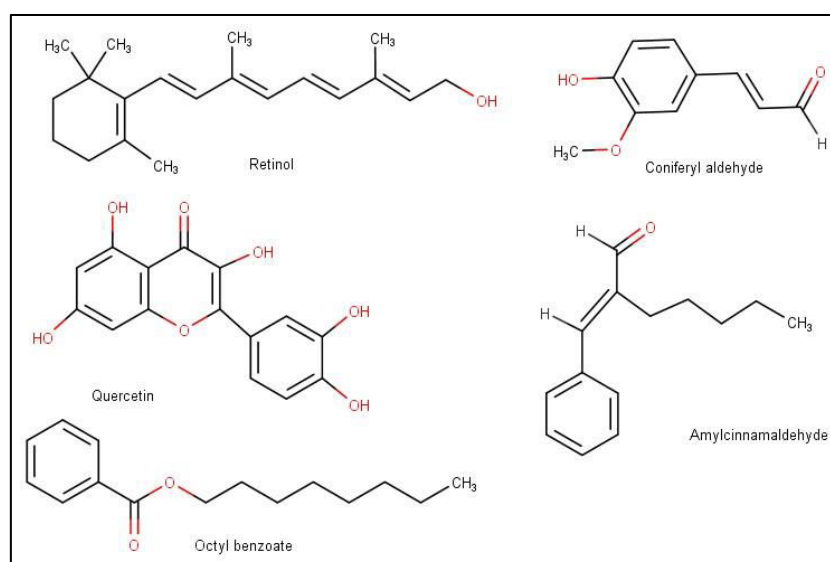


figure 71: chemical structures of the best ligands for the h-SAL protein HEIE mutant

table 30: parameters of the fitted curves in figure 70. The parameter A2 was fixed to a value of 100 % for all curves since these were the measurements without ligand present.

	Quercetin	Octyl benzoate	Coniferyl aldehyde	Amyl-cinnamaldehyde	Retinol
A1 [%]	35	53	41	45	42
$\sigma$ A1 [%]	6	4	6	5	2
Logx0 [ $\mu$ M]	4.7	3.6	5.4	4.7	1.6
$\sigma$ logx0 [ $\mu$ M]	0.9	0.9	1.1	1.0	0.4
p	-0.2	-0.2	-0.2	-0.2	-1.1
$\sigma$ p	0.1	0.1	0.0	0.1	1.2

table 31: calculated values for the dissociation constants.  $K_d = 6.6 \mu$ M (NPN) see table 29,  $[L] = 2 \mu$ M

	Logx0 = IC50 [ $\mu$ M]	$1+[L]/K_d$	Kdiss [ $\mu$ M]
Quercetin	4.7	1.3	3.6
Octyl benzoate	3.6	1.3	2.8
Coniferyl aldehyde	5.4	1.3	4.2
Amyl-cinnamaldehyde	4.7	1.3	3.6
Retinol	1.6	1.3	1.2

## 4. Discussion

### 4.1 Research question

The protein of interest, h-SAL, belongs to a group of proteins, known to carry pheromones in other mammals. Using the reverse chemical ecology approach, the aim of this thesis was to gain insight into the pheromonic communication of the hominids, in which the gene was still functional and how pheromonic communication in humans would work, if this gene had not been switched off. In order to do this, both protein sequences found have been expressed and their ligand binding properties have been studied.

The main problem that could not be solved is, that random mutation changed the genomic sequence in the approximately 400 000 years since the selective pressure on this gene has been relaxed. This is the main reason, why the results of this thesis have to be taken with caution. However, the best ligands from the binding studies might be taken as a starting point when studying other humanoid and ape species.

### 4.2 Improvements in h-SAL protein purification

Starting from the initial purification approach there were two main improvements made. The first one is the switch to purification from the pellet, instead of from the supernatant. The main advantage of this approach was already highly reduced levels of impurities in the pellet and therefore fewer purification steps needed to gain a pure fraction. The second improvement was the modification of the chromatographic program. The main improvement here was the switch to higher NaCl concentrations in Buffer B. This was necessary since the protein-peak eluted at 100 % buffer B and tailed considerably at the lower salt concentration. Another improvement was elongating the gradient between 27 and 80 % Buffer B (1M). With the higher salt concentration, the DNA peak eluted close to the protein peak, resulting in with DNA contaminated protein fractions. Elongating the gradient in this region lead to more and less concentrated fractions, but also less contaminated fractions.

As for the size exclusion chromatography, it was showed that this approach did not result in improved purity of the yielded fractions. A possible reason could be that the h-SAL protein might form high molecular weight aggregates, which can't be separated from the monomeric molecule. This theory will be tested in further experiments.

Ammonium sulphate precipitation proved to be successful since the protein precipitates between 33 and 39 % ammonium sulphate saturation. However, the following purification via anion exchange chromatography needs to be improved, since there was a substantial amount of protein not binding to the resin at all. A possible reason could be that the dialysis was not sufficient and dialysing for a longer time period could be worth trying. In the end, ammonium sulphate precipitation was successful and could, after some improvements in the subsequent chromatography, be included in future purification strategies.

After the purification it would be useful to verify the proteins identity by mass spectrometry and the proteins folding with circular dichroism. Due to financial and organisational constraints these measurements were not part of this thesis. Overexpression of a protein of the desired molecular weight was used as indication, that expression was successful.

### 4.3 Different approaches for introducing the HEIE mutation into h-SAL

Two different approaches were tested for introducing the mutation into the h-SAL gene. The approach, using the overlapping primers was not successful. In figure 56 precipitate in the

pockets of the agarose gel can be clearly seen. This might be due to the primer design since the overlapping region on both ends of the primer, which flank the desired insert, are only 3 and 12 bases long in case of the forward primer and 6 and 9 bases long in case of the reverse primer. Therefore, annealing of reverse primer to the forward primer instead of the plasmid template might have been a problem.

The two-step approach solved this problem, but also produced a lot of undesired non-mutated PCR product (the forward strand is synthesised without mutation in the first round of PCR and also to a large extent in the following rounds). This is why it is common to use two mutated primers. Although cleaning up the primers after the first steps of the PCR would not be necessary in theory, it served to purify them. In the end, this work proves that using only one mutated primer also leads to the desired product.

#### 4.4-NPN fluorescence quenching vs. tryptophan quenching

Two different quenching experiments were set up, using the h-SAL protein, in order to investigate if the usage of the fluorescence probe 1-NPN has an influence on the ligand binding results.

##### 4.4.1 Theoretical predictions and docking experiments

When comparing docking experiments with Retinol, Quercetin, 1-NPN and  $\beta$ -Ionone, only  $\beta$ -Ionone was found in the binding pocket on the inside of the protein. This could be due to the inflexibility of the protein model and there could be a change in the 3-dimensional structure of the protein when certain ligands bind. Since 1-NPN was also found outside the binding pocket in this docking experiments, tryptophan quenching experiments were set up to verify that Retinol does indeed bind the h-SAL protein better than  $\beta$ -Ionone and the results obtained can be considered trustworthy.

##### 4.4.2 Retinol

*table 32: comparison of fluorescence quenching experiments using 1-NPN or tryptophan quenching, with Retinol as the ligand*

	$K_{diss}$ [ $\mu$ M]
1-NPN fluorescence quenching	4.9
Tryptophan quenching	4.8

The Dissociation constant obtained from the tryptophan quenching experiment does not differ from the one obtained by 1-NPN fluorescence quenching when calculated with the same ligand concentration set as 100 % fluorescence intensity. However, when using ligand concentration 0  $\mu$ M as the reference point the dissociation constant for Retinol is found to be 1.3  $\mu$ M. This means the assumption that 2  $\mu$ M ligand concentration lies on the plateau of the quenching curve, which is used to enable comparable curve fitting, is fundamentally wrong and the results obtained from this evaluation are wrong. Very low, as well as very high concentrations of ligands should have been measured as this data would be needed for normalisation of the curve. This was, however, not possible due to experimental difficulties and a evaluation based on 1-NPN binding was chosen instead, since this method proved to provide more reliable information.

##### 4.4.3 $\beta$ -Ionone



table 33: comparison of the mean values ( $n=3$ ) for the  $\beta$ -Ionone 1-NPN and tryptophan fluorescence quenching experiments

Ligand concentration [ $\mu\text{M}$ ]	1-NPN fluorescence quenching [%]	Tryptophan fluorescence quenching [%]
2	100	100
4	90	90
6	-	84
8	80	74
12	74	67
16	66	62

Though this normalisation can't be used to gain any information about dissociation constants, it can be seen that the data is comparable.

#### 4.4.4 Conclusion

The aim of this experiment was to evaluate if the tryptophan quenching experiments and the 1-NPN fluorescence quenching experiments are comparable. Usually, it would have been necessary to measure the binding curves at very low and very high concentrations of ligand to be able to normalise and fit it correctly. In this case, this is not necessary since the normalisation, though it does not give any additional reliable information, is sufficient for this experimental question and shows that the measurements done with the 1-NPN fluorescence quenching can be trusted.

#### 4.5 h-SAL protein ligands

The aim of the binding experiments was finding a suitable ligand for the h-SAL protein. In order to do so, different groups of ligands were used, and the dose-response curves were measured by fluorescence quenching. Typically, the fluorescence decreases when the ligand binds, since it displaces 1-NPN into the aqueous environment. The increase of fluorescence at higher ligand concentration after the expected decline at lower concentrations, found in some of the dose-response curves, was found to be due to the formation of micelles. These provide a hydrophobic environment, in which encapsulated 1-NPN yields higher fluorescence intensity<sup>99, 88, 100</sup>.

The best ligand tested was Retinol. It binds with a  $K_{\text{diss}}$  of 1.3  $\mu\text{M}$ . This was not unexpected since Breustedt et al.<sup>98</sup> found that many human lipocalins bind retinoids. Since many proteins bind Retinol, it is likely that in its original form, the protein bound also another ligand, which is the pheromone.

The second-best ligand found was Octyl benzoate with a  $K_{\text{diss}}$  of 2.1  $\mu\text{M}$ . Octyl benzoate, being a synthetic compound, should share structural similarities with the unknown ligand and is similar to Retinol, in respect of possessing a ring-structure and a tail of similar length. Retinols ring would be more flexible, due to not being aromatic. Interestingly, 3,7-Dimethyloctyl benzoate which also shares the characteristic of having a phenyl group, as well as a chain of the same length (10 atoms) binds with a  $K_{\text{diss}}$  of 2.8. This could mean, that there the two additional methyl groups can be accommodated in the binding pocket. Shortening and bending (due to a cis-configured double bond) the hydrophobic chain, as in the case of Amylcinnamaldehyde lead to an dissociation constant of 2.9. Also this structural changes do not influence the dissociation constant significantly. Also, Hexyl benzoate does not bind as well as Octyl benzoate, which means a longer hydrophobic chain is favourable.

Nonivamides binding ability is also worse than Octyl benzoate, though having a longer chain. This could also be due to increased hydrophilicity on the phenyl ring.

Judging from the ligands, which did not display any notable binding capabilities with the h-SAL protein, there are some groups of chemicals, which are unlikely to be the pheromone. These are chemicals with only a linear chain of C10 or longer and some oxygen-containing group, also if they contain one cis-configured double bond. Small ligands, like the terpenoids Safranal,  $\alpha$ -Santalol or Carvone are also unlikely. Although Quercetin binds with a  $K_{diss}$  of 3.1 and possesses some similarities to Androstenone and  $5\alpha$ -Androstan-17 $\beta$ -ol-3-one, steroid compounds are also unlikely to be the sought-for ligand, since the latter two display no considerable binding to h-SAL.

#### 4.6 h-SAL protein vs. h-SAL HEIE mutant protein

Since two different amino acid sequences were found for the h-SAL protein and there was no indication on which would have been the original protein, both of them were analysed. It is of interest if the two proteins differ significantly from another.

##### 4.6.1 Comparison of protein models

When comparing the protein models, first it can be seen that their three-dimensional structure overlaps very well, apart from one site where the  $\beta$ -sheet is elongated, and the loop twisted. From this similar binding data would be expected.

##### 4.6.2 Comparison of dissociation constants

*table 34: comparison of dissociation constants for different ligands for the h-SAL protein and the h-SAL protein HEIE mutation*

	h-SAL protein $K_{diss}$ [ $\mu$ M]	h-SAL HEIE mutant protein $K_{diss}$ [ $\mu$ M]
Quercetin	3.1	3.6
Octyl benzoate	2.1	2.8
Coniferyl aldehyde	3.8	4.2
Amylcinnamaldehyde	2.9	3.6
3,7-Dimethyloctyl benzoate	2.8	-
Retinol	1.3	1.2

When comparing the Dissociation constants of h-SAL and h-SAL HEIE mutation (see table 34), it can be noted, that the values are all similar. Though there is a difference between those two proteins, it doesn't influence the binding much.

##### 4.6.3 Conclusion

Differing from another in four amino acids in the protein sequence, this is likely to account for the differences in dissociation constants found. It might lead to a difference in the secondary structure, possibly forming a bigger beta-sheet, which would be a plausible explanation for the slight differences found. In this case a circular dichroism measurement could be useful.

#### 4.7 Similarity to SAL1 (*sus scrofa*) and OBP3 (*Ailuropoda melanoleuca*)

The similarity between the h-SAL protein and the related proteins pigs SAL1 and giant pandas OBP 3 could give valuable clues on which ligands could be used in further

experiments and depict how the amino acids in the binding pocket could relate to potential ligands.

#### **4.7.1 Comparison of protein models**

No valuable information could be gained from comparing the protein models. It should be also noted, that the model for h-SAL is only based on calculation and might not correspond with the real protein folding, while the models for SAL1 and OBP3 have been obtained from crystal structure. It could also be useful to devise further experiments, in order to find out if the protein was folded correctly after expression and purification.

#### **4.7.2 Comparison of ligand binding experiments**

The pigs SAL1 protein binds its pheromones, the steroid compounds 5 $\alpha$ -androst-16-en-3-one and 5 $\alpha$ -androst-16-en-3 $\alpha$ -ol. The giant pandas OBP3 has a completely different set of ligands. It binds natural terpenoids and long-chain unsaturated aldehydes. Considering that the best ligands for h-SAL found are Retinol and Octyl benzoate, and the steroid compounds, as well as most terpenoids and long-chain unsaturated aldehydes, did not show significant binding, it can be noted that the approximately 50-60 % sequence identity results in a completely different binding behaviour.

#### **4.7.3 Conclusion**

From what was found in the experiments in this thesis, the three proteins are not similar enough to expect implications for the ligand of h-SAL.

#### **4.8 Outlook**

The main issue with this research was the question, if the MUPP gene was, apart from the splice junction mutation, still representative of the gene as it was in its functional form. There is a possibility, that the original pheromone does not bind the protein anymore. Studies on the pheromones of other ape species could give a clue on which amino acids have been mutated and if they are responsible for a difference in ligand binding and if these studies would reveal a putative pheromone, this would be a good candidate for further ligand binding studies on h-SAL.

## 5. Supplementary Data

### 5.1 h-SAL HEIE protein sequence source

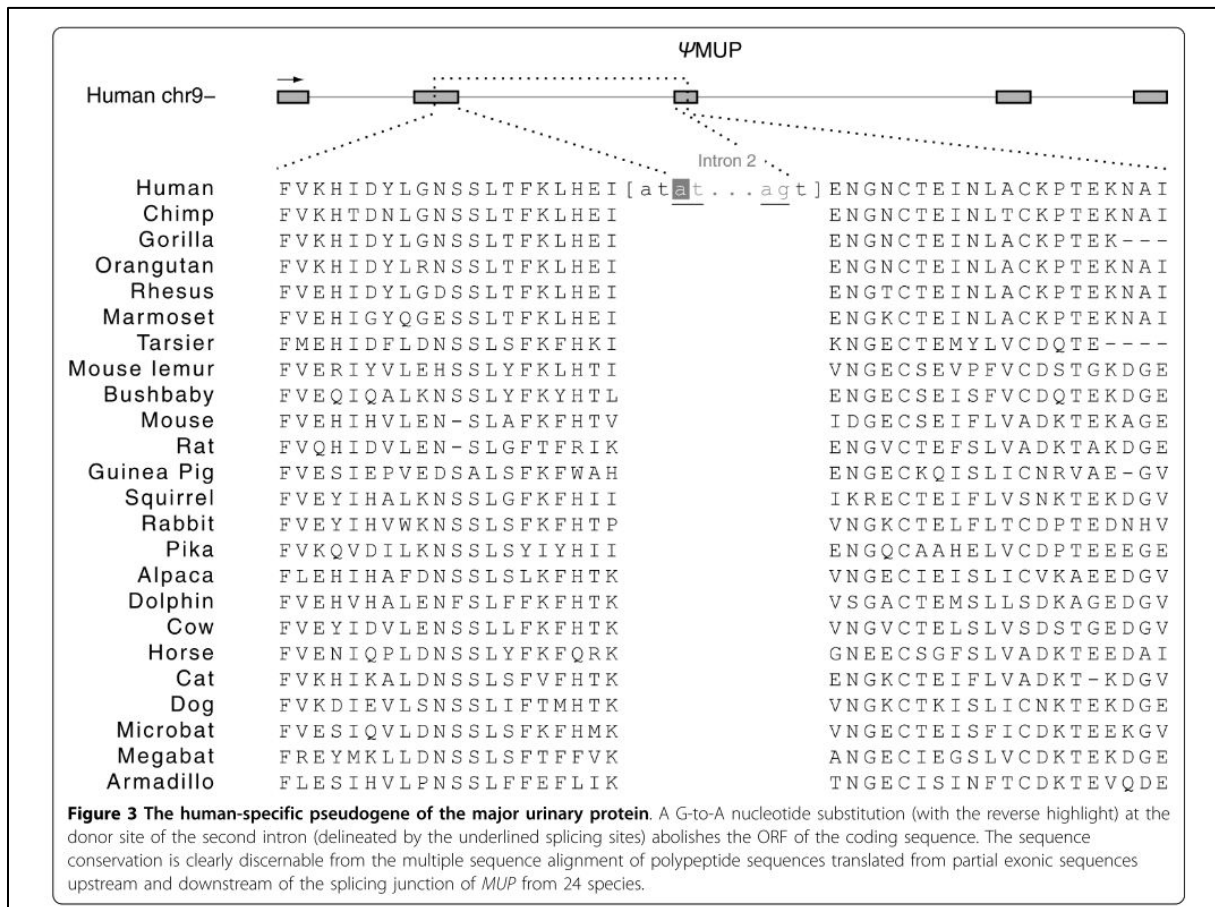


figure 72: figure 3 from the "Identification and analysis of unitary pseudogenes: Historic and contemporary gene losses in humans and other primates" paper by Zhang et al.<sup>75</sup>

## 5.2 Physical and chemical protein parameters by ExPASy

### 5.2.1 Human SAL

MQEEENNDVAV TSNFDLSKIS GEWYSVLLAS DCREKIEEDGSMRVFVKHID YLGNSSLTFFK  
 LNGNCTEINL ACKPTEKNAI CSTDYNGLNVIDILETDYDN YIYFYNKNIK NGETFLMLEL  
 YVRTPDVSSQ LKERFVKYCE EHGIDKENIF DLTKVDRCLQ ARDEGAA

**Number of amino acids:** 167

**Molecular weight:** 19,217.4

**Theoretical pI:** 4.54

**Extinction coefficients:**

Extinction coefficients are in units of  $M^{-1} \text{ cm}^{-1}$ , at 280 nm measured in water.

Ext. coefficient 19,285

Abs 0.1 % (=1 g/l) 1.004, assuming all pairs of Cys residues form cystines

Ext. coefficient 18,910

Abs 0.1 % (=1 g/l) 0.984, assuming all Cys residues are reduced

## 5.2.2 Human SAL HEIE mutation

MQEEENNDVAV TSNFDLSKIS GEWYSVLLAS DCREKIEEDG SMRVFVKHID YLGNSSLTFK  
LHEIENGNT EINLACKPTE KNAICSTDYN GLNVIDILET DYDNYIYFYN KNIKNGETFL  
MLELYVRTPD VSSQLKERFV KYCEEHGIDK ENIFDLTKVD RCLQARDEGA A

**Number of amino acids:** 171

**Molecular weight:** 19725.98

**Theoretical pI:** 4.54

**Extinction coefficients:**

Extinction coefficients are in units of  $M^{-1} \text{ cm}^{-1}$ , at 280 nm measured in water.

Ext. coefficient 19,285

Abs 0.1 % (=1 g/l) 0.978, assuming all pairs of Cys residues form cystines

Ext. coefficient 18,910

Abs 0.1 % (=1 g/l) 0.959, assuming all Cys residues are reduced

## 5.3 Sequencing results

### 5.3.1 h-Sal protein sequencing April 4, 2019

>hSALp30-9\_T7 -- 15..1051 of sequence

```
CTCTAGAATAATTTTGTAACTTTAAGAAGGAGATATACATATGCAGGAAGAGGAAAAC
AATGATGCGGTTACCAGCAATTTGACTTGTCCAAAATTAGCGGAGAATGGTATAGCGTT
CTGTTAGCGTCGGATTGTCGCGAGAAAATCGAGGAAGATGGGAGTATGCGCGTATTCTGTG
AAACACATTGACTATCTCGGTAATTCGTCACTTACGTTCAAACCTGAATGGCAATTGCACG
GAGATTAACCTGGCCTGTAAACCGACTGAAAAGAACGCGATTTGCAGTACAGACTATAAT
GGCCTGAACGTGATCGATATTCTCGAAACCGACTACGACAACCTACATCTACTTCTACAAC
AAGAACATTAAGAATGGTGAACCTTTCTGATGTTAGAACTGTATGTGCGTACTCCAGAT
GTCTCTTCCCAGCTGAAAAGAACGGTTTGTCAAATATTGCGAAGAACATGGCATCGATAAA
GAGAACATCTTTGATCTGACCAAAGTAGATCGCTGTTTGAAGCTCGTGATGAAGGTGCA
GCCTAAGAATTCGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCA
CCACTGAGATCCGGCTGCTAACAAAGCCCAGAAAGGAAGCTGAGTTGGCTGCTGCCACCGC
TGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCT
GAAAGGAGGAACATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAAGC
GCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCC
GCTCCTTTGCTTTTCTCCCTTCTTCTCGCCACGTTGCGCCGGCTTTCCCGTCAAGCT
CTAAATCGGGGGCTCCCTTAGGGTCCGATTTAGTGCTTTACGGCACCTCGACCCAAA
AACTTGATTAGGGTATGGTTACGTTAGTGGCCATCGCCCTGATAGACGGTTTTTTCGCC
CTTTGACGTTGGAGTCA
```

### Translated Sequence

**MQEEENND**AVTSNFDLSKISGEWYSVLLASDCREKIEEDGSMRVFVKHIDYLGNSSLTFKL  
NGNCTEINLACKPTEKNAICSTDYNGLNVIDILETDYDNYIYFYKNKIKNGETFLMLELYV  
RTPDVSSQLKERFVKYCEEHGHIDKENIFDLTKVDRCLQARDEGAA-

## 5.3.2 h-SAL protein HEIE mutation sequencing July 18, 2019

### Colony 5

>HSAL-M1-5\_T7 -- 5..1064 of sequence

GGGGAAATTCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGCAGG  
AAGAGGAAAACAATGATGCGGTTACCAGCAATTTTGACTTGTCCAAAATTAGCGGAGAAT  
GGTATAGCGTTCTGTTAGCGTCGGATTGTTCGCGAGAAAATCGAGGAAGATGGGAGTATGC  
GCGTATTCGTGAAACACATTGACTATCTCGGTAATTCGTCACTTACGTTCAAACCTGCATG  
AGATTGAGAATGGCAATTGCACGGAGATTAACCTGGCCTGTAAACCGACTGAAAAGAACG  
CGATTTGCAGTACAGACTATAATGGCCTGAACGTGATCGATATTCTCGAAACCGACTACG  
ACAACCTACATCTACTTCTACAACAAGAACATTAAGAATGGTGAAACCTTTCTGATGTTAG  
AACTGTATGTGCGTACTCCAGATGTCTCTTCCCAGCTGAAAGAACGGTTTGTCAAATATT  
GCGAAGAACATGGCATCGATAAAGAGAACATCTTTGATCTGACCAAAGTAGATCGCTGTT  
TGCAAGCTCGTGATGAAGGTGCAGCCTAAGAATTCGAGCTCCGTCGACAAGCTTGCGGCC  
GCACTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAACAAAGCCCCGAAAGGAA  
GCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAA  
CGGGTCTTGAGGGGTTTTTTTGTGAAAGGAGGAACCTATATCCGGATTGGCGAATGGGACG  
CGCCCTGTAGCGGCGCATTAAAGCGCGGGGGTGTGGTGGTTACGCGCAGCGTGACCGCTA  
CACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCCTTTCTCGCCACGT  
TCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTG  
CTTTACGGCACCTCGACCCCAAAAACCTTGATTAGGGTGATGGTTCACGTAATGGGCCATC  
GCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAATCCA

### Translated Sequence

**G**NSL-**NNFV-L-**

**E**GD**IHM**QEEENNDAVTSNFDLSKISGEWYSVLLASDCREKIEEDG**SM**RVFVKHIDYLGNSSLTFKLHEIENGNT  
EINLACKPTEKNAICSTDYNGLNVIDILETDYDNYIYFYKNKIKNGETFL**ML**LELYV**R**T**P**DVSSQLKERFVKYCEE  
HGIDKENIFDLTKVDRCLQARDEGAA-**E**FELRRQACGR**T**R

### Colony 6

>HSAL-M1-6\_T7 -- 4..1061 of sequence

TAGGGAAATTCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGCAG  
GAAGAGGAAAACAATGATGCGGTTACCAGCAATTTTGACTTGTCCAAAATTAGCGGAGAA  
TGGTATAGCGTTCTGTTAGCGTCGGATTGTTCGCGAGAAAATCGAGGAAGATGGGAGTATG  
CGCGTATTCGTGAAACACATTGACTATCTCGGTAATTCGTCACTTACGTTCAAACCTGCAT  
GAGATTGAGAATGGCAATTGCACGGAGATTAACCTGGCCTGTAAACCGACTGAAAAGAAC  
GCGATTTGCAGTACAGACTATAATGGCCTGAACGTGATCGATATTCTCGAAACCGACTAC

GACAACCTACATCTACTTCTACAACAAGAACATTAAGAATGGTGAACCTTTCTGATGTTA  
 GAACTGTATGTGCGTACTCCAGATGTCTCTTCCCAGCTGAAAGAACGGTTTGTCAAATAT  
 TGCGAAGAACATGGCATCGATAAAGAGAACATCTTTGATCTGACCAAAGTAGATCGCTGT  
 TTGCAAGCTCGTGATGAAGGTGCAGCCTAAGAATTCGAGCTCCGTCGACAAGCTTGCGGC  
 CGCACTCGAGCACCACCACCACCACCCTGAGATCCGGCTGCTAACAAAGCCCGAAAGGA  
 AGCTGAGTTGGCTGCTGCCACCCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAA  
 ACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGCGAATGGGAC  
 GCGCCCTGTAGCGGCGCATTAAAGCGCGGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCT  
 AACTTGGCAGCGCCCTAGCGCCCGCTCCTTTGCTTTCTTCCCTTCTTTCTCGCCACG  
 TTCGCCGGCTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGT  
 GCTTTACGGCACCTCGACCCCAAAAACCTTGATTAGGGTGATGGTTCACGTAGTGGGCCAT  
 CGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAAT

### Translated sequence

-GNSL-NNFV-L-

EGDIHMQEEENNDAVTSNFDLSKISGEWYSVLLASDCREKIEEDGSMRVFVKHIDYLGNSSLTFKLHEIENGCT  
 EINLACKPTEKNAICSTDYNGLNVIDILETDYDNYIYFYKNKIKNGETFLMLELYVRTPDVSSQLKERFVKYCEE  
 HGIDKENIFDLTKVDRCLQARDEGAA-EFELRRQACGRTRA

## 5.4 Original fluorescence data

### 5.4.1 h-SAL protein

1-NPN binding				
Perkin Elmer LS 55 Fluorescence Spectrometer				
slit Ex./Em [nm]	5			
excitation [nm]	337			
emission [nm]	380 - 450			
scan rate [nm/min]	200			
gain	low			
Blanc 1-NPN without protein				
[NPN/ $\mu$ M]	Emission	Emission	Emission	
2	3	4	4	
4	5	6	6	
6	8	8	8	
8	10	10	13	
12	14	13	16	
16	19	17	19	
1-NPN binding				
[NPN/ $\mu$ M]	Emission	Emission	Emission	
2	75	63	63	

4	125	105	101	
6	149	133	124	
8	174	154	152	
12	212	193	194	
16	242	220	224	
1-NPN binding	extracted pH 7.4	extracted pH 4.0		
[NPN/ $\mu$ M]	Emission	Emission		
2	83	81		
4	120	110		
6	142	139		
8	170	160		
12	206	190		
16	235	215		
ligand binding				
Perkin Elmer LS 55 Fluorescence Spectrometer				
slit Ex./Em [nm]	5			
excitation [nm]	337			
emission [nm]	410			
scan rate [nm/min]	200			
gain	medium			
Blanc 1-NPN without protein				
[NPN/ $\mu$ M]	Emission	Emission	Emission	
2	28	21	22	
4	44	34	35	
6	59	47	47	
8	72	69	72	
12	107	88	97	
16	156	112	114	
	(Z)-11-Hexadecenal	(+) Carvone	Dodecanal	Dodecanoic acid
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	436	473	500	480
2	394	411	430	480
4	371	407	431	482
8	366	395	417	487
12	351	377	416	482
16	365	367	412	466
	$\beta$ -Ionone	$\beta$ -Ionone	$\beta$ -Ionone	Citral
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission



0	538	410	412	460
2	475	369	380	441
4	418	338	351	416
8	371	306	319	404
12	332	282	287	384
16	307	267	264	384
	Retinol	Retinol	Retinol	Eugenol
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	451	492	413	517
2	176	241	205	479
4	142	185	156	451
8	111	132	116	430
12	97	113	103	417
16	92	103	95	403
	Safranal	Safranal	Safranal	Androstenone
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	399	288	263	460
2	358	251	237	413
4	335	231	222	388
8	301	207	199	390
12	276	188	184	421
16	259	174	170	450
	Quercetin	Quercetin	Quercetin	Quercetin
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	514	389	253	269
2	390	295	195	197
4	310	244	155	159
8	207	180	111	119
12	150	138	86	90
16	112	105	67	69
	1-Hexadecanol	1-Hexadecanol	1-Hexadecanol	Ethyl laureate
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	573	4299	4066	571
2	523	3936	3768	508
4	488	3783	3578	503
8	438	3415	3325	566
12	414	3018	3143	648
16	384	3121	3080	660
	(+) Butyl fenchol	(-) Butyl fenchol	(Z)-11-Hexadecenol	$\gamma$ -Undecalactone
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	464	448	440	463

2	421	402	381	421
4	402	382	368	389
8	378	361	338	373
12	364	346	338	362
16	354	334	338	351
	1-Dodecanol	1-Dodecanol	1-Dodecanol	Cyclamen aldehyde
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	568	4190	4150	502
2	498	3683	3897	498
4	459	3594	3564	502
8	428	3388	3407	525
12	404	3091	3162	555
16	385	3104	3094	566
	5 $\alpha$ -Androstan-17 $\beta$ -ol-3-one	Farnesol	Farnesol	Farnesol
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	484	3959	3969	504
2	457	3778	3700	440
4	443	3587	3358	403
8	427	3544	3195	378
12	412	3160	3139	352
16	407	3064	3057	340
	Geranyl acetone			
[Ligand/ $\mu$ M]	Emission			
0	541			
2	515			
4	507			
8	488			
12	486			
16	476			
tryptophan binding				
Perkin Elmer LS 55 Fluorescence Spectrometer				
slit Ex./Em [nm]	7			
excitation [nm]	295			
emission [nm]	337			
scan rate [nm/min]	200			
gain	medium			
	Retinol	Retinol	Retinol	
[Ligand/ $\mu$ M]	Emission	Emission	Emission	

2	714	795	782	
4	517	550	529	
6	442	459	432	
8	344	366	331	
12	293	310	275	
16	254	253	234	
	$\beta$ -Ionone	$\beta$ -Ionone	$\beta$ -Ionone	
[Ligand/ $\mu$ M]	Emission	Emission	Emission	
2	689	734	746	
4	628	647	673	
6	578	623	623	
8	516	536	556	
12	470	484	506	
16	425	454	461	
ligand binding				
Perkin Elmer Fluorescence Spectrometer FL6500				
slit Ex./Em [nm]	5			
excitation [nm]	337			
emission [nm]	418			
scan rate [nm/min]	240			
Photomultiplier voltage	low - 400			
Emission correction:	on			
Gain PMT	*1			
Accumulation number	1			
Blanc 1-NPN without protein				
[NPN/ $\mu$ M]	Emission	Emission	Emission	
2	261	317	429	
4	548	598	656	
6	964	932	982	
8	1030	1096	1315	
12	1466	1815	2128	
16	2073	2281	2705	
	$\alpha$ -Santalol	$\alpha$ -Santalol	$\alpha$ -Santalol	(Z)-13-Octadecenal
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	3232	3810	4145	3864
2	2796	3615	3801	3883
4	2743	3457	3513	3601
8	2524	3242	3297	3751

12	2284	3119	3190	3584
16	2259	2972	3079	3883
	Butyl benzoate	Butyl benzoate	Butyl benzoate	(Z)-9-Tetradecenal
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	2735	3649	4012	3744
2	2357	3597	3754	3249
4	2224	3196	3697	3053
8	1964	3106	3285	2738
12	2024	2810	3311	2954
16	1767	2691	3174	2897
	Hexyl benzoate	Hexyl benzoate	Hexyl benzoate	(Z)-9-Hexadecenal
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	4120	3568	3747	3131
2	3469	3515	3396	3142
4	3022	3127	3199	2745
8	2841	2755	2962	2909
12	2415	2634	2690	2802
16	2294	2330	2312	2770
	Octyl benzoate	Octyl benzoate	Octyl benzoate	Disparlure
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	4027	4371	4369	3907
2	2901	3525	3641	4840
4	2582	2952	3104	5117
8	2360	2668	2788	6579
12	1989	2338	2513	8393
16	1895	2235	2198	9512
	3,7-Dimethyloctyl benzoate	3,7-Dimethyloctyl benzoate	3,7-Dimethyloctyl benzoate	2-Methoxycinnam aldehyde
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	3554	3750	3856	4565
2	2784	3094	3325	4173
4	2523	2950	2943	4045
8	2202	2538	2631	3722
12	1988	2418	2374	3528
16	1692	2177	2220	3361
	Vanillin	Vanillin	Vanillin	Nonanoic acid
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	4759	4015	3915	4274
2	4212	3370	3560	3871
4	3904	3322	3183	3649

8	3410	2946	2608	3624
12	3137	2687	2445	3421
16	2634	2504	2321	3333
	Coniferyl aldehyde	Coniferyl aldehyde	Coniferyl aldehyde	Nonivamide
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	4290	3882	3751	4655
2	3714	3122	3218	4101
4	3013	2828	2890	4095
8	2660	2238	2269	3891
12	2220	1883	2014	3609
16	2037	1577	1770	3540
	Amylcinnamaldehyde	Amylcinnamaldehyde	Amylcinnamaldehyde	3,7,7-Trimethyloctanal
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	3847	3815	4133	4417
2	3356	3232	3296	4180
4	3133	2792	2894	3872
8	2559	2332	2260	3735
12	2195	2115	2101	3662
16	1957	1818	1703	3429
	Vanillylamine	Vanillylamine	Vanillylamine	Piperonyl alcohol
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	4771	4108	4365	4269
2	4303	4045	4081	3823
4	4005	3875	4164	3576
8	3711	3552	3612	3256
12	3836	3391	3448	3204
16	3787	3509	3411	3089
	Homovanillic acid			
[Ligand/ $\mu$ M]	Emission			
0	4167			
2	3699			
4	3826			
8	3456			
12	3333			
16	3294			
1-AMA binding				
Perkin Elmer Fluorescence Spectrometer FL6500				
slit Ex./Em [nm]	5			

excitation [nm]	375			
emission [nm]	516			
scan rate [nm/min]	240			
Photomultiplier voltage	low - 400			
Emission correction:	on			
Gain PMT	*1			
Accumulation number	1			
Blanc 1-AMA without protein				
[AMA/ $\mu$ M]	Emission			
2	103			
4	199			
6	260			
8	281			
12	552			
16	622			
1-AMA binding				
[AMA/ $\mu$ M]	Emission			
0	168			
2	347			
4	574			
8	729			
12	898			
16	1124			

#### 5.4.2 h-SAL protein HEIE mutation

ligand and 1-NPN binding				
Perkin Elmer Fluorescence Spectrometer FL6500				
slit Ex./Em [nm]	5			
excitation [nm]	337			
emission [nm]	416			
scan rate [nm/min]	240			
Photomultiplier voltage	low - 400			
Emission correction:	on			
Gain PMT	*1			
Accumulation number	1			
Blanc 1-NPN without protein				

[NPN/ $\mu$ M]	Emission	Emission	Emission	
2	261	317	429	
4	548	598	656	
6	964	932	982	
8	1030	1096	1315	
12	1466	1815	2128	
16	2073	2281	2705	
1-NPN binding				
[NPN/ $\mu$ M]	Emission	Emission	Emission	
2	3869	4649	4547	
4	5913	6712	6609	
6	7579	8398	8290	
8	8779	9630	9656	
12	11286	11729	11991	
16	13370	13322	13789	
	Nonivamide	Coniferyl aldehyde	Coniferyl aldehyde	Coniferyl aldehyde
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	4576	4035	3893	4288
2	4546	3521	3407	3552
4	4387	3182	3131	3110
8	4393	2565	2407	2566
12	4225	2155	2194	2241
16	4094	1987	1641	1744
	Amylcinnamaldehyde	Amylcinnamaldehyde	Amylcinnamaldehyde	Vanillin
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	4173	3690	3608	4124
2	3285	3233	3200	3573
4	2977	2826	2834	3309
8	2424	2297	2432	2944
12	2123	2004	2100	2560
16	1878	1715	1838	2359
	3,7-Dimethyloctyl benzoate	3,7-Dimethyloctyl benzoate	3,7-Dimethyloctyl benzoate	Butyl benzoate
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	3330	4242	3644	4816
2	2865	3373	3145	4418
4	2790	3007	2949	4174
8	2354	2691	2659	3854
12	2097	2527	2523	3735
16	1869	2134	2264	3667
	Octyl benzoate	Octyl benzoate	Octyl benzoate	$\alpha$ -Santalol

[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	4548	3597	3479	4247
2	3772	2902	2906	4066
4	3634	2541	2609	3964
8	3149	2158	2348	3534
12	2700	1937	2156	3359
16	2496	1806	1982	3260
	Retinol	Retinol	Retinol	Farnesol
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	4021	3983	3551	3748
2	2523	2277	2234	3469
4	2169	1995	1984	3235
8	1848	1701	1878	2969
12	1710	1712	1808	2771
16	1731	1592	1594	2982
	Quercetin	Quercetin	Quercetin	$\beta$ -Ionone
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	3693	3969	3659	3841
2	3067	3193	3244	3558
4	2711	2590	2687	3387
8	1927	2126	2316	3075
12	1586	1657	1791	2920
16	1387	1410	1605	2788
	Safranal	Safranal	Safranal	1-Hexadecanol
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	3775	4204	3595	3854
2	3432	3977	3317	3752
4	3222	3834	3043	3324
8	2912	3224	2843	3241
12	2716	3073	2407	3083
16	2531	2964	2270	3043
	Hexyl benzoate	Hexyl benzoate	Hexyl benzoate	1-Dodecanol
[Ligand/ $\mu$ M]	Emission	Emission	Emission	Emission
0	4124	3768	3975	3629
2	3606	3315	3463	3390
4	3264	3072	3148	3238
8	2933	2746	2810	3000
12	2734	2530	2621	2955
16	2537	2283	2430	2891





## 6. References

1. Karlson P, Lüscher M. "Pheromones": A new term for a class of biologically active substances. *Nature*. 1959;183(4653):55-56. doi:10.1038/183055a0
2. Anderbrant O. Pheromones. In: *Encyclopedia of Ecology*. ; 2008:2707-2709.
3. Scaloni A, Paolini S, Brandazza A, et al. Purification, cloning and characterisation of odorant- and pheromone-binding proteins from pig nasal epithelium. *Cell Mol Life Sci*. 2001;58(5-6):823-834. doi:10.1007/PL00000903
4. Rahmani R, Wallin EA, Viklund L, Schroeder M, Hedenström E. Identification and Field Assay of Two Aggregation Pheromone Components Emitted by Males of the Bark Beetle *Polygraphus punctifrons* (Coleoptera: Curculionidae). *J Chem Ecol*. 2019;45(4):356-365. doi:10.1007/s10886-019-01056-6
5. Pickett JA, Griffiths DC. Composition of aphid alarm pheromones. *J Chem Ecol*. 1980;6(2):349-360.
6. Cheseto X, Kachigamba DL, Bendera M, et al. Identification of Glutamic Acid as a Host Marking Pheromone of the African Fruit Fly Species *Ceratitis rosa* (Diptera: Tephritidae). *J Agric Food Chem*. 2018;66(38):9933-9941. doi:10.1021/acs.jafc.8b04481
7. Klewer N, Růžička Z, Schulz S. (Z)-pentacos-12-ene, an oviposition-detering pheromone of *Cheilomenes sexmaculata*. *J Chem Ecol*. 2007;33(11):2167-2170. doi:10.1007/s10886-007-9372-4
8. Castillo C, Maisonnasse A, Conte Y Le, Plettner E. Seasonal variation in the titers and biosynthesis of the primer pheromone ethyl oleate in honey bees. *J Insect Physiol*. 2012;58(8):1112-1121. doi:10.1016/j.jinsphys.2012.05.010
9. Pelosi P, Maida R. Odorant-binding proteins in insects. *Comp Biochem Physiol B - Biochem Mol Biol*. 1995;111(3):503-514.
10. BIGNETTI E, CAVAGGIONI A, PELOSI P, PERSAUD KC, SORBI RT, TIRINDELLI R. Purification and characterisation of an odorant- binding protein from cow nasal tissue. *Eur J Biochem*. 1985;149(2):227-231. doi:10.1111/j.1432-1033.1985.tb08916.x
11. Pelosi P. *Odorant-Binding Proteins*. Vol 29.; 1994.
12. Pelosi P. Perireceptor events in olfaction. *J Neurobiol*. 1996;30(1):3-19. doi:10.1002/(SICI)1097-4695(199605)30:1<3::AID-NEU2>3.0.CO;2-A
13. Tegoni M, Pelosi P, Vincent F, et al. Mammalian odorant binding proteins. *Biochim Biophys Acta - Protein Struct Mol Enzymol*. 2000;1482(1-2):229-240. doi:10.1016/S0167-4838(00)00167-9
14. Marchese S, Pes D, Scaloni A, Carbone V, Pelosi P. Lipocalins of boar salivary glands binding odours and pheromones. *Eur J Biochem*. 1998;252(3):563-568. doi:10.1046/j.1432-1327.1998.2520563.x
15. Cavaggioni A, Mucignat-Caretta C. Major urinary proteins,  $\alpha$ (2U)-globulins and aphrodisin. *Biochim Biophys Acta - Protein Struct Mol Enzymol*. 2000;1482(1-2):218-228. doi:10.1016/S0167-4838(00)00149-7
16. Briand L, Huet JC, Perez V, et al. Odorant and pheromone binding by aphrodisin, a hamster aphrodisiac protein. *FEBS Lett*. 2000;476(3):179-185. doi:10.1016/S0014-5793(00)01719-1
17. D'Innocenzo B, Salzano AM, D'Ambrosio C, et al. Secretory proteins as potential semiochemical carriers in the horse. *Biochemistry*. 2006;45(45):13418-13428.

doi:10.1021/bi061409p

18. Shahan K, Gilmartin M, Derman E. Nucleotide sequences of liver, lachrymal, and submaxillary gland mouse major urinary protein mRNAs: mosaic structure and construction of panels of gene-specific synthetic oligonucleotide probes. *Mol Cell Biol.* 1987;7(5):1938-1946. doi:10.1128/mcb.7.5.1938
19. Pelosi P, Baldaccini NE, Pisanelli AM. Identification of a specification olfactory receptor for 2-isobutyl-3-methoxypyrazine. *Biochem J.* 1982;201(1):245-248. doi:10.1042/bj2010245
20. Pevsner J, Trifiletti RR, Strittmatter SM, Snyder SH. Isolation and characterization of an olfactory receptor protein for odorant pyrazines. *Proc Natl Acad Sci U S A.* 1985;82(9):3050-3054. doi:10.1073/pnas.82.9.3050
21. Böcskei Z, Groom CR, Flower DR, et al. Pheromone binding to two rodent urinary proteins revealed by X-ray crystallography. *Nature.* 1992;360(6400):186-188. doi:10.1038/360186a0
22. Han L, Zhang YJ, Zhang L, et al. Operating mechanism and molecular dynamics of pheromone-binding protein ASP1 as influenced by pH. *PLoS One.* 2014;9(10):1-10. doi:10.1371/journal.pone.0110565
23. Flower DR, North ACT, Attwood TK. Mouse oncogene protein 24p3 is a member of the Lipocalin protein family. *Biochem Biophys Res Commun.* 1991;180(1):69-74. doi:10.1016/S0006-291X(05)81256-2
24. Devireddy LR, Gazin C, Zhu X, Green MR. A cell-surface receptor for lipocalin 24p3 selectively mediates apoptosis and iron uptake. *Cell.* 2005;123(7):1293-1305. doi:10.1016/j.cell.2005.10.027
25. Chakraborty S, Kaur S, Guha S, Batra SK. The multifaceted roles of neutrophil gelatinase associated lipocalin (NGAL) in inflammation and cancer. *Biochim Biophys Acta - Rev Cancer.* 2012;1826(1):129-169. doi:10.1016/j.bbcan.2012.03.008
26. Guglani L, Gopal R, Rangel-Moreno J, et al. Lipocalin 2 Regulates Inflammation during Pulmonary Mycobacterial Infections. *PLoS One.* 2012;7(11):1-12. doi:10.1371/journal.pone.0050052
27. Ertl NG, Elizur A, Brooks P, Kuballa A V., Anderson TA, Knibb WR. Molecular Characterisation of Colour Formation in the Prawn *Fenneropenaeus merguensis*. *PLoS One.* 2013;8(2). doi:10.1371/journal.pone.0056920
28. Kume S, Lee YH, Miyamoto Y, Fukada H, Goto Y, Inui T. Systematic interaction analysis of human lipocalin-type prostaglandin D synthase with small lipophilic ligands. *Biochem J.* 2012;446(2):279-289. doi:10.1042/BJ20120324
29. Marchal S, Marabotti A, Staiano M, et al. Under Pressure That Splits a Family in Two. The Case of Lipocalin Family. *PLoS One.* 2012;7(11). doi:10.1371/journal.pone.0050489
30. Flower DR. The lipocalin protein family: A role in cell regulation. *FEBS Lett.* 1994;354(1):7-11. doi:10.1016/0014-5793(94)01078-1
31. Kontopidis G, Holt C, Sawyer L. The ligand-binding site of bovine  $\beta$ -lactoglobulin: Evidence for a function? *J Mol Biol.* 2002;318(4):1043-1055. doi:10.1016/S0022-2836(02)00017-7
32. Flower DR, North ACT, Attwood TK. Structure and sequence relationships in the lipocalins and related proteins. *Protein Sci.* 1993;2(5):753-761. doi:10.1002/pro.5560020507

33. Buck L, Axel R. A Novel Multigene Family May Encode Odorant Receptors: A Molecular Basis for Odor Recognition. *Cell*. 1991;65:175-187.
34. Rodriguez I, Del Punta K, Rothman A, Ishii T, Mombaerts P. Multiple new and isolated families within the mouse superfamily of V1r vomeronasal receptors. *Nat Neurosci*. 2002;5(2):134-140. doi:10.1038/nn795
35. Dulac C, Axel R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell*. 1995;83(2):195-206. doi:10.1016/0092-8674(95)90161-2
36. Sauermost R, Freudig D. Jacobson's Organ. *Lexikon der Biologie*.
37. Cooper JG, Bhatnagar KP. Comparative anatomy of the vomeronasal organ complex in bats. *J Anat*. 1976;122(Pt 3):571-601.  
<http://www.ncbi.nlm.nih.gov/pubmed/1010789><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1231855>.
38. Bhatnagar KP, Meisami E. Vomeronasal organ in bats and primates: Extremes of structural variability and its phylogenetic implications. *Microsc Res Tech*. 1998;43(6):465-475. doi:10.1002/(SICI)1097-0029(19981215)43:6<465::AID-JEMT1>3.0.CO;2-1
39. Clancy AN, Coquelin A, Marcrides F, Gorski RA, Nobles EP. Sexual Behavior and Aggression in Male Mice: Involvement of the Vomeronasal System. *J Neurosci*. 1984;4(9):2222-2229.
40. Lomas DE, Keverne EB. Role of the vomeronasal organ and prolactin in the acceleration of puberty in female mice. *J Reprod Fertil*. 1982;66(1):101-107. doi:10.1530/jrf.0.0660101
41. Saito T. Maternal Behavior in Virgin Female Rats Following Removal of the Vomeronasal Organ. *Zoolog Sci*. 1988;5(5):1141-1143.  
<https://www.biodiversitylibrary.org/part/71796>. Accessed September 24, 2019.
42. Shapiro LS, Roland RM, Li CS, Halpern M. Vomeronasal system involvement in response to conspecific odors in adult male opossums, *Monodelphis domestica*. *Behav Brain Res*. 1996;77(1-2):101-113. doi:10.1016/0166-4328(95)00206-5
43. Meredith M. Vomeronasal, Olfactory, Hormonal Convergence in the Brain. *Ann New York Acad Sci*. 1998;855:349-361.
44. Halpern M, Martínez-Marcos A. Structure and function of the vomeronasal system: An update. *Prog Neurobiol*. 2003;70(3):245-318. doi:10.1016/S0301-0082(03)00103-5
45. Kondo Y, Sakuma Y, Tomihara K. Sensory requirements for noncontact penile erection in the rat. *Behav Neurosci*. 1999;113(5):1062-1070. doi:10.1037/0735-7044.113.5.1062
46. Wysocki CJ, Yamazaki K, Curran M, Wysocki LM, Beauchamp GK. Mice (*Mus musculus*) lacking a vomeronasal organ can discriminate MHC-determined odortypes. *Horm Behav*. 2004;46(3):241-246. doi:10.1016/j.yhbeh.2004.02.010
47. Hudson R, Distel H. Pheromonal release of suckling in rabbits does not depend on the vomeronasal organ. *Physiol Behav*. 1986;37(1):123-128. doi:10.1016/0031-9384(86)90394-X
48. Dorries KM, Adkins-Regan E, Halpern BP. Sensitivity and Behavioral Responses to the Pheromone Androstenone Are Not Mediated by the Vomeronasal Organ in Domestic Pigs. *Brain Behav Evol*. 1997;49:53-62.
49. Bhatnagar KP, Smith TD, Winstead W. The human vomeronasal organ: Part IV. Incidence, topography, endoscopy, and ultrastructure of the nasopalatine recess,

- nasopalatine fossa, and vomeronasal organ. *Am J Rhinol.* 2002;16(6):343-350.
50. Witt M, Georgiewa B, Knecht M, Hummel T. On the chemosensory nature of the vomeronasal epithelium in adult humans. *Histochem Cell Biol.* 2002;117(6):493-509. doi:10.1007/s00418-002-0407-1
  51. Knecht M, Witt M, Abolmaali N, Hüttenbrink KB, Hummel T. Das vomeronasale organ des menschen. *Nervenarzt.* 2003;74(10):858-862. doi:10.1007/s00115-003-1573-7
  52. Trotier D, Eloit C, Wassef M, et al. The Vomeronasal Cavity in Adult Humans. *Chem Senses.* 2000;25(4):369-380. doi:10.1093/chemse/25.4.369
  53. Giambanco I, Bianchi R, Ceccarelli P, et al. "Neuron-specific" protein gene product 9.5 (PGP 9.5) is also expressed in glioma cell lines and its expression depends on cellular growth state. *FEBS Lett.* 1991;290(1-2):131-134. doi:10.1016/0014-5793(91)81242-Z
  54. Smith TD, Bhatnagar KP. The human vomeronasal organ. Part II: prenatal development. *J Anat.* 2000;197 Pt 3:421-436. <http://www.ncbi.nlm.nih.gov/pubmed/11117628><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1468143>.
  55. Meisami E, Mikhail L, Baim D, Bhatnagar KP. Human Olfactory Bulb: Aging of Glomeruli and Mitral Cells and a Search for the Accessory Olfactory Bulb. *Ann N Y Acad Sci.* 1998;855:708-715. doi:10.1192/bjp.111.479.1009-a
  56. Rodriguez I, Mombaerts P. Novel human vomeronasal receptor-like genes reveal species-specific families. *Curr Biol.* 2002;12(12):409-411.
  57. Rodriguez I, Greer CA, Mok MY, Mombaerts P. A putative pheromone receptor gene expressed in human olfactory mucosa. *Nat Genet.* 2000;26(1):18-19. doi:10.1038/79124
  58. Young J, Trask BJ. V2R gene families degenerated in primates, dog and cow, but expanded in opossum. *TRENDS Genet.* 2006;23(5):212-215. doi:10.1371/journal.pgen.0010042
  59. Rajchard J. The steroids considered as human pheromones. *Ethol Ecol Evol.* 2010;22(3):311-314. doi:10.1080/03949370.2010.502327
  60. Savic I, Hedén-Blomqvist E, Berglund H. Pheromone signal transduction in humans: What can be learned from olfactory loss. *Hum Brain Mapp.* 2009;30(9):3057-3065. doi:10.1002/hbm.20727
  61. Bensafi M, Brown WM, Tsutsui T, et al. Sex-Steroid Derived Compounds Induce Sex-Specific Effects on Autonomic Nervous System Function in Humans. *Behav Neurosci.* 2003;117(6):1125-1134. doi:10.1037/0735-7044.117.6.1125
  62. Jacob S, McClintock MK. Psychological state and mood effects of steroidal chemosignals in women and men. *Horm Behav.* 2000;37(1):57-78. doi:10.1006/hbeh.1999.1559
  63. Morofushi M, Shinohara K, Funabashi T, Kimura F. Positive Relationship between Menstrual Synchrony and Ability to Smell 5alpha-Androst-16-en-3alpha-ol. *Chem Senses.* 2000;25(4):407-411. doi:10.1093/chemse/25.4.407
  64. Pause BM, Rogalski KP, Sojka B, Ferstl R. Sensitivity to androstenone in female subjects is associated with an altered brain response to male body odor. *Physiol Behav.* 1999;68(1-2):129-137. doi:10.1016/S0031-9384(99)00158-4
  65. Zeng C, Spielman AI, Vowels BR, Leyden JJ, Biemann K, Preti G. A human axillary odorant is carried by apolipoprotein D. *Proc Natl Acad Sci U S A.* 1996;93(13):6626-6630. doi:10.1073/pnas.93.13.6626

66. Wysocki CJ, Preti G. Facts, fallacies, fears, and frustrations with human pheromones. *Anat Rec*. 2004;1201-1211. doi:10.1002/ar.a.20125
67. McClinton MK. Menstrual synchrony and suppression. *Nature*. 1971;229:244-245.
68. Preti G, Wysocki CJ, Barnhart KT, Sondheimer SJ, Leyden JJ. Male Axillary Extracts Contain Pheromones that Affect Pulsatile Secretion of Luteinizing Hormone and Mood in Women Recipients. *Biol Reprod*. 2003;68(6):2107-2113. doi:10.1095/biolreprod.102.008268
69. Porter RH, Moore JD. Human kin recognition by olfactory cues. *Physiol Behav*. 1981;27(3):493-495. doi:10.1016/0031-9384(81)90337-1
70. Russell MJ. Human olfactory communication. *Nature*. 1976;260(5551):520-522. doi:10.1038/260520a0
71. Wendekind C, Seebeck T, Bettens F, Paepke AJ. MHC-dependant mate preferences in humans. *Proc R Soc Lond B*. 1995:245-249.
72. Wedekind C, Furi S. Body odour preferences in men and women: Do they aim for specific MHC combinations or simply heterozygosity? *Proc R Soc B Biol Sci*. 1997;264(1387):1471-1479. doi:10.1098/rspb.1997.0204
73. Mildner S. Human pheromones and body scents: Do they influence our behaviour? 2012.
74. Dulac C, Torello AT. Molecular detection of pheromone signals in mammals: From genes to behaviour. *Nat Rev Neurosci*. 2003;4(7):551-562. doi:10.1038/nrn1140
75. Zhang ZD, Frankish A, Hunt T, Harrow J, Gerstein M. Identification and analysis of unitary pseudogenes: Historic and contemporary gene losses in humans and other primates. *Genome Biol*. 2010;11(R26). doi:10.1186/gb-2010-11-3-r26
76. Meslin C, Brimau F, Meillour PN-L, Callebaut I, Pascal G, Monget P. The evolutionary history of the SAL1 gene family in eutherian mammals. *BMC Evol Biol*. 2011;11(148). doi:10.1186/1471-2148-11-148
77. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389-3402. doi:10.1371/journal.pone.0026263
78. Altschul SF, Wootton JC, Gertz ME, et al. Protein Database Searches Using Compositionally Adjusted Substitution Matrices. *Febs J*. 2005;272(20):5101-5109. doi:10.1016/j.freeradbiomed.2008.10.025.The
79. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep*. 2017;7(1):1-15. doi:10.1038/s41598-017-09654-8
80. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*. 2009;30(SUPPL. 1):162-173. doi:10.1002/elps.200900140
81. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303. doi:10.1093/nar/gky427
82. Bienert S, Waterhouse A, De Beer TAP, et al. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res*. 2017;45(D1):D313-D319. doi:10.1093/nar/gkw1132
83. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of

- individual protein structure models. *Bioinformatics*. 2011;27(3):343-350.  
doi:10.1093/bioinformatics/btq662
84. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612.  
doi:10.1002/jcc.20084
  85. Melrose DR, Reed HC, Patterson RL. Androgen steroids associated with boar odour as an aid to the detection of oestrus in pig artificial insemination. *Br Vet J*. 1971;127(10):497-502. doi:10.1016/S0007-1935(17)37337-2
  86. Guiraudie G, Pageat P, Cain AH, Madec I, Meillour PN Le. Functional characterization of olfactory binding proteins for appeasing compounds and molecular cloning in the vomeronasal organ of pre-pubertal pigs. *Chem Senses*. 2003;28(7):609-619.  
doi:10.1093/chemse/bjg052
  87. Swaisgood RR, Lindburg D, White AM, Hemin Z, Xiaoping Z. Chemical communication in giant pandas: Experimentation and application. In: *Giant Pandas: Biology and Conservation*. ; 2004:106-119.
  88. Zhu J, Arena S, Spinelli S, et al. Reverse chemical ecology: Olfactory proteins from the giant panda and their interactions with putative pheromones and bamboo volatiles. *Proc Natl Acad Sci U S A*. 2017;114(46):E9802-E9810. doi:10.1073/pnas.1711437114
  89. Leal WS. Reverse chemical ecology at the service of conservation biology. *Proc Natl Acad Sci U S A*. 2017;114(46):12094-12096. doi:10.1073/pnas.1717375114
  90. OpenWetWare Contributors. RbCl competent cell. OpenWetWare.  
[https://openwetware.org/mediawiki/index.php?title=RbCl\\_competent\\_cell&oldid=591507](https://openwetware.org/mediawiki/index.php?title=RbCl_competent_cell&oldid=591507). Published 2012. Accessed October 9, 2019.
  91. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol*. 2008;8:91.  
doi:10.1186/1472-6750-8-91
  92. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970;227(5259):680-685. doi:10.1038/227680a0
  93. Wingfield PT. Protein precipitation using ammonium sulfate. *Curr Protoc Protein Sci*. April 2001. doi:10.1002/0471140864.psa03fs84
  94. Matveeva E, Morisseau C, Goodrow M, Mullin C, Hammock B. Tryptophan Fluorescence Quenching by Enzyme Inhibitors As a Tool for Enzyme Active Site Structure Investigation: Epoxide Hydrolase. *Curr Pharm Biotechnol*. 2009;10(6):589-599. doi:10.2174/138920109789069260
  95. Tcatchoff L, Nespoulous C, Pernollet JC, Briand L. A single lysyl residue defines the binding specificity of a human odorant-binding protein for aldehydes. *FEBS Lett*. 2006;580(8):2102-2108. doi:10.1016/j.febslet.2006.03.017
  96. Spinelli S, Vincent F, Pelosi P, Tegoni M, Cambillau C. Boar salivary lipocalin: Three-dimensional X-ray structure and androstenol/androstenone docking simulations. *Eur J Biochem*. 2002;269(10):2449-2456. doi:10.1046/j.1432-1033.2002.02901.x
  97. Gasteiger E, Hoogland C, Gattiker A, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, ed. *The Proteomics Protocols Handbook*. Humana Press; 2005:571-607.
  98. Breustedt DA, Schönfeld DL, Skerra A. Comparative ligand-binding analysis of ten human lipocalins. *Biochim Biophys Acta - Proteins Proteomics*. 2006.  
doi:10.1016/j.bbapap.2005.12.006

99. Sun YF, de Biasio F, Qiao HL, et al. Two odorant-binding proteins mediate the behavioural response of aphids to the alarm pheromone (e)- $\beta$ -farnesene and structural analogues. *PLoS One*. 2012;7(3). doi:10.1371/journal.pone.0032759
100. Leal WS, Leal GM. Binding of a fluorescence reporter and a ligand to an odorant-binding protein of the yellow fever mosquito, *Aedes aegypti*. *F1000Research*. 2015;3:1-15. doi:10.12688/f1000research.5879.2
101. Grosdidier A, Zoete V, Michielin O. Fast docking using the CHARMM force field with EADock DSS. *J Comput Chem*. 2011;32(10):2149-2159. doi:10.1002/jcc.21797
102. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res*. 2011;39(SUPPL. 2):270-277. doi:10.1093/nar/gkr366
103. Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods*. 2011;8(10):785-786. doi:10.1038/nmeth.1701
104. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng*. 1997;10(1):1-6. doi:10.1142/S0129065797000537
105. Madeira F, Park Y mi, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019;47(W1):W636-W641. doi:10.1093/nar/gkz268
106. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019;37(4):420-423. doi:10.1038/s41587-019-0036-z



## 7. Acknowledgement

First, I want to thank Dr Paolo Pelosi for sharing his knowledge, his supervision and guidance during the course of this thesis. Special thanks to Wolfgang Knoll, PhD, for giving me the opportunity to work in this AIT – CEST collaboration. Additionally, I want to thank Philipp Fruhmann for his assistance in both administrative and scientific matters. I thank CEST for providing the funding for this thesis and my friends and colleagues from the BST group for always helping out and having fun together.

Secondly, I want to thank Professor Christian Becker for his patience and continuing support in supervising my thesis. I want to thank all professors at the University of Vienna and Lund University, which always offered knowledge, support and an amazing study environment. I want to thank my colleagues and friends from the University of Vienna and Lund University, which have studied and laughed with me and supported me during this great time of my life.

Thirdly, I want to thank my family and friends, which have always been encouraging and supporting me.

Fourthly, I want to state that Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.